A Survey on Container Processing in Railway Yards

Nils Boysen
Lehrstuhl für Allgemeine Betriebswirtschaftslehre/Operations Management, Friedrich-Schiller-Universität Jena,
D-07743 Jena, Germany, nils.boysen@uni-jena.de

Malte Fliedner
Institut für Betriebswirtschaftslehre, Fachgebiet Operations Research, Technische Universität Darmstadt,
D-64289 Darmstadt, Germany, fliedner@bwl.tu-darmstadt.de

Florian Jaehn, Erwin Pesch
Institut für Wirtschaftsinformatik, Universität Siegen, D-57068 Siegen, Germany
{florian.jaehn@uni-siegen.de, erwin.pesch@uni-siegen.de}

In spite of extraordinary support programs initiated by the European Union and other national authorities, the percentage of overall freight traffic moved by train is in steady decline. This development has occurred because the macroeconomic benefits of rail traffic, such as the relief of overloaded road networks and reduced environmental impacts, are counterbalanced by severe disadvantages from the perspective of the shipper, e.g., low average delivery speed and general lack of reliability. Attracting a higher share of freight traffic on rail requires freight handling in railway yards that is more efficient, which includes technical innovations as well as the development of suitable decision support systems. This paper reviews container processing in railway yards from an operations research perspective and analyzes basic decision problems for the two most important yard types: conventional rail–road and modern rail–rail transshipment yards. Furthermore, we review the relevant literature and identify open research challenges.

Key words: railway system; railway yard; container processing; decision support; survey

History: Received: January 2011; revisions received: July 2011, December 2011; accepted: January 2012.
Published online in Articles in Advance June 21, 2012.

1. Introduction

From a macroeconomic perspective, shifting freight traffic from the road network to the railway system is certainly desirable for several reasons. Increased rail usage for mid- to long-distance freight can provide an opportunity to relieve the often congested roads of major trading countries, not only in central Europe, where freight traffic (and road freight in particular) is projected to increase considerably over the next few decades (see, e.g., Progtrans 2007), but also in the United States (Frittelli 2003), Canada (TC 2004), and Australia (Meyrick and Associates 2006). Moreover, rail traffic is typically preferred on the basis of its reduced environmental impact, for instance with regard to CO₂ emissions, which are estimated to be more than four times higher per ton-kilometer in the case of road traffic (Allianz-pro-Schiene 2008). In spite of extraordinary support programs of the European Union (see, e.g., Tsamboulas, Vrenken, and Lekka 2007) and other national authorities (e.g., U.S. DOT 1991, 1998), railway systems still seem to be considerably less attractive for shippers, especially when compared with freight traffic by truck. Within the last 25 years, the fraction of overall freight traffic moved by train declined from 20% (1970) to 10% (2005) (EU 2007a). This development is mainly because of the lack of investment in railway infrastructure over the last decade. As a consequence, the modest absolute increase in rail freight volumes led to an overutilization of critical resources and thus to severe competitive disadvantages, from a shipper’s perspective, compared with road traffic. According to recent studies, only 53% of freight trains reach their destination with fewer than 30 minutes’ delay (EU 2007b), and the average delivery speed of a freight train is estimated to lie between 10 km/h (VDA 2006) and 20 km/h (EU 2001), predominantly because of long waiting times in rail yards.

In addition to necessary investments in the network infrastructure, a second critical driver to increase the market share of rail freight is therefore to establish more efficient freight handling processes in existing railway yards employing, for example, suitable optimization approaches and decision support systems. This paper surveys rail yard operations at conventional rail–road and modern rail–rail terminals from
an operations research perspective by characterizing important decision problems and solution approaches published in the scientific literature. On the basis of this analysis, future research challenges are identified.

The remainder of the paper is structured as follows. Section 2 defines the scope of this review by distinguishing different types of rail yards and briefly describing the associated decision problems. The two most important yard types—conventional rail–road terminals and modern rail–rail transshipment yards—are then studied individually in dedicated sections with regard to the core decision problems, existing literature, and future research challenges. Finally, §5 concludes the paper.

2. Scope of Review

A railway yard is a special transshipment node in a rail network where loads for trains are processed, i.e., collected, rearranged, unloaded, intermediately stored, loaded, and/or picked up. Our survey exclusively treats freight railway yards where containerized loads are processed. In spite of the large variety of such load units, we simply use the term “container” throughout this paper for any cargo carrier that is separable from its railcar and moveable by an alternative mode of transportation, for instance by truck or ship. The variety of containers processed in rail yards is typically much higher than that of seaports, where only a handful of different container types are transshipped. For instance, the German rail network distinguishes between 23 different container types (Kombiverkehr 2008) and a comparable number has been reported for North America (Muller 1999).

A selection of these containers includes, for instance, tank containers (a), standardized intermodal containers (b), swap bodies, railroaders, and trailers (c) loadable on some flatcar (a and b) or pocket wagon (c) as depicted in Figure 1(a)–(c). Excluded from our survey are therefore passenger railway systems (see, e.g., Freling et al. 2005; Kroon, Lentink, and Schrijver 2008) or railway yards where nonseparable cargo carriers are processed. Examples of the latter yard type are those handling bulk cargo (d), packaged goods (e), or liquids (f), which are loaded directly on a railcar (e.g., in the case of tree trunks) or in a nonseparable load carrier (see Figure 1(d)–(f)).

Usually a freight yard serves at least one of two main purposes in a railway network:

(i) On the one hand, a terminal may serve as an interface in intermodal transport so that shipments can be interchanged between the rail system and an alternative mode of transportation, such as trucks or ships. Typically in such a system, trucks serve customers on the last mile, whereas trains operate the long-haul routes. Alternatively, an intermodal yard might be located in (or nearby) a seaport for moving freight to and from the hinterland.

(ii) On the other hand, a transshipment yard might also serve as a hub node in a hub-and-spoke network so that containers or even railcars themselves are exchanged between different trains. This allows a consolidation of trains (i.e., several short trains with loads for multiple destinations are reduced to fewer long trains) so that economies in transportation are generated. Without hub nodes, rail freight is predominantly executed as point-to-point traffic. Because fixed costs for train traffic are high, point-to-point shipment is only profitable if long trains travel over long distances. Studies have estimated the break-even
Receiving tracks

Hump

Classification tracks

Departure tracks

Figure 2 Outline of a Shunting Yard

point for a container being moved either by truck or by a full train over a range of 400 km (Williams and Hoel 1998) or 500 km (van Klink and van den Berg 1998). Thus, hub-and-spoke systems have been identified as a promising starting point to attract rail freight traffic for small freight flows over shorter distances (e.g., Trip and Bontekoning 2002).

