CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

The Load Planning Problem for Double-Stack Intermodal Trains

Serena Mantovani
Gianluca Morganti
Nitish Umang
Teodor Gabriel Crainic
Emma Frejinger
Eric Larsen

September 2017

CIRRELT-2017-60

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

The Load Planning Problem for Double-Stack Intermodal Trains[ł]

Serena Mantovani[1,2], Gianluca Morganti[1,2], Nitish Umang[1,2,], Teodor Gabriel Crainic[1,3], Emma Frejinger[1,2,*], Eric Larsen[1,2]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Department of Computer Science and Operations Research, Université de Montréal, P.O. Box 6128, Station Centre-Ville, Montréal, Canada H3C 3J7

[3] Department of Management and Technology, Université du Québec à Montréal, P.O. Box 8888, Station Centre-Ville, Montréal, Canada H3C 3P8

Abstract. This paper presents a general methodology that addresses the load planning problem for intermodal trains. We propose a model that can deal with single- or double-stack railcars as well as arbitrary containers-to-cars matching rules. Moreover, we model weight and center of gravity constraints, stacking rules and technical loading restrictions associated with specific container types and/or contents. We propose an integer linear programming (ILP) formulation whose objective is to choose the optimal subset of containers and the optimal way of loading them on outbound railcars so as to minimize the resulting loading cost. An extensive numerical study is conducted. It shows that ignoring center of gravity constraints and containers-to-cars matching rules may lead to an overestimation of the train capacity and to select load plans that are not feasible in practice. We also show that we can solve realistic instances to optimality in reasonable computational time using a commercial ILP solver.

Keywords: Transportation, freight, double-stack train loading, load planning, intermodal railway terminals.

Acknowledgements. We gratefully acknowledge the close collaboration with personnel from the Canadian National Railway Company (CN). This research was funded by the CN Chair on Intermodal Transportation at Université de Montréal, Mitacs and a Collaborative Research and Development Grant from the Natural Sciences and Engineering Research Council of Canada (CRD-477938-14). We also acknowledge the support of Fonds de recherche du Québec through their infrastructure grants. While working on this project, the fifth author was holding the CN Chair on Intermodal Transportation, the third was a postdoctoral fellow with the Chair, and the fourth was Adjunct Professor with the Department of Computer Science and Operations Research, Université de Montréal. The authors present their most sincere thanks to Mr. Serge Bisaillon, analyst at CIRRELT, for his invaluable support. Finally, we are thankful to Jean-François Cordeau for helpful discussions.

[ł]Revised version of the CIRRELT-2016-68.

---

* Corresponding author: Emma.Frejinger@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2017

1 Introduction

Nowadays, an essential ingredient of a competitive economy is a cost-effective freight transportation system. Intermodal transportation is an important component of this system in which different transport modes are linked in order to move freight from a point of origin to a point of destination. Taking advantage of economies of scale, low volume demands are first shipped to an intermediate point, a consolidation terminal or hub, where traffic is sorted (classified) and grouped (consolidated). Then, the consolidated traffic is moved between hubs by efficient transport modes. In this paper we deal with intermodal railway transportation where containers are consolidated and transported by trains on the long-haul part of their trip. We focus on the North American market and on double-stack trains.

Intermodal transportation relies heavily on containerization because it ensures faster and safer handling and transfer between transport modes and decreases transportation costs. Intermodal containers are steel frame boxes designed to move goods across the world using different transport modes without any re-handling of the cargo. Since 2005, the containerized worldwide traffic has increased from 382 to 684 million of TEU (Twenty Foot Equivalent Unit) (CBRE Research, 2015) and, since 1990, North American ports have seen container traffic grow by an annual average of 5.3% (International Association of Ports and Harbors, 2015). This growth is placing a heavy burden on the entire consolidation-based transportation system, which must provide efficient, reliable and cost-effective services.

Terminals are major components of any intermodal transportation system and thus are critical to the entire international trade. They are special transshipment nodes that provide equipment and space where containers are processed, loaded, unloaded and stored to ensure a seamless transfer between different modes. Carriers, in our case railways, face a number of challenging planning issues, which may be examined according to the classical categorization with respect to the planning horizon, that is strategic, tactical, operational. In this study, we focus on the *load planning problem*, which is an operational problem arising at intermodal railway terminals.

Given a set of containers stored in a terminal and a sequence of railcars on an outbound train, the problem is to determine the optimal subset of containers to load and the exact way of loading them on an optimal subset of railcars while minimizing cost. We address this problem for double-stack trains. This is a challenging problem because the load plan must satisfy a number of complex loading rules that depend on specific container and railcar characteristics. For example, stacking rules depend on container sizes, weights, and contents and on center of gravity (COG) restrictions. While the methodology expounded in this paper is general, the North American market is the main focus of our attention because it is particularly challenging. Indeed, there are in North America a large number of railcar types and several more container types and containers than the standard 20 ft (feet) and 40 ft.

As we detail in Section 3, with one exception, the existing literature does not address the load planning problem for double-stack trains. Moreover, the simplifying assumptions that are adopted may lead to load plans that violate important loading rules and hence cannot be used in practice. For example, none of the studies model the COG restrictions. The objective of this paper is to propose a general methodology that addresses the load planning problem of double-stack trains taking into account all the different loading rules encountered in actuality.

There are a large number of possible ways – so-called *loading patterns* – in which containers of different sizes may be loaded onto a railcar of a particular standardized type. The multitude of railcar types and the very large cardinalities of several of the associated sets of loading patterns is a key issue. We refer to this problem as *containers-to-car matching*. In connection with this problem, we make a number of contributions. First, we propose a general model that can deal with single- and double-stack trains where railcars can be of different types and be therefore subject to different loading rules. Second, our model accounts for additional loading constraints related to the specific container types, contents and weights as well as to COG restrictions. Third, we present an extensive set of numerical results based on a case study focusing on the North American market.

The numerical results indicate that our model provides an appropriate framework for solving very large instances of the load planning problem in reasonable time using a commercial solver. They also demonstrate that failing to account for containers-to-cars matching as well as COG and stacking restrictions may lead to overestimations of the usable train capacity and to suggesting load plans that are not applicable in practice.

The remainder of the paper is structured as follows. Section 2 describes the load planning problem in detail. Section 3 is dedicated to a review of the existing literature on the assignment of containers to railcars and to highlighting our main contributions. Section 4 presents the ILP formulation of the load planning problem. Section 5 describes the content of the empirical study and examines its results. Finally, Section 6 draws conclusions and discusses possible directions for future research.

2 The double-stack train loading problem

This section presents a detailed description of the load planning problem for double-stack trains. We examine the ways in which containers and railcars can be physically matched together and explain how these loading possibilities depend on the exact characteristics of the containers and railcars. We start by successively describing the intermodal containers and the rules for stacking them as well as the intermodal railcars. We then present the rules governing the loading of containers onto railcars.

2.1 Intermodal containers

Intermodal containers are characterized by (i) their size (length and height) (ii) their type (iii) their contents and (iv) their weight, filling level and weight distribution. In order to facilitate their handling, sizes are standardized. There are four ISO standard sizes used worldwide: 20 ft high cube, 40 ft low and high cube, 45 ft high cube (the height of low cube containers is 8 ft 6 in / 2.6 m whereas it is 9 ft 6 in / 2.9 m for high cubes). This paper focuses on the North American market, where there are two additional sizes of high cubes: 48 ft and 53 ft.

For each size, containers are available in several standardized types. Some are illustrated in Figure 1. Ninety percent of the global fleet consists of general purpose containers, called "dry containers", that are steel frame boxes with 6 solid sides (upper left in the figure). Several other types of containers are designed to transport goods for which dry containers are not suitable. For instance, reefers (refrigerated containers) or heated containers are designed to carry goods needing temperature control (bottom left in the figure). During transport, the reefers can either be connected to a genset (power generator set) supplying electrical power to a number of them or can have individual power units. Figure 1 also shows an open/soft top container without a roof (upper right), an open-side container and a tank container for the transportation of liquids (bottom right). While the designs of these containers are different, their sizes remain standard. Containers can carry hazardous materials in which case special restrictions usually govern their storage and transport.

