CHALMERS

Capacity factors in intermodal road-rail terminals

Master of Science Thesis
in the Master Degree Programme Industrial Ecology

Adriana Huelsz Prince

Department of Technology Management and Economics
Division of Logistics and Transportation
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2015
Report No E2015:116

MASTER’S THESIS E 2015:116

Capacity factors in intermodal road-rail terminals

ADRIANA HUELSZ PRINCE

Examiner: Violeta Roso,

Supervisor: WSP: Pehr-Ola Pahlén

Department of Technology Management and Economics

Division of Logistics and Transportation

CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2015

Capacity factors in intermodal road-rail terminals

Subtitle

ADRIANA HUELSZ PRINCE

Master's Thesis E 2015:116

Department of Technology Management and Economics
Division of Logistics and Transportation
Chalmers University of Technology

SE-412 96 Göteborg, Sweden
Telephone: + 46 (0)31-772 1000

Cover:

[http://www.terex.com/port-solutions/en/solutions/solutions-for-terminals/intermodal-terminals/index.htm.]

Chalmers Reproservice

Göteborg, Sweden 2015

Preface

This thesis was carried out in collaboration with WSP Group Sverige, whose researchers identified the complexity of rail node capacity determination, and the current lack of a simple and generic tool for this calculation. As the development of an adequate method requires an in-depth analysis of terminal capacity, this thesis was started in order to perform this task.

I would like to thank my supervisor in WSP Pehr-Ola Pahlén and my examiner at Chalmers, Violeta Rosso for their guidance and useful feedback. I would also like to thank Anna Elias and Jonas Emanuelsson for a very enlightening study visit, as well as various WSP consultants for some very interesting conversations and great *fikas*. Finally, I send greetings to my family away from home, my dear international friends, who have made my experience in Sweden an amazing one.

Summary

Intermodal road-rail transportation is a key strategy for the European Union to promote much-needed mobility while reducing emissions from the transport sector. Nevertheless, the railway has increasingly lost market share due to the higher lead times and lower reliability it can offer in comparison to road transport. Due to its dependence on economies of scale, a correct management of infrastructure capacity at the strategic, tactical and operational levels is necessary for the railway to offer a good service quality while being profitable.

In general, while the capacity in railway lines is relatively simple to calculate with the many tools available, the calculation of node capacity is more complex. Furthermore, railway administrators and planners tend to focus, perhaps in consequence, on managing the capacity in lines. This thesis was initiated as a means to find out the factors involved in rail node capacity and to put more light on the subject. Intermodal road-rail terminals were chosen as the specific node of analysis due to the importance of intermodal transport for sustainability.

Therefore, the purpose of this thesis is to explore the internal and external factors influencing the capacity of road-rail intermodal terminals, and to describe the tools that could be used to calculate this capacity. Furthermore, a set of analytical methods, which were adapted from those found in literature, are proposed for estimating theoretical, nominal and practical capacity. Nevertheless, these are only to be used for a rough approximation, and not when performing a capacity analysis involving investment plans. Instead, a simulation method is to be used in this case. An existing simulation model described in the reviewed literature, which was previously developed by a research consortium, is proposed as a possible tool to be used for thorough capacity analyses. Additionally, a set of modifications is suggested for said model to be more generic and precise; so to say, to include the various operational parameters of different terminals, and to avoid simplifications which may overlook relevant capacity factors.

Besides the proposition of calculation tools, perhaps the most important contribution of the thesis is the differentiation between the different types of capacities in intermodal road-rail terminals. While theoretical capacity is a function of the fixed resources (i.e. the infrastructure and equipment), practical capacity is mostly determined on how these fixed resources are used. In turn, the use of resources is highly dependent upon coordination between actors at all levels, which is related to the use of IT tools. Such statements may sound obvious in a first approximation, but a closer analysis reveals that these concepts are hardly applied in practice, perhaps due to the increasing specialization and the complexity of business configurations in the current intermodal transport sector.

Table of contents

List of figures

List of abbreviations

CO_2	Carbon Dioxide
DSS	Decision Support System
EC	European Commission
EU	European Union
GHG	Greenhouse Gas
ILU	Intermodal Loading Unit
IM	Infrastructure Manager
IRRT	Intermodal Road-Rail Transport
OM	Operations Management
OR	Operations Research
PPH	Pre- and Post- haulage
TEU	Twenty-Foot-Equivalent Unit
TMS	Terminal Management System
TRB	Transportation Research Board
UIC	International Union of Railways
UIRR	International Union of Road-Rail Combined Transport

1 Introduction

This section will provide the motivation for the study regarding sustainability and the consequent need to promote road-rail intermodal transport. Next, a background of the main concepts of intermodal in general, and road-rail transportation in particular, is provided. Afterwards, the importance of capacity management for the competitiveness of this transport scheme is presented in order to serve as an introduction to the research problem and purpose of the thesis.

1.1 Background

1.1.1 Transport and sustainability

According to the International Transport Forum, the transportation sector accounts for 25% of global carbon dioxide (CO_2) emissions, with road transport being the dominant source. Furthermore, it is responsible for 60% of the world's oil consumption and 70% of carbon monoxide emissions (OECD, 2010a, p. 5; Tstita & Pilavachi, 2013, p. 444). Particularly, the expected increase of road freight transport presents a significant problem, and the technological improvements for road freight vehicles will not be sufficient to prevent this increase (OECD, 2010a, pp. 25, 26). Within the EU, despite the great efforts carried out by the Commission to lower overall emissions, transportation is the sole sector where emissions are still rising (EC, 2015). In particular, freight and long-distance passenger transport are key factors to address if the EU is to meet its emission targets (EEA, 2014, p.35).

The economic sustainability of the present transport system is also an issue. There is concern that the current capacity of the transport networks will not be able to accommodate the rising demand, so that congestion could hinder economic development and cause important external costs (Golinska & Hajdul, 2012, p. 8; UIC-GTC, 2004, p. 2).

On the other hand, mobility is fundamental for the functioning of the world economy as well as for the general well-being of citizens. Transportation is not only a driver but also a consequence of economic development, in terms of increased flows of knowledge, goods and services (OECD, 2010b, p. 13; Smith, 2003, p. 244). Furthermore, transport creates jobs, generates wealth, and facilitates the inclusion of developing countries on global trade by narrowing the gap between core and peripheral areas (EC, 2011, p. 4; EEA, 2014, p. 57; EUROSTAT, 2009, p. 121).

In order to provide mobility without increasing the environmental impact, the European Comission created the 'White Paper for a competitive and resource transport system'. One of the ten goals described in this papers is "to shift 30% of road freight over 300 km to other modes such as rail and inland waterways by 2030, and 50% by 2050" (EC, 2011, p. 9). Furthermore, the Commission declares that "rail transport is literally the strategic sector, on which the success of the efforts to shift the balance (of transport modes) will depend, particularly in the case of goods" (EC, 2001, p. 13).

It is expected for most of the modal shift volumes to be served by intermodal road-rail transport (UIC-GTC, 2004, p. 3), which allows for combining the economic and environmental benefits of the railway, while providing the flexibility and reach of road transport.

1.1.2 Intermodal transport

Intermodal transport is defined by the United Nations Economic Commission for Europe as the "multimodal transport of goods, in one and the same intermodal transport unit by successive modes of transport without handling of the goods themselves when changing modes" (UNECE, 2010, p. 157). An intermodal transport unit, also called ***load unit*** or *unit load,* may be a container, a swap body or a semi-trailer, which is adapted to be handled and transported by more than one type of vehicle in an intermodal transport chain.

This scheme is based on bundling many individual goods into one standardized load unit, which is then transported and transferred between modes as an individual item (Gudehus & Kotzab, 2012, p. 39). The load carried may consist of finished or semi-finished goods, bulk cargo (e.g. grains, coal) and fluids (e.g. natural gas, chemicals), among others. The operation of loading/unloading load units in order to transfer them from one mode to another in a terminal is called ***transshipment.***

A term which may overlap that of intermodal transport is ***co-modality***, defined as the practice of using each transport mode by its own merits, whether in a unimodal or a multimodal transport chain (EUROSTAT, 2009, p. 6). The efficiency of intermodal transport is case-dependent, since it lies on factors such as the product type and the location of intermodal terminals relative to the sender and the receiver of goods. For example, road transportation is the best option for goods with low shelf life which must be transported over relatively short distances.

In an ***intermodal road-rail transport*** (IRRT) chain, load units containing goods are collected from various sources via road vehicles and transported to a road-rail intermodal terminal (*pre-haulage*). Once in the terminal, the units are unloaded from the trucks and loaded onto a train which performs the long distance transportation (*main haulage/long haulage*). When reaching its destination terminal, the load units are transshipped back to road vehicles, which perform the transport to the final destination (*post-haulage*). Figure 1 illustrates such concept and allows for the visual comparison of unimodal road transport and intermodal road-rail transport. In the case of unimodal toad transport, the routes might not always be as in the diagram, since consolidation of cargo into longer vehicles for the long-haul might also take place to certain extent. The term *drayage* is used generally to refer to transportation over short distances, which are in this case the collection and distribution activities (pre- and post-haulage, or PPH).

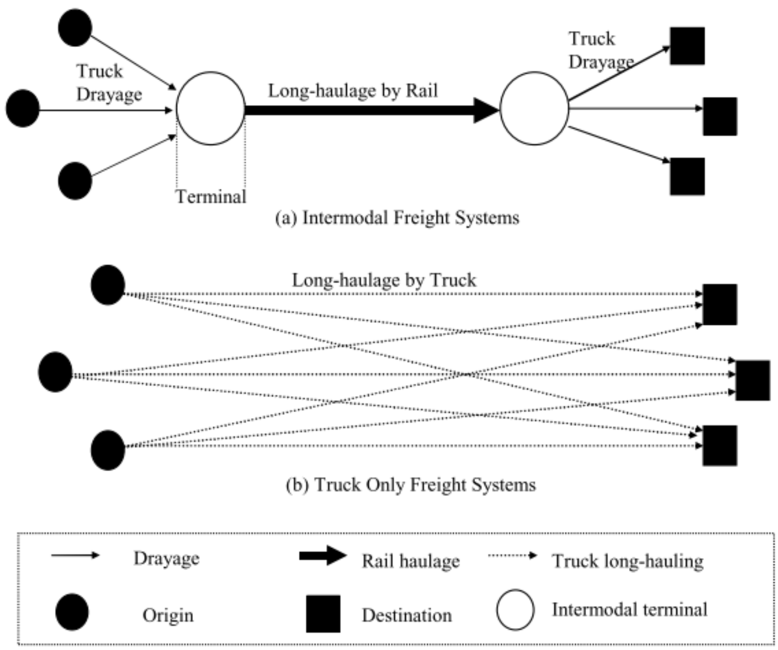


Figure 1: Intermodal road-rail transport (a) and unimodal road transport (b)
(Kim, 2010, p. 34)

1.1.3 Railway freight terminals

Railway freight terminals (also called *railway yards* or *stations*) present very different layouts, operations and uses, according to the market needs and technological capabilities present at each moment within the railway sector (N. Boysen, Fliedner, Jaehn, & Pesch, 2012a). ***Marshaling and shunting yards***, constitute the hubs of the traditional railway system. A hub can be described as a node in a network through which all cargo travels between its departure and destination. In these terminals, incoming trains are split apart, their wagons rearranged and combined, and outgoing trains are formed (see Section 2.2). These flows passing through these hubs have their origin and destination at feeder terminals which can be ***private sidings*** (private yards) or ***free-loading areas***. Afterwards, with the rise of intermodal freighting, ***intermodal terminals*** appeared in order to provide a connection between the different transport networks. They may receive many different names, such *intermodal yards* and *inland/port container terminals*, among others. Within the railway network, they do not constitute hubs, but rather feeder nodes. A more recent concept in transportation research is that of third generation terminals, called ***new hub terminals*** (also called *mega-hubs*) which provide mass-scale rail-rail vertical transshipment of intermodal load units. Their aim is to function as hubs, substituting the role of marshalling yards, which are too time-consuming for intermodal transport, and to provide instead a fast transfer (Bontekoning, 2006). These terminals are expected to combine the functionality of marshalling yards with the transshipment technology of intermodal terminals.

A ***road-rail intermodal terminal*** may be defined as "a place equipped for the transshipment and storage of intermodal loading units (ILUs) between road and rail" (Woxenius, Roso, & Lumsden, 2004). The basic and complimentary operations carried out at intermodal terminals are enlisted in Figure 2.

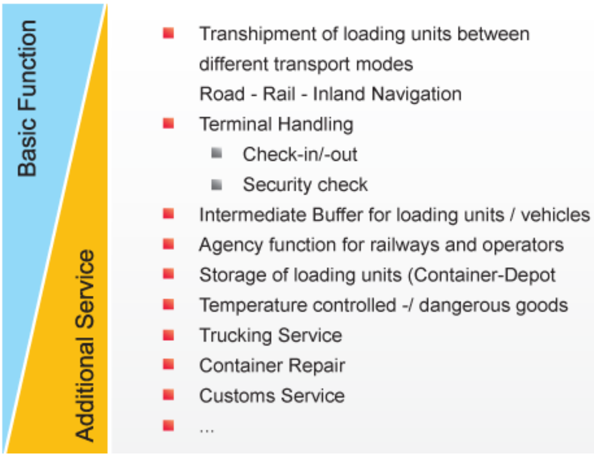


Figure 2: The activities carried out at intermodal terminals
(DIOMIS, 2007)

1.1.4 Capacity management

Due to its large fixed costs, the railway depends on economies of scale for its competitiveness. This implies a need for large resource utilization rate, comprised by both large load factor of trains *and* infrastructure utilization rate. Nevertheless, an inadequate capacity management leads to bottlenecks in the system, rising average total costs and outweighing the benefits brought by economies of scale (Prentice, 2003). Such costs comprise many elements, including congestion, opportunity cost (inability to establish new train services), increased traffic interactions, decreased reliability, and loss of market share.

Infrastructure under-investment has been described as an important chronic bottleneck for intermodal transport (Prentice, 2003). Particularly, the growth in intermodal traffic has been larger than the intermodal terminals are able to manage. Despite investment efforts in terminal capacity, some terminals remain chronic bottlenecks.

1.2 Research problem

The capacity management of railway networks performed by infrastructure managers is often inadequate. Capacity planning activities currently focus on the network´s lines, an outlook that is incomplete since it disregards a crucial part of the system: the railway nodes. The deficiency of information about the nodes might mislead the efforts made to increase the system's efficiency, wasting resources in the process (see for example observations by Lindner, 2011; Marinov & Viegas, 2011; A. G. Woodburn, 2008).

At the present, there is a lack of an adequate method to calculate the capacity in freight railway nodes, which can be applied readily in different scenarios. Current methods are either simple but providing unsuitable results, or adequate, but too complex and case-specific. According to the Research Center VECTOR, "the most challenging task of building an intermodal network is estimating terminal capacity and delay" (2015). In order to incentivize the development of intermodal transport, there is a need for an adequate yet simple methodology which can be conveniently used to calculate said capacity.

It is still unclear how such a method should look like and which is the most optimal way of approaching the issue. Therefore it seems convenient to have a clear theoretical description before its development. This involves determining the optimal scope as well as exploring the aspects involved in resource utilization. The factors affecting the capacity in railway nodes shall be thoroughly analyzed in order to describe the relations between them.

1.3 Purpose and research questions

The purpose of this thesis is to explore the internal and external factors influencing the capacity of road-rail intermodal terminals, and to describe the tools that could be used to calculate this capacity.

Research questions:

- Which are the existing models for line and node capacity calculation? Which are their principles and assumptions?
- Which factors should be taken into account when analyzing intermodal road-rail terminal capacity? Which are the most important parameters, and the relationships between them?

1.4 Scope

The literature research will not be limited to a specific region or time frame. This is because excluding geographical regions would limit the amount of already scarce resources even further and thus leave out valuable input. Sources related to intermodal road-rail terminals are not abundant; therefore literature from all backgrounds was included in the preliminary review. Industry reports as well as reports from the public sector and research centers will be included. Literature from the European Union will be mostly consulted for describing the state of the current system. Nevertheless, for technical aspects whose basic features are shared among all systems, all literature is consulted.

The type of rail node which is the focus of attention in this thesis will be intermodal road-rail freight terminals. Most intermodal trains are not marshalled since they function as shuttle trains, and only a small percentage travel through marshalling yards (Bärthel, Östlund, & Flodén, 2011; Bontekoning, 2006). These hubs will only be examined in order to determine in which extent their operations can affect intermodal transport operations, as well as to get ideas from the models used in these types of rail nodes.

Additionally, passenger terminals are excluded from the analysis. According to Lumsden (2007, p. 88), rail transport for passenger and goods have very different settings. In most EU countries, they are separated through different schedules, infrastructure, and rolling stock features. Passenger transport, which mostly operates during the day, is prioritized over freight transport in shared lines, forcing freight trains to operate mostly during the night. According to an industry practitioner, this trend is not expected to change in the short and medium term[1]. Furthermore, due to the rolling stock and track characteristics, some tracks may be used exclusively for passenger transport, while others may be used only for goods transport.

[1] Interview with Pär Svensson. Logistics development manager, *Eskilstuna Logistik*. May 2015.

However, an interaction that was observed between the passenger and freight systems in the case of shared tracks, is described in the Section 3.6.

In terms of the reach within the modelling process, this thesis will provide a framework and establish the preliminary description of a model of road-rail terminal capacity. Such framework will include theoretical and empirical input for describing the current state of the system, as well as to compare the effectiveness of existing models in the described system. This process is carried out in order to assure that the analysis of capacity is adequate for the current state of the rail and intermodal systems, and that it can provide useful results for capacity planning and management. Therefore, it is not within the reach of this thesis to carry out a modelling process or to develop a readily usable method for capacity analysis.

All of the tools found for calculating capacity and modelling intermodal terminals, whose principles are stated explicitly, are reviewed in sections 2.4.3 and 2.4.4. Reports or papers which do not provide a complete explanation of the methodology used are mentioned but excluded from the analysis. Therefore, the reviewed tools mostly cover research papers and public organization´s reports which propose methods and strategies for capacity planning, but there is no evidence of whether these were actually utilized in a practical case; so to say, there is no guarantee that the proposed models were used as primary input in a real infrastructure development.

The scope of this thesis covers the railway lines as well as the nodes, and does not examine road transportation networks, as it considers only the effect of road transport in terminal operation. A more comprehensive analysis would include the rail system, the terminal, and the road network. Furthermore, within the terminal, this thesis does not analyze deeply the operations of storage and gate reception, but rather focuses on transshipment and short-term storage. It is also outside the scope of the thesis to thoroughly explore the trade-offs between costs and capacity, and the related optimization problem.

The development or readjusting of a simulation method itself is out of the scope of this thesis, and the intention is rather to provide preliminary insights into the topic, which could be later used for the adjustment of the reviewed simulation tools. Furthermore, the data needed for this development would not be easily available and would require a much larger amount of time.

1.5 Method

Firstly, a literature research was carried out in order to have an overview of the system that surrounds an IRRT terminal, including the involved actors and management activities, the railway network operations and the railways´ historical background. Furthermore, a review of the existing models for line and terminal capacity calculation was performed in order to present their assumptions, parameters and structure.

In parallel, empirical input was collected through seven semi-structured exploratory interviews and two study visits, the first performed in Gothenburg intermodal terminal (Combiterminal) and the second in Hallsberg marshalling yard. The questionnaire that was loosely followed in the exploratory interview performed to terminal operators and managers is shown in **Annex 1**. The persons interviewed were Bosse Eriksson, Terminal Operations Manager at Eskilstuna intermodal terminal (company: M4 Grupper AB, logistic service provider, haulier, and terminal operator); Pär Svensson Logistics Development manager of Eskilstuna Intermodal Terminal (company: Eskilstuna Logistik, owned by Eskilstuna municipality); Anna Elias, Terminal Operations Manager at Gothenburg Combiterminal (company: GreenCargo, rail operator and terminal operator); Jonas Emanuelsson, systems manager in Jernhusen (state-owned company, terminal owner); as well as the logistics and transportation consultants at WSP Group (consultancy company) Sverige Dag Hersle, Fredrik Bärthel and Lennart Hammarbäck.

These study visits and exploratory interviews served a two-fold objective. Firstly, they were used to clarify the functioning of the system and provide a context which made it possible to correctly analyze literature. This was especially useful for the development of the framework; as the literature was being reviewed, several information gaps were encountered, and practical knowledge was necessary for connecting the concepts provided by different sources.

Second, the empirical input was useful to describe details of information and material flows in terminals, which were lacking in the reviewed literature. Furthermore, results from interviews provided the opinions of several industry actors regarding the most relevant capacity and performance factors to be considered. Particularly, information of the activities carried out at the Gothenburg Combiterminal was used to provide real-life examples of the concepts elaborated in Section 3.