To carry out these two tasks, different types of yards have been established over the years, which, in accordance with the chronology of their appearance, can be generally grouped into three terminal generations (Boysen, Fliedner, and Kellner 2010):

- **First generation:** In traditional shunting (or classification) yards, trains arrive at a set of receiving tracks, where railcars are decoupled and pushed over a hump or ramp. The cars are then redirected via a system of track switches and classification tracks to departure tracks, where they are assembled to outbound trains (see Figure 2). Most of these yards are dedicated to function (ii), of rearranging either wagons with nonseparable cargo carriers or railcars whose container assignment is not altered to outbound trains; however, loading operations (function (i)), such as those of bulk cargo, might be carried out. Shunting yards have a long history dating back to the beginnings of rail transport, even though shunting operations have always been very time-consuming, especially when compared with the exchange of containerized loads between trains without altering the railcar composition (see second- and third-generation terminals). It is estimated that in railway systems that employ shunting yards, 10% to 50% of a train’s total transit time is required for shunting (Bontekoning and Priemus 2004). Owing to the growth of containerized transport, shunting yards have lost their former importance, and many have been put out of service in recent decades (see, e.g., Rhodes 2003). Nevertheless, there are still several operational shunting yards in different railway networks; in some regions (in particular in China), some have even been newly constructed, mainly because of the comparatively low investment cost of technical equipment.

- **Second generation:** At today’s conventional rail-road terminals, trains usually keep their railcars and only containers are actually transshipped, typically by means of huge gantry cranes that span multiple parallel railway tracks. Such yards often accommodate additional elements, such as storage areas for the intermediate stacking of containers and adjacent truck lanes for immediate transshipment from trains to trucks and vice versa. Rail–road terminals have become one of the cornerstones of intermodal freight; their main purpose is to serve as an interface between different modes of transportation (function (i)). The German railway network, for example, features 24 rail–road terminals spread over the country (see DUSS 2010). More recently, these yards have also been applied as part of hub-and-spoke networks (function (ii)), for instance between Germany (with hubs in Ludwigshafen, Munich-Riem, and the Port of Nuremberg) and Italy (with hubs in Bologna, Busto Arsizio, Milan, and Verona; see Kombiverkehr 2009).

- **The third generation** of modern rail–rail transshipment is dedicated to a rapid consolidation of trains (function (ii)). The layout of these yards is similar to that of second-generation terminals. However, for further acceleration of container transfers, a fully automated sorting system is employed instead of conventional floor storage. Such a sorter consists of shuttle cars that receive containers close to their initial positions on inbound trains and move them alongside the yard to their target positions. Only then does a gantry crane pick up the containers and transport them to their dedicated outbound trains. Most of these terminals are still in the design phase; however, some of these novel hub yards have already been constructed in the European Union (e.g., Port-Bou; see Martinez et al. 2004) and others are currently under development. For instance, construction of the so-called German Mega Hub in Hannover-Lehrte is expected to finally start in 2014 after a tedious design phase, which has been documented in more detail by Alicke (2002) and Rotter (2004).

This study exclusively focuses on second- and third-generation terminals because first-generation shunting yards are fundamentally different in structure and operation and traditionally dedicated to processing railcars with nonseparable cargo carriers. Therefore, shunting yards seem to be a subject suited to exclusive treatment. They are, moreover, rarely part of modern container-based rail networks, which are expected to be the main driver of future rail freight; shunting yards are therefore of minor relevance in the context of container transshipment.
For an introduction to shunting yard processes, see Gatto et al. (2009) or the valuable classification of Hansmann and Zimmermann (2008). Furthermore, all freight terminals that are not explicitly designed to transship container units as well as innovative terminal concepts that have not yet passed through the purely conceptual phase are excluded from the scope of this survey. The former group includes, for instance, special terminals dedicated to automobile transshipment (see, e.g., Mattfeld and Kopfer 2003; Fischer and Gehring 2005) or company-owned railway sidings (e.g., Lübbecke and Zimmermann 2003), whereas the latter consists of concepts such as automated shunting terminals (see, e.g., Hansen 2004) or moving train techniques (see, e.g., Ballis and Golias 2004). Instead, it is the aim of this study to review scientific approaches that tackle the long- to mid-term decision problems of the design phase with regard to the layout and resource allocation of the terminal and short-term decision problems that are solved as part of the daily operations of conventional rail–road and modern rail–rail terminals. The problems are studied exclusively from the perspective of the terminal operator, and macroeconomic effects are thus not considered.

Owing to the confinement of (isolated) terminal operations, decision problems with regard to the complete railway network consisting of multiple terminals and the interconnecting track system are further excluded. This exclusion comprises location planning (e.g., see Klincewicz 1998; Arnold, Peeters, and Thomas 2004), performance estimation of a railway network with respect to the capacity of nodes and connections (see, e.g., Ballis and Golias 2002, 2004), distributing empty wagons within a network (see, e.g., Nozick and Morlok 1997), and train scheduling (e.g., see Newman and Yano 2000). Moreover, the smallest load unit considered in this survey is the container, and the stowage planning of containers (see, e.g., Geng and Li 2001; Pisinger 2002) is thus not covered. To conclude, Figure 3 schematically defines the scope of this review.

Yard operations have already been discussed in review papers with a wider scope, which cover rail transshipment yards as one part of a broader topic. For instance, surveys on general railway optimization (Assad 1980; Bussieck, Winter, and Zimmermann 1997; Ferreira 1997; Cordeau, Toth, and Vigo 1998; Newman, Nozick, and Yano 2002), intermodal transport (Macharis and Bontekoning 2004; Bontekoning, Macharis, and Trip 2004; Crainic and Kim 2007; Caris, Macharis, and Janssens 2008), and seaport terminal operations (Vis and de Koster 2003; Steenken, Voß, and Stahlbock 2004; Stahlbock and Voß 2008) also briefly elaborate on rail yards. However, the extended scope of these surveys prevented an in-depth discussion of decision problems, existing optimization approaches, and future research challenges of rail yard operations.

3. Rail–Road Transshipment Yards

3.1. Yard Layout, Transshipment Process, and Decision Problems

Rail–road terminals mainly serve as interface nodes in intermodal transport, where gantry cranes transship containers between trains and trucks and vice versa. A rail–road terminal is schematically depicted in Figure 4.

Freight trains are parked on parallel transshipment tracks of the terminal. Typically, a terminal segment consists of between two and four parallel tracks, and a maximum gantry span of six tracks is possible (e.g., Steenken, Voß, and Stahlbock 2004). Because freight trains in Europe have a typical length of 600–750 m (Ballis and Golias 2002), the track area accessible by cranes for container processing is typically of about the same length. Larger terminals, such as Köln–Eifeltor and Hamburg–Billwerder in Germany, consist of multiple parallel terminal segments. Trucks...
to be determined because these interdependent layout factors heavily influence yard performance and are not easily reversible. Hence the job of the yard planner is to carefully trade off investment cost of a specific layout against estimated operational performance. Operations research methods are especially suited to quantifying the latter part of this tradeoff, and an accurate estimation of the performance of a specific yard layout is thus an essential task for supporting decisions of the design phase. Note that, although not in the scope of this review, the results for yard performance need to be further evaluated with respect to the yard’s network integration because an expansion of capacity in a non-bottleneck yard does not necessarily increase overall network performance. Existing approaches for performance estimation of rail–road terminals are reviewed in §3.2.