Figure 1: Examples of container types

Containers can be stacked one on top of another. In addition to rules governing the weights and the positioning of containers loaded onto railcars, the stacking of containers must conform to rules prescribing their relative position. In essence, the containers must

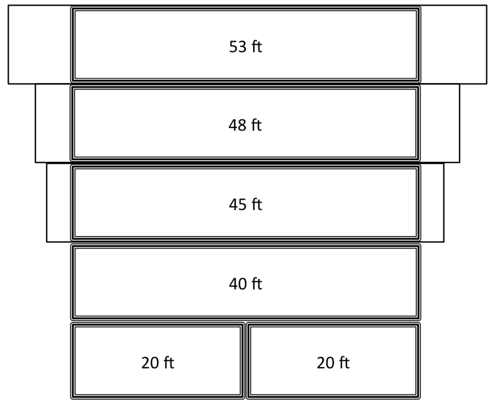

Figure 2: Container stacking at 40 ft distance

be positioned so as to ensure that their load is transferred in accordance with the design of their steel frames. Specifically, the container above can be connected to the container(s) below with four inter box connectors (IBC) designed for this purpose and the standard lengthwise distance between the connecting points where these couplings can be installed is 40 ft. This is illustrated in Figure 2 where the thick lines indicate this 40 ft distance. Hence, a 40 ft container can be loaded on top of two 20 ft but a 20 ft container cannot be loaded on top of a 40 ft. Since the connecting points are symmetrically located from the mid-length of the containers, a longer container (45, 48, 53 ft) must be centered on top of a shorter one (40, 45, 48 ft) or on top of a pair of 20 ft containers.

Lastly, we assume that there exists a per container cost associated with the failure to load an available container standing for, e.g., customer penalties for late arrival and storage costs in the terminal.

2.2 Intermodal railcars

Intermodal trains consist of a sequence of railcars designed to carry single- or double-stacked containers. Intermodal railcars are characterized by their number of platforms and by the length, weight-carrying capacity and tare weight of each one. Figure 3 illustrates a five-platform double-stack railcar. In accordance with the North American industry standard, the front platform is named A, the rear B and the other platforms C to E from front to rear. Similarly, the platforms of a three-platform railcar are named A, C, B from front to rear, and so forth. Each double-stack platform has two slots: bottom and top.

Costs are associated with the operation of a train and are incurred in the acquisition and maintenance of the locomotives and railcars, in purchasing fuel and employing crews

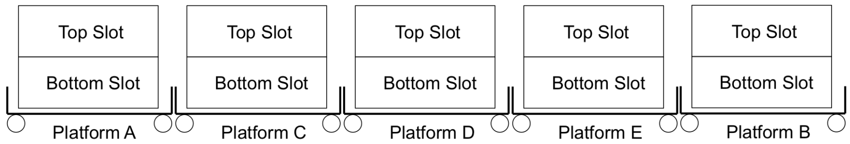

Figure 3: Five-platform double-stack railcar

(see e.g. Bouzaiene-Ayari et al., 2014). We hence assume that there is a cost associated with leaving slots empty on outbound railcars.

2.3 Loading containers on railcars

Trains are composed of blocks where a block in this context is a group of railcars that move between the block origin and destination (OD) pair of terminals without being reclassified. The purpose of grouping railcars (with different OD terminal pairs) into blocks is to minimize the transfer from one train to another or the classification of individual railcars at intermediate terminals.

Containers arrive to an intermodal rail terminal by trucks or by vessels. Upon their arrival, the containers are either classified according to the block on which they will travel and stored in the yard, or directly loaded on outbound railcars of this block. Since containers can arrive shortly before, or even during the loading operations, load plans must be computable in a short time (preferably within a few minutes).

The assignment of containers to slots must conform to a number of rules that depend on the characteristics of the railcars and the containers. We start by describing the rules that pertain to *container size* only. We refer to them as *containers-to-cars matching rules*. For the North American market, the *AAR Guide* (Association American Railroads, 2014) provides for each listed series of railcars a complete description of the combinations of container lengths that can be loaded in the bottom and top slots of each platform. Except for the bottom slots that can generally accommodate up to a pair of 20 ft containers placed end-to-end, slots can receive at most one container. Table 1 provides an illustration for a five-platform railcar series. The second block of rows is excerpted from the AAR Guide. It prescribes for each one of the five platforms (A-E) which sizes of container can be loaded in the bottom and top slots respectively. Each row in the third block states one particular loading possibility, i.e. a *loading pattern*, satisfying the prescriptions of the second block.

The platform length for this series of railcars is 40 ft whence the bottom slots can accommodate one or two 20 ft containers ($2 - 20'$ in the table) or one 40 ft ($1 - 40'$ in the table). The load in the top slot must conform to the stacking rules and to the space available. The space between platforms can in some cases be sufficient to allow the loading in the top slot of a container exceeding the length of the platform. This is examplified in Table 1 where each top slot can accommodate a 40 ft, a 45 ft or a 48 ft container and where the space between the platforms allows to load 53 ft containers in the top slot of platforms A, D and B, provided there is a 40 ft container or no container at all in top slots of platforms C and D (see table footnote). Crucially, these joint requirements imply that the containers-to-cars matching rules cannot be described for each platform separately. Slots may also be left empty. However, an upper slot can be filled only provided the slot below is filled and a top slot cannot be filled if there is a single 20 ft container loaded in the lot below. There are clearly several different ways in which to load a five-platform railcar so as to satisfy the loading capabilities stipulated in the AAR Guide and the lower rows of Table 1 exemplify a very small number of them. For example, the last row describes a loading pattern where a 40 ft container is placed in every slot except in the top one of platform E. The latter is left empty.

Bottom slot					**Top slot**				
A	*C*	*D*	*E*	*B*	*A*	*C*	*D*	*E*	*B*
AAR Guide									
$2-20'$ $1-40'$	$2-20'$ $1-40'$	$2-20'$ $1-40'$	$2-20'$ $1-40'$	$2-20'$ $1-40'$	$1-40'$ $1-45'$ $1-48'$ $1-53'^{(*)}$	$1-40'$ $1-45'$ $1-48'$	$1-40'$ $1-45'$ $1-48'$ $1-53'^{(*)}$	$1-40'$ $1-45'$ $1-48'$	$1-40'$ $1-45'$ $1-48'$ $1-53'^{(*)}$
Some examples satisfying AAR Guide									
$2-20'$	$2-20'$	$2-20'$	$2-20'$	$2-20'$	$1-48'$		$1-40'$		$1-45'$
$1-40'$	$2-20'$	$1-40'$	$2-20'$	$1-40'$	$1-45'$	$1-40'$	$1-53'$	$1-40'$	$1-53'$
$2-20'$	$1-40'$	$2-20'$	$1-40'$	$1-40'$	$1-48'$	$1-45'$	$1-48'$	$1-45'$	$1-48'$
$1-40'$	$1-40'$	$1-40'$	$1-40'$	$1-40'$	$1-40'$	$1-40'$	$1-40'$		$1-40'$

Table 1: Example of AAR Guide railcar series BN 63900 - 63909 type IBC 100 tons (*): 53 ft containers in top slot of platforms A, D and B only when a 40 ft container or none is loaded in top slot of platforms C and E.

The assignment of the containers to the slots of the railcars is conditioned by the weights of the containers and the weight-carrying characteristics of the railcars. There are two main loading restrictions with respect to the weight of the load on a platform. First, the total weight of the containers loaded on a platform must be smaller than the weight capacity of the platform. Second, a condition pertaining to the height of the center of gravity must be satisfied. This expression is used in the North American railway industry with a meaning identical to that of the expression *center of mass*. Although it designates more generally the mean location of a distribution of mass in space, it is defined in the context of railway operations as the mean location of mass along the vertical axis of a platform. The AAR Guide states "*...The COG for a double-stack car and the load in the platform must be less than or equal to 98 inches at top of rail. Reference Rule 89, Section C.2.e. in the AAR Field Manual*". A failure to obey this rule would imply practically that the container placed in the top slot of the platform is too heavy in comparison with the weight of the container(s) placed in the bottom slot. This situation would be viewed as a risk factor to a derailment. While the actual COG depends on the filling level and the load distribution in the containers, the COG restriction stated in the AAR Guide relies on the assumption of a uniform weight distribution. In the case of a solid body with uniform weight distribution, the center of mass is the same as the centroid of the body. In this paper we follow the AAR definition of COG.