Within the description of the terminal system in Section 2.3, the reader might notice that some sources correspond to the same as the sources for the reviewed models. This is because in the system description of this thesis, only the framework provided by the introductory sections of those papers was used. Comparatively, for the literature review of models, only the section of each paper which abstracts and imitates the real system was utilized.

The information will be analyzed from the point of view of Operations Management (OM). As will be observed in Section Fel! Hittar inte referenskälla., current methods and models utilize tools which originate in the Operations Management and Operations Research (OR) disciplines.

Operation management theory is a solid area that has been applied for years in industrial processes and can therefore provide insight into the analysis of terminal capacity and operations. It is defined by Slack et. al. as “the activity of managing the resources which are devoted to the production and delivery of products and services” (2013, p. 4), while Brown et.al. determine that it “is concerned with those activities that enable an organization (and not just a part of it) to transform a range of basic inputs (materials, energy, customers ‘requirements, information, skills, finance, etc.) into outputs for the end customer” (2013, p. 4). Likewise, this discipline analyses several problems faced when calculating and planning capacity, which are also present in the railway system, such as demand variability. OM is the practical equivalent of OR, a research discipline which provides optimization and decision-support tools such as inventory theory, queuing theory, network modelling, optimization and simulation techniques (Taha, 2007).

The application of Håkansson's three element approach which theorizes industry networks, as carried out by Woxenius´s (1994) was especially useful to understand the road-rail intermodal business structure. As explained by H.A. Taha, expert in OR, it´s "more than just mathematics". In fact, gaining an understanding of the system, its actors, responsibilities and interactions was the biggest challenge in the development of this thesis.

As is explained in Section 3, Findings and analysis, a pure technical approach would not be sufficient for the author to adequately analyze and understand both the real system and theoretical models, having furthermore no previous practical experience in intermodal or railroad transportation.

2 Framework development

This section has the aim of developing a framework which will aid towards the determination of the scope, and a general description of the system surrounding the terminal. Furthermore, the terminal itself is described is described, followed by the general concepts related to capacity. Afterwards, the tools found for the analysis of line and terminal capacity are reviewed.

2.1 System description: Railway and road-rail intermodal transport

2.1.1 Networks

The railway system is often represented as a network, defined as a set of interconnected nodes and links (Mayhew, 2009). Depending on the perspective taken, links and nodes represent different components. When railways are analyzed at a macroscopic level, a node is any point in a network in which two or more lines meet, aggregating one terminal or a group of interconnected terminals situated in the same location; while the links represent main lines composed by running tracks. The infrastructure components that form part of the network are described in more detail in **Annex 2**.

In a mesoscopic view, each terminal, line junction, and station represents a node, and lines represent the links (this level is not frequently addressed in literature and lacks a more formal definition). From a microscopic view, tracks—including running tracks and sidings—form the links, while each junction forms a node (see Armstrong et al., 2010; Marinov, Sahin, Ricci, & Vasic-Franklin, 2013). The common definition for a railway node is as any point in a network in which two or more lines meet, including junctions, passenger stations and rail freight yards (UIC, 2004, p. 6).

The term *node* will be used in this thesis as viewed from the mesoscopic perspective, where links correspond to rail lines and nodes correspond to terminals and yards. In turn, a microscopic perspective is utilized when describing the nodes and their components.

In intermodal transport systems, nodes may be regarded as gateways which link different transport networks, and are classified as intermodal or intramodal gateways. Intermodal gateways link networks which are based on different transport modes, such as intermodal terminals, while intramodal gateways link different transport networks of the same mode. According to Woxenius, Roso and Lumsden (2004), international rail services today still use intramodal gateways between bordering countries, since national rail networks remain highly incompatible regarding technical, operative and legal requirements.

Rail freight companies normally carry out *bundling* or consolidation of shipments through the network in order to achieve the best combination of economies of scale and transport quality (i.e. lead time, reliability and frequency). This decision is carried out according to:

- Market requirements (i.e. the desired transport quality and cost)
- Transport distance
- The volumes that can be attracted (i.e. the size of flows)
- Characteristics of the products transported (e.g. shelf life, fragile goods)
- Presence of intermediate markets along routes

(Woxenius, 2007)

The type of bundling chosen is also called the ***train production system***, and can be divided between full-trainload and less-than-trainload systems. Full-trainload systems are employed whenever there are large enough volumes in a route to fill up an entire train, forming direct links, shuttle trains, or block and system trains. Meanwhile, less-than-trainload systems are used when the volume of each origin/destination (O/D) relation is not enough to fill a train, and there is need to consolidate cargo. They include hierarchic networks, hub-and-spoke networks, and corridors (COSMOS, 2015), see Figure 3. A detailed account of each production system can be found in **Annex 3**, based on the descriptions by (Bärthel et al., 2011; Bergqvist, 2008).

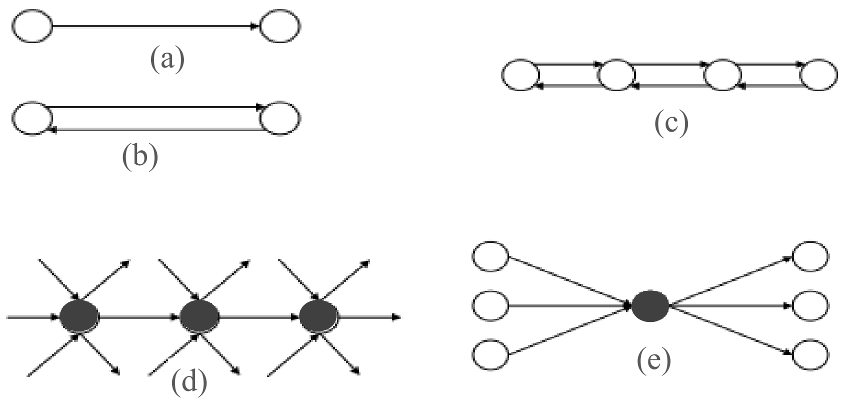


Figure 3: Bundling network designs
Full-trainload systems: (a) direct links, (b) shuttle train. Less-than-trainload systems: (c) corridor network, (d) hierarchic network, (e) hub-and-spoke network. Dark circles represent hub terminals. *(Woxenius and Bärthel, cited by Bärthel et al., 2011, p. 84)*

Additionally, the EU COSMOS project (2015) mentions other train production systems for less-than-trainload traffic such as Y-shuttle train (in which two trains coming from different terminal are merged), gateway traffic (combining different railway networks), and mixed intermodal/conventional traffic (in which intermodal traffic is mixed with wagonload).

Full-trainload systems are the best option in the case that high volumes can be achieved; hence it is mostly used for long distances and high-volume routes. It is the simplest way to operate and provide the highest transport quality. Nevertheless, the need for large volumes makes such routes only profitable to operate between big markets; additionally the amount of origin-destination relations served is limited, as well as the service frequency.

Less-than-trainload systems have the advantage of achieving large economies of scale, a convenient service frequency, relatively low costs, and are able to serve low-volume routes. Nevertheless, their service quality is lower: a large amount of handling activities in the nodes increase transport times, while reliability is decreased because of the higher number of possible interferences. The bigger the amount of nodes, the larger are the lead times and the expected time unreliability.

Traditional bulk transport

The traditional freight railway system is centered on the transport of bulky goods such as metals, coke, coal, aggregates (e.g. sand, gravel, crushed stone), and semi-finished industrial products (A. Woodburn, 2014, p. 3). In the case of full-trainload traffic, *block trains* or *system trains* are utilized, in which by definition the load of the whole train belongs to just one customer, and the rail system is used as "conveyor belt" between big manufacturing sites of an industry. In turn, less-than-trainload traffic travels through the so-called ***wagonload*** or *carload* system, whose physical organization can be described as a hub-and-spoke or hierarchical network. In this scheme, each wagon forms part of the composition of two or more trains during its journey (Marinov et al., 2013).

In the single wagon hierarchical system, wagons are loaded in goods yards and then carried by local or regional trains into marshalling yards (i.e. the hubs). Afterwards, they are transported between yards by interregional trains, with marshalling and shunting operations in-between. In the end, they are delivered by regional trains into their destination yards (Bärthel et al., 2011, p. 84).

The hubs of the wagonload network are called marshaling or shunting yards (also classification yards) where classification and consolidation activities take place. Incoming trains from different locations are disassembled, and their wagons (or wagon groups) are classified and grouped together with wagons with the same destination, forming new outgoing trains (see Section 2.2).

Intermodal system

Before the surge of intermodalism, the transport of downstream products[2] was mostly a task for road transport. With the rise of IRRT, the railway has incorporated these loads into its share, but it is necessary to provide a combination of lower costs, and/or lower lead times to remain competitive in this market. In this respect, several rail operators who offer IRRT services have abandoned consolidation practices and opted for direct links and shuttle trains. This provides a better service quality and avoids the costly and time-consuming marshalling process (Bontekoning, 2006, p. 3). Only a reduced number of operators combine intermodal cargo with wagonload.

[2] Downstream products refer to manufactured or semi-manufactured materials, which are closer to the customer in a supply chain and thus have more embedded value, as opposed to raw materials (located upstream).

2.1.2 Business structure

During the planning process of a project aimed at developing new technologies for combined transport, Woxenius (1994) identified the need to study the industrial organization of the combined transport industry and its production systems. In order to develop a technology, it is necessary to discern who the intended user is, as well as the environment that surrounds it. However, a general lack of knowledge about the industry´s structure was identified. For this reason he developed a three-element approach based on previous models of industry networks, describing the system according to its elements: actors, resources and activities.

The actor categories mentioned in his work in 1994 are later modified in 2002. In the later work, the actors are classified as shippers, forwarders, haulers, intermodal operators, terminal companies, rail operators, and equipment leasing companies (Woxenius, 2002).

As Woxenius (1994, p. 4) comments, the combination of different transport systems can often cause confusion within the intermodal transport terminology. Moreover, the various combinations of tasks that companies perform makes it complicated to categorize them as one actor or the other. He also mentions that the sector is traditionally divided between companies focused on rail, and companies focused on road transport.

Shippers are the final users of the intermodal transport system, and their interest and knowledge of the system´s functioning depends on the size of their shipments. Generally, the larger the load, the more involved the shipper is in the system. Sometimes, shippers who fill out load units perform their own local road haulage, cargo terminal transfer[3], cargo terminal services, and/or supply of load units. Shippers of general cargo who fill less than one load unit normally do not know about the system´s workings. Most shippers do not have direct contact with the transport operators and instead contract the services of one or many forwarders (Woxenius, 2002).

The ***forwarder*** acts as intermediary between the shippers and the transport operators. It must satisfy criteria established by the shipper, such as the transport volume, frequency and prices. Often, the forwarder is the one to decide which transport mode will be used and has the ability to control the transport chain. The shipper normally gives little importance to the transport mode and provider chosen by the forwarder, as long as price and quality requirements are met (Flodén, 2007, pp. 9, 53).

Forwarders also offer other services such as consolidation of consignments, storage, documentation, customs and supply of load units. There are many types of forwarding companies; the large, traditional forwarders are usually inclined towards haulers and use intermodal only when there is need for extra capacity (Woxenius, 1994).

Haulers perform the carriage of the load units by road. They might be specialized to one unit load or have the capability to handle various types, and their organizational size varies widely across the sector (Woxenius, 1994).

[3] Cargo terminal in this context refers to any place within the transport chain in which transported materials are handled. Therefore, does not make a direct reference to an intermodal terminal.

According to Woxenius (2002), ***intermodal operators*** are the companies offering terminal-to-terminal services, which include the rail haulage and transshipment. Others describe the ***intermodal transport coordinator*** as the company in charge of coordinating intermodal transport services, but not necessarily performing them itself. It serves as an intermediary between the different actors in the intermodal system. In some cases, it has a direct communication which shippers, being able to perform door-to-door transport. In other cases, the intermodal transport operator only communicates with forwarders and carriers; and performs terminal-to-terminal transport (Behrends et.al., 2011; Flodén, 2011, p. 9; Woxenius, 1994, p. 77). It is not declared explicitly whether these two roles are or not equivalent.

Terminal companies offer the transshipment services, and may belong to rail operators, infrastructure managers or dedicated terminal companies. Flodén (2011) mentions that the ownership of terminal companies may be outsourced, which is why this thesis will make a differentiation between terminal owner and terminal operator, since this separation is not explicit in literature but seems necessary as they have different roles in capacity planning. A ***terminal owner*** will be defined as the organization in charge of planning, building and maintaining a railway freight terminal, and may be a private or public actor. In the case that it is publicly owned, this role normally corresponds to a public rail infrastructure manager (IM). When it is private, the owner might be the rail IM, a railway operator, or a dedicated company with such core business. A ***terminal operator*** is the company in charge of performing the operations and management of the terminal. The possible ownership configurations are equal to those of the terminal manager. Additionally, one organization may act as both terminal manager and operator, but the rising trend is for a terminal owner to subcontract the operations to a dedicated terminal operator (Bergqvist, 2014, p. 27; Flodén, 2011, p. 3).

A ***railway operator*** (also called *railway company* or *railway undertaking*), provides traction and performs the actual transportation of goods or passengers by running trains through the network. These companies normally focus either on passenger rail or freight rail, but sometimes operate both. The term *railway undertaking* is sometimes used to identify a company that performs solely the traction function, i.e. is appointed by another company to provide locomotives for hauling rail wagons (Liilennevirasto, 2012; SNCF, 2007; Trafikverket, 2012). An operator that organizes rail traffic but does not provide traction is called a *traffic organizer*. It does sell transport services, but does not perform the traction function itself. Instead, it is able to appoint this service from a railway operator (Trafikverket, 2012, p. 7).

Equipment leasing companies own locomotives, wagons and/or load units and lease them to the other actors in the system. Due to the increasing deregulation of the sector, ownership structures become ever more divided through time.

Other actors include the railway infrastructure manager, public authorities and international organizations, mentioned in the following paragraphs.

The railway ***infrastructure manager (IM)***, also called *infrastructure owner*, is the organization in charge of maintaining, constructing, managing and planning the fixed assets of a network (RFF, 2009; SNCF, 2007; Trafikverket, 2012). It owns the infrastructure in regard to the lines, and may own also terminals and yards in some systems. This organization can be either private or public, a condition that is different along several rail systems.

In a vertically integrated system, the infrastructure is owned and operated by one same organization, i.e. the infrastructure owner and the transport operator are the same organization. Oppositely, in vertically separated systems, the railway operators are separate entities from the IM, and pay fees to the IM to have the right to use tracks in certain time slots (Beck, Bente, & Schilling, 2013, p. 30).

International transport organizations and also play a relevant role in the sector, through the diffusion of best practices, standards, and coordination initiatives. In particular, the International Union of Railways provides equipment and rolling stock standards, as well as practical reports on combined transport. With the cooperation of European transport consultancies and operators from several countries, interdisciplinary projects have been developed with the objective of boosting the competitiveness of intermodal transport in a European-wide scale, in cooperation with the EC. Furthermore, Woxenius (1994) mentions the International Union of Road Rail Combined Transport (UIRR), which is formed by combined transport operators and intermodal terminal owners and operators.

2.1.3 Historical background

According to Rodrigue et al. (2013), containerization is without a doubt the main driver of intermodal transport. The practice was born in 1955 as a strategy to achieve a streamlined flow of goods in intercontinental shipping, where goods are delivered via road vehicles to cargo vessels, constituting a strictly multimodal chain (World Shipping Council, 2015). It is relevant to mention that the practice of containerization does not necessarily mean that the only load units used are containers; swap bodies and semi-trailers are also widely utilized.

Only 50 years after the start of containerization, not only an approximate of 60% of shipment by sea in terms of value would be done via maritime containers, but it has also become an important part of inland transport. This is because the concept of loading units has made it possible for the transport industry to improve operations in many aspects (World Shipping Council, 2015).

As intermodal transport allows for materials to be managed as a single mass within a load unit, it brings important economies of scale due to reductions in handling, labelling and storage costs (Gudehus & Kotzab, 2012, p. 39; Tompkins et.al., 2010, p. 186). Load units also serve as a protective mechanism for the goods, which decrease need for various small-scale protective packages, saving material resources and avoiding waste. Moreover, reduced handling operations lowers the risk of contamination and damage of goods, in addition to improving traceability (Lumsden, 2007, p. 409; Prentice, 2003).

Before the 1990´s, all European railways were arranged as separate state-owned, vertically integrated national networks. One governmental authority would both own the infrastructure and manage its operations. As such, each country´s network evolved separately, developing different technical standards and operational methods (Hilmola et.al., 2007; Vrenken et.al., 2005). Countries such as the UK and Sweden set a precedent in Europe by liberalizing their railway (Hilmola, et.al., 2007). Afterwards, the European Commission achieved a generalized privatization of transport networks, through regulations and reports such as the 2001-2004 Rail Infrastructure Packages and the 2011 White Paper. The overall objective is to boost competition among transport companies in order to create more efficient transport services, and to enhance European economy though a Trans-European Transport Network (TEN-T) (EC, 2013).

In several countries passenger rail operators are still public and vertically integrated, while the IM is still a public monopoly (Liilennevirasto, 2012). The EC is currently preparing the 4th railway package, which aims to de-regulate and separate passenger operations (Katsarova, 2014).

Nevertheless, in spite of the efforts by the EC and other organizations, the competitiveness of rail freight is still hindered by the past national nature of rail networks, which shape its infrastructure and operations. According to the EC, an important reason for the loss of share of the railway is its lagging position in terms of opening and de-regulation in comparison to other modes of transport (EC, 2001, p. 7). The influence of historical factors on the current system is explained in the following paragraphs.

Firstly, different technical characteristics and legal constraints decrease interoperability and complicate cross-border crossings. For example, it is sometimes necessary to perform extra operations such as changing locomotive and other equipment when passing through a border, since not many locomotives are equipped for working within different systems. Technical differences include variances in train gauge, clearance parameters, power source, and signaling and security systems between country networks (and within a network). This makes the transport process more costly, increases transport lead times and increases the risk of unexpected delays (EUROSTAT, 2009; Vrenken et al., 2005).

Second, the past national nature of railway systems provided no competitive incentives, and hindered collaboration between different national operators. These factors have slowed down innovation and coordination between actors in the sector. The situation is reflected on the lack of a common terminology within the sector (and with other transport sectors), no standardized planning and operational methods, low resource standardization, and an extraordinarily low implementation of information technologies (in specific parts of the rail system). Furthermore, the transport consultancy companies CombiConsult and Kassal + Parters report than in the EU: "there is (…) a lack of dissemination and mutual learning" (UIC-GTC, 2004, p. 16).

Interestingly, the difference in track gauge between different countries is also related to strategies aiming to protect national security, according to historic geographical configurations. For example, while Finland was occupied by Russia (1809-1917), Russia built tracks with a gauge specifically different from that of Germany (Poot, 2004, p. 244). A contemporary example external to the EU, can be found in the article "Mongolia's railway to China turns into national security threat" (Dettoni, 2014).

Third, as deregulation is still a recent occurrence, management structures within the rail and intermodal transport are still unclear. As a result, European intermodal transport has been described as disorganized, with actors performing ad-hoc roles and partnerships (Wichser et.

al., 2007). Additionally, as a consequence of liberalization, there is more and more division of tasks among a bigger number of actors. Information sharing is essential for coordination and optimization of a transport chain, but normally, for competitive purposes, private actors in the EU prefer to keep information regarding traffic volumes and future strategies confidential (UIC-GTC, 2004, p. 19). Furthermore, according to Anderson & Walton (1998, p. 4), companies within an intermodal chain may have incompatible interests, such as different schedules, which prevent the overall chain to function efficiently.

2.1.4 Market loss

Until three decades ago, freight transport in Europe was strongly regulated. In certain countries where intermodal volumes were low and/or trade routes comprised short distances, rail operators used traditional wagonload operations for intermodal transport. Hence, intermodal traffic was either combined with wagonload, or was carried by dedicated intermodal trains which were marshalled in hubs. Due to the strong transport regulations and fees for road transport, this could be carried out for rail while still being competitive with road (UIC, 2012, p. 52).

But as road transport was gradually liberalized following the EC´s first white paper in 1992, the lower service quality offered by wagonload bundling systems undermined rail freight competitiveness. To assess this problem, rail operators gradually adopted more direct routes. As a result, most of the intermodal traffic today happens through dedicated shuttle trains (UIC, 2012, p. 52).

The concentration on dedicated shuttle trains implies a reduction of the intermodal transport network. The inability to serve intermediate markets shrinks volume flows between certain origin and destination relations. As traffic volumes lower, some routes and terminals are closed down since fixed costs are no longer covered. This decreases service quality, since the network becomes less flexible, more costly, and has services with lower frequency. Service quality is also hindered due to increased congestion, as rail operators increasingly focus on high-volume corridors and terminals (Bontekoning, 2006, Chapter 2.5.2; Koenig, 2001; UIC-GTC, 2004, p. 86).

As a result, inland combined transport services in the EU, especially for domestic transportation, are not performing well economically (UIC-GTC, 2004, p. 4). The vicious circle of market loss and financial problems is depicted in Figure 4.