For a given yard layout, the operational process of container transshipment at a rail–road terminal is now described in more detail. On the basis of a given timetable of trains, container moves need to be processed periodically subject to arrival and departure times of trains. Typically, all trains arrive in the morning, are processed over the course of a day, and leave the terminal in the late evening. Owing to the general right of way of passenger trains in many European countries and, for example, Australia, freight trains are often bound to travel during the night exclusively. Once a train arrives at the transshipment yard (after a potential interim stay on a holding track of the yard), the train is first assigned to a vertical and horizontal parking position of the yard. The vertical parking position relates to the actual track on which the train enters the yard, but the notion of a horizontal parking position requires some additional explanation. Typically, the yard area is subdivided horizontally into slots of equal size measured in units of the length of a
standard railcar (or any other unit). The resulting grid is hence used to identify the coordinates of any given container in the yard, and the horizontal parking position of a train refers to the slot in which the traction vehicle is positioned. For rail–road terminals, the problem of assigning parking positions to trains has not been studied in detail thus far. This is because its impact on yard performance is usually considered to be rather small. Because trucks can be parked directly next to the respective container of the train, cranes need to move only vertically for the most part, and the time required compared with the time-consuming pick and drop operations of cranes is often negligible. The horizontal parking positions do determine the accessibility of individual containers with respect to the different gantry cranes of the yard. Owing to the immense fixed cost, however, freight transport by train is often only profitable if full trains (i.e., trains used to capacity) are moved. Therefore, train and yard lengths are about the same and the degree of freedom for varying horizontal parking positions is often not significant enough. In practice, parking positions are therefore typically assigned according to a simple first-come-first-served policy (Kozan 1997).

As soon as a train is parked, the unloading of all inbound containers can commence. To avoid double-handling, a container is preferably transshipped directly from the train to its dedicated truck. In the following, this form of container transfer is referred to as a direct move. Clearly, a direct move requires the simultaneous presence of a respective train and truck in the yard. The target truck is then called up from the holding area and is assigned a free parking position in the parking lane next to the respective railcar. If the target truck has not yet arrived and is therefore not directly available, the container is moved to the intermediate storage yard. This kind of double-handling is referred to as a split move (see Boysen, Jaehn, and Pesch 2011), where a storage location close to the respective railcar is sought so that crane operating times are reduced. However, because containers are usually stacked on top of each other, split moves to and from storage are subject to additional restrictions, such as those relating to weight, stability, and estimated departure times of containers, to avoid subsequent reshuffling. Analogous to stacking logistics in seaports (see Steenken, Voß, and Stahlbock 2004), three interrelated decision problems are associated with split moves. First, a suitable storage position is to be identified that minimizes the risk of container blockages. Second, stored containers might need to be pre-marshaled on the basis of updated information with regard to arrival times of trucks, which is especially reasonable during idle times of cranes. Finally, containers need to be efficiently retrieved from the storage area as soon as the respective truck arrives, which might in turn require additional handling of any blocking containers. Typically, the frequencies of direct and split moves vary over time (see Bose 1983; Ballis and Golias 2002). Shortly after a train’s arrival, direct moves in particular are processed, i.e., moves from a train to trucks that are already waiting. In a second phase, wagon-to-storage moves (i.e., from a train to the storage area) predominate, and in the final phase, storage-to-truck moves are processed.

At most terminals, outbound operations are executed only after inbound operations are completed. However, intermixed operations of inbound and outbound containers are certainly possible. The processing of outbound operations is carried out analogously to that of inbound containers. Whenever trucks have deployed containers prior to the train’s arrival, a split move occurs, and containers from the storage yard are loaded onto the train; deliveries that arrive during the loading process of the target train can be processed as direct moves. Therefore, prior to a train’s arrival, truck-to-storage moves prevail, which are then superseded by direct moves processed after a train’s arrival. In the final phase, mainly storage-to-wagon moves are executed. In some yard settings, outbound containers are moved by special container trailers rather than directly by customers’ trucks. The containers are then carried to a separate storage area where customer trucks pick up and deliver containers. Clearly, this concept avoids double-handling of containers in the yard at the price of an additional transshipment in the storage area and higher investment cost of the many different container trailers required for the wide range of possible containers. Nonetheless, this practice is often applied in North American yards (see Ferreira and Sigut 1993; Kozan 1997).

During the loading operations of an outbound train, there exist some degrees of freedom with regard to the exact position of each container on a train. Therefore, a load plan that determines the loading pattern of containers on wagons is required. A typical terminal faces a high variety of container types and multiple different wagons, which can vary in length between 40 and 104 feet (see Bruns and Knust 2012). Given a specific setting of outbound containers and railcars, the loading problem has to consider several hard constraints, i.e., wagon length, separation of dangerous goods, weight restrictions, and train height. Furthermore, the quality of a load plan can be determined by different conflicting objectives, such as the utilization of trains, setup time and/or cost for changing a railcar’s pin configuration, or processing times for moving a container from its current position to the respective wagon. From the overall network perspective, load planning is also heavily interdependent on the distribution of wagon types across yards.
This last aspect has been investigated by Powell and Carvalho (1998), for instance.

Once the load plan is determined, the set of container moves is finally fixed and the planning can focus on determining transshipment schedules for each crane. Because gantry cranes principally work in parallel, it seems especially desirable to split the overall workload evenly among cranes so that train processing is accelerated. However, in most yard settings, gantry cranes share a dedicated track for their horizontal movement along the yard, which prevents them from passing by one another. Clearly, these noncrossing constraints of gantry cranes need to be integrated into a suitable decision support tool as hard modeling constraints. Two distinct policies have been developed to avoid such crane interferences (see Boysen and Fliedner 2010). On the one hand, the assignment of container moves to cranes can be static, which means that each crane receives a disjoint area of operations, where all container moves falling into the area are exclusively processed by the respective crane. On the other hand, containers can be assigned dynamically on the basis of the actual positions of cranes and the set of moves that needs to be executed. Clearly, the latter policy offers more degrees of freedom for crane scheduling. However, the coordination of cranes becomes more complex and requires real-time crane scheduling procedures to rule out any crane interference. In real-world yards, a static crane split with equally sized yard areas is the most widespread choice (see Boysen and Fliedner 2010). The sequence of moves falling into a crane’s area is typically not optimized with the help of a sophisticated decision support tool. Instead, the crane operator simply chooses among those moves currently being displayed on the monitor of the steeple cab on the basis of some nearest-neighbor decision rule (Boysen, Fliedner, and Kellner 2010).

To summarize, the operational process needs to address the following essential decision problems.

(I.i) Decide on storage positions of containers handled by split moves.

(I.ii) Assign each truck a parking position.

(I.iii) Determine the positions of outbound containers on trains.

(I.iv) Assign container moves to cranes.

(I.v) Determine the sequence of container moves per crane.