The COG restriction is expressed as an upper bound on the on the weight of the container in the top slot, given the characteristics of the container in the bottom slot. Figure 4 provides an illustration for a single platform. There are three solid bodies: the platform p, the bottom container i and the top container i'. Their centroids are illustrated with black dots and the associated heights from the top of the rail are denoted m_p, m_i and $m_{i'}$, respectively. The bottom and top containers are connected with IBCs. Under the assumption of a uniform weight distribution, the height of the COG m for the three solid bodies is

$$m = \frac{m_p g_p + m_i g_i + m_{i'} g_{i'}}{g_p + g_i + g_{i'}} \quad (1)$$

where g_p is the platform tare weight and g_i and $g_{i'}$ are the weights of the bottom and top containers respectively. According to the AAR Guide, $m \leq M$ where M equals 98 in (2.5 m). Using (1) and M it is possible to compute a maximum weight c for the top container. By rearranging (1) and using M instead of m and c instead of $g_{i'}$ we obtain

$$c = \frac{g_p(M - m_p) + g_i(M - m_i)}{m_{i'} - M}. \quad (2)$$

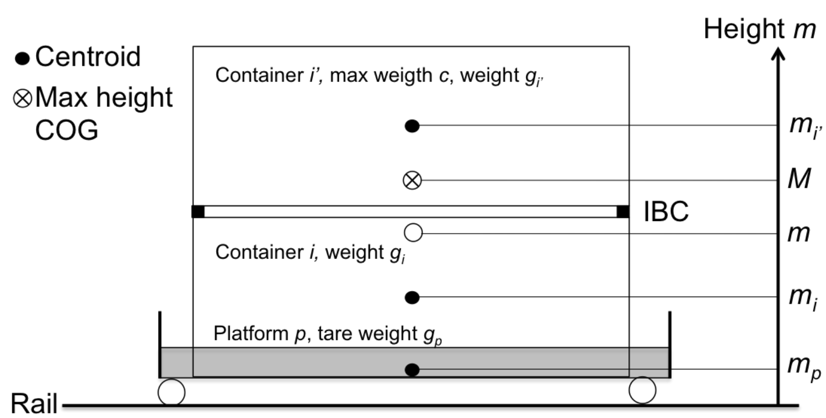

Figure 4: COG restriction

Containers exist in a diversity of types carrying a diversity of contents and rules are attached to particular combinations of types and contents. These rules give rise to a number of additional constraints in the container loading problem. For the North American market we have identified six *technical loading restrictions* that apply to certain types of containers and contents or to combinations thereof:

1. Loading is restricted to railcars having a given minimum weight-carrying capacity (independently of railcar series). This restriction applies to containers whose weight is above a certain threshold and needs to comply with additional restrictions not captured by the weight-carrying capacity of a platform.

2. Loading is restricted to certain positions in the sequence of railcars (e.g., hazardous material).

3. Loading is restricted to high weight capacity railcars (only certain railcar series).

4. Loading in top slot is forbidden.

5. Loading in top slot and double stacking is forbidden.

6. Loading must be on a platform within a maximum distance from a specific container (e.g., for the reefers that must be connected to a genset).

This set of technical loading restrictions is sufficiently general to cover the specificities that we have identified thus far in relation with the North American railways. Additional technical loading restrictions might have to be defined in order to reflect new or presently unknown railway policies or country regulations.

In summary, we focus on the load planning problem: Given a set of containers stored in a terminal, a sequence of railcars, and the relevant constraints, determine the subset of containers to load and the exact way of loading them. The objective is to minimize the cost of unloaded containers and the cost of empty slots. A key performance indicator currently used to measure the efficiency of a load plan is the *slot utilization*, which measures the percentage of the available slots on the railcars that are occupied in the load plan (Burriss, 2003). We note that we focus on a deterministic setting, and that we do not model the different handling costs associated with retrieving containers in the terminal. Our goal is to develop a general methodology, which can be used within a decision support tool that provides load plans to decision makers. We deal with all the loading rules and restrictions that arise for double-stack trains, by taking into account the multitude of containers and railcars types that exist in the North American market.

3 Literature review

The load planning problem may be viewed as a special case of the packing-cutting-knapsack problems (Martello and Toth, 1990; Dowsland and Dowsland, 1992; Dyckhoff et al., 1997). The goals and the associated models are different, however. For example, in two- (Lodi et al., 2002) and three-dimensional packing (Crainic et al., 2008) one faces a much larger number of items than the number of available (or desirable) loading units (bins) and the dimensions of the items span a broad range of values from tiny to almost as large as the bin. One then focuses on identifying the "best" bin and the "best" position in the bin to load all items in as few bins as possible. In the rail load planning problem, on the other hand, bins - the railcar platforms - and the items - the containers - are fundamentally of similar dimensions, the positioning being determined by the physical configurations of both. The goal is then changed from packing as many items as possible into as few bins as possible to identifying the best combination (assignment) of given container dimensions and weights to the available railcars given technical loading constraints (e.g., total weight and COG).

COG and load balancing concerns also arise when planning the loading of vehicles for other freight transportation modes, e.g., trucking, sea and air transportation. Each transportation mode has its own vehicle and operation characteristics, resulting in particular forms of these general restrictions. For example, the axle weight restriction for trucks may result in particular requirements for weight distribution when loading the containers before even the ocean segment of their trip (Lim et al., 2013). The distribution of weight, and thus of containers, is of capital importance for the stability of ships and airplanes. The COG of the vehicle thus becomes a hard safety constraint in ship stowage (Steenken et al., 2004; Stahlbock and Voß, 2008) and airplane (Mongeau and Bes, 2003), but while the number of container re-handles (at intermediate stops) is generally not relevant in the latter case, it is an element to be taken into account in the former case (Imai et al., 2006).

We open this overview of the literature relevant to the rail load planning problem by pointing to two surveys whose scopes extend to intermodal freight transportation activities in general: See Crainic and Kim (2007) for the planning of intermodal carrier and terminal operations, and Carlo et al. (2014) for transportation activities in container terminals. Several studies focus on the train blocking problem (e.g., Bodin et al., 1980; Newton et al., 1998; Barnhart et al., 2000). For general views on the rail load planning problem per se, see Heggen et al. (2016) for a recent classification of the existing literature and Boysen et al. (2013) for a comprehensive overview of the planning issues that arise specifically in railway yards, including the load planning problem.

Specifically in connection with the rail load planning problem, Feo and Gonzalez-Velarde (1995) made the first contribution and, later on, Powell and Carvalho (1998) dealt with the problem of balancing the flat cars over a network from a load planning perspective. Similarly to these two studies, most contributions in the literature addressed the simpler single-stack load planning problem where the set of matches between container and railcar combinations is smaller than the double-stack one. The existing literature examined simpler settings than that of this paper, accounting mainly for limits on axial and total train weight. It has generally focused on objectives related to, e.g., minimizing handling costs in the yard (e.g., Corry and Kozan, 2006, 2008; Ambrosino et al., 2011; Ambrosino and Siri, 2015) or train set-up costs (Bruns and Knust, 2012;

Bruns et al., 2014), rather than optimizing the capacity made available by a given train or block as in this paper.

The authors focusing on single-stack loading deal with load planning at different degrees of detail. For example, Corry and Kozan (2008) consider matching different container and railcar types, while Corry and Kozan (2006) do not. Bruns and Knust (2012) extend the former work by considering both the matching problem between containers and railcars and the weight constraints. Heggen et al. (2016) build on the latter and integrate a number of practical loading constraints. Ambrosino et al. (2011) and Ambrosino and Siri (2015) minimize re-handling in the yard and unproductive movements of cranes. Anghinolfi et al. (2014) consider several container lengths and possible railcar (platform) loading combinations to accommodate them, combined to axial and train weight restrictions. For single-stack loading problems, Dotoli et al. (2015) consider issues often addressed during previous planning processes (e.g., block planning) such as the positioning of the loaded cars within the train and their transfer from one train to another. Besides the weight-related restrictions, the authors also address the so-called commercial value of the train measured by the priority and, possibly, the value of the containers. Dotoli et al. (2017) extends this work to a decision-support system, which also addresses the management of the containers in the yard. Finally, Bruns et al. (2014) consider several sources of uncertainty (regarding, e.g., weights, lengths and equipment failures) in a robust optimization approach.

The aforementioned studies focus on single-stack trains, and on the main challenges associated with optimizing yard or transport operations rather than on train loading. We study the operational problem of loading double-stack trains in a context where the tactical train and block plan, i.e., what trains are operated, what sequence of trains moves each block, and what block moves each container (loaded on a car) from its origin to its destination, were previously constructed. Loading double-stack trains is a difficult problem taking place in a complex setting. On the one hand, it requires considering the matching between a multitude of different railcar (platform types and configurations) and container types, while putting containers on top of other containers. On the other hand, we enforce a good number of technical constraints, in particular the COG restrictions and stacking rules. We therefore focus on the train (block) load planning problem, assuming the cars making up the train (block) are given, as well as the containers to load.