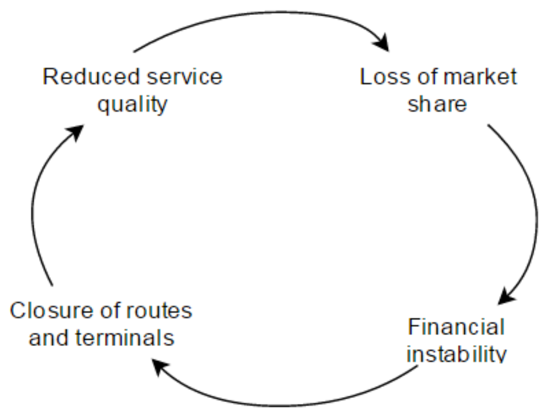


Figure 4: The vicious circle of railway and rail-road intermodal transport
(based on UIC-GTC, 2004)

2.1.5 Planning and management

2.1.5.1 Strategic level management

This management level has a long term perspective, usually 10 to 20 years, and involves decisions related to the strategic goals of a company. It comprises changes in the company´s structure, as well as investment choices regarding the geographical location of new infrastructure, system capacity, technology acquisition, and the resources to be utilized (Jonsson, 2008, p. 305; Macharis & Bontekoning, 2004, p. 402; Marinov et al., 2013, p. 60; Steadieseifi et. al. 2014, p. 2). Strategic decisions are usually influenced by a company´s surroundings in a wider scope, i.e. market and economic forces, policies, industry regulations, society, and the environment (Ahrens et al., 2009, p. 1070).

The actors who own infrastructure (i.e. both the IM and terminal owner) have the task of planning the physical features of the transport network, i.e. the ***physical network design***. This involves deciding on the configuration as well as capacity of the infrastructure (Macharis & Bontekoning, 2004, p. 27). In particular, the *IM* is in charge of the design of the transport network configuration in regard to the links. Decisions comprise the construction of new lines on previously unconnected areas (construction), as well as the addition or removal of tracks for existing routes (reconstruction). Meanwhile, the *terminal owner* decides on the localization, number and design of new terminals (including infrastructure and superstructure), as well as infrastructure adjustment for existing terminals. Ideally, the terminal owner and the IM are in close collaboration for this matter so to create a network which allows for a streamlined flow of goods.

Meanwhile, the carrier must plan the total capacity of the vehicle fleet (and/or rolling stock) (Behrends & Flodén, 2012, p. 5). Additionally, it may make a draft plan about which routes to serve (Bontekoning, Macharis, & Trip, 2004a, p. 14).

Decisions for the service and physical network design are commonly performed by utilizing network models in order to compare between different configurations (Crainic & Bektas, 2007, p. 7; Macharis & Bontekoning, 2004, p. 409). Such network models may be adapted by establishing different objective functions according to each actor´s interests. Cooperation activities between all actors should occur at this level so to align strategic decisions, but this is rarely the case (see Section **3.5**).

2.1.5.2 Tactical level

Tactical planning involves medium term decisions which aim to optimally allocate the existing resources, in order to meet the required transport demand, service performance, and the strategic objectives. In this level analysis are conducted regarding congestion, capacity, and system performance (Marinov & Viegas, 2011, Chapter 363). The time scope for decision making at this level is of months or weeks (Macharis & Bontekoning, 2004).

In this level, the *carrier/rail operator* performs the activities of vehicle scheduling, crew scheduling and ***service network design*.** The latter includes the setting of the transport service schedules and itineraries (including which links and nodes will be used), which are related to speed, frequency and service level. Such decisions also involve the selection of direct connections or bundling/consolidation systems (Behrends, 2006, p. 6; Steadieseifi et al., 2014, p. 6), i.e. the train production system (see Section **2.1.1**).

The itineraries are developed by the rail operators, often using modelling tools such as RailSys, and serve as a base for the crew and vehicle scheduling (Marinov et al., 2013, p. 69). They are then used to apply for slots in the yearly timetable, for which they might compete against other operators. The rail operator might not receive the exact slots it applied for, since the final timetable developed by the IM must reconcile a several overlapping applications. Therefore, the tasks of rail operator at this level also include the re-design of the service network according to the final timetable, which in turn causes the need to adjust crews and vehicles/rolling stock scheduling.

The IM is in charge of developing the ***timetable*,** which contains all the data related to the scheduled movement of trains in a rail system during one calendar year (SNCF, 2007). Around 10 to 8 months before the start of the next calendar year, rail companies present their applications for the desired paths in the timetable. A path is described as the "the infrastructure capacity required to run a given train from one point to another at a given period of time" (SNCF, 2007). The application for paths must include information such as departure time and location, arrival time and location, and the priority according to the rail operator. It might also contain intermediate stops which don´t overpass a certain time limit, the earliest acceptable departure time, and the latest acceptable arrival time (Trafikverket, 2012, Chapter 5). Rail operators apply for paths in the form of *access rights*, which give the operator the right to periodically run a number of trains on certain times on specific parts of the network (Gibson, 2003, p. 41).

A preliminary version of the timetable is published around a semester before the start of the timetable´s year, after which the applicants have the chance to appeal.

The IM must coordinate the needs for capacity of several rail operators and solve several conflicts (e.g. many applicants applying of overlapping paths). When there is congestion, it must prioritize the applications in order to allocate capacity in the fairest and most efficient way possible, a process that is different for every system. In several European countries, passenger trains are prioritized over freight trains. Next, the IM takes into account the priority given by the train operator himself, besides utilizing predetermined criteria according to the IM or another organization´s discretion (Ballis & Golias, 2004, p. 423; Trafikverket, 2012, Chapter 5).

As the final timetable is published, the IM issues several related documents. One is the ***train-graph*** or *train diagram,* which illustrates the train paths in one section of a line during a certain time period travelling in both directions, as in Figure 5. The time is represented in the x-axis, while the y-axis shows the distance travelled. In this graph the different speeds of trains can be visualized: the steeper a line is, the larger the speed of a train, while horizontal lines represent complete stops performed at stations, and curves represent braking and acceleration. It is also visible that the speeds of each train vary according to the line section it is traversing, becoming lower as the trains approach the beginning and end stations.

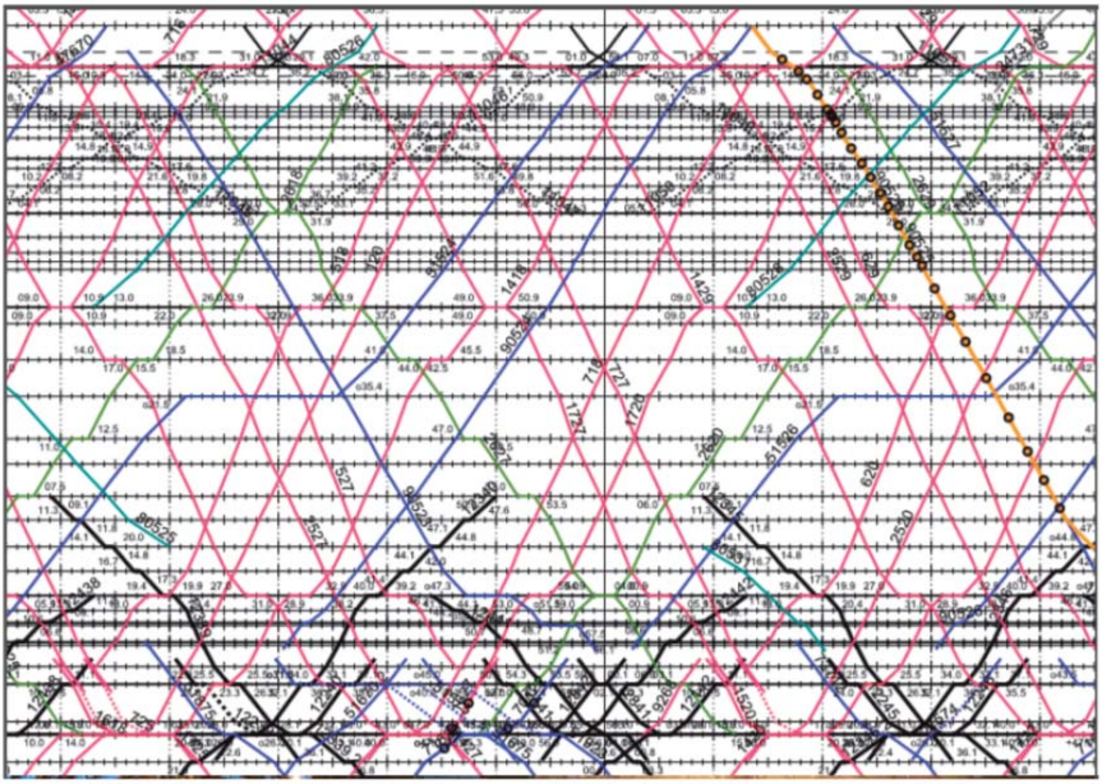


Figure 5: Train-graph representing the movement of trains during a certain period of time for a given line section
(Gibson, 2003, p. 41).

At this level, the terminal owner makes decisions regarding the capacity levels of the terminal´s superstructure capacity, while the terminal operator plans labor and equipment capacity, and redesigns operational routines and layouts (Macharis & Bontekoning, 2004).

2.1.5.3 Operational level

At the operational level, short term decisions are made with the objective of ensuring that the service is delivered to the customer in a safe and reliable way through an efficient use of resources. This level includes all the day-to-day operations, as well as the adjustment of tactical plans to current conditions, i.e. real-time decision making (Macharis & Bontekoning, 2004).

At this level, the haulers´ and rail operators´ role is to perform the transport service and manage empty units and vehicles. Decisions include adjusting its operations to sudden changes in demand, regarding volumes, times and service types (Macharis & Bontekoning, 2004).

The Infrastructure Manager´s main role at this level is to perform the traffic control. This is done by ***train dispatchers*** (also called traffic controllers), the staff responsible for the following activities, within a local or regional part of the network (Marinov et al., 2013, p. 62; Reinach & Gertler, 1998, p. 10):

- Monitor and coordinate train movement
- Initiate and stop train movements by controlling signals and switches, and communicating with the staff along the tracks.
- Adjust the situation to unplanned events (such as unscheduled delays) and emergencies
- Train rescheduling, involves forecasting of possible traffic conflicts and resolving them before they occur.

The traffic decisions made by train dispatchers are based on years of experience, and commonly no decision-support systems are used. This is because their roles involve several interrelated tasks which cannot be modeled as a single system due to their complexity. Certain chores might be automatized (but only separately), while others can´t be easily simulated, such as real-time rescheduling while solving train conflicts.

The activities within intermodal terminals and marshalling yards are presented in sections **2.2** and **2.3**.

2.2 Marshalling yards

Information presented in this section has been obtained from and Khoshniyat (2012), Bontekoning (2006), as well as a visit to Hallsberg marshalling yard, Sweden, on July 10th, 2015.

As mentioned in Section **1.1.3**, in marshalling yards incoming trains are separated and wagons are arranged according to their destinations, forming new trains. They most normally service wagonload trains but in some cases also process mixed wagonload/intermodal trains, a case which happens in the Hallsberg yard.

2.2.1 Resources

The infrastructure of marshalling yards is generally divided into three areas (see Figure 6):

- *arrival tracks* (also called arrival yard, arrival group, receiving area), in which trains are received and prepared for shunting into the marshalling tracks.
- *marshalling tracks* (also classification area, classification tracks), in which wagons are classified and grouped.
- *departure tracks* (also departure yard, departure group), in which trains are prepared and wait for their departure time.

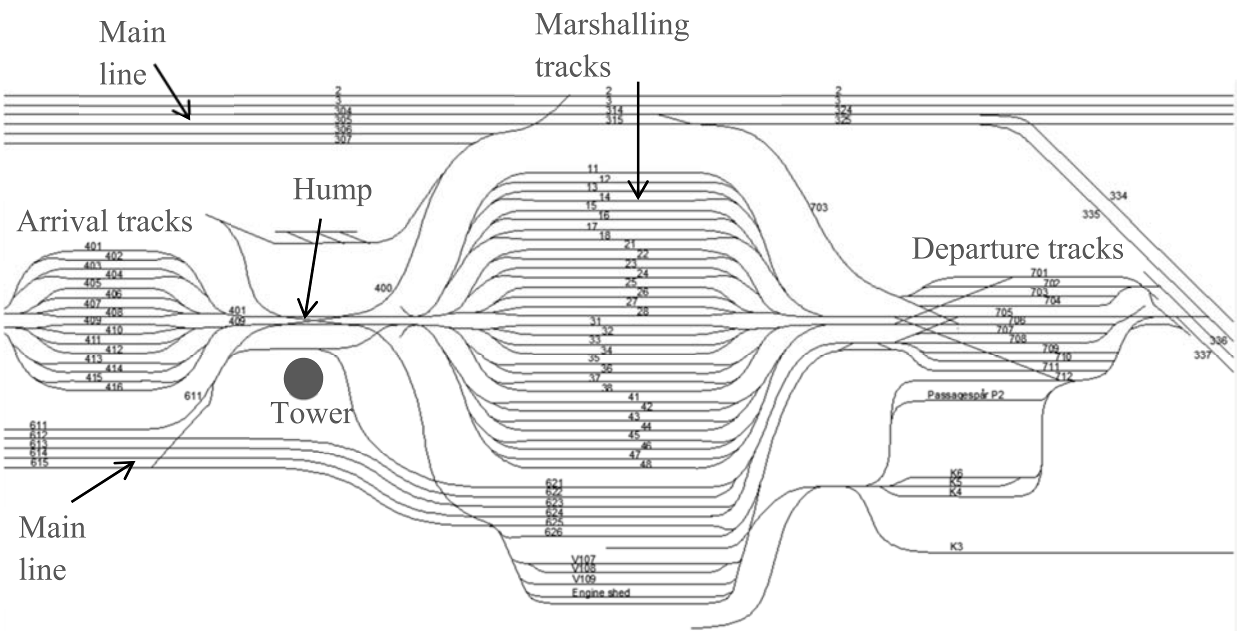


Figure 6: Layout of Hallsberg marshalling yard
(own creation upon plans provided by Trafikverket to Dag Hersle)

Additionally, most yards have a tower situated next to the hump, in which train dispatchers supervise and control the switches, brakes, and communicate work instructions. The infrastructure also includes the catenary and overhead contact wires in the case of electrified yards. In terms of the superstructure, the terminal uses shunting locomotives which can be powered by electricity or diesel, and have technical characteristics different from the network locomotives.

Some yards do not have all three tracks groups, and in such cases some activities may be consolidated in one same area (N. Boysen, Fliedner, Jaehn, & Pesch, 2012b). According to their layout, marshalling yards may be described as flat yards, gravity yards and hump yards. In a flat yard, shunting locomotives perform all the work for wagon shunting, while in gravity and hump yards, the movement is aided by the force of gravity. The most common layout in the EU is the hump yard, in which a small elevation, or hump, is located in the neck between the arrival tracks and the classification tracks (Christoph, 2010; Marinov et al., 2014).

The superstructure of a marshalling yard consists on shunting locomotives, which are normally owned and operated by the terminal operator.

2.2.2 Operations

As a train arrives to an arrival track, the network locomotive is uncoupled and the train set is prepared for being shunted, which includes brake release, wagon inspection, and loosening the couplings between wagons. Next, it waits for its turn to use the hump; as the time approaches, a shunting locomotive is coupled to the end of the wagons set (the left side according to the layout in Figure 6). In the tower, train dispatchers observe the operation and prepare the switches (which are controlled remotely) in order to direct the train into the hump, and then each wagon group into the correct classification track. While being instructed via radio by the dispatchers at the tower, the shunting locomotive slowly pushes the wagons over the hump. As each wagon group arrives to the top, its couplings are manually detached from the rest of the train by an operator, so they can roll with the force of gravity into the classification tracks (*roll-in).* Fixed brakes along these tracks, called *buffer stops*, prevent the wagons from colliding with previous wagons or reaching the next track group. Same as the switches, these buffer stops are controlled by the dispatcher at the tower. After the switches in the classification yard are re-set, the next wagon group is pushed and the operation continues.

Once a group of wagons are located in a classification track, they must wait until all of the wagons assigned to the same outgoing train have arrived, which might take many hours long. Once the complete set is in the track, the wagons and brakes are coupled, and the buffer stops are released. Next, a shunting locomotive is attached to the wagons from the right and pulls them out into the departing tracks (*pull-out*), where the line locomotive is attached, and the brakes are prepared and inspected. Finally, the train waits until its departure time.

Ideally, the number of marshalling tracks is the same as the number of destinations that the yard serves. Nevertheless, this is not always the case, since it might not be cost efficient and the area is constrained by the available land. In this case, when early wagons must wait to have a marshalling track assigned, the operator is forced to combine wagons from several trains on the same track. In the Hallsberg Marshalling yard these are called *scrap* or *mixing* tracks, and are also located within the classification yard. When one of these wagons is assigned to a classification track, it must be shunted out from the classification yard into the arrival yard (*pull-back operation*). Next, the shunting locomotive must change sides and push the group over the hump. **Annex 4** provides a detail diagram of the operations performed in each track group of the marshalling yard.

It is relevant to mention that the names given to these operations (roll-in, pull-back, pull-out) vary widely across literature and regions, and thus it cannot be expected that the same terminology is used in different sources.

2.3 Intermodal terminal

The main role of road-rail intermodal terminals (also called *intermodal rail yards or transshipment yards*), is to function as an interface between the road and rail networks by performing the transshipment of load units.

The *intermodal terminal owner* is responsible for planning, maintaining and managing the infrastructure and superstructure (i.e. the transshipment equipment). The *intermodal terminal operator* manages the day-to-day operations, contact with customers, staff, and determines operational procedures and policies (sometimes in collaboration with the terminal owner). In turn, the terminal service is contracted by an intermodal operator (DIOMIS, 2007).

2.3.1 Resources

The key infrastructure components include the following components (see Figure 7):

- *Transshipment tracks*; are used to park trains while they are being loaded and unloaded.

- *Side tracks* (also called *holding tracks, waiting tracks* and *parking tracks*); are used for generic purposes. These include train inspection, snow removal, parking of wagons, and in rare cases shunting and maintenance. A group of side tracks is called a *side yard*.
- *Buffer storage* (also called *parking lanes*); serve as parking places for trucks whose cargo is directly transshipped, and for keeping units that need to be placed next to the track for a short period of time.
- *Storage yard*; is a large, flat, often open space, used for short-term (usually 24 hours) storage of load units.
- *Terminal gate*; where incoming vehicles are received and registered, and units are inspected and logged into the terminal system. It is often located near the terminal office.
- *Terminal office*; The staff in the terminal office perform various functions, including registration of load units in the system, management of customer bookings, as well as general communications with the IM, rail operators, hauliers and/or shippers.
- *Driving lanes* (or open space) for truck and equipment circulation

(Bontekoning, 2006, p. 49; Boysen et.al., 2012b, p. 316; Rizzoli, Fornara, & Gambardella, 2002, p. 2)

Furthermore, a relevant feature is the infrastructure connecting the terminal to the rail network. Nevertheless, it is generally not considered as part of the terminal, since it is commonly owned and/or operated by external actors (Bärthel et al., 2011, p. 118).

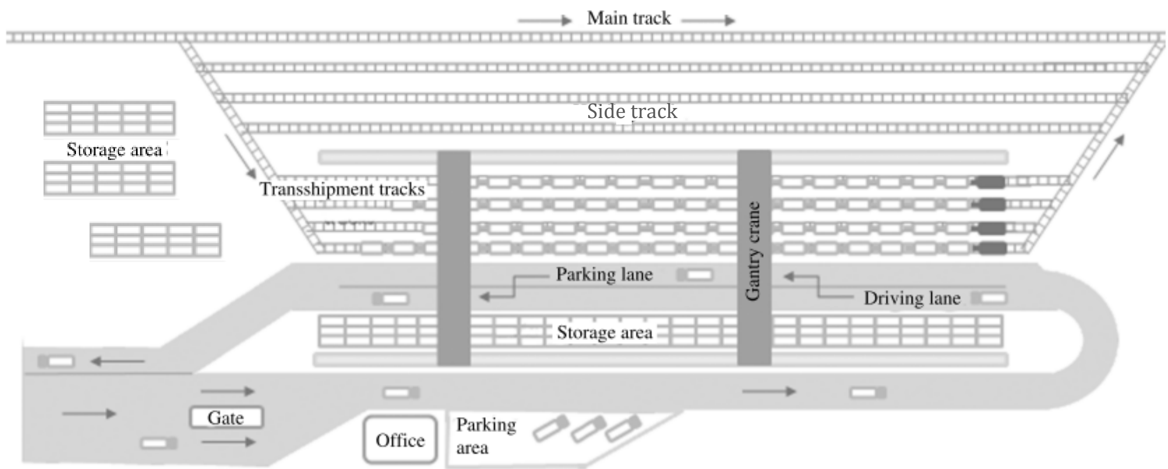


Figure 7: Top view of a road-rail intermodal terminal
(N. Boysen et al., 2012a, p. 316)

Tracks and lanes are normally parallel to each other and might appear in any order and combination, including mixed with one another. Some terminals have additional features such as warehouses for long-term storage of load units, resources for wagon and locomotive maintenance, as well as infrastructure and equipment for complimentary logistic services (DIOMIS, 2007). Figure 7 depicts the top view of a generic intermodal terminal with two gantry cranes, four transshipment tracks and three holding tracks.