In the following sections, the literature on layout planning (§3.2) and operational container processing (§3.3) is summarized.

### 3.2. Literature on Layout Planning

Existing literature on layout planning exclusively consists of simulation studies. These simulations are applied to anticipate yard performance for different terminal layouts.

A discrete-event simulation study including both a macro (network) and micro (terminal) perspective was carried out by Rizzoli, Fornara, and Gambardella (2002). In that study, different technologies and operational policies were compared with regard to their effect on terminal and network performance. A similar simulation model was described by Kondratowicz (1990). Lee et al. (2006) presented a simulation study that was designed to support decisions on the number and locations of rail terminals at a Korean container port. Basic mathematical expressions were derived to calculate the number of tracks and cranes required for a specific number and locations of rail terminals. The authors simulated different train and truck arrival patterns as well as container move settings by applying a simple crane scheduling rule; i.e., every crane processes containers successively while continuously traveling in a specified direction as long as a receiving truck is available (if not, the crane changes direction for the next container). The study was carried out for different numbers and locations of terminals.

Ferreira and Sigut (1993, 1995) compared the resulting performance of container handling between a conventional rail–road terminal and a roadrailer terminal. A roadrailer is a special trailer, which is provided with a detachable bogie or a single rail axle so that they are capable of being hauled on road and rail without requiring a wagon. The application of alternative wagon types was compared as part of a simulation study for an Australian terminal. The results indicated the more efficient handling of containers compared with roadrailers.

Another simulation tool dedicated to model a single terminal was introduced by Benna and Gronalt (2008). Terminal layout, arrival patterns of trains and trucks, and container settings were specified as part of the input data. Simple priority rule-based approaches were applied to determine crane schedules and intermediate storage positions of containers. As quality measures, the study evaluated lifting performance, system capacity, and service level. A similar tool was described by Gronalt, Benna, and Posset (2007).

The results of a large European Union research project, which aimed to increase rail terminal performance, were presented by Ballis and Golias (2002, 2004) and Abacoummik and Ballis (2004). The authors developed an extensive expert system consisting of a macro model that covers a complete railway network and a micro model simulating train processing in a single yard. A general overview of the macro and micro model was provided by Ballis and Golias (2004). They tested the micro model for 17 different terminal layouts with varying numbers of tracks and cranes as well as lifting technologies. Ideal terminal layouts for a given transshipment volume were determined by calculating the total cost per container.
The macro model was employed to anticipate the market share of rail freight over a longer planning horizon for a specific network structure and terminal configuration. They included a case study of the railway corridor from the large North Sea harbors to Switzerland. A more detailed description of the micro model was presented by Ballis and Goliás (2002) and Abacoumkin and Ballis (2004).

Kozan (2006) used a simulation model for a terminal, in which gantry cranes can be supported by additional lifting equipment (e.g., reach stackers and forklifts). For joint loading and unloading operations over multiple days, the author compared different crane settings to identify a suitable crane configuration that provides a reasonable tradeoff between investment cost and operational performance. Sequencing of container moves for different arrival patterns of trucks and trains was guided by simple first-come-first-served policies. By means of simulation, Vis (2006) compared the use of manned straddle carriers with that of automated stacking cranes. The total travel time required to handle all container moves was applied as a performance measure to determine the yard layout for the landside of a seaport terminal.

Although mainly dedicated to seaport container operations, a helpful paper for generating representative simulation scenarios was provided by Hartmann (2004). The paper features a data generator for deriving diverse transshipment scenarios. It can be directly applied to simulate rail-road terminals by generating arrival patterns of trucks and the different container types to be (un)loaded.

### 3.3. Literature on Operational Planning

Thus far, none of the literature has explicitly treated the determination of storage positions in a rail-road yard (problem (I.i)). However, this decision problem is closely related to problems arising in the stacking logistics of seaports. In this context, the problem has attracted much research, for instance by de Castillo and Daganzo (1993); Kim (1997); and Kim, Park, and Ryu (2000). The subproblem of re-sorting containers during the idle time of cranes (so-called pre-marshaling) has been investigated, e.g., by Lee and Hsu (2007), Choe et al. (2011), and Lee and Chao (2009). Finally, Kim and Hong (2006) and Caserta, Voß, and Sniedovich (2009), for instance, provided solution procedures for determining a suitable predetermined sequence of crane moves to remove containers from intermediate storage. In addition to these static problem settings, online stacking rules were investigated by Dekker, Voogd, and van Asperen (2006) and Borgman, van Asperen, and Dekker (2010). A more detailed review of these approaches was given by Steenken, Voß, and Stahlbock (2004) and Stahlbock and Voß (2008). However, the extent to which these approaches are directly applicable to rail-road terminals remains to be studied.

A first basic version of the train loading problem (I.iii), where load patterns of containers are determined, was presented by Feo and González-Velarde (1995). Given a predetermined matrix defining which container can be assigned to any railcar on the basis of the pin configuration, the approach seeks to minimize the number of wagons per train. However, the model and solution approaches are restricted to at most two containers per wagon. For the solution of the basic train loading problem, a simple branch-and-bound approach based on LP-relaxation and a heuristic GRASP (see Feo and Resende 1995) procedure have been introduced, where initial solutions are locally improved by a two-opt search. The procedures are shown to be efficient in the case of real-world data for a North American terminal. Corry and Kozan (2006) optimized load planning with respect to handling times and the weight distribution within a train. The authors considered only one type of container and leave out weight restrictions for wagons. Furthermore, they assumed that each container can be loaded onto any wagon. The problem was formulated as an integer linear program and solved with an off-the-shelf solver. In a subsequent paper, Corry and Kozan (2008) aimed to derive a loading plan considering multiple objectives. On one hand, the train length and on the other hand the total handling time of containers is minimized. Multiple container types were modeled. Load pattern restrictions were considered for the container-to-wagon assignment, but neither weight restrictions with regard to the maximum load per wagon nor that with regard to the whole train were integrated. The model was formulated as an integer linear program, and solutions for real-world problem instances were generated by a local search procedure.

Recently, Bruns and Knust (2012) investigated another version of the loading problem (I.iii) of trains that are subject to weight and length restrictions of wagons. In the objective function, three weighted objectives are considered: maximizing the utilization of trains, minimizing the setup cost for changing the existing pin configuration, and minimizing the transportation cost from storage position to railcar. Two different mixed-integer programs were introduced for this problem setting and were shown to be solvable even for real-world instances.