To the best of our knowledge, the first contributions to the double-stack loading literature aimed for automatic heuristic rules and procedures. Pacanovsky et al. (1995) embedded such procedures into a simulation-based decision-support system. Lai et al. (2008a) is the first optimization study on the double-stack load planning problem. Similarly to this study, they also ignore handling costs. Their focus is on minimizing the aerodynamic drag of double-stack trains that depends on the gaps between containers and the location of these gaps along the train. They present an integer linear programming formulation, but they make a number of simplifying assumptions. First, they address the matching among containers and railcars types, deriving the loading patterns without considering the possible platform dependencies. This implies that loading rules can be defined for platforms independently (Table 1 shows an example where this assumption is invalid). Second, they ignore the possible dependencies between the loadings of the individual railcars in the sequence (such dependencies are introduced by technical restrictions, e.g., the requirement that reefers must be loaded in close proximity of the genset supplying the required power). These dependencies make it inappropriate to define loading rules over each platform or each railcar independently. Third, they study the problem without accounting for COG and technical restrictions. The authors extend the model to a rolling horizon setting and show that one could improve the loading by considering several trains at a time. Lai et al. (2008b), Lang et al. (2011) consider COG concerns within the study of a limited number of containers-to-cars configurations based on the case of Chinese rail. Detailed formulas are developed for each configuration and are embedded into a multi-objective formulation.

As this literature survey illustrates, there currently does not exist a comprehensive optimization model for the double-stack train loading problem considering a realistic set of constraints and a broad range of container and railcars types. We present such a model is the next section.

4 Mathematical formulation

A realistic load plan must comply with the applicable set of loading patterns, weight and COG restrictions, stacking rules and technical restrictions. This section presents an integer linear programming (ILP) formulation of the load planning problem whose objective is to maximize slot utilization. This is accomplished by minimizing an appropriately weighted sum of the cost of containers that are not loaded and the cost of railcars that are used for loading. We open this section with detailed explanations, first, of the mathematical structure describing the containers-to-cars matching rules and, second, of the COG constraints. Next, we provide a detailed description of the full ILP formulation.

4.1 Modeling containers-to-cars matching

We model the containers-to-cars matching through *loading patterns*. A loading pattern describes a feasible assignment of container lengths to the slots of a railcar. Whereas Corry and Kozan (2008) and Lai et al. (2008b) also use loading patterns, the main difference here lies in the fact that we account for dependencies between the loadings of the platforms on a railcar. As illustrated by the example shown in Table 1, accounting for these dependencies is important but leads to an exponential growth in the number of loading patterns as the number of platforms increases.

Let H be the set of standard container lengths in feet. In our case, $H = \{20, 40, 45, 48, 53\}$. A loading pattern $k \in K_j$ is a n-tuple that specifies the total number of containers of each length $h \in H$ that can be loaded on each platform of a given railcar $j \in J$. We show an example of a one-platform railcar in Table 2. Each row corresponds to a loading pattern and there is a total of 11 possible patterns $|K_j| = 11$, including empty slots but excluding an empty railcar. The first row shows, for example, that the first platform can hold one 20 ft container, the second, two 20 ft containers, the third, one 40 ft container and so forth. Notice that the loading patterns do not indicate the slots in which the containers can be loaded, only the number per platform. This information may be inferred from the content of the platform based on stacking rules. When a railcar consists of several platforms the n-tuples are concatenated from left to right. For example, a particular pattern for a railcar comprising three platforms is described by a 3-n-tuples concatenation. The set of loading patterns K_j is composed of all feasible loadings as described in Association American Railroads (2014) and discussed in depth in Section 2.3. The set of railcars J can be divided into subsets J_t ($\bigcup_{t \in T} J_t = J$) where each type $t \in T$ has a unique set of loading patterns K_t. The loading patterns can be generated and stored a priori for all railcar types.

k / h	*20*	*40*	*45*	*48*	*53*
1	1	0	0	0	0
2	2	0	0	0	0
3	0	1	0	0	0
4	2	1	0	0	0
5	2	0	1	0	0
6	2	0	0	1	0
7	2	0	0	0	1
8	0	2	0	0	0
9	0	1	1	0	0
10	0	1	0	1	0
11	0	1	0	0	1

Table 2: Example set of loading patterns K_j for a one-platform railcar j (the rows correspond to one pattern $k \in K_j$ and columns to container lengths $h \in H$)

The number of loading patterns increases exponentially with the number of platforms. However, in rather general circumstances, there may exist redundancies among them, in the sense that, for a given railcar, a number of distinct loading patterns may accommodate exactly the same set of containers. As an example of redundancy that justifies a reduction in the number of loading patterns without loss of generality, consider the three distinct loading patterns assigning one 40 ft container to a three-platform railcar: the 40 ft container could be placed on platform A, B or C. However, in view of the independence of the platforms it would be sufficient to consider only one of these three patterns.

We remove redundancy between loading patterns by defining equivalence classes over the loading patterns and selecting a single representative loading pattern for the class. We define the equivalence classes as follows: loading patterns for a given railcar type are deemed equivalent if they can be obtained one from the other through a permutation of the individual n-tuples describing the loadings of each platform. We call this *equivalence with respect to platform permutations*. Proceeding in this fashion, we can achieve important reductions in the cardinalities of the sets of loading patterns without loss of generality (we present descriptive statistics for the North American railcar fleet in Section 5).

It is important to note that the classes of equivalence for the loading patterns must be defined in accordance with the characteristics of the load planning problem at hand, if their use is not to cause a loss of generality in the description of the loading possibilities. An example where equivalence with respect to platform permutations may not hold is the model presented in Lai et al. (2008a) where the aerodynamic efficiency of the load plan is optimized and where the longitudinal position of the containers is of importance. Notice, however, that the key aspect of optimizing aerodynamic efficiency resides in choosing the location of empty slots/platforms, and that this aspect becomes significantly less important in situations with excess demand. In this case, the equivalence classes may not impact the quality of the solution.

4.2 Modeling the COG restriction

We described the COG restriction in Section 2.3 as an upper limit on the weight of container i', $g_{i'}$, loaded in the top position on a platform p. Stated as an inequality the weight limit (2) is

$$g_{i'} \leq c = \frac{g_p(M - m_p) + g_i(M - m_i)}{m_{i'} - M}. \tag{3}$$

Notice that it depends non-linearly on the characteristics of the container loaded in the bottom slot. In the following we indicate how we can express the COG restriction using linear constraints.

While (3) depends on the size of the containers through the height of their centroids (m_i and $m_{i'}$), we only need to consider four height configurations. Indeed, as illustrated in Figure 5, containers are either low cube (LC) or high cube (HC). Hence, for a given container in the bottom slot i, $m_{i'}$ can take two values depending on i' being HC or LC. Let m_i^{LC} and m_i^{HC} denote these two possible values of $m_{i'}$. The platform height and tare weight (m_p and g_p) are constants. We can now write (3) with the following two inequalities:

$$g_{i'} \leq c_i^{\text{LC}} = \frac{g_p(M - m_p) + g_i(M - m_i)}{m_i^{\text{LC}} - M} \quad \text{if } i' \text{ is LC}, \tag{4}$$

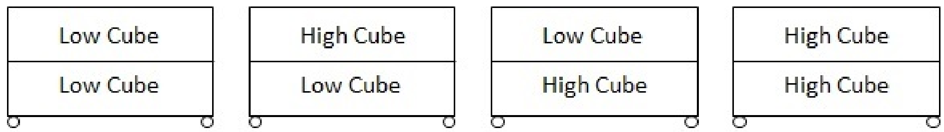

Figure 5: Four height configurations relevant to the COG constraints

$$g_{i'} \leq c_i^{\text{HC}} = \frac{g_p(M - m_p) + g_i(M - m_i)}{m_i^{\text{HC}} - M} \quad \text{if } i' \text{ is HC.} \tag{5}$$

The COG restriction is always satisfied when the container in the bottom slot has the same weight or is heavier than the one in the top slot ($g_i \geq g_{i'}$). Finally we note that we can compute c_i^{LC} and c_i^{HC} for all containers $i \in N$ a priori. We take into account that two 20 ft containers can be loaded in the same bottom slot by considering this pair of 20 ft containers as the bottom load and compute the weight limit on the upper container accordingly, based on the total weight of the pair.

4.3 ILP formulation

A container of length $h \in H$, $i \in N_h$, $N = \bigcup_{h \in H} N_h$, is characterized by its weight g_i, length l_i, cost if left on the ground π_i, and, possibly, by a particular technical loading restriction. Let N_{LC} and N_{HC} denote the sets of low-cube and high-cube containers. Let $N_s \subseteq N$ be the set of containers affected by technical loading restriction $s \in S$. Then, for the $s = 1, \ldots, 6$ classes of technical loading restrictions identified for the North American market (Section 2.3), we have: D_W, the minimum weight-carrying capacity of a railcar that can receive container $i \in N_{s_1}$; pre-processed parameter $F_j = 1$ when, given the sequence of railcars, one cannot load on railcar j containers $i \in N_{s_2}$ (0, when one can); indicator $\alpha_j = 1$ when railcar $j \in J$ with high-weight capacity U_j may receive containers $i \in N_{s_3}$; R, the maximum number of consecutive platforms on the train between a refrigerated container and the source of electric power.