Regarding the superstructure, which involves the transshipment equipment, Woxenius (2007, p. 35) states that transshipment methodologies have remained similar throughout the last 40 years despite the development of a large number of new equipment technologies. The most common equipment in use today involves vertical transshipment, including reach stackers, fork lifts and gantry cranes, which can be seen in Figure 8.

Figure 8: Transshipment equipment: gantry crane (left), reach stacker (middle) and a forklift (right).

Gantry cranes (also called *overhead cranes*, or *rail-mounted cranes*) are utilized in medium to large terminals, and can span from two to four rail tracks, plus buffer and parking lanes. Gantry cranes are classified as fixed equipment since they are constrained to travel along the top of the tracks via rails or rubber tires, which limits their flexibility but gives them the advantage of using less terminal space. The most commonly used gantry cranes are electrically powered and mounted in rails. Smaller terminals normally have mobile cranes such as *reach stackers* (also called *side lifts*), *fork lifts* and *straddle carriers*, which are more flexible but have the disadvantage of needing more space for movement. Fork lifts and straddle carriers are used mostly in maritime terminals, and not so commonly in road-rail terminals. Due to their flexibility they are often used as complement to gantry cranes in bigger terminala. Nevertheless, mobile cranes might present disadvantages in big terminals with high volumes and congested layouts (Ballis & Golias, 2002; Vrenken et al., 2005).

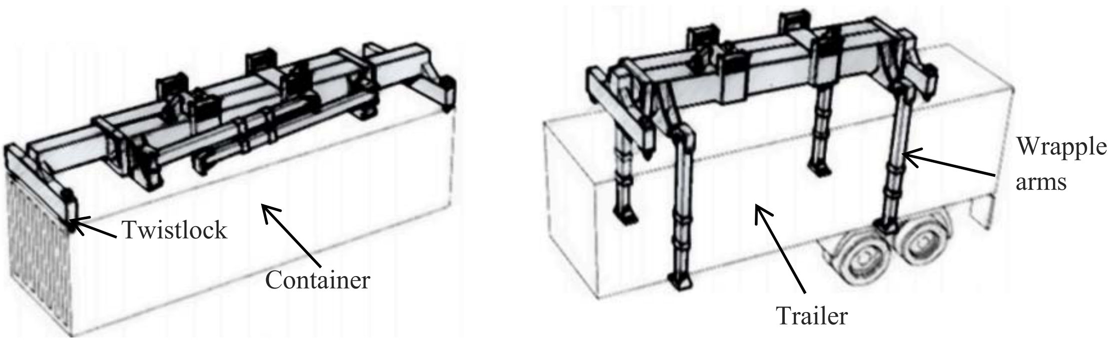


Figure 9: A spreader for container loading with twistlock attachments, equipped with grapple arms for swap body and trailer loading.
(ELME in Lowe, 2005, Chapter 13)

The *spreader* is the device that allows for transshipment equipment to lift containers. These are lifted from the top through special opening in the corners called *open-corner fittings* (also called corner mounts), and they are fixed to the spreader by standardized *twistlocks*. Since containers have different lengths, spreaders are able to adjust to the length of each container. Normally, the corner fittings are located only in the corners of the containers, but some containers have them also in the middle to allow for more flexibility. Comparatively, trailers and swap bodies lifted from their *gripping arm lifts* with *wrapple arms*. Some load units might also have *lifting pockets* in the bottom to be handled with forklifts when empty

(Bromma, n.d.; CargoNet, 2012; Lowe, 2005 Ch. 13). Since each load unit requires a different type of handling, transshipment equipment might be either specialized for one unit, or flexible and able to carry several unit types, such as in Figure 9.

2.3.2 Activities

When a train arrives, is directed towards a side yard where the electric network locomotive is uncoupled and substituted by a shunting locomotive. If a transshipment track is available, it is directly shunted out from the side yard and into the transshipment track. Else, the train is held at the side yard until it is assigned to one (Bontekoning, 2006).

The shunting operations are normally the responsibility of the train operator, who will either do it itself or appoint the service from a local company. Additionally, it is possible that this operation is performed in a separate shunting yard, i.e. without using the terminal's tracks (DIOMIS, 2007, p. 15).

The arrival procedure changes if the network locomotive is a diesel locomotive, or if the terminal´s tracks have overhead contact wires on one end and the access to the terminal is electrified. In this case, there is no need to change the locomotive and the train might enter the transshipment yard as long as there is at least one empty track available. If the transshipment tracks are full, the train must also wait at the side yard (Bontekoning, 2006). When trains cannot enter in reverse, the network locomotive must be detached at the terminal´s entrance and leave, while a shunting locomotive carries the wagon cut into the transshipment tracks (WSP, 2014).

According to Boysen et. al. (2012a), the area of the yard is divided in a grid formed by vertical and horizontal sub-division, which allows for identifying units within a yard. The length of the vertical divisions corresponds to the measurements of a wagon or a load unit. Incoming trains are directed to a horizontal position, which corresponds to a transshipment track, and a vertical position, which indicates which part of the track they shall occupy. The assignment of trains to a vertical position is not very relevant as there are few degrees of freedom, while the horizontal position is of more importance since it affects the reach of cranes. According to the authors, trains are normally assigned a position on a first-come, first-served basis.

Once the train is parked in the terminal, the train starts to be unloaded by cranes. Unit transshipment may be carried out as a direct move (also called direct transshipment), in which the units is unloaded from the train and directly loaded to a truck, or as a split move (also called indirect transshipment), which is done via the storage area. Direct moves are more efficient for the terminal, but they require that the road vehicle is present at the same time as the train. If the road vehicle has not arrived, the unit is located close to the railcar in the storage area (N. Boysen et al., 2012a).

Loading operations often start once unloading operations have finalized, but combined operations are also possible. As with unloading, transshipment may be perform in direct or split moves. According to Ballis & Golias (2002), when the number of trains arriving and already located in the terminal surpasses the number of transshipment tracks, trains in the transshipment tracks must be unloaded quickly and the empty wagons switched to the side tracks (an action termed ***to clear the train***), in order to allow for the unloading of the other incoming trains. This requires for the cleared train to be completely unloaded before being set aside, so that the units are ready to be picked up by the road vehicles.

Ballis & Golias (2002) mention that in industry, the utilization of transshipment tracks is classified as '***static capacity***' or '***dynamic capacity***'. In a static capacity situation, one train set is served per day, so to say, one incoming and one outgoing train. In a dynamic capacity situation the terminal serves more than one train set per day, which causes the need to *clear* one or more trains.

Once a train is fully loaded it is prepared for departure, which is usually the responsibility of the rail operator. For departure, the same conditions for shunting and changing of locomotives as in the arrival are applied. Additionally, brake tests and final inspections are performed. In the end, the train waits for its departure time (Bontekoning, 2006).

Lifting operations in intermodal terminals are characterized by peak times in specific parts of the day, as a function of train arrival and departure schedules. Therefore, some equipment is only used during peak time, while operators might perform different tasks depending of the time of day (Flodén, 2007, p. 143). Such peak times are formed because the dominant train production scheme in Europe consists on shuttle trains that perform *night-leaps* between large origin and destination terminals (Behrends, 2006, p. 3), due to the fact that passenger trains are prioritized during the day (N. Boysen et al., 2012a). In this system, trains are loaded in the evening at the departure terminal, and arrive early the next morning to their destination terminal (Sommar & Woxenius, 2007, p. 164). Train schedules depend on the timetable, market demands, and opening times of the terminal (Ballis & Golias, 2002).

Ballis and Golias (2002) define three phases of crane utilization along the day, see Figure 10. The first phase starts in the morning after arrival of a train. At this time, several road vehicles may be already waiting and many others are arriving for unit pick-up, so almost all crane movements correspond to direct transshipments. In the second phase, indirect and direct transshipments are combined. Trucks continue arriving and picking up units, which might be on the train of previously stored, while other units are placed in buffer storage before they are picked up. In the third phase, all remaining units are unloaded from wagon to storage, in order to allow for shunting operations. In the fourth phase, all incoming trucks are loaded from storage.

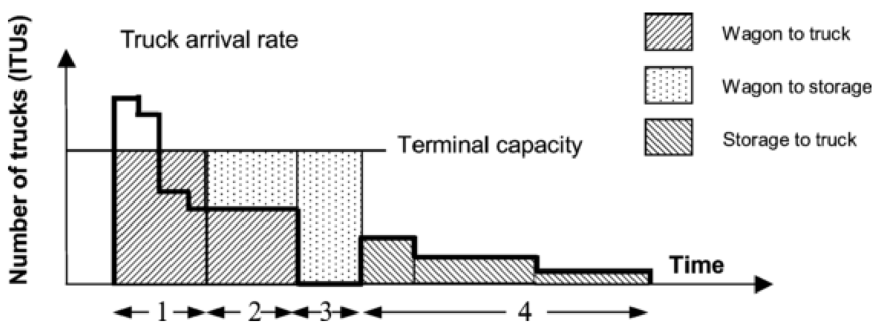


Figure 10: Phases of intermodal crane movements
(Ballis & Golias, 2002)

Regarding the storage yard, Ballis and Golias (2002) state that the space requirements for the storage yard depend upon the volume of load units, the average unit dwell time (or total time an average unit is in the terminal) and characteristics of the units. Containers are stackable, while semi-trailers and most European-type swap bodies are not. Therefore, almost all road-rail terminals present a stacking height in their yards of one or two rows.

2.4 Capacity

2.4.1 Principles of capacity in OM and OR

An understanding of capacity is crucial for managing operations, since it is a valuable concept in decision making processes such as whether to take new business commitments, and whether or not to increase capacity levels to meet changing customer demand, which might require large investments. However, according to Slack et. al.: "for such an important topic, there is surprisingly little standardization in how capacity is measured. Not only is a reasonably accurate measure of capacity needed for operations planning and control, it is also needed to decide whether it is worth investing in extra physical capacity" (2013, p. 330).

In operations management, the capacity of a system from a strategic perspective is defined as "the potential output of a system that may be produced in a specific time, determined by the size, scale and configuration of the system´s transportation inputs" (Brown et al., 2013, p. 398). From an operational perspective, it is defined as "the maximum level of value-added activity over a period of time that a process can achieve under normal operating conditions" (Slack et al., 2013, p. 324).

Many classifications of capacity are encountered in literature. For example, Slack et.al. (2013, p. 324) classify capacity as theoretical capacity, design capacity, and effective capacity, while Brown et.al. (2013, p. 398) only mention design and effective capacities. On the other hand, Jonsson (2008, p. 309) makes the classification of theoretical maximum capacity, nominal capacity and net capacity; and Damij & Damij (2014, p. 79) mention only theoretical capacity. **Annex 6** provides some definitions of the types of capacity as provided by different authors, and the next lines are aimed to reconcile and compare them.

The definitions found for theoretical, nominal and design capacity, differ on whether they explicitly exclude/include actual operating times. In general, ***theoretical capacity*** is related to the maximum capacity assuming ideal conditions and non-stop operations (i.e. operating round-the-clock), without interruptions of any kind (see Damij & Damij, 2014, p. 79; Jonsson, 2008, p. 309; Slack et al., 2013, p. 324). Meanwhile, ***nominal capacity*** and ***design capacity*** seem to be equivalent terms, referring to the maximum capacity assuming ideal conditions, but taking into account actual operating times (see Brown et al., 2013; Jonsson, 2008; Slack et al., 2013).

On the other hand, definitions for practical and effective capacity change according to which are the time losses included. Such times losses can be classified as scheduled (unavoidable) time losses (such as setup times, maintenance, lunch breaks, etc.) which are necessary for the process, and unscheduled (avoidable) times losses due to avoidable events (such as quality problems, machine breakdowns, employee absenteeism).

Both Brown et.al. and Slack et.al. agree in that *effective capacity* is the maximum output once time losses have been included, but while the first include both scheduled and unscheduled time losses, the second explicitly exclude unscheduled time losses. Jonsson´s (2008, p. 309) definition of *net capacity* seems to correspond to effective capacity: "the capacity estimated to be at the company´s disposal to carry out planned production activities", and includes both scheduled and unscheduled time losses, since it includes "indirect unavoidable downtime and non-planned manufacturing such as repairs, urgent orders".

Meanwhile, practical capacity is defined by as Mc.Nair & Richard Vangermeersch (1998, p. 28) as "the level of output generally attainable by a process; it is theoretical capacity adjusted downwards for unavoidable nonproductive time, such as setups, maintenance and breakdowns". It is relevant to point out that they classify breakdowns as unavoidable time losses.

As there is no clear distinction between the terms ***practical capacity***, ***effective capacity*** and *net capacity*, they may be regarded as equal terms, and may be defined according to Slack´s (2013, p. 324) description: "the maximum output that can be achieved once time losses are accounted for". In which the time losses included may change from process to process, according to what is more useful in each case. Specifically in the context of transport systems, practical capacity is defined as "the number of operations that can be accommodated with no more than a given amount of delay, usually expressed in terms of maximum acceptable average delay" (U.S. Congress, 1984, p. 47). This explanation seems more appropriate for the purpose of this thesis, as corresponds better to the ones for railway capacity (Section 2.4.2).

The ***actual output*** of an operation refers to the amount actually produced during an established period of time (Slack et al., 2013, p. 329). According to the definitions given above, actual output is equal to effective capacity in the case that demand is greater than practical capacity (see Figure 11).

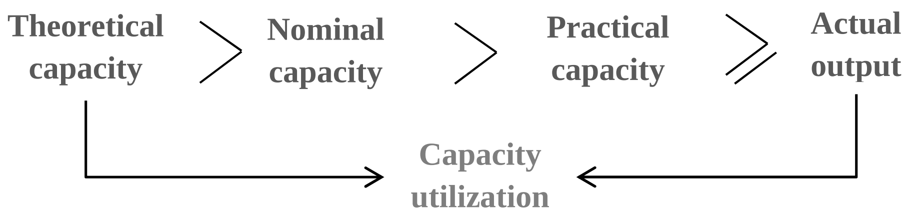


Figure 11: The relation between different types of capacity
(own creation)

The way in which production potential is utilized in an operation can be measured by two factors, namely ***capacity utilization*** and *efficiency*. Capacity utilization refers to the ratio of the actual output to the theoretical capacity, while efficiency is the ratio of actual output to practical capacity (with *practical capacity* as defined above) (Slack et al., 2013, p. 329).

Capacity determination in OM

In general, achieving an unambiguous determination of a process´s capacity is problematical due to the complexity of most processes. Only in the case of processes which are highly standardized, consuming a constant set of inputs and outputs, it is possible to easily and accurately calculate capacity (Slack et al., 2013, p. 328).

When planning the capacity for a process, it is necessary to match capacity to demand in order to avoid a very high or very low capacity utilization. A process with over-capacity (i.e. low capacity utilization) is not cost-efficient, since fixed costs and capital investment are being covered by a lower sales income. Meanwhile, a process with under-capacity (whose demand is higher than capacity) might be forced to reject certain orders (and therefore to lose income), and/or accept higher inventory levels and throughput times (thus higher inventory costs and lead-times) (Slack et al., 2013, p. 328).

An important challenge when planning capacity is the ability to meet variable demand. The more variation there is in the demand and/or capacity of a process, the lower its effective capacity. Furthermore, the larger the variation in arrival times, the longer throughput times and lower capacity utilization. Therefore, for any process:

"The greater the variability, the more extra capacity will need to be provided to compensate for the reduced utilization of available capacity"

(Slack et al., 2013, p. 351)

According to Slack et.al. (2013, p. 351), processes with higher variability intend to have a higher *base capacity* (or *static capacity*), and adjust *variable capacity* units (e.g. staff and opening hours) to demand levels. This raises investment costs, but minimizes throughput times and waiting times.

There are three main methods to respond to demand variability, which can be utilized alone or in combination with each other: the level capacity plan, the chase demand plan, and the manage demand plan (i.e. yield management). The first plan consists on maintaining stable production levels throughout a time period. This is based on building up inventory in periods with over-capacity in order to satisfy demand in periods when demand overpasses capacity. Such method is useful only for non-perishable products which can be easily stored. The second plans consists on adjusting variable capacity units to actual demand by outsourcing operations, varying the size of the workforce, adjusting working hours, etc. It is mostly used by operations whose output can´t be stored, such as in the service industry; it is more complicated than the level capacity plan, and brings higher costs. The third plan consists on changing demand through price, which is a usual approach in the service industry. It consists on stimulating demand in off-peak times with lower prices, and to discourage it during the peak times (Greasley, 1999; Slack et al., 2013, pp. 338–340).

Demand management, in turn, forms part of ***yield management***, which is a group of methods utilized to maximize the profit of certain processes that have these characteristics:

- relatively fixed and high capacity
- clearly segmented market
- non-storable outputs
- sales are made in advance
- low marginal sales cost

(Slack et al., 2013, p. 340)

Capacity planning methods in OM

Capacity planning consists in determining the most cost-effective and feasible method to meet capacity requirements. In other words, to assess each capacity plan according to the nature of the expected demand and chose the most appropriate. This is carried out by using one of three methods to predict the consequences of a determined capacity plan, i.e. cumulative representations, queuing theory, and simulation modelling (Greasley, 1999, p. 126).

Cumulative representations consist on graphing the expected aggregated demand for a certain time period, together with the planned capacity levels. In this sense, the process is able to satisfy the expected demand if the total under-capacity is greater than total over-capacity (area B). This method is only applicable for processes where output can be stored whenever supply capacity exceeds demand (Greasley, 1999, p. 126).

Queuing theory methods are best suited for service operations where output can´t be stored and waiting times appear. In ideal situations, there would be no waiting times since customers arrive at exact, pre-booked times and processing times are constant. Nevertheless, queues arise due to variability, i.e. when arrivals rates and service times are variable, which is true for almost all processes. Queuing theory allows analyzing the best trade-off between waiting times and the amount of capacity, in order to avoid under- and over-capacity. The aim is to determine a level of capacity that provides acceptable waiting times. Queuing theory may be utilized through analytical or simulation approaches; for a more thorough explanation of the principles of queuing theory see (Greasley, 1999, p. 128).

The main actors in queuing systems are the customers and the servers, where a customer is any entity which enters the system (e.g. an ILU in the case of intermodal transport), and a server is any resource which performs an action on the customer (e.g. a machine). Queuing systems are normally represented with Kendall´s notation A/B/m/b, where:

$$A = time\ between\ arrivals, i.e. inter - arrival\ times$$

$$B = time\ to\ process\ each\ customer$$

$$m = number\ of\ servers\ or\ service\ facilities$$

$$b = maximum\ number\ of\ items\ allowed\ in\ the\ system$$

(Slack et al., 2013, p. 361)

With the constraint that only one unit can be served in each served at a time. When the factor b is omitted, it means that the number of units allowed in the system is unlimited, which implies that there is no maximum limit to the queue. Inter-arrival times can be deterministic (denoted as D), such as when following fixed schedules, or probabilistic (Taha, 2007, p. 551).

The most common probability distributions for inter-arrival and processing times are the Poisson distribution and the general distribution (such as the normal distribution). While the general (depicted as G) distributions are widely known, the Poisson probability distribution function (depicted as M) is used when interarrival times are random. In this setting, the estimated time for the arrival of the next customer cannot be estimated given the time of the previous arrival, nevertheless, the number of arrivals in a certain time period is known. In a In this system, time between arrivals is represented with an exponential distribution (Taha, 2007, p. 476).

Simulation methods consist on using software to imitate the operations of a system, in order to observe the system´s behavior under different "what if" scenarios. They aim to reflect the complexity of real-world systems arising from randomness and process inter-dependence (Greasley, 1999, p. 128), and to estimate the performance measures of a system (Taha, 2007, p. 605).

Simulation models are widely based on the Monte Carlo method, since they have an inherent statistical nature. This method allows for replacing empirical data with simulated input, which is generated using probability distributions which fit to the empirical data. The utility of this, is to be able to have large amounts of input data for the simulation, without having to actually retrieve this data from real life.

Taha (2007) classifies current simulation models as continuous or discrete. The first group is based on continuous change, and normally uses differential equations. In the second, also called ***discrete event simulation*** models, each change in the system occurs in a particular instant in time, as a result of an event which modifies the system. These models are used primarily to analyze waiting times according to arrival and departure patterns, and are linked to queuing theory (Taha, 2007, p. 605). A more recent approach is agent-based simulation, in which autonomous agents act independently based on individual objectives (i.e. without a centralized control), and whose interactions shape the properties of the overall system (Macal & North, 2010).