An additional aspect of the train loading problem (I.iii) was first investigated by Lai and Barkan (2005). Intermodal trains often contain larger segments of empty wagons (e.g., segments consisting of flatcars without a container), which lead to aerodynamic characteristics that are much worse than those...
of full trains with close spacing, e.g., the case for hopper cars. Note that unloaded wagons are transported whenever diverging wagon demands and supplies at the rail yards need to be balanced within a rail network (see, e.g., Nozick and Morlok 1997). Therefore, considerable savings in fuel cost can be achieved if train planning considers the additional objective of generating long chains of loaded railcars alternating with long chains of unloaded railcars. Lai and Barkan (2005) quantified the aerodynamic and energy penalties of specific load and car combinations under idealized conditions by assuming that containers can be assigned to wagons without restrictions. Lai, Barkan, and Önal (2007) described a wayside machine vision system that automatically monitors the gap lengths between intermodal loads on passing trains, which allows automatic evaluation of the aerodynamic efficiency of loading patterns. In a subsequent paper, Lai, Barkan, and Önal (2008a) presented a mixed-integer model for determining fuel-efficient train loads considering weight and length restrictions of wagons, which was solved with an off-the-shelf solver. It turned out that the estimated savings calculated with real-world data amount to a remarkable potential of annual fuel savings of US$28 million. The joint optimization of multiple trains’ load plans (with identical destination) and uncertain information on future trains and incoming loads were incorporated into the aforementioned mixed-integer model in another paper by Lai, Ouyang, and Barkan (2008b). They iteratively solved the model on a rolling horizon scheme, where exponentially decreasing weights were assigned to the objective functions with regard to the fuel efficiency of future trains.

Kozan (1997) provided a simple heuristic decision rule for determining the crane split (I.iv) and a simple dispatching rule for the assignment of trains to railway tracks. He employed simple analytical measures to anticipate the processing times of current trains and thus identify the track that would enable the earliest expected departure of a current train. These simple heuristic rules were then applied in a simulation study, where the resulting throughput times of containers for different train arrival and loading patterns were compared for different yard layouts. Boysen and Fliedner (2010) also investigated decision problems (I.iv) and introduced a polynomial dynamic programming approach to determine static and disjoint crane areas such that the workload is evenly shared among cranes. In a simulation of real-world yard operations, they showed that their approximate surrogate objective for determining the crane split is strongly positively correlated with actual processing times, and simple real-world policies are clearly outperformed.

Souffriau et al. (2009) proposed a holistic approach, which jointly determines load plans (I.iii) and crane schedules (I.iv) and (I.v). They used a decomposition approach to determine follow-up destinations of trains, such that the number of containing container moves is minimized. This problem was solved as a linear assignment problem. The load plan was hence determined by minimizing the transportation cost of container moves in a mathematical model with an off-the-shelf solver. Only three different container types as well as length restrictions for the wagons were considered. Finally, the crane schedule, which distributes container moves among cranes and sequences moves per crane, was modeled as a sequential ordering problem and solved by a variable neighborhood search. Another holistic approach for train processing at the landside of a seaport was provided by Froyland et al. (2008). They treated an intermodal terminal in Australia, where five successive gantry cranes transship containers between trains (two tracks), trucks (60 slots), straddle cranes (serving ships), and an intermediate storage area with a maximum capacity of 2,100 twenty-foot equivalent units. They jointly investigated decision problems (I.i), (I.ii), and (I.v) and determined container positions in intermediate storage, parking positions of trucks, and crane schedules, respectively. The problem was decomposed into three stages, where problems (I.i) and (I.ii) were solved by mixed-integer programming and cranes (I.v) were scheduled on the basis of simple priority rules. Finally, Montemanni et al. (2009) modeled the sequencing of a given set of container moves per crane (I.v) as a sequential ordering problem and provided local search and ant colony optimization (Dorigo and Stützle 2004) as solution procedures.

Moreover, dynamic crane scheduling as defined by decision problems (I.iv) and (I.v) bears some similarities with quay crane scheduling at seaports, where a given number of huge quay cranes is employed to (un)load container vessels at their berths. As at rail terminals, quay cranes share a track and may not interfere with nor cross each other during container operations. Typically, within quay crane scheduling, a vessel is separated into holds (or bays), which are exclusively served by a dedicated crane, and noncrossing constraints need to be considered whenever cranes change holds. If, analogously, a rail yard is separated into small horizontal areas, such as slots of container length comprising all containers of the parallel tracks within the respective slot, then the solution procedures developed for quay crane scheduling could in principle be applied to solve crane scheduling in a rail yard. The first optimization approaches for quay crane scheduling stem from Daganzo (1989) and Peterkofsky and Daganzo (1990). These studies, however, did not consider noninterference constraints of cranes. Kim and Park (2004) considered
noncrossing constraints and presented a model formulation along with exact and heuristic solution procedures. Alternative solution methods were presented by Lee, Hui, and Miao (2008b), who also provided an NP-hardness proof. Related contributions stem from Lim et al. (2004); Zhu and Lim (2006); Lim, Rodrigues, and Xu (2007); and Lee, Hui, and Miao (2008a). A comprehensive review on quay crane scheduling was provided by Bierwirth and Meisel (2010). Of particular interest with regard to crane scheduling in rail–road yards are the results of Lim, Rodrigues, and Xu (2007), who showed that under given noncrossing constraints—and some additional simplifying assumptions—optimal crane schedules are unidirectional, in the sense that each crane can move from left to right while processing container moves without ever changing direction. This finding is especially relevant to dynamic schedules because it reduces the real-time effort for collision detection to a minimum and thus might make static crane bounds expendable. However, this property only holds whenever containers can be processed in an arbitrary sequence, an assumption that many other quay crane scheduling approaches equally make. This does not hold for the majority of rail–road terminals because dynamic arrival times of trucks need to be considered in the transshipment plans, and therefore the aforementioned approaches cannot be directly applied and need to be adequately extended.

### 3.4. Future Research Challenges

Although there have been plenty of studies on conventional rail–road terminals, there are still many open questions for future research.

With regard to layout planning, it can be stated that all existing simulation studies apply comparatively simple priority rules to solve subordinate operational decision problems when estimating yard performance. Notice that this might introduce the bias that yard layouts are systematically underrated with regard to performance because sophisticated scheduling procedures would allow for more efficient resource utilization than anticipated by simple rules of thumb. Consequently, existing studies might recommend more efficient layouts accompanied by higher investment cost than actually required. It follows that simulation studies incorporating sophisticated scheduling procedures could provide a valuable contribution for promoting the success of intermodal transport.

Furthermore, several potential yard layouts have not yet been evaluated. For instance, crossover cranes operating on different tracks could be employed, as in intermediate block storages at modern seaports, where even triple crossover gantry cranes are currently in operation (Dorndorf and Schneider 2010). Here, a pair of twin cranes running on the same tracks is supported by a large crossover crane on its own rails. Evaluating the performance of these alternative crane layouts in comparison to existing layout configurations would be valuable decision support for future terminal projects.

An important decision, which at the same time heavily influences the yard layout, is the question of whether the European policy of allowing customer trucks to directly enter the transshipment area or the North American policy of applying container trailers that transfer containers in the holding area is actually the better choice. Although the latter policy reduces the load of gantry cranes by avoiding additional split moves and allows for better (deterministic) planning of crane schedules, it extends delivery times for customers. A detailed comparison of both policies and their impact on yard layout is a challenging subject for future research.