A railcar $j \in J$ is characterized by its weight-carrying capacity G_j and a utilization (by at least one container) cost τ_j. Let P represent the set of platforms of all railcars, and P_j the set of platforms of railcar $j \in J$. Each platform p is characterized by its length L_p, its weight-carrying capacity G_p, and a sequence number γ_p, numbered from head to tail of the train. Let Q be the set of all slots, Q_p the set of slots of a given platform p, and μ_q be a binary parameter equal to 1 if $q \in Q$ is a bottom slot, 0 otherwise. Furthermore, let $c_i^{\text{LC}\ p}$ and $c_i^{\text{HC}\ p}$ be the low cube and high cube weight limit, respectively, of the top slot for container $i \in N$ loaded in the bottom slot of platform $p \in P$, calculated using (4) and (5).

Railcars are defined by their type as presented in Section 4.1. For the sake of notational simplicity we let K_j be the set of loading patterns for railcar $j \in J$, with $n^h_{k(p)}$, the number of containers of length $h \in H$ on platform p in loading pattern $k \in K$.

We define two main sets of *decision variables*. First, $v_{iq} = 1$, if container $i \in N$ is assigned to slot $q \in Q$, and zero otherwise. Second, $w_{jk} = 1$ if railcar $j \in J$ is assigned loading pattern $k \in K_j$, and zero otherwise. We also define two sets of auxiliary binary variables linking the container assignment variables v_{iq} to platforms and railcars. More precisely, let $y_{ip} = 1$, if container $i \in N$ is loaded on platform $p \in P$, 0 otherwise, and $x_{ij} = 1$, if container $i \in N$ is loaded on railcar $j \in J$, 0 otherwise. The model then becomes:

$$\min \quad \sum_{i\in N}\pi_i(1-\sum_{q\in Q}v_{iq})+\sum_{j\in J}\tau_j(\sum_{k\in K_j}w_{jk}) \tag{6}$$

$$\text{s.t} \quad \sum_{q\in Q}v_{iq}\leq 1 \qquad \forall i\in N \tag{7}$$

$$y_{ip}=\sum_{q\in Q_p}v_{iq} \qquad \forall i\in N,\ \forall p\in P \tag{8}$$

$$x_{ij}=\sum_{p\in P_j}y_{ip} \qquad \forall i\in N,\ \forall j\in J \tag{9}$$

$$\sum_{k\in K_j}w_{jk}=1 \qquad \forall j\in J \tag{10}$$

$$\sum_{k\in K_j}n^h_{k(p)}w_{jk}=\sum_{i\in N_h}y_{ip} \qquad \forall p\in P_j,\ \forall j\in J,\ \forall h\in H \tag{11}$$

$$\sum_{i\in N}\sum_{q\in Q_p}\mu_q v_{iq}l_i\leq L_p \qquad \forall p\in P \tag{12}$$

$$\sum_{i\in N}y_{ip}g_i\leq G_p \qquad \forall p\in P \tag{13}$$

$$\sum_{i\in N_{\text{LC}}}\sum_{q\in Q_p}(1-\mu_q)v_{iq}g_i\leq\sum_{i\in N}\sum_{q\in Q_p}\mu_q v_{iq}c_i^{\text{LC}\ p} \qquad \forall p\in P \tag{14}$$

$$\sum_{i\in N_{\text{HC}}}\sum_{q\in Q_p}(1-\mu_q)v_{iq}g_i\leq\sum_{i\in N}\sum_{q\in Q_p}\mu_q v_{iq}c_i^{\text{HC}\ p} \qquad \forall p\in P \tag{15}$$

$$\sum_{j\in J}x_{ij}(G_j-D_W)\geq 0 \qquad \forall i\in N_{s_1} \tag{16}$$

$$\sum_{j\in J}x_{ij}\ F_j=0 \qquad \forall i\in N_{s_2} \tag{17}$$

$$\sum_{j\in J}x_{ij}(\alpha_jU_j-g_i)\geq 0 \qquad \forall i\in N_{s_3} \tag{18}$$

$$\sum_{q\in Q}v_{iq}(1-\mu_q)=0 \qquad \forall i\in N_{s_4} \tag{19}$$

$$y_{ip}+\sum_{q\in Q_p}v_{i'q}(1-\mu_q)\leq 1 \qquad \forall i\in N_{s_5},\ \forall i'\in N\setminus i,\ \forall p\in P \tag{20}$$

$$\sum_{p\in P}\gamma_p y_{ip}-\sum_{p\in P}\gamma_p y_{i'p}\leq R+(|P|-R)(1-\sum_{p\in P}y_{i'p}) \quad \forall i,i'\in N_{s_6},i\neq i' \tag{21}$$

$$v_{iq}\in\{0,1\} \qquad \forall i\in N,\ \forall q\in Q \tag{22}$$

$$y_{ip}\in\{0,1\} \qquad \forall i\in N,\ \forall p\in P \tag{23}$$

$$x_{ij}\in\{0,1\} \qquad \forall i\in N,\ \forall j\in J \tag{24}$$

$$w_{jk}\in\{0,1\} \qquad \forall j\in J,\ \forall k\in K \tag{25}$$

The objective (6) of the ILP model seeks to minimize the total of the cost of the containers left on the ground and the cost of the railcars used to load at least one container. We note that $\sum_{k \in K_j} w_{jk} = 0$ when railcar j is not used. Under certain cost parameterizations this generalized cost leads to the maximization of the slot utilization.

There are five sets of loading constraints in the model. The *assignment constraints* (7) ensure that each container $i \in N$ can be assigned to at most one slot $q \in Q$. For a given container $i \in N$, $\sum_{q \in Q} v_{iq} = 0$ implies that the container is not assigned to any slot and thus is left on the ground. Constraints (8) and (9) define the auxiliary assignment variables y_{ip} and x_{ij} of containers to slots and platforms, respectively.

The *loading pattern constraints* (10) ensure that exactly one loading pattern $k \in K_j$ is assigned to each railcar $j \in J$. Constraints (11) link variables w_{jk} and y_{ip}, enforcing that the number of loaded containers of length $h \in H$ on platform $p \in P$ equals $n^h_{k(p)}$. Constraints (12) ensure that the length of the container(s) loaded in the bottom slot of platform $p \in P$ does not exceed the length of the platform.

The *weight capacity constraints* (13) ensure that the total weight of the loaded containers does not exceed the maximum allowable weight limit of the platform. The *COG restrictions* are modeled by constraints (14) and (15).

In addition to dimensional and weight restrictions, there are also a variety of *technical loading restrictions* imposing or forbidding the loading of certain types of containers on specific railcars or slots. Constraints (16) - (21) correspond to the six classes of technical loading restrictions we identified for the North American market. Notice, however, that these can be easily extended to describe other company policies or country regulations.

Constraints (16) state that containers $i \in N_{s_1}$ can only be loaded on railcars that have the minimum weight-carrying capacity D_W. Constraints (17) restrict the loading of containers $i \in N_{s_2}$ to a particular railcar in the given sequence of railcars, while constraints (18) restrict the loading of containers $i \in N_{s_3}$ to railcars with a suficiently high weight capacity.

There are two types of stacking constraints. First, containers $i \in N_{s_4}$ cannot be loaded in the top slot (Constraint (19)). Second, containers $i \in N_{s_5} \subseteq N_{s_4}$ cannot be loaded in a top slot and cannot be double stacked (20).

Constraints (21) concern the storage of refrigerated containers that need a source of electrical power (the genset is a container $i \in N_{s_6}$), limiting the distance between any two loaded containers belonging to the set N_{s_6}. Finally, expressions (22)-(25) define the domain of the decision variables.

5 Numerical results

Two numerical studies have been conducted. The first one assesses the effects of containers-to-cars matching and COG restrictions on load planning solutions. The second one examines the relationship between the particular sets of characteristics presented by the load planning problems and the computation times required for their solution. Without loss of generality, and to simplify the discussion, we assume for all instances that container costs $\pi_i = \pi,\ \forall i \in N$ and railcar cost $\tau_j = \tau,\ \forall j \in J$.

The Java programming language was used for processing the data and for running and post processing the solutions on an Intel(R) Core(TM) i5-5300U, 2.30 GHz CPU processor equipped with 24 GB of RAM. The ILP optimization model was solved using a 32-bit version of the IBM ILOG CPLEX 12.6 solver, with a preset computational time limit of 10 hours. The reported computational times only account for the solver's CPU time, since the computations associated with pre-processing the data and post processing the solutions required negligible time.

The sets of loading patterns were generated one time, a priori, using the Python programming language. The generation for every railcar series found in the North American fleet required less than 45 minutes of CPU time.