2.4.2 Railway capacity

In the railway context there seems to be no widely accepted definition of capacity, since it is not easy to define or quantify due to its large amount of inter-related sub-systems, the complex track layouts, as well as the large amount of terminology involved (Burdett & Kozan, 2006, p. 617; Mattsson, 2007, p. 132).

Annex 7 presents the definitions of rail system and/or line capacity found in literature, which are classified according to their equivalence to the different general capacity definitions given above. Some authors have been repeated since they give first their general impression of the definition in literature, and then provide one for the particular paper, which are both valuable.

The International Union of Railways establishes that there is no ultimate definition of capacity, since the capacity of the railway infrastructure depends on the way it is used (UIC, 2004, p. 2). They explain that it is not possible to establish a generally valid definition of theoretical capacity, let alone a method for its calculation (p. 5). Accordingly, the only factor that can be objectively measured is capacity utilization (p. 9).

In spite of the UIC´s declaration, Abril et al (2008, p. 776) give useful descriptions of the types of railway capacity and the way they are calculated, in relation to the UIC´s capacity utilization analysis method. A line section´s ***theoretical capacity*** is defined as the maximum amount of traffic that can safely circulate through that section at a certain time interval under ideal conditions. The word *safely* refers to minimum headway times, while *ideal conditions* refers to the assumption that all traffic is homogeneous (i.e. with same train type, weight, length, and speed), that there are no time delays, no interference between vehicles. It also involves the assumption of minimum running times, and minimum blocking distances. It is important not to confuse the concept of line section with that of a block section, as the second is only present in fixed block signaling.

On the other hand, a line section´s ***practical capacity*** is defined as the maximum amount of traffic that can safely circulate through that section at a certain time interval with a certain level of reliability, given actual timetables, traffic mix, and running times (Abril et al., 2008, p.776). Landex & Kaas (2006, p. 3) provide a similar definition, but using the term *punctuality*.

A common approach for determining the overall capacity of a line is to set line capacity as the capacity of the most constrained line section, i.e. the bottleneck approach. Nevertheless, the influence of interfering lines must also be taken into account; a line cannot be regarded as an independent sub-system, due to the presence of crossings and junctions with other lines (Abril et al., 2008, p. 778).

According to Poole (1962), the capacity of a line depends on the capacity of its bottleneck, which can be either a line section or a yard. He mentions that in the case that the capacity of a line section is increased, the bottleneck may be shifted to another line section or to a yard.

The timetable and capacity

After the path applications from the rail operators are received, the IM is in charge of developing the timetable according to infrastructure, safety, and traffic factors. The first step consists on modelling the actual infrastructure, i.e. the layout of tracks, lines and nodes, average transit times, signaling system, etc. either with a macroscopic or microscopic approach. In this step, rail modeling software such as RailSys is utilized (Marinov et al., 2013, pp. 67–69). See **Annex 5** for a description of the functioning of signaling systems.

Next, the line is divided into line sections, and the ***running time*** (the time required for a train to transverse a certain line section) is calculated for each line section according to train configurations (length, weight, traction) and technical characteristics of the tracks (i.e. slopes, gradients, curves) (Marinov et al., 2013, pp. 67–69). According to Abril et.al. (2008, p. 792), running times are a function of the section length and train speed.

Afterwards, the required headway is determined. The ***headway*** is defined as the time distance between two consecutive trains in one same line and direction, measured from the front of one train to the front of the next train. Next, the scheduled time is calculated for each train, which corresponds to the time between arrival and departure. The timetable is created by fixing the paths of the trains with higher priority according to their schedules arrival and departure times. The trains with lower priority are added subsequently, with the minimum separation established by the headway (Marinov et al., 2013, pp. 67–69).

The ***minimum headway*** is the shortest possible distance that can be allowed by the signaling and safety systems (Landex & Kaas, 2006, p. 2). In the US, for example, it must be long enough to allow for a train running at full speed to come to a complete stop (TRB, 2003, p. 3).

Furthermore, IMs also include buffer times and running time supplements according to local traffic conditions, and therefore the headway tends to be larger than the minimum headway. Landex & Kaas define ***buffer time*** (also called *slack time*) as the "difference between the actual headway and the minimum allowed headway", and ***running time supplement*** as "the difference between planned running times and the minimum running time" (2006, p. 2). Mattsson (2007, p. 131), define *slack time* or *buffer time* as the times inserted in the timetable headway with the purpose of "absorbing" the effect of unscheduled train delays, as well as to account for the fact that trains do not always have an actual running time equal to their minimum running time (Mattsson, 2007, p. 131). Meanwhile, Abril et.al. (2008, p. 792) state that the headway is the sum of the running times, braking times and release times (time to release a block section after it has been occupied), plus a constant operating time. The slack time can also be called *scheduled delay* (Mattsson, 2007), so that total delay equals the amount of scheduled and unscheduled delay. A train´s unscheduled delay depends on the stochastic variations present in day to day operation, and is a function of the trains´ scheduled delay, as well as other trains' delays (schedules and unscheduled).

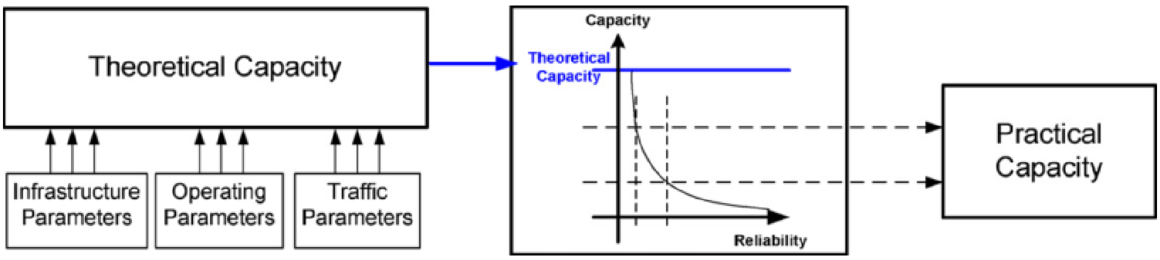


Figure 12: The relationship between theoretical capacity, practical capacity, and reliability
(Abril et al., 2008, p. 805)

IMs must define a headway that provides an acceptable level of reliability, so to say, that does not exceed a certain amount of unscheduled delay. As the slack time increases, the system becomes more stable or reliable, but practical capacity decreases (Arcot, 2007, p. 6) (see Figure 12). A timetable is said to be robust or stable when it is more reliable, and when the effect of a single train's delay upon other paths is low. A congested timetable is robust depending on the way the headway was determined: the larger the headway, the more reliable the timetable is, but the less amount of paths that can be accommodated (Landex & Kaas, 2006, p. 2; Marinov et al., 2013, pp. 67–69).

Additionally, in the case of lines share with passenger traffic, the headway must include station dwell times. In these situations line capacity is mostly constrained by the station with longest dwell time. This is because passenger lines normally contain transit stations in which trains stop along the main line. In sections next to passing loops, meeting loops, and flat (*at-grade*) junctions, the headway must include an operating time for the switches to move and for both trains to come to a stop (depending on siding length) (TRB, 2003, p. 4).

Components of railway capacity

According to the UIC (2004), railway capacity is based on the interrelationship between the following factors:

- Average speed
- Number of trains
- Heterogeneity: i.e. train mix, the sequence between train categories, speed differences, and the presence of opposing directions in single track lines.
- Stability: i.e. the stability of the timetable, which influences train reliability.

In turn, the organization mentions that the parameters that heavily constrain capacity are the priority given to trains, the configuration of the timetable, the process for capacity allocation, the time allowances (buffers inserted in the timetable), environmental regulations (e.g. the prohibitions to operate at certain times due to noise), safety measures, and technical design of tracks and lines (UIC, 2004).

An additional constraint (in the case of electrified lines), are the power supply limitations of the catenary, which determine the number of trains running at the same time (TRB, 2003, p. 9).

Lines with more than one track have more capacity than single-track lines. Furthermore, the amount of train meets, and the number and length of passing sidings in a single track line are also important capacity determinants. The more passing and meeting sidings a line has, the more trains with different characteristics that can run on a line. Nevertheless, more passing and meeting operations increase scheduled and unscheduled delays. Likewise, as passing sidings become larger, the trains have more chance to maintain their running speed in passing and meeting operations, and therefore the capacity of the line increases (TRB, 2003, p. 82).

How signaling and safety systems affect railway capacity

In general, establishing a safe separation between trains, results in reduced capacity utilization. This is because each train takes up more capacity than the actual track space it physically occupies due to the headway and/or the block distances. In a fixed block signaling system, the longer the block sections in a line and the larger the headway, the lower the amount of trains that can the a line in a certain time interval. Furthermore, when a train is passing a boundary between two blocks, both block sections are technically occupied (TRB, 2013b, p. 11).

The effect on capacity is further amplified when taking into account traffic mix; to achieve the greatest safety possible, signaling systems are often established according to the "worst case scenario". The minimum distance between signals and of block sections starting points is determined based on the biggest allowed train length, loads and speeds along a line. These two last factors significantly influence braking distances. Likewise, maximum allowable speeds, loads and train lengths are established according to distance between signals, nature of signals and block length (TRB, 2013b, p. 11).

According to the TRB, vehicle capacity in a railway system depends on the minimum possible headway. In turn, the headway is calculated according to the parameters of the signaling system, the station dwell time (in case of passenger transit stations), and the interactions with other vehicles (traffic mix) (TRB, 2013a, p. 6).

A thorough explanation of the way different signaling systems affect capacity has been developed by the US Transportation Research Board (TRB, 2013b) and Abril et al. (2008, p.

777), while Goverdeet.al. (2013) explore the implications on line capacity of the European Train Control System (ETCS), as well as of several types of ATP Systems.

Capacity analysis methods

The approaches that have been developed to evaluate capacity in the railway sector can be divided in analytical, modelling and optimization methods (Abril et al., 2008, p. 79).

Analytical methods constitute the simplest approach; they are based on mathematical formulas and are usually applied for the calculation of theoretical capacity. To obtain practical capacity, the result is in turn adjusted downwards, or obtained directly with the original formula using adjustment factors. These methods provide a good first approximation of capacity and are useful to identify major capacity constraints, but the quality of their result is vastly dependent on the input parameters. Therefore, they are best used as a preparatory task for simulation methods (Abril et al., 2008, p. 79). According to (Clarke, 1995), numerical models can be divided into empirical and analytical functions. Empirical functions contain several constants that should be determined by fitting empirical data; therefore, the model should be calibrated to each specific case. Analytical functions, on the other hand, can be generically applied without need for model calibration.

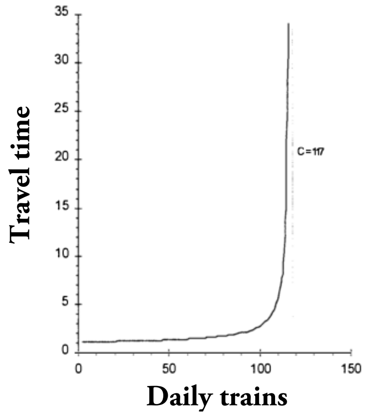
Optimization methods provide more useful results for planning and decision-making, as they give an optimal solution within given constraints. They can be based on numerical models or dedicated software (Abril et al., 2008, p. 79).

Simulation methods aim to imitate, mostly through the use of a computer model, the behavior of a system in the real world and the dynamic interrelations between the system components, in order to explore the system under different scenarios. Simulation methods are often hybrid and may be used along with optimization methods, such as when validating a timetable developed with optimization methods (Abril et al., 2008, p. 79).

2.4.3 Existing methods for determination of railway line capacity

Analytical tools

A great number of analytical tools have been developed for calculating rail line capacity, time and delay. Clarke (1995) examines several analytical tools for the calculation of railway line capacity and travel time, which is linked to running times and delay. Within the empirical analytical formulae, he mentions the formula by Mosher, 1963, used to calculate travel time according to capacity and traffic volume, which yields a travel time-volume graph similar to Graph 1. A similar formula is the *Polynomial Delay Function*, which, according to the author, is widely used. It links the variables of travel time at zero volume, travel time with a certain amount of congestion, and the maximum capacity, along with empirical constants. Both formulas need to be calibrated with empirical data in order to fit to the real system, which is normally carried out using simulation models.



Graph 1: Graph of travel time vs traffic volume
(Clarke, 1995, p. 27)

Within the purely analytical tools, Clarke (1995) presents the Poole model from 1962 (see Section **2.4.3.3**), the *Petersen Model for Single Track Lines* of 1974, the *Janic model* from 1988, and the *Greenberg model*, also from 1988. Petersen´s model may be used to calculate average travel times in single-track lines with passing and meeting loops, assuming uniform arrivals of trains within a certain class, and equally spaced passing and meeting loops. The Janic and Greenberg models are also used to estimate travel times in single track lines, and are based on queuing theory. Greenberg´s model estimated single-track line delay of slow speed trains in busy periods, in lines where passing and meeting loops are widely spaced between each other. Unlike the Poole model, this does not need the assumption of equally spaced sidings, uniform traffic, and meets of only two trains at a time. The model uses an M/D/∞ queue; where the time between arrivals corresponds to a Poisson process, the service time is equal to the train´s running time on a given line section, and the number of servers is unrestricted. In this sense, the servers correspond to the sections together with passing loops and are placed in serial order (Clarke, 1995). Poole's model is presented in Section **2.4.3.3** as part of the extended review, and Petersen´s model is not further explained since it is regarded as highly similar to Poole´s.

Furthermore, Mattson (2007) presents a review of several other analytical methods exploring the relationship between capacity and delay (both scheduled and unscheduled, which is related to reliability) under different traffic and infrastructure conditions). Additionally, Rudjanakanoknad et. al. (2013) developed several analytical formulae, which used together with time-space diagrams, may be easily applied to calculate capacity of a line containing passenger transit stations between two terminal stations[4]. They provide tools for several scenarios, including: homogenous and non-homogenous traffic in double tracks, as well as in single tracks with our without sidings.

Optimization tools

[4] Passenger stations are divided in transit stations or terminal stations. *Transit stations* are located along a main line, where local trains stop for a very short period to embark and disembark passengers. Passing and meeting operations can be done at these stations. Meanwhile, terminal stations are located at the ends of a line (Marinov et al., 2014).

These are mostly used in the tactical level to produce optimally saturated timetables[5]. Abril et.al. (2008) regard the timetable compression method utilized by the Capacity Utilization Index (CUI, see section 2.4.3.4) and the UIC 406 (see section 2.4.3.5) as an optimization tool, since it may be used to produce an optimally saturated timetable. Nevertheless, it may also be utilized as an analytical tool when applied for the calculation of capacity utilization per se. Other tools for optimizing line planning and transit will not be reviewed, as the number of analyses would be too large for the scope of this thesis.

Simulation tools

General simulation software that has been used within railway transport include Minitab and Arena. Nevertheless, many tools have been developed which are specifically designed for use in railway link modelling. Barber et. al. (2007) provide a comprehensive and comparative review of several simulation software tools used for railway management, including Railsys, Dons, OpenTrack, SiSYFE, Demiurge, etc. They are normally used to plan, model and simulate infrastructure, timetables and operations in links. As they aid in planning, they are not only used for modelling and simulation but also contain optimization tools (Abril et al., 2008). Pouryousef et.al. (2015) mention also the software "RTC, MultiRail, RAILSIM and CMS", while Bärthel et. al. (2011) add to the list the simulation models SIMONE, SIMU, and IS SENA. They indicate that these tools are used to analyze single links, and in general, their input parameters include: "track layout, signaling system, maximum speed, type and characteristics of rolling stock, and timetable" (p. 161). Most include in the modelling passing and meeting loops, junctions, passing passenger stations, and similar components.

RailSys will be the only tool analyzed, since it widely used, and because it is assumed that the mentioned tools have similar properties and principles.

2.4.3.1 Mussone & Calvo´s formula

Mussone & Calvo (2013) describe the simplest analytical method for calculating single line capacity that was found in literature, which consists on the following formula:

$$C = \frac{T}{h + \Delta h_t} \tag{1}$$

(Mussone & Calvo, 2013, p. 13)

Where:

$C = capacity\ in\ trains\ per\ time\ period$
$T = time\ period\ studied$
$h = average\ headway\ of\ trains$
$\Delta h_t = variations\ from\ the\ average\ headway$

This formula is very simplified and may not be so optimal, but it allows for a clear visualization of the role of the headway in line capacity.

[5] They also used for the crew and vehicle scheduling, and selection of routes according to network models, although not directly related to capacity analysis.

2.4.3.2 Scott´s formula

Scott´s formula is another simple, and rather old analytical method for rail line capacity determination, also described as a rule of thumb used within railway engineers and operators (PPIAF, 2015). It is apparently more used in Indian railways (see Indian Railway Employee, 2015). It is applicable in a single line which has passing sidings as a very rough estimate of the amount of trains that can transverse one line per day.

$$Daily\ Line\ Capacity = \frac{1440}{(T_1 + T_2)} \times E \tag{2}$$

(Vanichkobchinda, 2007, p. 9)

Where:

1440 refer to minutes per day

$T_1 = running\ time\ [min]$

$T_2 = block\ operations\ time\ \ [min]$

$E = efficiency\ factor\ (0.7\ to\ 0.9)$

According to Vanichkobchinda, (2007, p. 9), the running time refers to the average time to transverse the section with the longest travel time, the block operation time refers to the signaling and safety headway time to allow for the next train to enter the section, and the efficiency factor accounts for unexpected events and maintenance.

According to the association PPIAF (2015), the formula may be used to calculate capacity in the line for trains travelling in two directions, and line capacity depends on the longest travel time between sidings. In their "toolkit", the formula is written by summing up T_1 and T_2 into a sole factor T, the longest travel and stopping time between passing sidings, which shall correspond to the longest block section. They do not give a definition of the efficiency factor.

It is not mentioned whether Scott´s formula is utilized for theoretical or for effective capacity. Due to the factor E, and the fact that T_1 is calculated for the slowest train, the result of this formula can be deemed as an equivalent of effective capacity, as long as T_2 is related to actual timetables. Likewise, it could represent theoretical capacity in the case that E is equal to 1, T_1 is calculated with a homogeneous traffic mix and minimum running times, and T_2 is equal to the minimum headway. As the value depends majorly upon the travel time on a critical section, this formula follows the bottleneck approach.

Due to the fact that it requires to calculate the travel times in a section, and that T_2 is calculated according to the time that another train can occupy a section, it seems to be more appropriate for lines with fixed block signaling, which is also one of the earliest types of safety systems in the railway.

2.4.3.3 Poole´s formula

Poole (1962) described a similar analytical tool to calculate a line's capacity within his book on railroad costing. It is rather equivalent to the model described above, but contains more detail and possibilities to be adjusted to different scenarios. According to him, capacity depends on the running times between sidings, the number and capacity of sidings in each loop, the running time through switches and sidings, the regularity between train arrivals into the line, and the type of communications and signaling systems. The author describes the formula as one for the calculation of theoretical capacity.

The model is to be applied in a line section between meeting sidings, as has the following assumptions:

- there is an equal number of trains running in each direction
- trains arrive at equally spaced intervals
- trains are prioritized according to direction
- meeting loops are evenly spaced within a line and each siding may contain only one train.
- the delay for each meet is only equal to the scheduled delay,
- only two trains meet at each given line section at a time.

$$Daily\ Line\ Capacity = \frac{1440}{t+\frac{t}{2}+\frac{m}{2}} \tag{3}$$

(Poole, 1962, p. 147)

Where:

$t = time\ to\ clear\ main\ line\ between\ siding\ switches\ at\ full\ speed\ \ [min]$
$m = (scheduled) delay\ for\ each\ meet\ [min]$

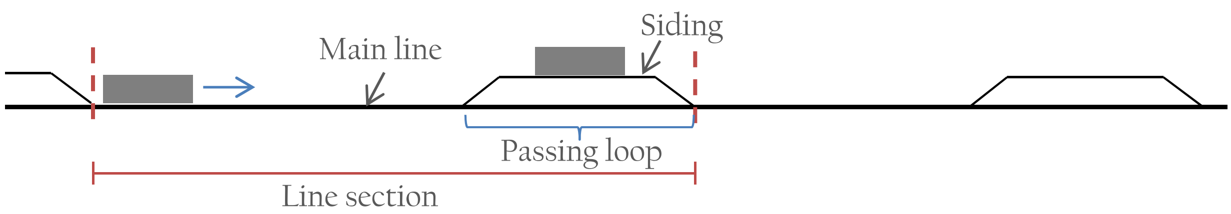


Figure 13: Depiction of a line section with one meeting loop
(own creation)

Parameter m refers to the delay undergone by the train which enters the siding, and is composed by the times for deceleration, entering the siding, running the siding (if it is longer than train length), and leaving the loop into the main line (see Figure 13). It excludes the time spent waiting for the running train to clear the section, which is included in the factor $t/2$. In this formula, $t/2 + m/2$ represent the average delay per meet and per train; even if only one train has to wait, both trains are involved in the operation. It becomes evident in the formula that the larger the (scheduled) delay, the larger the capacity. For the case of centralized traffic control (CTC), in which trains are prioritized according to different factors other than direction, the term $t/2$ is to be substituted by $t/4$.