Furthermore, additional research in the operational area is required with regard to each of the decision problems defined in §3.1:

1. (i) Existing research on identifying appropriate stacking positions of containers in intermediate storage is mainly dedicated to operations in seaports. Although in principle the developed procedures can be used in both fields of application, the dimensions of the storage areas in rail terminals are much smaller. The mean stacking height, for instance, in many rail yards is merely between 1 and 1.5 containers (Ballis and Golias 2002). Therefore, it would be a valuable contribution to test whether applying the sophisticated procedures developed for seaports in fact accelerates container processing sufficiently to justify the investment cost of the required information system.

2. (ii) The assignment of parking positions to trucks is a widely unexplored field of research. Only Froyland et al. (2008) have integrated this problem into a holistic planning approach. Clearly, this problem is of minor importance if only a few trucks enter the yard simultaneously because each truck can be parked directly next to its respective container location. However, directly after a train’s arrival, typically, many trucks are already waiting for container processing. These trucks compete for scarce parking positions close to their respective containers. In such a setting, the (dynamic) assignment of parking lots to trucks such that congestion in the yard is reduced requires sophisticated decision support and is thus a challenging problem for future research.

3. (iii) The train loading problem has attracted the largest number of research contributions thus far. However, a versatile model integrating all real-world weight and loading constraints of wagons (as, e.g., defined by Bruns and Knust 2012) with aerodynamic aspects (Lai, Barkan, and Önal 2008a) along
with suitable solution procedures is still missing. Furthermore, the degrees of freedom for train loading are diminished by the diversity of trailers and wagons to be processed. A further standardization of containers promises reduced effort in changing pin configurations of railcars and thus load plans that are more efficient. Therefore, quantifying such standardization effects could help further encourage standardization agreements for rail transport.

(I.v) With regard to the assignment of container moves to cranes, two basic policies are distinguished in this survey: static assignment, where each crane operates in a distinct yard area, and dynamic assignment, where the obstruction of cranes is to be avoided in real time. Clearly, the dynamic approach promises crane schedules that are more efficient but comes at the price of a suitable information system. A thorough quantitative comparison of both policies would offer valuable decision support for real-world yards.

(I.v) Once all container moves are specified and assigned to cranes, the sequence of container moves per crane resembles a sequential ordering problem (Montemanni et al. 2009). However, in reality, truck arrivals in particular are subject to uncertainties, and sequential ordering thus needs to be executed in an online environment. It remains an open question as to how to integrate sequential ordering in a rolling planning horizon. Moreover, it should be tested whether such an approach is indeed able to considerably outperform the common policy of experienced crane operators making decentralized scheduling decisions.

Until now, freight traffic has only been profitable if full trains are moved over comparatively long distances (see §2), and therefore, the degrees of freedom for parking trains in the transshipment area are limited. However, with the realization of hub-and-spoke systems, smaller trains might also become profitable, which in turn affects the operational planning environment. For instance, horizontal parking positions might be used to evenly balance the workload among cranes if the lengths of trains vary sufficiently, which gives rise to a parking problem (similar to the parking problem of rail–rail terminals; see §4.1). A careful examination of trends in intermodal transport might yield interesting insights with regard to upcoming challenges of yard planning.

Finally, in addition to an isolated investigation of the above decision problems, holistic approaches seem an especially promising field for future research. Currently, there are limited proposals for hierarchical procedures, e.g., those presented by Froyland et al. (2008) and Souffriau et al. (2009). The high degree of interdependence and relatedness of the discussed decision problems makes determining the right sequence of decisions and hierarchical integration of all or at least some decisions a challenging task.

4. Rail–Rail Transshipment Yards

4.1. Yard Layout, Transshipment Process, and Decision Problems

Modern rail–rail terminals mainly serve as hub nodes in a hub-and-spoke rail network. Containers are transshipped among trains without the need to exchange railcars, and inbound trains are thus consolidated to a (reduced) set of full outbound trains. In some implementations a terminal might additionally process rail–road operations, but these are usually of minor relevance in comparison to the hub operations. A pure rail–rail terminal is schematically depicted in Figure 6.

The main technological innovation that rail–rail terminals introduce in comparison with conventional rail–road terminals is that the simple floor storage area is replaced by a fully (or partially) automated sorting system. The sorter consists of moving and buffer lanes, where automated guided vehicles (Bostel and Dejax 1998) or shuttle cars (Alicke 2002) receive a container from a gantry crane close to the container’s initial position on the train and move it alongside the yard toward its dedicated container position on the outbound train. A fully automated system can, for instance, employ rail-mounted shuttle cars, which use a rotation mechanism for changing tracks (Franke 2002) and are propelled by contact-free linear synchronous motors with an electronic position detection system that is able to direct shuttle cars with an accuracy of ±3 mm (Bauer 1998). Simulation studies for the Megahub in Hannover-Lehrte indicate that such a sorting system increases container processing by up to 45 container moves per hour and crane (Rotter 2004).

The impact of the high-performance sorting system also explains the critical importance of the system during the design phase. The choice of an appropriate drive technology and the dimensioning of the system with regard to storage space and shuttle cars are two of the most important decisions in this context, in addition to the choice of general yard layout determined by the number of tracks and cranes.

Figure 6 Schematic Representation of a Rail–Rail Transshipment Yard
In §4.2, operations research tools that support the design phase of a rail–rail terminal are reviewed.

The operational process of container consolidation at a rail–rail terminal is similar in principle to that at a rail–road terminal. Because the hub visit constitutes a time-consuming additional step in the distribution process, some organizational changes are necessary to speed up transshipment and avoid a tedious stay for an extra day. Generally, trains are required to exchange containers within a few hours so that the consolidation process is rapid and trains can depart to their next destinations the same day (or night). Therefore, a rail–rail terminal is operated in distinct so-called pulses (Bostel and Dejax 1998) or bundles (Aliche 2002; Rotter 2004) of trains. This means that all tracks are occupied by trains, which are simultaneously served and jointly leave the system only after all container moves of the respective bundle are processed. Whenever the total number of incoming trains exceeds the number of tracks, a first decision problem assigns each train to a bundle. This decision is subject to release dates and departure times of trains as given by the train schedule and might consider several objectives that conflict in part, such as minimizing the number of containers dedicated to a train of an earlier bundle or maximizing the number of direct moves among trains of the same bundle. The former objective reduces the number of containers that do not arrive at their outbound trains (and thus need to be held until the arrival of the next train that serves the respective destination), whereas the latter objective accelerates train processing by reducing the amount of double-handling.

Once the pulses are determined, vertical and horizontal parking positions need to be assigned to each train of a bundle. Trains that exchange a large number of containers should be assigned to neighboring tracks to reduce the total distance that a loaded crane moves. Because hub terminals also tend to be visited by shorter trains, appropriate horizontal parking positions of trains can positively affect yard performance, for instance, by evenly spreading container moves among cranes or by moving start and target positions of a container to the same area of operation so that a single crane can process the job instead of relying on the sorting system.