Table 3 reports descriptive statistics for the sets of loading patterns associated with all railcar types in the North American fleet. Each row corresponds to single- (S) or double-stack (D) railcars with a given number of platforms. The third column gives the number of unique sets of loading patterns (there is a total of 60 sets). The next block of three columns reports the average, minimum and maximum cardinalities of the original sets of loading patterns and the last three columns the same figures for the reduced sets. The latter are obtained by defining equivalence classes as we describe in Section 4.1. Whereas the cardinalities of the sets of loading patterns increase with the number of platforms and become very large, the use of equivalence classes results in important reductions.

5.1 Effects of containers-to-cars matching and COG restrictions

In order to isolate and measure the effects of containers-to-cars matching and COG restrictions on load planning solutions, we designed a stylized experiment. It is based on 396 generated instances, differentiated with respect to the main container and railcar characteristics, namely length and weight of containers and length and weight capacity of railcar platforms. The goal is to examine the changes in the use of train capacity resulting from changes in the characteristics of the containers and railcars. In the following, we first describe the instance generation, and then we present the results.

# platforms	Single-/ double-stack	# sets	Cardinality original set Avg.	Min	Max	Cardinality reduced set Avg.	Min	Max
1	S	1	7	7	7	7	7	7
1	D	6	18	6	27	15	6	21
2	S	1	25	25	25	15	15	15
3	S	3	171	125	245	56	35	80
3	D	10	4,741	1,000	9,261	940	220	1,771
4	D	2	106921	83,521	130,321	6,080	4,845	7,315
5	S	6	3,803	32	7,776	194	6	371
5	D	31	485,664	1,024	4,084,101	10,915	56	53,130

Table 3: Descriptive statistics for the sets of loading patterns for railcars in the North American fleet

In all generated instances, we keep the length of the railcar sequence as fixed. Yet, the capacity in terms of number of slots can still vary since platforms have different lengths. We define four railcar scenarios that involve either one- or five-platform cars with either 40 ft or 53 ft platforms. We assume that it is possible to include 25% more 40 ft slots than 53 ft ones and fix the capacity to 250 40 ft slots or 200 53 ft slots. For each railcar scenario, we choose one railcar type whence the set of loading patterns is the same for all railcars in a given scenario.

We consider 18 different scenarios for the container sets. The number of containers in each set is equal to the number of slots on the train. They have different characteristics in terms of mix of container lengths and weights. There are five different length mixes: 50% 40 ft containers and 50% 53 ft containers, 75% 40 ft containers and 25% 53 ft containers and vice versa, 100% 40 ft and 100% 53 ft. The containers are assigned weights in three different ways, two deterministic and one random. The deterministic cases present favorable weight distributions, i.e. where one can use the maximum capacity because there are no restrictions related to the COG. This holds when either all containers have equal weight, or half of the containers are light and half heavy. We draw weights at random for 40 ft and 53 ft containers from uniform distributions, respectively in [8,000;62,000] lb (equivalent to [4;31] tons) and [11,000;72,000] lb (equivalent to [5.5;36] tons), and we generate 20 instances for each length mix. We define light and heavy to be the first and third quartile, respectively. The 18 different scenarios defining the container sets are denoted S1–S18 and each scenario comprises 22 instances (20 random and 2 deterministic) whence a total of 396 instances are solved.

Tables 4 and 5 report the results for the scenarios with one- and five-platform railcars, respectively. In both tables, the first two columns show the number of loaded containers and the number of used railcars in the optimal solution. The third column shows CPLEX solution time. The gap is not reported because all the instances are solved to optimality. Note that in case of random weights, we report an average over the 20 instances.

250 CONTAINERS	125 ONE 40ft PLATFORM RAILCARS		
INSTANCE DESCRIPTION	**LOADED CONTAINERS**	**USED RAILCARS**	**TIME [s]**
S1 : 250 40ft containers			
1) Containers same weights	250	125	17.8
2) Containers half low and half high weights	250	125	22.07
3) Containers random weights	250	125	46.30
S2: 200 40ft containers and 50 53ft containers			
1) Containers same weights	250	125	14.61
2) Containers half low and half high weights	250	125	22.02
3) Containers random weights	244	123	51.54
S3: 150 40ft containers and 100 53ft containers			
1) Containers same weights	250	125	11.21
2) Containers half low and half high weights	250	125	14.49
3) Containers random weights	238	120	76.0
S4: 125 40ft containers and 125 53ft containers			
1) Containers same weights	250	125	10.22
2) Containers half low and half high weights	250	125	11.49
3) Containers random weights	234	118	200.82
S5: 100 40ft containers and 150 53ft containers			
1) Containers same weights	200	100	9.28
2) Containers half low and half high weights	200	100	12.81
3) Containers random weights	200	100	35.62
200 CONTAINERS	**100 ONE 53ft PLATFORM RAILCARS**		
INSTANCE DESCRIPTION	**LOADED CONTAINERS**	**USED RAILCARS**	**TIME [s]**
S6: 200 40ft containers			
1) Containers same weights	200	100	8.10
2) Containers low and high weights	200	100	16.22
3) Containers random weights	200	100	15.63
S7: 125 40ft containers and 75 53ft containers			
1) Containers same weights	200	100	9.69
2) Containers low and high weights	200	100	13.55
3) Containers random weights	200	100	23.47
S8: 75 40ft containers and 125 53ft containers			
1) Containers same weights	200	100	9.45
2) Containers low and high weights	200	100	15.77
3) Containers random weights	200	100	28.07
S9: 0 40ft containers and 200 53ft containers			
1) Containers same weights	200	100	9.84
2) Containers low and high weights	200	100	8.74
3) Containers random weights	200	100	27.09

Table 4: Effects of matching problem and COG restrictions: number of loaded containers, number of used railcars and solution time for one-platform railcars

250 CONTAINERS	25 FIVE 40ft PLATFORM RAILCARS		
INSTANCE DESCRIPTION	**LOADED CONTAINERS**	**USED RAILCARS**	**TIME [s]**
S10: 250 40ft containers			
1) Containers same weights	250	25	111.69
2) Containers half low and half high weights	250	25	167.99
3) Containers random weights	250	25	178.14
S11: 200 40ft containers and 50 53ft containers			
1) Containers same weights	250	25	126.59
2) Containers half low and half high weights	250	25	132.83
3) Containers random weights	233	24	935.88
S12: 150 40ft containers and 100 53ft containers			
1) Containers same weights	250	25	119.32
2) Containers half low and half high weights	250	25	120.38
3) Containers random weights	219	24	764.72
S13: 125 40ft containers and 125 53ft containers			
1) Containers same weights	200	25	116.75
2) Containers half low and half high weights	200	25	127.73
3) Containers random weights	200	25	331.79
S14: 100 40ft containers and 150 53ft containers			
1) Containers same weights	175	25	113.67
2) Containers half low and half high weights	175	25	125.89
3) Containers random weights	175	25	329.88
200 CONTAINERS	**20 FIVE 53ft PLATFORM RAILCARS**		
INSTANCE DESCRIPTION	**LOADED CONTAINERS**	**USED RAILCARS**	**TIME [s]**
S15: 200 40ft containers			
1) Containers same weights	200	20	514.81
2) Containers low and high weights	200	20	539.24
3) Containers random weights	200	20	733.83
S16: 125 40ft containers and 75 53ft containers			
1) Containers same weights	200	20	574.13
2) Containers low and high weights	200	20	636.86
3) Containers random weights	200	20	799.33
S17: 75 40ft containers and 125 53ft containers			
1) Containers same weights	200	20	471.43
2) Containers low and high weights	200	20	844.02
3) Containers random weights	200	20	932.13
S18: 0 40ft containers and 200 53ft containers			
1) Containers same weights	200	20	429.88
2) Containers low and high weights	200	20	513.44
3) Containers random weights	200	20	859.54

Table 5: Effects of matching problem and COG restrictions: number of loaded containers, number of used railcars and solution time for five-platform railcars

The results show that the solution time is less than 200 seconds for one-platform railcars, while it increases to a maximum of 935 seconds for five-platform railcars. This results from the increased cardinality of the sets of loading patterns K_j. In the case of 40 ft one-platform railcars, the maximum number of containers that can be loaded is equal to the number of slots, that is 250. However, 53 ft containers can only be loaded in the top slot since their length exceeds the platform length. Hence, as long as there are less than 125 53 ft containers in the instances (S1-S4 in Table 4), all slots can be used under the favorable weight values of the deterministic instances. In Table 4, S5 is an example of a scenario where the number of 53 ft containers exceeds the number of top slots, and where, as a result, even though the weight values are favorable, some of the containers cannot be loaded.

Loading patterns may impose additional restrictions for five-platform railcars. In particular, as seen in Section 4.1, 53 ft containers cannot be loaded in the top slots of contiguous 40 ft platforms. This is illustrated in scenarios S13 and S14 of Table 5 where regardless of the favorable deterministic weight values, some slots must be left empty because of the high proportion of 53 ft containers.