In the same book, Poole provides formulas to calculate capacity and total (scheduled) delay in the case of trains with different classes, speeds and prioritization rules. He states that the maximum permissible delay is usually the most important factor for determining practical track capacity, rather than the maximum number of trains than can pass a section. Finally, he also provides a formula to evaluate the capacity in double-track lines.

2.4.3.4 Capacity Utilization Index

The Capacity Uilization Index (CUI) is employed in the UK to measure capacity utilization in lines. The following description of the method is based on the report performed by consultancy Sinclair Knight Merz Group (2012) for the British government, which analyzes current capacity constraints in the network.

The capacity utilization index (CUI) is used in the UK to measure the amount of used and available space for a given timetable. As the CUI grows, the amount of delay of trains grows exponentially, and a threshold of 80% is deemed as the limit at which performance becomes unacceptable.

First, the paths of trains running through a section during a one hour interval are compressed, which means that they are placed as close together as the headway allows, but without changing the order and path characteristics (i.e. stops, meeting and crossing operations, and travel times stay constant). After the compression, the remaining amount of time represents the available capacity, while the occupied part of the hour represent the actual capacity used. In Figure 14, after the compression of the paths, 45 minutes of the total hour are occupied, meaning that the CUI equals 75%.

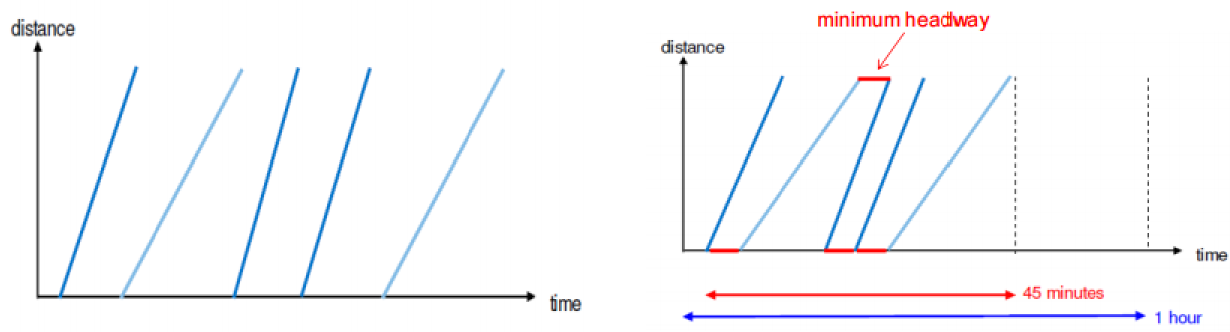


Figure 14: The timetable compression in the CUI
(Sinclair Knight Merz Group, 2012, p. 23)

2.4.3.5 UIC 406 (timetable compression method)

The International Union of Railway´s capacity analysis method (UIC, 2004), developed in 2004 and denominated UIC 406, is deemed as an appropriate way to analyze capacity unitization in railway lines. This method has related predecessors which are based on analytical calculations, such as the UIC 405-1 from 1983 and UIC 405 OR from 1996 (Abril et al., 2008). According to (Landex & Kaas, 2006), it has been applied in at least five studies within the EU rail systems since its creation, in addition to the works by (Abril et al., 2008; Goverde et al., 2013; Lai, Huang, & Chu, 2014; Mussone & Calvo, 2013; Pouryousef et al., 2015). According to Sameni et.al. (2005, p. 3), the UIC method is widely used in continental Europe.

For a thorough explanation of the method see (UIC, 2004), (Khadem Sameni et al., 2005), and (Landex & Kaas, 2006); as well as (Lindner, 2011), who provides clarification and criticism on the method.

The UIC recommends using timetable compression for the calculation of capacity utilization, which is the same principle as utilized by the CUI. The difference between the two methods is that the UIC defines a specific scope:

- Time delimitation: the time interval to study must be at least two hours, during the peak period of a representative day.
- Physical delimitation: the study must be situated in the most constrained section of a line. In single tracks, the measuring points should be situation at passing and meeting loops.

Furthermore, the compression of paths must exclude buffer times, and only take into account minimum safe headway and infrastructure occupation times. Nevertheless, the structure of paths, meeting and crossing operations (overtakings), stops, and speed must remain constant. The method may be done by graphical analysis or analytical calculations.

The capacity utilization equals to the ratio of total occupation time relative to the size of the time interval studied:

$$K = \frac{k}{U} \cdot 100 \quad (4)$$

(UIC, 2004, p. 406)

Where:

$K = capacity\ consumption\ [\%]$
$U = time\ interval\ chosen\ [min]$
$k = total\ time\ consumed\ [min]$

In turn, the total time consumed is equal to the sum of several parameters:

$$k = A + B + C + D \quad (5)$$

(UIC, 2004, p. 406)

Where:

$A = Infrastructure\ occupation\ time.$ Is the total time that the line section is occupied by train paths after the compression. It is formed by the sum of the actual physical occupation time of each block section, plus the time needed to “prepare” the block section before train arrival, and the time needed to “clear” the section and release the route after the train leaves the section.

$B = Buffer\ time.$ Is the time added to the minimum running distance of a block section (related to minimum safe headway, also called minimum theoretical headway), in order to help eliminate or lower the effect of the delay of one train (i.e. running time larger than minimum running time) on subsequent train paths.

$C = Supplement\ for\ single - track.$ Refers to the crossing buffer time, or the time between train paths with opposite directions, measured in sections with meeting loops.

$D = Supplement\ for\ maintenance.$ Covers the possibility of lower speed due to maintenance works.

The UIC establishes a threshold for capacity utilization of mix-traffic lines of 75% in peak hours, and an overall of 60% during the day. Nevertheless, they mention that percentage can be higher in the case of lines with rather heterogeneous traffic, and it might, in theory, reach the level of 100% in cases where buffer times are sufficiently large to account for a robust timetable.

If the index is a number below these levels, additional train paths are inserted in the leftover space in order to calculate available capacity. If no additional paths can be inserted, the leftover capacity is considered as lost capacity. Otherwise, the compression process is repeated iteratively until the threshold is reached, as shown in Figure 15.

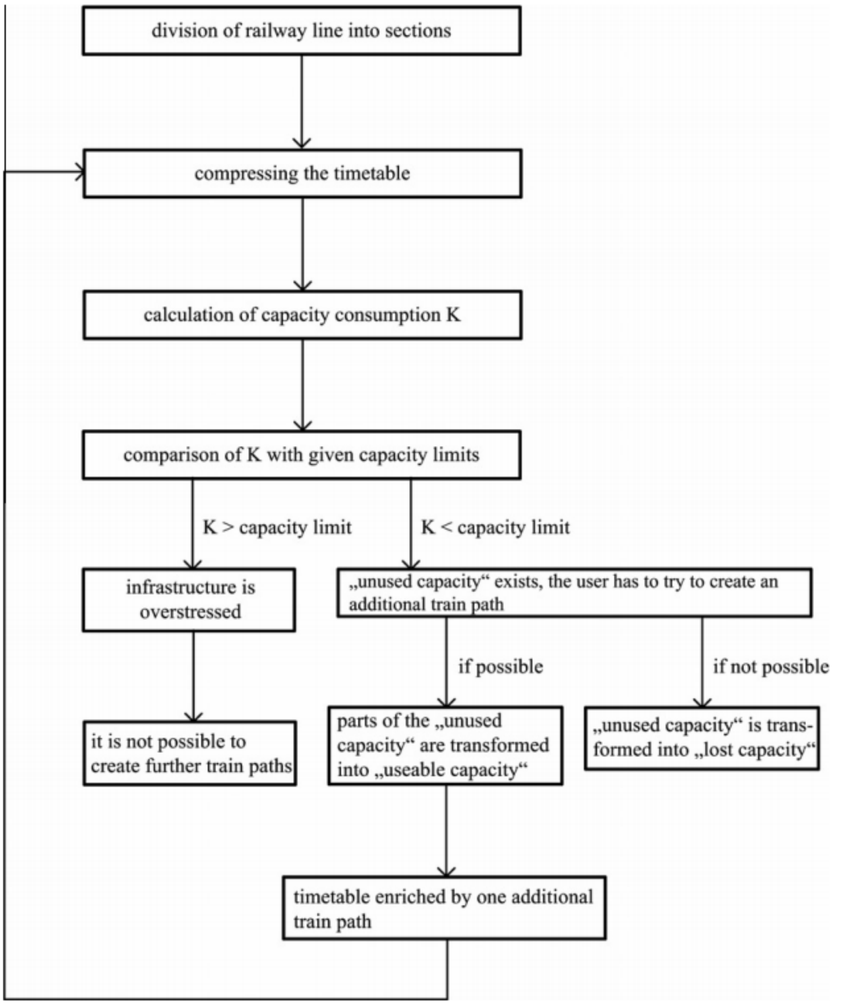


Figure 15: Steps for the UIC Capacity evaluation process
(Lindner, 2009 as cited by Lindner, 2011)

2.4.3.6 RailSys

This software tool was developed by the University of Hannover and RMCon, Rail Management Consultants, and has been applied fruitfully in several projects within Europe and Australia, among other countries. It consists of four modules, namely the infrastructure manager (STED), timetable manager, simulation manager, and the evaluation manager (Barber et al., 2007).

The infrastructure manager allows for a microscopic simulation of the infrastructure, including the tracks, passing loops, switches, signals, (passing) stations, and speed indicators. It also contains technical characteristics of the tracks. Barber et. al. (2007) describe that the infrastructure is represented as a succession of nodes and links, where the links are the tracks and the nodes are either junctions, signal locations, or other connecting points.

The timetable manager is used to construct the timetable, and can determine running track occupation times according to the infrastructure and signaling systems. It also helps detect and solve conflicts between train paths within a network, testing different train configurations and routes in order to achieve a conflict-free timetable (Barber et al., 2007).

The simulation manager uses the model created in the previous module to test the reliability and robustness of the timetable. One part of the module allows for simulating the spread of unresolved conflicts in a timetable, while the second part is an operational simulation which introduces random delays in train paths (Barber et al., 2007).

The evaluation manager analyzes the results of the simulation and displays the information using several graphs which inform about several factors regarding delay in the network, such as the distribution of delay through the network and expected additional delay (Barber et al., 2007).

2.4.4 Existing methods for determination of road-rail terminal capacity

Analytical methods

Only four analytical methods for calculating terminal capacity were found in total, one of which utilizes an empirical analytical approach. These are reviewed in sections **2.4.4.1** to **2.4.4.4**.

Simulation methods

Within a report carried out for the project "MINT-Modelling and decision support system for evaluation of intermodal terminal networks", Bärthel et. al. (2011) reviewed existing simulation models for the railway and IRRT, and carried out a survey in order to know the extent to which these tools were used by different actors. The sample included private and public organizations within the countries studied in the MINT project, which were actual or possible users of these software tools. Most organizations turned out to use models in a high degree, except for authorities and dedicated consultants (Bärthel et al., 2011, pp. 127–161).

The authors state that the use of models for the design and development of intermodal terminal networks (including terminal design models) is small. While authorities use them to some extent, commercial companies (including terminal owners and operators) almost don´t utilize them, and mostly make decisions based on "tacit knowledge and simplified calculations". In the case that private companies do use models for IRRT planning, they tend to be local models, used by only one or few organizations, and not publically available. Furthermore, dedicated models for intermodal terminals majorly focus on maritime container terminals (Bärthel et al., 2011).

All simulation models found were developed by researchers and/or consortiums formed by universities, authorities and private actors, and normally share the same modelling principles, i.e. discrete-event simulation with queuing theory. Bontekoning (2006)'s model, and the terminal module of the PLATFORM (Rizzoli et al., 2002) project are reviewed in sections 2.4.4.6 and 2.4.4.5. Other models include Birdsell's model (1985), and SimCont (Gronalt, Benna, & Posset, 2006), which are not further described in the work.

2.4.4.1 Method by Kassel+Partner and CombiConsult

The following method for the calculation of terminal capacity was used by Kassel+Partner and CombiConsult in a report prepared for the UIC with the name "Infrastructure Capacity Reserves For Combined Transport By 2015" (UIC-GTC, 2004), which had the objective of analyzing rail transport flows and infrastructure, in order to predict and resolve future capacity constraints. **Annex 8** describes the context in which this method was used within the report.

The report defines terminal transshipment capacity as the "the technical-operational capability of handling intermodal transport units in a certain period of time" (p.89). The result of the method is in terms of TEU per year.

The terms *transshipment volume* and *rate of employment* are also presented, the first of which is equivalent to *actual output*, and the second to *capacity utilization*, as were defined in this thesis. Further on, the concept of *practical capacity* is used only one time in the whole text, where it is established as equivalent to 80% of the (nominal) capacity; it is also mentioned that capacity is saturated at a level of 80% of utilization (UIC-GTC, 2004, pp. 103, 104)(p.103,104).

The authors also mention that this method for determining terminal's capacity is "widely acknowledged".

Terminal capacity is established as the minimum between two sub-systems' capacities: the length of the transshipment tracks, and the capacity of the handling equipment. Therefore, capacity is calculated with a bottleneck approach.

Capacity according to length of transshipment tracks:

$$\boldsymbol{C_{rail} = \frac{L_{track}}{L_{wagon}} \cdot LF \cdot FF \cdot 2 \cdot n_{track} \cdot t_{year} \quad \left[\frac{TEU}{year}\right]} \tag{6}$$

(UIC-GTC, 2004)

Where:

$L_{track} = length\ of\ transshipment\ track\ [m]$

$L_{wagon} = length\ of\ average\ wagon\ [m]$

$LF = load\ factor\ \ [TEU/wagon]$

$FF = flow\ factor, use\ of\ a\ track\ during\ the\ day$

$t_{year} = terminal\ operating\ days\ per\ year$

$n_{track} = number\ of\ transshipment\ tracks$

Capacity according to handling equipment

$$C_{equipment} = C_{gantry} + C_{mobile} \cdot UF_{mobile} \qquad \left[\frac{TEU}{yr}\right] \tag{7}$$

$$C_{gantry} = N_{gantry} \cdot \frac{P_{gantry}}{MH_{gantry}} \cdot Ot_{gantry} \cdot t_{year}$$

$$C_{mobile} = N_{mobile} \cdot \frac{P_{mobile}}{MH_{mobile}} \cdot Ot_{mobile} \cdot t_{year}$$

(UIC-GTC, 2004)

Where:

$UF_{mobile} = utilization\ factor\ of\ mobile\ cranes\ (e.g. reach\ stackers)$
$N_{gantry} = number\ of\ gantry\ cranes$
$N_{mobile} = number\ of\ mobile\ cranes$
$MH = factor\ for\ management\ handlings$
$Ot = operating\ hours\ per\ day$
$P = performance\ of\ crane\ [TEU/hr]$

Overall terminal capacity:

$$C_{terminal} = \begin{cases} C_{rail} & if\ C_{equipment} \geq C_{rail} \\ C_{equipment} & if\ C_{equipment} < C_{rail} \end{cases} \tag{8}$$

(UIC-GTC, 2004)

In **Equations 6** , the load factor refers to the average amount of full units per wagon, with values around 70% and 80% (p.108).

The flow factor in refers to the amount of trains handled in each track per day, and is mostly dependent on demand. It is equal to one if only one train is handled or turned-around per track per day (*static operational concept*), two if each track was used by two trains on the same day, etc. The flow factor is larger when a terminal serves tight-scheduled shuttle trains and achieves turnaround times of 3 to 6 hours (DIOMIS, 2007).

It is not clear what the utilization factor for mobile cranes used in Equations 7 stands for, nor the factors for management handlings.

2.4.4.2 Method by Silvio Nocera

The paper "Terminal Kapazität in Combinierter Verkehr" (Terminal Capacity in Combined Transport), by Silvio Nocera (2008), has the objective of developing a heuristic method that can be applied to a majority of terminals, in order to provide a realistic estimation of yearly capacity. A terminal is considered as composed by different subsystems that interact with each other dynamically, and which in turn depend on several resources. Terminal capacity is defined as the maximum amount of load units moved in a certain period of time. Containers are considered as the only type of units transported, and measured in terms of TEUs. This article does not mention a specific time perspective (i.e. strategic, tactical or operational).

According to the author, under the constraint that the interactions between sub-systems are neglected, the total terminal capacity can be stated as the capacity of its most constrained sub-system. Nevertheless, he mentions that the application of this assumption must be applied with care to avoid important errors. The sub-systems are classified as: (1) road vehicle sub-system; (2) train sub-system; (3) equipment sub-system; (4) storage sub-system. Each one could be a potential critical factor. In this review, only the calculation of train and equipment handling sub-systems will be presented. The sub-indexes present in the original paper were translated into English, otherwise the formulas remain the same.

Formulas for track sub-system capacity

Assumptions:

- The length of the trains that can be processed by the terminal is constrained by transshipment track length, but is depend on the network´s maximum allowed length.
- Two types of wagons: 18 m (can accommodate 3 TEU) and 25 m (can accommodate 4 TEU)

$$C_{rail} = n_{track} \cdot c_{train} \cdot n_{train} \cdot Ot_{year} \cdot f \quad \left[\frac{TEU}{year}\right] \tag{9}$$

$$c_{train} = c_{18} \cdot n_{18} + c_{25} \cdot n_{25} \quad \left[\frac{TEU}{train}\right]$$

$$n_{train} = \frac{Ot_{day}}{t_a + t_b + t_c + t_d + t_e} \quad \left[\frac{train}{track \cdot day}\right]$$

Under the constraint:

$$3\, n_{18} \cdot L_{18} + 4\, n_{25} \cdot L_{25} \ \leq L_{max,train}$$

(adapted from Nocera, 2008)

Where:

$n_{track} = number\ of\ tracks$

$c_{train} = capacity\ of\ train$

$n_{train} = number\ of\ trains$

$t_{year} = operating\ days\ per\ year$

$t_{day} = operating\ hours\ per\ day$

$L_{18}, L_{25} = length\ of\ 18\ and\ 25\ m\ wagons$

$n_{18}, n_{25} = number\ of\ 18\ and\ 25\ m\ wagons$

$c_{18}, c_{25} = capacity\ of\ 18\ and\ 25\ m\ wagons\ \left[\frac{TEU}{wagon}\right]$

$t_a = time\ between\ arrival\ and\ unloading\ \ [hr]$

$t_b = unloading\ time\ [hr]$

$t_c = time\ waiting\ for\ departure\ [hr]$

$t_d = time\ for\ departure\ [hr]$

$t_e = min.time\ between\ trains\ \ [hr]$

$f = coupling\ coefficient\ to\ cover\ for\ empty\ unit\ runs$

$L_{max,train} = maximum\ lenght\ of\ train$

Formulas for handling sub-system capacity

Assumptions:

- Some cranes are assigned for internal movements (within the yard only), while others are assigned for external handling (i.e. road and train vehicle loading/unloading)
- The assignment of tasks among each type of equipment depends upon the size of the terminal and distances between subsystems.

The author mentions that the two types of equipment are mobile cranes, which are used for internal and external handling, and straddle carriers. However, this is not quite clear. More specifically, the terms used are *pneukran*, which refers to a mobile crane, and *portahubswagen*, which, literally translated, means a straddle carrier. Nevertheless, he could as well mean gantry cranes instead of straddle carriers, since straddle carriers are classified as mobile equipment, and since it would not be logical to disregard gantry cranes.

External sub-system capacity:

$$C_{eq,ext} = C_{sc} + C_{mobile} \quad \left[\frac{TEU}{year}\right] \tag{10}$$

$$C_{sc} = n_{sc} \cdot N_{mov/h} \cdot t_{day} \cdot t_{year}$$

$$C_{mc} = U \cdot n_{mc} \cdot N_{mov/h} \cdot t_{day} \cdot t_{year}$$

(adapted from Nocera, 2008)

Where:

$C_{eq,ext} = capacity\ of\ equipment\ for\ external\ handling$
$C_{sc} = capacity\ of\ straddle\ carrier$
$C_{mc} = capacity\ of\ mobile\ crane\ for\ loading/unloading$
$n_{sc} = number\ of\ straddle\ carriers$
$n_{mc} = number\ of\ mobile\ cranes$
$N_{mov/h} = number\ of\ lifts\ per\ hour$
$U =$ Utilization of mobile cranes ($percentage\ of\ effective\ loaded\ movements$)
$n_{mc} = number\ of\ mobile\ cranes$

The author explains that the factor U must take into account that some of the movements are already reserved for the storage of units which have been moved by the mobile cranes, and that some cranes are busy with internal handling of containers.