The problem of determining an appropriate load pattern of containers on trains is similar to that arising at rail–road terminals. A load plan, which minimizes the overall distances that containers move, seems to have a somewhat smaller effect on the yard performance if the sorting system is not a bottleneck, in which case containers can be prepositioned next to their intended target positions. However, whenever inbound and outbound loads are exchanged simultaneously and the parking positions of trains and operating areas of cranes are fixed, the load plan alone determines the target position of the outbound container, and therefore the number of split moves can be reduced by moving the target position closer to the starting position.

The assignment of container moves to cranes can be executed under a static or dynamic policy of distinct or variable crane areas, respectively (see §3.1), and should generally aim to avoid split moves and to share the workload evenly among cranes.

Finally, the schedule of container moves per crane is to be determined. The resulting problem constitutes an extension to crane scheduling at rail–road terminals and is similar in structure to a sequential ordering problem. One important aspect is that crane moves are asymmetric in distance; i.e., executing move A before B results in a distance different from that of the reverse order. This is because any two loaded moves need to be connected by an unloaded move of the crane and the distance between the end position of move A and the start position of move B, which will typically differ from the distance between the end position of move B and the start position of move A. Furthermore, container moves are subject to precedence constraints whenever a container is blocking another container’s target position. In addition, split moves via the sorting system need to be considered. This leads to heavily interdependent crane schedules because the release date of a container transported by the sorter depends on when another crane has fed this container into the sorting system. This problem becomes even more complex if the sorter is a bottleneck and the availability of shuttle cars over time needs to be considered.

The basic decision problems of container processing in rail–rail yards can be summarized as follows.

(I.I) Schedule the service slots of trains by assigning them to bundles.

(I.II) Assign each train a parking position.

(I.III) Determine the positions of containers on trains.

(I.IV) Assign container moves to cranes.

(I.V) Schedule the shuttle cars in the sorter.

(I.VI) Determine the sequence of container moves per crane.

Existing literature on decision support with regard to these decision problems is reviewed in §4.3.

4.2. Literature on Layout Planning

There have been very few studies investigating suitable layouts of modern rail–rail terminals.

Meyer (1999) investigated the layout problem of a rail–rail terminal that successively processes bundles of six trains. In addition, the terminal is assumed to handle a limited volume of rail–road container exchanges. An animated computer simulation based on Petri nets (see Peterson 1981) was developed to
They presented a genetic algorithm for solving the problem of trains operating in disjoint crane areas. Split moves via the sorting system were considered to minimize the makespan of train processing. Furthermore, an assignment of trains to bundles and the number of containers that could not be transported were presented. The study of Boysen, Jaehn, and Pesch (2011) considered an additional objective function that minimized costs while accelerating train processing using synchronized crane operations. Suggesting a portfolio matrix, Wiegmans et al. (2007) concluded that a modern rail–rail yard is the appropriate choice for a hub node whenever fast operations take priority, whereas shunting yards and rail–road terminals are favorable for cost-efficient operations. The results presented were mainly based on the thesis of Bontekoning (2006).

4.3. Literature on Operational Planning
The assignment of trains to bundles (II.i) was first investigated by Boysen, Jaehn, and Pesch (2011), who formulated a basic transshipment yard scheduling problem that minimizes a weighted objective function considering split moves between those trains that are assigned to different bundles and the number of revisits by trains that could not receive all containers during their first visit to the yard. The problem was shown to be NP-hard in the strong sense, and different heuristic and exact solution procedures were presented. The study of Boysen, Jaehn, and Pesch (2012) extended this research. The transshipment yard scheduling problem was modified to consider an additional objective function that minimized the number of containers that could not be transported to their target trains in time. Additional complexity results were presented, and more efficient exact branch-and-bound procedure and a very efficient heuristic ejection chain approach were provided.

Kellner, Boysen, and Fliedner (2012) presented a solution approach to solve decision problem (II.ii), i.e., the parking problem of trains. In their approach, the assignment of a given bundle of trains to tracks (vertical position) and horizontal parking positions along the spread of the yard aimed to minimize the makespan of train processing. Furthermore, split moves via the sorting system were considered between trains operating in disjoint crane areas. They presented a genetic algorithm for solving the resulting problem and tested their approach in a simulation study. An even simpler approach for determining parking positions (II.ii) was presented by Alicke and Arnold (1998). They merely modeled the track assignment of trains (vertical position) in a very basic fashion without, for example, considering horizontal parking positions, sorter operations, and multiple cranes. Instead, they developed a simple priority value weighting the number of container moves with their total horizontal distance to approximate the resulting workload between two trains. These weights were then applied in a quadratic assignment problem to determine the track assignment.

Bostel and Depjay (1998) treated problem (II.iii) and aimed to jointly determine load plans for inbound and outbound containers. Start and target positions of a container move on inbound and outbound trains were to be brought as close together as possible so that the travel distances of cranes and the resulting costs of train processing were minimized. However, with regard to load planning, the underlying assumptions were rather limiting. It was assumed that each container can be stored on any wagon, and weight restrictions were not considered. Four different models for this problem were derived by additionally considering container transfer by shuttle cars and storage constraints in the sorting system. Each model requires different solution approaches, and therefore, several procedures based on the linear assignment problem, the minimum flow problem, and different start and improvement heuristics were developed. A computational study using real-world data for a French railway company showed a huge potential for accelerating train processing using simultaneously optimized load plans.

Boysen, Fliedner, and Kellner (2010) assigned static and disjoint crane areas (II.iv) to a bundle of trains with given parking positions to minimize the makespan of train processing. The sorting system was assumed to be activated whenever start and target positions of a container move fell into the different crane areas. They presented a polynomial dynamic programming procedure for solving the resulting problem and tested the solutions against typical real-world policies in a simulation of yard operations.

Alicke (2002) jointly treated decision problems (II.v), (II.v), and (II.vi). A given set of crane moves is assigned to cranes with overlapping areas of operation, which are blocked whenever a crane enters an area. Whenever a start or target position falls in an overlapping area, the procedure dynamically decides which of two neighboring cranes processes the move. This decision also influences whether or not a container move uses the sorting system in a split move. The model takes the movement speed and availability of shuttle cars into account. The overall problem was modeled as a constraint satisfaction problem.
and tested on data sets of the German MegaHub in Hannover-Lehrte. Different heuristic rules for fixing variables of the constraint satisfaction problem were compared.

At the border of two countries and railway systems, rail–rail terminals are also used to bridge different track gauges. This requires a special yard setting where complete train loads are transshipped by cranes onto a train with the gauge width of the destination railway system. Martinez et al. (2004) investigated two simple rules for crane scheduling (II.vi) at a terminal on the border between France and Spain. Both rules were compared by means of a simulation study. The same terminal was investigated by Gonzalez et al. (2008), who provided a mixed integer model to jointly determine the load plan of outbound trains (II.iii) and crane schedules (II.vi). Their objective was to minimize crane distances while observing the weight and length restrictions of wagons. The model was solved with an off-the-shelf solver that was shown to be suitable for real-world instances of small size.