The results obtained with differing container weights clearly illustrate that the maximum capacity of 40 ft railcars can only be reached under favorable weight settings. In the case of random container weights where COG restrictions play a role, the results indicate a decrease in the average number of loaded containers, even when their lengths are well-matched to the railcars. The 53 ft platform railcars are more flexible because they can accommodate 53 ft containers also in the bottom position. For every set of weight values, it is possible to load all containers. However, since 53 ft railcars are longer, there are only 200 slots compared to 250 for the 40 ft platform railcars.

In order to load as many containers as possible, there is a trade-off between using 53 ft and 40 ft platforms and this trade-off depends on both the size of the containers and their weights. For example, the 250 slots on the 40 ft platforms can only be used under the most favorable settings. As the share of 53 ft containers increases (in particular for the random weight setting), the number of containers loaded decreases towards 200 (and might possibly reach less than 200 if weights are unfavorable as in S14).

This stylized numerical study demonstrates that ignoring COG restrictions and containers-to-cars matching (i.e., assuming favorable container length and weight settings) may lead to an overestimation of the usable capacity offered by the railcars in a block or in a train. Indeed, in the case of 40 ft platforms, the number of loaded containers varies between 175 and 250 while the capacity is 250.

5.2 Computational times

This section examines the computation times required to solve realistic instances of the load planning problem. For this purpose, we generate instances of diverse sizes and characteristics. Sets of railcars are sampled at random from the types available in the North American fleet. We also generate sets of containers with different cardinalities comprising containers with diverse characteristics.

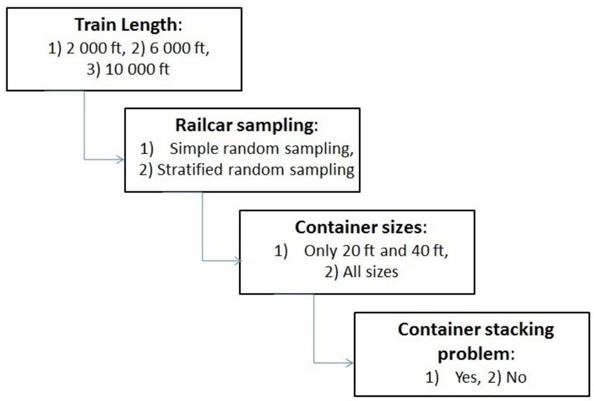

Figure 6: Overview of the instance generation process

Figure 6 overviews the generation process. Four block lengths are considered: 2,000 ft (0.6 km), 6,000 ft (1.8 km), 10,000 ft (3 km) and 14,000 ft (4.3 km). Recall that the load planning problems are solved over blocks of railcars and that the latter are groups of railcars identified according to the common terminals of origin and destination of the containers they transport. Loading and unloading a block is highly dependent on the layout of the terminals, in particular on the length of the tracks and the location of the container storage in relation to the tracks. In practice, block lengths of 10,000 or 14,000 ft may therefore be considered unreasonably long. We include them for the sake of comparison. For the same reason, 6,000 ft is considered very large and 2,000 ft a realistic length.

For each train length we generate 20 sequences of railcars by sampling the North American distribution of railcar types. We apply two different sampling protocols: simple random (10 sequences) and stratified random (10 sequences). We classify the railcars in the North American fleet according to their flexibility in accommodating a diversity of load patterns. This flexibility is indexed over the railcar types by calculating for each one the average number of loading patterns per platform. Figure 7 shows a histogram of the share of railcar types over the values of the flexibility index. Since a large proportion of the railcar types exhibit high flexibility indices, the railcar sequences generated by simple random sampling present a greater share of railcars with high flexibility index values than the sequences generated by stratifying the sample over the values of the flexibility index.

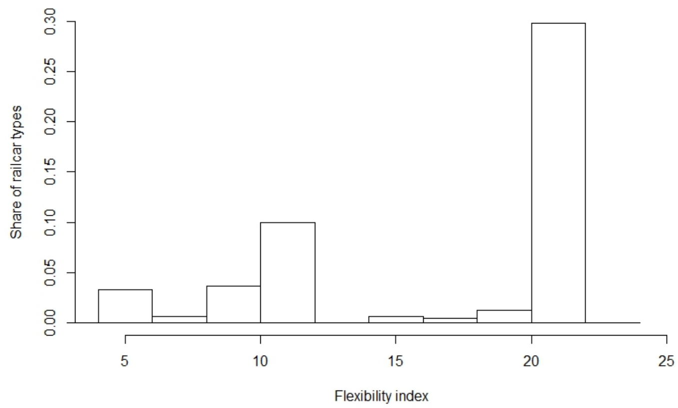

Figure 7: Share of North American railcar types over flexibility index values

Containers are selected as follows. For each block length, and for each railcar sequence, we consider four sets of containers. The sizes of these sets are equal to 1.5 times the number of slots in the railcar sequence. This ensures that solutions achieving a slot utilization close to 100% are possible. There are two mixes of container lengths: one with only 20 ft and 40 ft containers and one with all lengths (20,40,45,48 and 53 ft). In a manner similar to that of the previous section, we assign weights to containers by drawing from a weight distribution that is conditional on container size. In order to assess the influence of the technical loading restrictions on computational time, some instances include containers affected by stacking restrictions. In these instances, each container is assigned a stacking restriction at random which results in a share of 3%.

We solve a total of 240 instances, that is, six railcar sequence scenarios with 10 sequences each, and four container sets per railcar sequence. Table 6 displays the average computational time. The results show that we can find an optimal solution for the instances of 2,000 ft in less than 24 seconds on average, for all scenarios. The stratified random sampling contains a higher share of railcars with low flexibility index than the other, which results in longer computational time. The average computational time is longer for instances of 6,000 ft. Still, all instances can be solved in less than 17 minutes on average. We note that the instances of 10,000 ft can be solved to optimality but in the most complex case (i.e., including containers of all sizes and featuring technical restrictions) requires on average 3.5 hours which is not reasonable for the problem at hand. The cardinality of the sets of loading patterns has an important impact on computational time. This can clearly be seen by comparing the computational time for the instances including only 20/40 ft containers with those including all container sizes. The number of variables ranges from some 45,000 for the simplest instances (2,000 ft and 20/40 ft container) to some 1.9M in the more complex settings (14,000 ft and all container sizes with stacking restrictions). Similarly, the number of constraints ranges from some 19,000 to 2.6M.

Summing up, this numerical study shows that we can solve realistic instances in short computational time and very large instances (6,000 ft) in reasonable time. It is also possible to solve to optimality instances of 10,000 ft and the simpler settings of the 14,000-ft instances. However, in the most complex settings, this cannot be accomplished in reasonable time on the single core machine used here. In the most complex setting (14,000 ft, all container sizes and stratified sampling) the optimal solution is not found for 6 out of 10 instances in 10 hours. The average gap for those instances is 0.43.

Block length / Sampling protocol	Containers without technical loading restrictions				Containers with technical loading restrictions			
	20 & 40 ft		All sizes		20 & 40 ft		All sizes	
	Time [s]	Slot U [%]	Time [s]	Slot U [%]	Time [s]	Slot U [%]	Time [s]	Slot U [%]
2,000 ft								
Simple random	7.11	100.00	13.10	100.00	7.95	100.00	14.36	100.00
Stratified random	11.97	99.54	21.20	99.78	12.92	99.54	24.74	99.78
6,000 ft								
Simple random	184.59	100.00	450.96	100.00	661.05	100.00	639.37	100.00
Stratified random	209.15	99.58	576.12	99.54	377.63	99.50	1,077.98	99.58
10,000 ft								
Simple random	967.42	100.00	4,010.52	100.00	1,963.78	100.00	8,266.35	100.00
Stratified random	1,653.16	99.50	4,217.13	99.54	2,755.56	99.50	13,254.41	99.54
14,000 ft								
Simple random	4,714.59	100.00	17,510.18	100.00	18,220.68	100	22,860.32	98.78
Stratified random	5,677.45	99.64	22,828.71	99.67	15,295.58	99.64	**27,508.56**	**99.59**

In bold: statistics based on 4/10 instances that were solved to optimality in a time limit of 36,000 s.

Table 6: Average computational time and average slot utilization for instances with diverse characteristics

6 Conclusions and directions for future research

In this paper we studied the load planning problem for double-stack intermodal trains. Given a set of containers stored in a terminal, a sequence of railcars, and the relevant constraints, determine the subset of containers to load and the exact way of loading them. The objective is to minimize the cost of unloaded containers and the cost of empty slots. Under certain cost parameterizations this generalized cost leads to the maximization of the slot utilization, a key performance indicator in the industry. Previous studies in the literature either do not address the load planning problem for double-stack trains or make simplifying assumptions that may lead to load plans that violate important loading rules. The problem related to double-stack trains is challenging because the load plan must respect a number of loading rules that depend on container and railcars characteristics such as containers-to-cars matching and COG restrictions.