Internal sub-system capacity:

$$C_{eq,int} = N_{mc} \cdot k_1 \cdot N_{mov/h} \cdot t_{day} \cdot t_{year} \cdot k_2 \quad \left[\frac{TEU}{year}\right] \tag{11}$$

(adapted from Nocera, 2008)

Where:

$C_{eq,int} = capacity\ of\ equipment\ for\ internal\ handling$

$N_{mc} = number\ of\ mobile\ cranes$
$k_1 = percentage\ of\ lifts\ for\ internal\ handling$
$k_2 = coefficient\ to\ take\ into\ account\ empty\ trips\ of\ the\ cranes$

The author mentions that both of the last formulas are a simplification of the real system, since an accurate model for the behavior of cranes would consist on an optimization or vehicle-routing problem, which are very complex to solve and whose computation requires large amounts of time. Therefore, he states that these formulas provide a sufficiently good approximation.

2.4.4.3 Method by Lee, Jung, Kim, Park, & Seo

The conference proceeding under the name "A simulation study for designing a rail terminal in a container port" (Lee, Jung, Kim, Park, & Seo, 2006), has the objective in aiding in the design of a new road-rail terminal within a container area. The authors state that in this case the most important design parameters are the number of transshipment tracks and cranes, which can be initially determined by using analytical formulas. The aim is to determine an initial design which can serve a required number of TEUs per year. The formulas shown below have been adjusted with clearer variable names and reducing the number of formulas, therefore they are not equal to the cited research paper.

Number of required cranes:

$$n_{cranes} = \frac{D_{ctr} \cdot f_p}{C'_{crane} \cdot Wd \cdot Wy} \qquad number\ of\ cranes\ required \qquad (12)$$

$$D_{ctr} = \frac{D_{TEU}}{f_{TEU}} \qquad transport\ requirement\ [ctr/yr]$$

$$C'_{crane} = C_{crane} \cdot U_{crane} \qquad effective\ handling\ capacity\ of\ a\ crane\ [ctr/hr]$$

(adapted from Lee et al., 2006)

Where:

$f_p = peak\ factor \ \ \geq 1$
$Wd = working\ hours\ [hr/day]$
$Wy = working\ days\ [day/yr]$
$D_{TEU} = transport\ requirement\ [TEU/yr]$
$f_{TEU} = factor\ to\ convert\ containers\ to\ TEU$
$C_{crane} = handling\ capacity\ of\ a\ crane\ [ctr/hr]$
$U_{crane} = Utilization\ of\ the\ crane$

Number of required tracks:

$$n_{tracks} = n_{trains} \cdot \frac{t_{service}}{Wd} \qquad required\ number\ of\ tracks \qquad (13)$$

$$n_{trains} = \frac{D'_{ctr}}{c_{train,ctr}} \qquad number\ of\ trains\ that\ arrive\ per\ day$$

$$D'_{ctr} = \frac{D_{ctr} \cdot f_p}{Wy} \qquad number\ of\ containers\ unloaded/loaded\ per\ day$$

$$c_{train,ctr} = \frac{c_{train,TEU}}{f_{TEU}} \qquad Load\ capacity\ per\ train\quad [ctr/train]$$

$$c_{train,TEU} = NR \cdot 2 \qquad Load\ capacity\ per\ train\quad [TEU/train]$$

$$t_{service} = \frac{c_{train,ctr}}{f_{TEU} \cdot C'_{crane}} + t_{loco} \qquad time\ to\ service\ a\ train\quad [hr/train]$$

(adapted from Lee et al., 2006)

Where:

$NR = average\ number\ of\ wagons\ per\ train$

$t_{loco} = time\ for\ locomotive\ operations$

2.4.4.4 Frontier envelope analysis

The report "Evaluating intermodal terminals: framework for government participation" (Anderson & Walton, 1998) was developed by the Center for Transportation Research in the University of Texas at Austin for the analysis of intermodal terminal capacity. The study had the objective of ranking terminals according to their eligibility for improvement-oriented government sponsorship, according to their current capacity needs. After the facilities are been ranked, they are asked to deliver an improvement plan, which is then analyzed with a cost-benefit analysis. The overall aim is to determine which terminals might have a greater benefit from government funds, in order to optimize spending directed towards the enhancement of intermodal transport.

Methodology

The methodology for calculating terminal capacity followed in this study was based on a simple application of the Data Envelopment Analysis (DEA). DEA is currently a leading tool in efficiency and productivity management, which provides information on the capabilities of different types and quantities of resources across a large pool of units. Originally, it is used as a benchmarking technique as it "compares service units considering all resources used and services provided, and identifies the most efficient units or best practice units (…) and the inefficient units…" (Sherman & Zhu, 2006).

The study's method consisted of plotting available data of terminal capacity versus different resources for 299 terminals in the U.S. and Canada, and establishing a formula for the upper envelope of capacity as a function of the availability of each resource type. Outliers were removed using a technique called the "four spread" method[6].

The data on reported terminal capacity and available resources at each terminal was collected from a national terminal database (from 1997) as well as surveys sent to terminals in Texas. The state of Texas was used in order to exemplify the utilization of the method. A particular

[6] This method is used to provide a data analysis that is resistant to outliers, which are located in the top and bottom 25% of the studied variable. The *fourth spread* refers to the difference between the median of the higher half minus the median of the lower half of the data set (Devore, 2014).

problem encountered by the authors, was the lack of information on the procedure followed for the calculation of the reported capacity.

The database provided the numbers for lifting equipment, parking spaces and reported capacity for 299 terminals. For the calculation regarding the rest of the resources, i.e. gate operations, tracks and hustler trucks, the data had to be obtained from surveys in the state of Texas, which covered a sample of 111 terminals. In these analyses, the actual volume served per terminal, instead of the capacity, was plotted versus the resources,

The report defines *ultimate capacity* as the maximum volume that can be processed, assuming a 24-hour utilization, while the term *realistic capacity* is defined as the likely maximum output under realistic normal conditions. The systems analyzed were the lifting equipment, gate operations, parking area, working tracks, and hustling tractors (used for internal horizontal transportation of units).

Lifting equipment:

According to the interviews with operators made by the study, side lift (mobile cranes, reach stackers) productivity was of around 2.5 to 3 minutes per lift, while gantry cranes would take only 2 to 1.5 minutes per lift. From this data, theoretical capacity could be calculated, but a more realistic measure was needed by carrying out the regression. The reported terminal capacity was plotted against the number of mobile cranes and the number of gantry cranes, respectively, in order to calculate the capacity upper limit. To provide better results, the regression was done separately for the two types of equipment, and only terminals which had the studied type of crane were included in each regression. For example, for calculating the slope (or denominator) for gantry cranes, terminals with side lifts were excluded. This gave a total sample of 46 terminals studied for gantry cranes and 115 for side lifts. The final result is the following formula:

$$C_{equipment} = 63\,300\;SL + 77\,400\;GC \quad \left[\frac{lifts}{year}\right] \tag{14}$$

(Anderson & Walton, 1998, p. 70)

Where:

$SL = Number\ of\ side\ lifts$

$GC = number\ of\ gantry\ cranes$

With an R^2 of 0.9036 for GC and of 0.9391 for SL, see Figure 16.

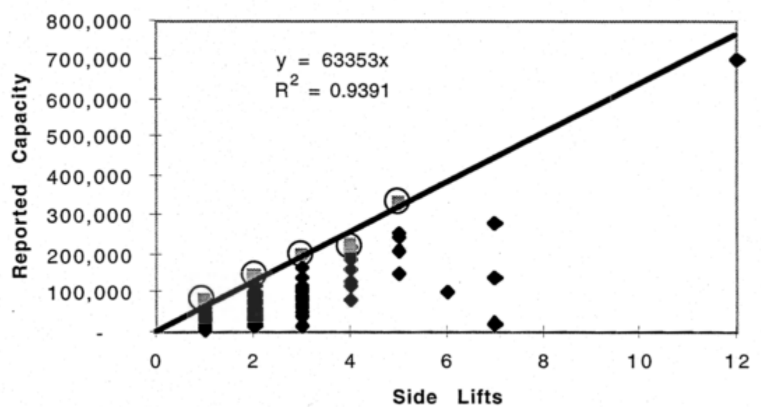


Figure 16: regression of side lifts vs. reported capacity
(Anderson & Walton, 1998, p. 70)

In this sense, for example, if a terminal has only one side lift, it could handle 63,000 units per year under realistic conditions. The authors mention that this number is around 34% of the ultimate capacity (calculated according to lifts per minute, aggregated for the whole year), which indicates that each lift operates about one third of the time within a complete year.

Gate operations:

According to the data from the operators' interviews performed by the authors, the average time to serve a truck at the gate was of 5 to 7 minutes. Nevertheless, this data is described in the report as low quality data, since it is mostly based on operators' opinions. Likewise, the ultimate capacity that could be calculated from it is not representative, since traffic is not constant throughout the operation, and due to the possibility for the formation of queues due to random arrivals.

As the database did not contain the information needed, the number of gates was taken from the surveys at the Texas terminals. Therefore, this analysis was done with a sample of 111 terminals. The regression analysis performed for this resource graphs annual volume against the number of checkpoints at terminal gates. The final formula is:

$$C_{Gate} = 58{,}000\ G\ \ [transactions/year] \tag{15}$$

(Anderson & Walton, 1998, p. 71)

Where:

$G = number\ of\ gates$

With an R^2 of 0.997.

Parking area:

According to the report, the annual truck parking capacity is a function of the number of parking spaces, and the average dwell time. It is also influenced by the number of empty units occupying space in the yard. It is not very clear how the term of parking places is used in the report, but it is most probably used to refer to storage spaces, while dwell time probably refers to unit dwell time. Therefore, the sub index *storage* will be used in the equation.

With data from the terminal database, the authors calculated that terminal capacity is in average 200 times the number of parking spaces. Since the volume handled depended on peak times, which are in turn different for containers and trailers (5 and 15% of the time, respectively), there was no frontier visible in the regression. Therefore, a simple linear regression of reported capacity versus parking spaces was performed, instead of the calculation of the upper frontier. The final equation is as follows:

$$C_{storage} = \frac{p(365/DT)}{1.05+0.1\ F_T} \tag{16}$$

(Anderson & Walton, 1998, p. 72)

Where:

$DT = average\ unit\ dwell\ time\ [days]$

$F_T = fraction\ of\ trailers$

$P = number\ of\ storage\ spaces$

With no R^2 recorded.

Working tracks

According to the report, track productivity depends on train schedules, the layout of the yard, and length of tracks. The data for working tracks was obtained from the surveys, as it was not included in the database. Furthermore, while some tracks are utilized many times per day, others are used just once. The authors introduce the difficulties posed by variable sizes of flatcars, which is a reason why the plot of annual volume versus track length does not provide a relationship. Therefore, the factor of track capacity was excluded from the terminals´ prioritization.

Hustling tractors

Hustling tractors are vehicles used for horizontal, internal unit handling through the yard. The annual volume of the terminal was plotted against the number of husting tractors, giving a formula of:

$$C_{Hustling} = 14\,830\ T \tag{17}$$

(Anderson & Walton, 1998, p. 74)

Where:

$T = number\ of\ husting\ tractors$

With an R^2 of 0.9555.

Interestingly, the points have almost no deviation from the regression line. The authors explain that this is due to the fact that this equipment requires very low investment and can be readily acquired according to changes in demand volumes.

2.4.4.5 PLATFORM Project

A tool for simulating IRRT terminals was developed by Rizzoli, Fornara, & Gambardella (2002), as part of the EC-funded PLATFORM project. This project had the aim of simulating a road-rail environment in order to assess the impacts of different technologies and managements policies on the performance of intermodal terminals. The motivation for such study was the concern that current infrastructure would be insufficient to cover the future growth of intermodal flows. According to the authors of the executive summary for the PLATFORM project, “the simulation environment can be used to analyze how to make combined transport competitive for long and even medium distance and thus lead to a substantial reduction of road-based transport” (IT Ingegneria dei Trasporti, 1999, p. 9). For further explanation of the context of the study as well as the results of the simulation under different scenarios see the report by IT Ingegneria dei Trasporti (1999).

The PLATFORM project is modelled using agent-based simulation, and is formed by two sub-systems; one is the Intermodal Transport Planner, and the second consists of three simulation modules: the intermodal terminal module, the road simulation module, and the (rail) corridor simulation module. In particular, the terminal module is based on discrete-event simulation. In the following paragraphs a brief overview of the intermodal simulation model´s principles is presented; for a more extended description of this module see (Rizzoli et al., 2002). In the next paragraphs only the terminal model will be described.

The paper mentions agents which manage information within the terminal:

- Platform planner: is in charge of managing platform operations, manages trains, and knows the position and schedules for all units in the platforms. In this model, a platform is formed by a transshipment track and its corresponding parking lane.
- Yard planner: has the knowledge of the availability and position of each unit within the terminal, as well as the spaces where units can be stored.
- Road gate: registers arrival and departure of load units by truck, and serves incoming trucks.
- Platform scheduler: “assigns single operations to available cranes” (one per platform). It schedules the unloading and loading of units in the platform, according to their priority.

Furthermore, the processes that are modelled are divided as follows:

- unit arrival by truck
- unit departure by train
- unit arrival by train
- unit departure by truck.

These include unloading and loading operations, as well as temporary storage.

For modelling ITU (Intermodal Transport Unit) arrival by truck in respect of the status of the outgoing train, three cases are specified: “The ITU arrives well ahead of the deadline”, “The ITU arrived just before the deadline”, and “The ITU arrives late with respect to the deadline”. Depending on the case, the vehicle is sent to a parking lane or the storage yard. In the case that the unit arrives after the deadline, it is stored and booked on another train.

The time to serve a unit, t, is calculated as the average time that it takes for a crane to travel along the track in order to load/unload a unit. If the planning of the trucks is perfect, the cranes would not need to move back and forth, making only small movements. However, if there is no coordination, the crane will be constantly travel back and forth along the tracks and t becomes higher. When the crane serves more than just one track at the time, a small quantity is added to the time.

The prioritization of units that are unloaded from a train depends on their expected time for departure (ETD). Units with highest priority are the ones whose truck has already arrived. Next, the units whose time for departure is lower to the parameter *k* are unloaded, and the units which are expected to be picked up at last, are unloaded at the end. It is not specified what would happen in the case the ETD is not known.

Trains are not modeled as such but as a set of ITUs that must be moved. Each individual move is an operation, and a sequence of jobs (e.g. loading/unloading a train), is a job. Each operation is assigned to cranes by the platform scheduler according to priority. Operations with equal priority are scheduled according to their physical position in a circular manner, or "round robin" policy.

In this model, the shunting area is considered to be outside the terminal, since shunting is the responsibility of the rail operator, and the rail gate is the feature connecting both areas. Nevertheless, the model does take into account the time for trains to wait in the shunting area unit a track is available, and the time to take them from the shunting area into the track.

The parameters set by the user include:

- Time parameters:
    - Length of the simulation
    - Work shifts specifications
    - *K* parameter: refers to the amount of hours for departure, depending on which the unloading jobs are prioritized
    - Train timetable.
    - Truck arrival patterns for unit delivery and pickup. They can be determined based on historical data or generated with probability distributions. In the overall mode, an Intermodal Transport Planner is able to schedule the movements of vehicles in the road network.
    - The time before the deadline when a train is told to leave by the platform planner.
- Structural parameters:
    - Yard layout. Involves defining the number of platforms (including capacity of their parking lanes), as well as the capacity for the storage yard.
    - Rail track layout: involves assigning a track to each platform
    - Equipment: number and type of cranes, their performance (max. moves/hour), and operating costs. In the case of gantry cranes, they must be assigned to platforms.
    - List of the cranes active in each work shift
    - Road gate structure
    - Rail gate structure. Includes the number of shunting tracks, the number of tracks linking the shunting area and the terminal, and the time required to move a train between the shunting yard and the terminal.

In the end of the paper, the authors apply the model to one intermodal terminal. In this example, a logistic distribution is used for unit delivery by truck, and a normal distribution for unit pick-up. Furthermore, arrivals of delivery of units were modeled within two scenarios, in the first, 90% of trucks arrived within an hour, and in the second, 90% arrived within four hours before train arrival. Due to the shape of the distribution functions, the results showed that

43% of train unloading moves were direct transshipments, while the number was 97% for loading moves (for the first scenario). In the second scenario, the unit dwell time increases, as well as the number of direct transshipments, but on the other hand, the waiting times for trucks at the gate decrease as their arrivals are more distributed over time.

As said before, the shunting operations are not considered within the intermodal terminal simulation. Nevertheless they are included in the overall combined road/rail transport simulation tool.

The performance indicators for the terminal include average waiting and dwell time of trucks, average lifts per load unit, percentage of direct unit transshipments, average dwell time in the terminal for the load units, average and maximum utilization rate of buffer (parking) lanes, average and maximum utilization rate of storage yard, and average loading or unloading time per train.

2.4.4.6 Method by Ivonne Bontekoning

For her doctoral dissertation "Hub exchange operations in intermodal hub-and-spoke terminals" (2006), Bontekoning modelled the operations of new hub terminals, shunting yards and road-rail terminals. The aim of the work was to evaluate the performance of the different exchange facilities under various operational conditions, and to identify the best operational conditions for new hub terminals. The term of operational conditions is used to refer to demand and facility capacity. The motivation of the work was to identify opportunities and challenges for the establishing of new hub terminals, which are seen as a solution for intermodal transport to serve low-volume routes, offer more service frequency, and be competitive against road transport in shorter distances.

The modelling was carried out based on queuing theory and using Arena, which is a discrete-event simulation tool using queuing theory. As the project was oriented towards new hub terminals, the elements taken into account for road-rail terminals were simplified. The model considers no road transport system, so the units that enter via rail are not considered to leave in a road vehicle but in a later outgoing train. This makes the operations quite different from a normal road-rail terminal.

The road-rail terminal was modelled as one work station with multiple servers, where the overall train handling operations resemble the terminal operations described in Section **2.3**, except for the lack of road transport. In this model, the "clients" or to be processed correspond to the load units, and each arriving or departure train is modelled as a batch of units which enter and leave together via rail.

The demand components, i.e. the parameters setting the "clients" of the process are:

- Train length
- Type of traction (it is not considered afterwards)
- Type, size of wagons and load units
- Number of wagon groups
- Size of wagon group
- Load units and wagons per train
- Load order

- Arrival schedule,
- Departure schedule

The main resources to be considered are:

- Type and dimensions of infrastructure
- Type and amount of equipment (nevertheless, the model only considers overhead cranes)

Main indicators (from Arena Simulation):

- Train and wagon group sojourn times (total dwell time)
- Train and wagon group service time (time in the terminal area)

Train arrivals are deterministic, the arrival schedule contains train arrival time, number of load units, number of wagons and wagon groups. Different arrival patterns, volumes and delays may be inserted by varying the deterministic data, but these files shall be manually created. Arriving trains are served in first-come-first-serve basis, and are represented by a batch of units which is broken up after the arrival to the terminal. Trains wait at a side yard if all tracks are full, until a tracks is free, and it is assumed that all trains must change locomotive at a side yard. When a train leaves the terminal, the next train in the queue can enter.

Dimensions of infrastructure and terminal layout, wagon position, train position and unit position in the yard are not modelled directly, but are expressed through the service time distributions of the servers. Infrastructure elements such as tracks, side yard and storage yard as modelled as queues. The side yard and the storage yard are considered as queues with unlimited capacity.

Waiting time after arrival in the side yard is not considered, because it mostly depends on the availability on train paths and it would make the modelling very complex.

The size and type of each load unit is not modelled, but rather stated indirectly in the number of units per wagon. It is also assumed that each train in a batch carries the same amount of load units and wagons. Train length then depends on the number and length of wagons.

Train departures as modelled as the departure of a batch of load units, which happen when the terminal operations are completed. Since departure schedules are not considered, priority rules based on due dates or delay are not modelled. The only way delay may be investigated is not through deviation from schedule but according to service times, but it is not the aim of the study to examine punctuality.

Since there is no horizontal transport system, the gantry cranes do all the lifts. It is assumed that all cranes can carry all types of load units, and that each crane can only process load units in its own section. It also assumes 100% balances utilization of cranes, which is not achievable in real life.

An important contribution is the details of the crane cycle times, which include handling times per unit, depending on the number of cranes. The tasks considered within the cycle time are:

- Driving crane to right flow of load units

- Adjusting spreader
- Unfolding grapple arms (for swap body and trailer)
- Letting down (empty) spreader
- Positioning and attaching load unit
- Lifting (loaded) spreader
- Driving crane to right row of load units
- Turn container to get doors on the correct side
- Letting down spreader
- Positioning and detaching load units
- Lifting empty spreader

The cycle time equals the sum of all times, except for the activities that can be done in parallel, which are: crane driving, adjusting of spreader, turning containers and folding grip arms. In this case only the maximum time is counted. It is assumed that adjustment of spreader, (un)folding grip arms and turning container take less time than the other activities. Therefore, the model does not consider different handle times depending of unit type.

3 Findings and analysis

This section will present the findings regarding the functioning of the terminal and surrounding system, which do not appear in the general description provided in the framework. Especially, the importance of coordination between actors for capacity utilization and reliability is stressed. According to the capacity concepts and methods reviewed, several methods are proposed according to the type of capacity to calculate and the adequacy needed.