4.4. Future Research Challenges

Because consolidating containers at a rail–rail terminal is still an emerging technology for railway systems (Bontekoning, Macharis, and Trip 2004), there remain many open fields for future research.

Similar to the situation for rail–road terminals, the few existing simulation studies merely apply very simple myopic decision rules when evaluating the performance of terminal layouts. However, the threat of underestimating the performance of rail–rail terminals seems even more imminent because many terminals are currently in the conceptual evaluation phase. If poor scheduling rules lead to a poor forecast of yard performance, it is to be expected that some projects (including new hub-and-spoke systems) will not be constructed, which in turn might further deteriorate the market share of rail freight traffic.

The following open research challenges in operational planning are identified:

(II.i) Assigning trains to bundles is currently performed in a deterministic setting. However, because train timetables are bound to change, transshipment yard scheduling might be more appropriately modeled on rolling horizons to account for unplanned train cancellations or arrivals.

(II.ii) Currently, there are only two approaches for determining parking positions of trains (Alicke and Arnold 1998; Kellner, Boysen, and Fliedner 2012), both of which are heuristic in nature. Efficient exact-solution procedures might yield valuable insights into the solution structure of this specific assignment problem.

(II.iii) Bostel and Dejax (1998) introduced simultaneous load planning for inbound and outbound trains to minimize the distances for crane moves. However, they assumed that only a single container can be loaded onto a wagon and further that any container can be assigned to any car. Real-world weight constraints and length restrictions were not considered, and neither were split moves resulting from container moves across different crane areas. Therefore, future research should seek to integrate more of the diverse real-world constraints of train planning, which have already been widely explored for conventional rail–road terminals (see §3.1).

(II.iv) Analogously to rail–road terminals, studies on the performance of static versus dynamic crane areas have not yet been undertaken (see §3.1).

(II.v) Existing research mostly assumes that the sorting system is not a bottleneck. However, whenever the availability of shuttle cars is not guaranteed, crane scheduling needs to be complemented by sophisticated scheduling of shuttle cars. An interesting yet unexplored problem in this context is to further the real-time control of sorting vehicles to generate deadlock-free travel routes.

(II.vi) Analogously to rail–road terminals, the scheduling of crane moves can be modeled as a sequential ordering problem as soon as all crane moves have been specified and assigned to cranes. However, the problem becomes somewhat more complicated in rail–rail yards because split moves processed by the sorting system need to be considered. Some containers become available only after they have been fed into the sorter by another crane, and crane schedules thus cannot be decomposed. Suitable crane scheduling procedures have not yet been developed.

A further field for future research is to integrate the above decision problems into a holistic procedure. For instance, parking positions of trains and load plans of inbound and outbound trains both influence crane moves and determine whether or not a container move needs to use the sorting system. Integrating both problems in a simultaneous planning procedure or developing a hierarchical framework that incorporates the two decision problems would be a valuable contribution.

5. Conclusion

This paper surveyed layout planning and operational decision problems arising in rail freight yards. The core decision problems of rail–road terminals and modern rail–rail transshipment yards were characterized and existing research was reviewed. To provide a more concise overview of existing research, Table 1 lists the literature of the field along with the problem treated and the methodology employed. By contrasting the structure of decision problems with the scope of existing research, several avenues for future research are identified.
Table 1  Summary of the Literature

<table>
<thead>
<tr>
<th>Source</th>
<th>Terminal</th>
<th>Decision problem</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benna and Gronalt (2008)</td>
<td>Rail–road</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
<tr>
<td>Boysen and Fleider (2010)</td>
<td>Rail–road</td>
<td>(II.i)</td>
<td>Linear assignment, network flow, mixed-integer programming</td>
</tr>
<tr>
<td>Boysen, Jaehn, and Pesch (2011)</td>
<td>Rail–road</td>
<td>(II.i)</td>
<td>Dynamic programming, beam search</td>
</tr>
<tr>
<td>Corry and Kozan (2008)</td>
<td>Rail–road</td>
<td>(II.i)</td>
<td>Branch and bound, GRASP</td>
</tr>
<tr>
<td>Feo and González-Velarde (1995)</td>
<td>Rail–road</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
<tr>
<td>Ferreira and Sigut (1993)</td>
<td>Rail–road</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
<tr>
<td>Ferreira and Sigut (1995)</td>
<td>Rail–road</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
<tr>
<td>Gonzalez et al. (2008)</td>
<td>Rail–rail</td>
<td>(II.i.i), (II.v)</td>
<td>Mixed-integer programming</td>
</tr>
<tr>
<td>Kondratowicz (1990)</td>
<td>Rail–road</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
<tr>
<td>Lai and Barkan (2005)</td>
<td>Rail–road</td>
<td>(II.i)</td>
<td>Mixed-integer programming</td>
</tr>
<tr>
<td>Lai, Barkan, and Önal (2008a)</td>
<td>Rail–road</td>
<td>(II.i)</td>
<td>Mixed-integer programming</td>
</tr>
<tr>
<td>Lai, Ouyang, and Barkan (2008b)</td>
<td>Rail–road</td>
<td>(II.i)</td>
<td>Mixed-integer programming</td>
</tr>
<tr>
<td>Lee et al. (2006)</td>
<td>Rail–road</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
<tr>
<td>Meyer (1999)</td>
<td>Rail–road</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
<tr>
<td>Montemanni et al. (2009)</td>
<td>Rail–road</td>
<td>(I.v)</td>
<td>Local search, ant colony optimization</td>
</tr>
<tr>
<td>Souffriau et al. (2009)</td>
<td>Rail–road</td>
<td>(II.i.i), (II.v), (I.v)</td>
<td>Linear assignment, mixed-integer programming, variable neighborhood search</td>
</tr>
<tr>
<td>Vis (2006)</td>
<td>Rail–road</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
<tr>
<td>Wiegmans et al. (2007)</td>
<td>Rail–road, rail–rail</td>
<td>Layout</td>
<td>Simulation</td>
</tr>
</tbody>
</table>

Clearly, there are future research challenges not only in the context of each individual terminal type but also with regard to their integration into an existing railway network. Each terminal type varies in investment cost and operational performance, and choosing the right terminal type with a proper layout and efficient operational transshipment processes is thus a challenging task. Furthermore, individual performance assessment needs to be complemented by a network analysis that takes the relations to all other nodes of the rail network into account. It follows that a concerted research effort is required to successfully promote an efficient use of rail freight in the future.

Acknowledgments

The authors appreciate the careful reading and insightful comments of the referees and the editor. This work has been supported by the German Science Foundation (DFG) with the grant “Optimierung der Containerabfertigung in Umschlagbahnhöfen” [BO 3148/1-1 and PE 514/16-1].

References


Boysen et al.: Survey on Container Processing in Railway Yards
Transportation Science 47(3), pp. 312–329, © 2013 INFORMS