We formulated an ILP model and made a number of contributions. First, we proposed a general methodology that can deal with double- or single-stack railcars with arbitrary loading patterns. The patterns account for loading dependencies between the platforms on a given railcar. Second, we modeled COG restrictions, stacking rules and a number of technical loading restrictions associated with certain types of containers and/or goods.

We presented two numerical studies. On the one hand, we show that we can solve realistic size instances in reasonable time using a commercial ILP solver. On the other hand, we illustrate that failing to account for containers-to-cars matching as well as COG restrictions may lead to an overestimation of the available train capacity. The results showed that the computational time varies with the size and characteristics of the instances. For example, it is more time consuming to solve instances with five-platform railcars and several container sizes compared to fewer platforms and only 20 and 40 ft containers. This is due to the cardinality of the sets of loading patterns. It is also more time consuming to solve instances with containers having technical loading restrictions than those without.

On the one hand one may use the proposed methodology in decision-aid tools for terminal managers in charge of the load planning. On the other hand, in a more tactical or strategic planning setting to assess railcar fleet management decisions. We also note that we can extend the model to plan several trains ahead under perfect information, similar to Lai et al. (2008b).

There are several possible directions for future research. One could extend the model in order to consider handling costs in the yard or the potential penalties of not delivering on time, for example, by selecting containers according to their location in stacks or their priority and time left before the due date, respectively. Furthermore, several aspects of the problem may be subject to uncertainty, for example, the availability and characteristics of containers and railcars. Modeling this uncertainty is another topic of future research.

Acknowledgements

We gratefully acknowledge the close collaboration with personnel from the Canadian National Railway Company (CN). This research was funded by the CN Chair on Intermodal Transportation at Université de Montréal, Mitacs and a Collaborative Research and Development Grant from the Natural Sciences and Engineering Research Council of Canada (CRD-477938-14). We also acknowledge the support of Fonds de recherche du Québec through their infrastructure grants. While working on this project, the fifth author was holding the CN Chair on Intermodal Transportation, the third was a postdoctoral fellow with the Chair, and the fourth was Adjunct Professor with the Department of Computer Science and Operations Research, Université de Montréal. The authors present their most sincere thanks to Mr. Serge Bisaillon, analyst at CIRRELT, for his invaluable support. Finally, we are thankful to Jean-François Cordeau for helpful discussions.

References

D. Ambrosino and S. Siri. Comparison of solution approaches for the train load planning problem in seaport terminals. *Transportation Research Part E: Logistics and Transportation Review*, 79:65–82, 2015.

D. Ambrosino, A. Bramardi, M. Pucciano, S. Sacone, and S. Siri. Modeling and solving the train load planning problem in seaport container terminals. In *Automation Science and Engineering (CASE), 2011 IEEE Conference on*, pages 208–213. IEEE, 2011.

D. Anghinolfi, C. Caballini, and S. Sacone. Optimizing train loading operations in innovative and automatedcontainer terminals. In *Proceedings of the 19th World Congress The International Federation of Automatic Control Cape Town, South Africa. August 24-29, 2014*, pages 9581–9586, 2014.

Association American Railroads. Loading Capabilities Guide, 2014. available at: https://www.aar.org/StatisticsAndPublications/Publications?title=LoadingCapabilitiesGuide.

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

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

B. Bouzaiene-Ayari, C. Cheng, S. Das, R. Fiorillo, and W. B. Powell. From single commodity to multiattribute models for locomotive optimization: a comparison of optimal integer programming and approximate dynamic programming. *Transportation Science*, 50(2):366–389, 2014.

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

F. Bruns and S. Knust. Optimized load planning of trains in intermodal transportation. *OR Spectrum*, 34(3):511–533, 2012.

F. Bruns, M. Goerigk, S. Knust, and A. Schöbel. Robust load planning of trains in intermodal transportation. *OR Spectrum*, 36(3):631–668, 2014.

C. Burriss. Eastbound statistical analysis: slot utilization and mixed cars. In *OASIS Users' Conference, Austin, TX*, 2003.

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

CBRE Research. 2015 North America Ports and Logistics Annual Report, 2015. available at: http://www.cbre.us/research/2015-US-Reports/Pages/2015-North-America-Ports-Logistics-Annual-Report.aspx (Accessed on September 25, 2016).

P. Corry and E. Kozan. An assignment model for dynamic load planning of intermodal trains. *Computers & Operations Research*, 33(1):1–17, 2006.

P. Corry and E. Kozan. Optimised loading patterns for intermodal trains. *OR Spectrum*, 30(4):721–750, 2008.

T. Crainic, G. Perboli, and R. Tadei. Extreme Point-Based Heuristics for Three-Dimensional Bin Packing. *INFORMS Journal on Computing*, 20(3):368–384, 2008.

T. G. Crainic and K. H. Kim. Intermodal transportation. In C. Barnhart and G. Laporte, editors, *Handbooks in Operations Research and Management Science*, volume 14, chapter 8, pages 467–537. Elsevier, 2007.

M. Dotoli, N. Epicoco, and C. Seatzu. An improved technique for train load planning at intermodal rail-road terminals. In *20th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2015)*. IEEE, 2015.

M. Dotoli, N. Epicoco, M. Falagario, C. Seatzu, and B. Turchiano. A decision support system for optimizing operations at intermodal railroad terminals. *IEEE Transactions on Systems Man and Cybernetics: Systems*, 47(3):487–501, 2017.

K. A. Dowsland and W. B. Dowsland. Packing problems. *European Journal of Operational Research*, 56(1):2–14, 1992.

H. Dyckhoff, G. Stheithauer, and J. Terno. Cutting and packing. In F. M. M. Dell'Amico and S. Martello, editors, *Annotated Bibliographies in Combinatorial Optimization*. John Wiley & Sons, Chichester, U.K., 1997.

T. A. Feo and J. L. Gonzalez-Velarde. The intermodal trailer assignment problem. *Transportation Science*, 29(4):330–341, 1995.

H. Heggen, K. Braekers, and A. Caris. Optimizing train load planning: Review and decision support for train planners. In *International Conference on Computational Logistics*, pages 193–208. Springer, 2016.

A. Imai, K. Sasaki, E. Nishimura, and S. Papadimitriou. Multi-objective Simultaneous Stowage and Load Planning for a Container Ship with Container Rehandle in Yard Stacks. *European Journal of Operational Research*, 171(2):373–389, 2006.

International Association of Ports and Harbors. World Port Data 2015, 2015. available at: http://www.iaphworldports.org/statistics (Accessed on September 20, 2016).

Y.-C. Lai, C. P. Barkan, and H. Önal. Optimizing the aerodynamic efficiency of intermodal freight trains. *Transportation Research Part E*, 44(5):820–834, 2008a.

Y.-C. Lai, Y. Ouyang, and C. P. Barkan. A rolling horizon model to optimize aerodynamic efficiency of intermodal freight trains with uncertainty. *Transportation Science*, 42(4): 466–477, 2008b.

M. Lang, J. Przybyla, and X. Zhou. Loading containers on double-stack cars: Multi-objective optimization models and solution algorithms for improved. http://www.civil.utah.edu/~zhou/container_scheduling.pdf, 2011.

A. Lim, H. Ma, C. Qiu, and W. Zhu. The Single Container Loading Problem with Axle Weight Constraints. *International Journal of Production Economics*, 144(2):358–369, 2013.

A. Lodi, S. Martello, and M. Monaci. Two-dimensional packing problems: a survey. *European Journal of Operational Research*, 141:241–252, 2002.

S. Martello and P. Toth. *Knapsack Problems - Algorithms and Computer Implementations*. John Wiley & Sons, Chichester, UK, 1990.

M. Mongeau and C. Bes. Optimization of Aircraft Container Loading. *IEEE Transactions on Aerospace and Electronic Systems*, 39(1):140–150, 2003.

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

D. L. Pacanovsky, C. T. Jahren, R. N. Palmer, R. R. Newman, and D. Howland. A decision support system to load containers to double-stack rail cars. *Civil Engineering Systems*, 11(4):247–261, 1995.

W. B. Powell and T. A. Carvalho. Real-time optimization of containers and flatcars for intermodal operations. *Transportation Science*, 32(2):110–126, 1998.

R. Stahlbock and S. Voß. Operations Research at Container Terminals: - A Literature Update. *OR Spectrum*, 30(1):1–52, 2008.

D. Steenken, S. Voß, and R. Stahlbock. Container Terminal Operation and Operations Research - A Classification and Literature Review. *OR Spectrum*, 26(1):3–49, 2004.