3.1 Implicit knowledge

It was found that there is scarce relevant literature describing in detail the operations within intermodal terminals. Of the works encountered, only the paper by Boysen et.al. (2012a) has the specific objective to provide an account of these activities. Nevertheless many details are missing, especially regarding information flows. The descriptions provided by the models are excluded in this declaration, since their role is not to describe operations in detail but to abstract them.

Comparatively, there are many works dedicated to container maritime ports, which have different properties from road-rail terminals (Gronalt et al., 2006). The fact that there is more literature regarding container ports than inland intermodal terminals might be because container ports require a great amount of standardization, coordination and use of communication technologies in order to serve intercontinental markets. Meanwhile, inland intermodal terminals differ greatly between each other in terms of material and information flows, and their characteristics are more determined by regional situations and demand. The consequence is that there is a great deal of implicit knowledge in this area, as no formalized knowledge would be capable of covering all operations of the various intermodal terminals in detail. As mentioned in section 2.4.4, in their review of models Bärthel et. al. (2011) explain that many IRRT models are based on tacit knowledge.

The analytical models for terminal capacity which were found, contain a certain amount of implicit assumptions that may not be easily identified and analyzed without knowledge of the operational details in terminals and the terminology used. No academic sources were encountered, which could provide useful descriptions of how road-rail intermodal terminal capacity is analyzed and planned in the industry in practice. In fact it was rather the opposite; in a report from the DIOMIS project (in UIC, 2008) it is mentioned that many terminal developers do not tend to share information regarding new developments and layout re-designs.

3.2 Definitions and terminology

While reviewing the literature, an important problem which was identified was the lack of a standardized terminology regarding actors, infrastructure, resources and operations. One may find one concept which is named using several different terms, as well as one same term used for naming different concepts. It is not within the reach of the thesis to present all these variations, but it is important to remark that there is a lack of standardized fundamental knowledge upon which the intermodal research area can be built. In this section, only the most relevant examples are presented.

The most basic concepts are termed and described differently depending on each researcher´s background and nationality. This supports Bontekoning´s (2004b) statement that this discipline has a number of research groups working rather independently, and with less cross-references than in other areas of knowledge. In this sense, the same basic concepts may be shared within members of a research circle seamlessly, but as long as the same basic knowledge is not shared among the international research community belonging to the research area, the cross-analysis of works from different backgrounds is hindered.

Actors

It was observed that several works classify actors differently, and roles are mixed and overlapping across literature. In some academic works, the infrastructure owner is regarded as the same actor as the rail operator or carrier (see for example Crainic & Bektas, 2007). In others, the roles of the IM, the rail operator and the terminal operator are assigned to one same actor (see Macharis & Bontekoning, 2004). This might be an appropriate method for grouping similar actors into a theoretical structure, but in reality these activities may be done by various actors which in turn have different strategies and planning problems. Academic works which intend to provide a review of the industry, operations and decision making, disregard certain concepts that are used implicitly in the more industry-oriented reports.

One example is that when examining terminal operations, the UIC and CombiConsult (in DIOMIS, 2007), which may be classified as industry actors, regard the terminal operator and the terminal owner as different actors which are in control of separate resources and operations. The same is observed in Bergqvist (Bergqvist, 2014), who examines the consequences of different contract configurations in IRRT. On the other hand, researchers Macharis & Bontekoning (2004) englobe and overlap the roles of rail operators, terminal owners, IMs and terminal operators.

Perhaps the lack of differentiation between several types of actors in certain parts of the literature is due to the fact that the market deregulation and the vertical disintegration is a recent occurrence. Probably, some theories and actors' descriptions were based in the previous vertical integrated systems with less dynamic business relations. Nevertheless, this no longer represents the actual system.

The ownership structure of actors within the intermodal and rail systems has a significant effect on its performance and competitiveness. Objectives are not the same between public and private organizations, as well as between different private companies. Therefore, a challenge for achieving an efficient network is the lack of agreement between state and private organizations about an appropriate way for reaching efficiency, as well as on the definition of efficiency itself (Beck et al., 2013, p. 9; 2007, p. 52; Woxenius, 1994). In theory, when an organization within the railway industry (i.e. IM and/or transport operator) is publicly owned, its objective is to maximize welfare. Beck et al. (2013, p. 3) explain that a public organization defers from a private one in that it normally provides a network coverage above what a free market would offer, coupled with lower costs than the market rate.

Operational terms

Regarding operational terms, the most relevant observation is related to the terms for marshalling and shunting. In marshalling and shunting yards, marshalling operations are carried out, i.e. the breaking up and regrouping of wagons sets. These terms are used in literature as having absolutely the same meaning. Nevertheless, Fredrick Bärthel mentions that the two concepts have a slightly different meaning. This is because shunting yards have a rather local nature, while marshalling shards have an interregional scope. Furthermore, marshalling yards possess dedicated resources for train disassembly and wagon rearranging, and most of the times present a hump. On the other hand, shunting yards are used for train formation by joining wagon groups. In this sense, shunting yards serve the regional trains closer to their departure yards, and in marshalling yards inter-regional and international trains are marshalled.

The meanings of the verbs *marshalling* and *shunting* are quite different, and their distinct applications are especially unclear. Marshalling is defined as "the breaking up of freight train formations and the subsequent sorting of wagons into train loads for final destination, carried out at a marshalling or shunting yard" (Bärthel et al., 2011, p. 12) (Figure 17, a). The use of the verb *marshalling* is relatively homologized across literature; however the use of the term *shunting* is vague.

In theory, the verb *shunt* is defined as the "operation of moving a rail vehicle or set of rail vehicles inside a railway station or other railway installations" (UNECE, 2010, p. 25) (Figure 17, b). In this sense, shunting denotes a short-scale movement of wagons along sidings, which can be performed nearby intermodal terminals, in goods yards, in sidings along a main line, in maintenance workshops, etc. Furthermore, shunting can be used to move wagons into a storage track, into unloading and unloading sites, for putting them aside along a line, for transferring them into a terminal, etc.

(a) (b)

Figure 17: Depiction of shunting (a) and marshalling (b) operations

Nevertheless, in several works *shunting* is actually used as a synonym for *marshalling,* i.e. referring to the train reassembly and classification carried out strictly at marshalling and shunting yards (see Ahrens et al., 2009; Arcot, 2007; Bontekoning & Trip, 2004; Christoph, 2010; Ferreira & Kozan, 1992; Gatto, Maue, Mihalák, & Widmayer, 2009; Macharis & Bontekoning, 2004; Marinov et al., 2014, 2013; Wezel & Riezebos, 2011). It is sometimes not clear from the context if the author is making reference to the act of marshalling or shunting when the word *shunting* is used (e.g. Nilsson, 2011, p. 10). The misuse of the word causes confusion for the reader and may create misinterpretations of the author´s message, as both concepts are close enough to be undistinguishable in certain contexts, but different enough to significantly alter the message. It is therefore necessary to use these two terms only according to their formal definition.

Furthermore, to facilitate research, it is necessary to create a common framework for logistics transport which makes all business relations, actors, and actor sub-categories as explicit as possible. As well, the relationship between each actor, its corresponding resources and activities must be unambiguously stated. For achieving an optimal capacity management, coordination is key. This is true in any industrial network, but more in intermodal and railway transport, where components within the rail network are so highly interdependent. Many scholars have thoroughly analyzed industry structures and business relationships (e.g. Behrends et al., 2011; Bergqvist, 2014; Flodén, 2011), but there is still much to be done, especially in concealing the concepts related to the rail and road sectors.

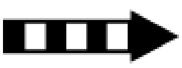
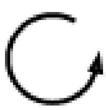
For a framework related to actor categories, Woxenius´s (1994) three-part model may be utilized and expanded further. Such framework could first state all the type of actors sub-categories, including their alternative names in different transport networks and their implications. Furthermore, the word *actor* may not be the most appropriate for classification, since the term may be related to an organization or company. Actors shall be defined not as organizations but according to the most basic roles.

To avoid confusion, the word *role* will be used instead of *actor*, and for each basic role in the industry, the corresponding resources and activities shall be assigned. In turn, the word *actor* shall be used to refer to the organization performing one or more roles in the system, which can be private or public.

3.3 Material and information flows within terminals

In Figure 18, a flowchart with Value Stream Map (VSM)[7] symbols is used to give a general depiction of the material and information flows in an intermodal terminal in daily operations. The information has been obtained from the visit to the intermodal road-rail terminal in Gothenburg (Gothenburg Combiterminal), an interview with Anna Elias and Jonas Emanuelsson, as well as the literature sources and interviews individually specified.

Table 1: Description of symbols used in Figure 18

Symbol	*Meaning in VSM*	*Significance*
	*Push arrow*: represents that material is ”pushed” from one process to the next, regardless of the state of the downstream process.	Arriving vehicles ”push” their load units into the terminal; their arrival schedule determines the time in which they are unloaded.
	*Pull arrow*: represents the physical removal of material by a downstream process.	Departing vehicles ”pull” the units from the terminal; their arrival schedule determines the time in which they are unloaded.

[7] A Value Stream Map is a type of flow chart utilized to depict the flow of information and inventory. This tool is used in Lean Manufacturing, which is an industry practice designed to eliminate all sources of waste in a production process. Although VSM symbols are used, it must be noted that figure 18 is not a proper VSM, since it is adjusted to fit this thesis´s informational needs.

	Supermarket: an upstream process replenishes a supermarket or stock point, where material is stored. Downstream from such stock.	When vehicles are unloaded, the units enter the yard, where they might be stored for some time (in buffer storage or parking lane). They are later withdrawn in order to be loaded onto another vehicle.
	Two-way information flow	Booking + confirmation *or* work order + confirmation of task completion
	One-way information flow, inconstant	The rail operator rarely communicates the status and manifest of the incoming trains to the terminal.
	One-way information flow	
	Material flow / transportation	

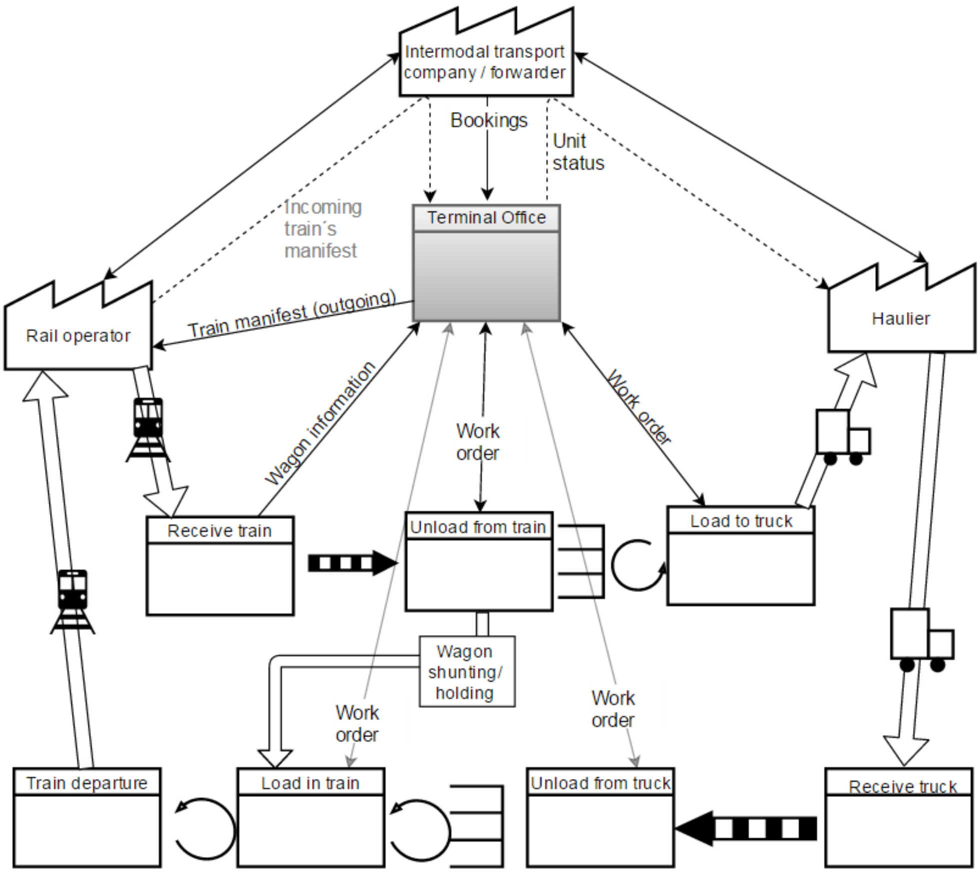


Figure 18: Flowchart of an intermodal terminal's operations with VSM symbols.

The top row of Figure 18 represents the main actors of the system and the information flows between them (it is not certain that this configuration holds true in most cases; as mentioned in the previous section, it is difficult to identify the actors of the system and their roles). The operations regarding the flow of import units are presented in the middle row, while in the bottom, the flow of export units is shown. The operations of receipt, unloading, shunting and keeping wagons, loading and dispatching a train, which are shown in the four left boxes of the bottom rows, form the action named ***turning around*** a train. The overall sequence and nature of the operations that were observed in the terminal, match the description of Boysen et.al. (2012a) and the authors cited in Section 2.3. The time to process a wagon set (shunting, unloading and loading) is called ***turnaround time***, and the total time that a train set remains at a terminal is the ***dwell time***.

A list of all the factors observed during the visit, which could affect capacity are given in **Annex 9**.

3.3.1 Loading plan

In most occasions the terminal operator is not informed about the wagon composition of incoming trains before their arrival. As the train arrives the terminal staff must register wagon order, wagon types, and unit characteristics. Afterwards, they start preparing the ***loading plan***, i.e. the loading sequence of the outgoing train, including pairing up each unit with the correct kind of wagon.

This activity is done manually, i.e. without any dedicated Decision Support System (DSS) software. Nevertheless, it represents a complex decision-making problem since several constraints must be met:

- Corresponding unit and wagon type
- Corresponding unit(s) and wagon length
- Corresponding unit(s) weight and wagon load capacity
- Correct wagon cuts: wagons going to the same destinations must be grouped together, in the case that trains are not operated as shuttle services.
- Dangerous goods: some goods should not be put next to each other, and others should not be placed in the front or rear of a train.
- The location of a unit in the yard, in relation to the train

According to the staff in the Gothenburg Combiterminal, a train with many wagon cuts and varied unit mix may take around 6 hours to turn-around, while an "uncomplicated train", may take as little as two hours.

It can be inferred that the amount of complexity for train turnaround operations depends on the type of train production system. Shuttle trains will have a less complicated turnaround as no wagon groups have to be arranged in groups.

In the Gothenburg Combiterminal, the decisions for loading a train are performed by the crane operators at the time they are performing the loading operations. When an incoming train arrives, they observe the location and type of wagons it contains. At the same time, the crane's screen shows the load work orders to be performed. The operator then proceeds to find the unit in the yard, which corresponds to the work order selected, and decides in which wagon to place it. When he/she has performed the loading operation, the work order is marked as confirmed and the last four numbers which identify the wagon are written down.

Lifts are normally assigned to the cranes and reach stackers through work orders (unless the case of automatic gantry cranes). In the Gothenburg Combiterminal, they are communicated to the crane operator automatically through the Terminal Management System (TMS), while in other terminals they are communicated by radio or walkie-talkie. Work orders for loading and unloading road vehicles are formed when a vehicle carrying a container or a swap body checks in at the gate, while work orders for loading a train are made following a loading plan.

According to the loading plan and the actual loading pattern of the train (which might present variations from the plan), a document called ***train manifest*** is made. This document is delivered to the train operator, and is required for legal purposes. It specifies the actual characteristics of the departing train including: which wagon is carrying which load unit (including empty units), the specifications of the load units, and the wagon arrangement. Nevertheless, this document is not normally delivered to the destination terminal operators.

In the case of terminals with gantry cranes, their scheduling must be planned after the loading plan is ready. Most normally, when two or more gantry cranes are located in the same track group, they are assigned specific areas of equal size within which they can maneuver. Therefore, each crane handles only the units falling into its designed area (N. Boysen et al., 2012a).

When trains are not operated as shuttle services, it might be necessary to change the wagon arrangement according to the loading plan. As in the case of train arrival, the shunting operations performed for wagon rearrangement are normally the responsibility of the train operator and in some cases may be performed in a separate shunting yard, i.e. without using the terminal´s tracks. In the Gothenburg Combiterminal, this service is offered by the terminal operator using the terminal´s side tracks.

3.3.2 Bookings and pricing

Customers are required to book transshipment services in advance. In the Gothenburg Combiterminal, customers must book the transshipment service from the terminal through an online tool. Such booking must contain the details of the rail trip, which is defined by the train number, arrival/departure date and time of day. For each trip, the characteristics of the load units corresponding to the train are written down, including the dimensions, type, weight, and if it contains dangerous goods.

Normally, the time in which the road vehicles will pick up or deliver the units is not specified by the customer, and truck drivers do not communicate with the terminal. If the time of arrival of a truck is not coordinated with the arrival/departure time of the train, the unit must be placed in temporary storage. No rent is charged for temporary storage unless the time overpasses a certain limit. When this happens a marginal rent is charged, which does not bring big revenue to the terminal. Terminals offer this service because the customer often needs to have a buffer, but they would prefer to have as less units as possible in such storage. This is also because their main revenue comes from lifting units, and it is preferable to conduct only one lift per unit.

Most terminals charge *per lift*, regardless of the season, the time of day, type of customer and unit type [8]. Unloading and loading of road vehicles is not priced, and for each unit booked, only one lift is charged regardless of the amount of actual lifts performed. Therefore, the ideal scenario is when trains and road vehicles are coordinated, so that the unit is directly

[8] Interview with Lennart Hammarbäck, Consultant for society logistics. *WSP Group*. May 2015.

transshipped. In this case there is no need for the lifting required to put and remove unit from temporary storage. Nevertheless, the terminal operator might perform improvised deals with the customer in some specific situations.

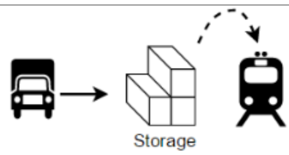
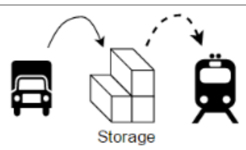
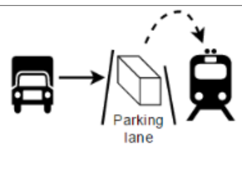
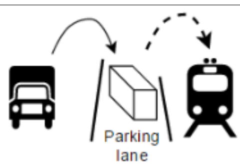
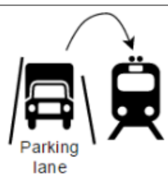
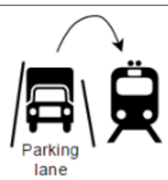
3.3.3 Flow of load units

In the Gothenburg Combiterminal, the internal information regarding unit lifts is managed through work orders. In turn, an IT system is used to administrate work orders and keep track of units within the terminal. Additionally, each crane is equipped with a screen. As work orders appear on screen, each operator “chooses” a work order to perform, in order to avoid the overlapping of activities between two cranes. After an activity is performed, they confirm the completion of the corresponding order.

When road vehicles arrive to the gate the overall process goes as already described in section **2.3**. Next, the truck driver is given the location in which the unit must be left inside the terminal. If the outgoing train is ready (or soon ready) at the transshipment track, the truck will be indicated to go to a parking lane next to the track. Else, it will be directed towards the storage yard.

If the load unit is a trailer, the driver is able to detach the chassis manually, without the need of a transshipment device (Table 2, second column). In this case, the driver can leave the terminal right away (if the workload is low, a transshipment device might help unloading the unit). If the truck carries a container or swap body, it must be unloaded by a device (Table 2, third column). Sometimes this implies that the truck must wait to be serviced (delay). The fact that trailers do not need to be unloaded by cranes was not found anywhere in the literature, therefore it is not clear whether this happens in all terminals and with all kinds of trailers.

Table 2: Depiction of the trip of incoming load unit inside the terminal depending on outgoing train status and unit type. (own creation)

Train status	*Trailers*	*Containers and swap bodies*	*Symbol key*	
Train will arrive later or another day			Direct unload (no lift)	→
Train will be ready to load soon			Lift with delay	
Train is ready to be loaded on transshipment track			Lift, possible delay	

The nature of the re-location of empty units may vary depending on who the owner of the wagons is. In the Gothenburg Combiterminal, the units are normally property of the train operator, so if there are empty wagons left, empty units are loaded in the empty wagons of the operator's outgoing train, which balances the amount of units in each location.

When a road vehicle arrives to pick up a unit, it is first registered at the gate, and the booking checked. Staff in the terminal office then checks if the unit is active in the system. If it is, the driver is directed to the site of the yard where the unit is located, and if it isn´t, and the corresponding train has arrived, a work order for transshipment with high priority is generated. Work orders for loading/unloading trucks normally have a higher priority, in order to service trucks as quickly as possible.

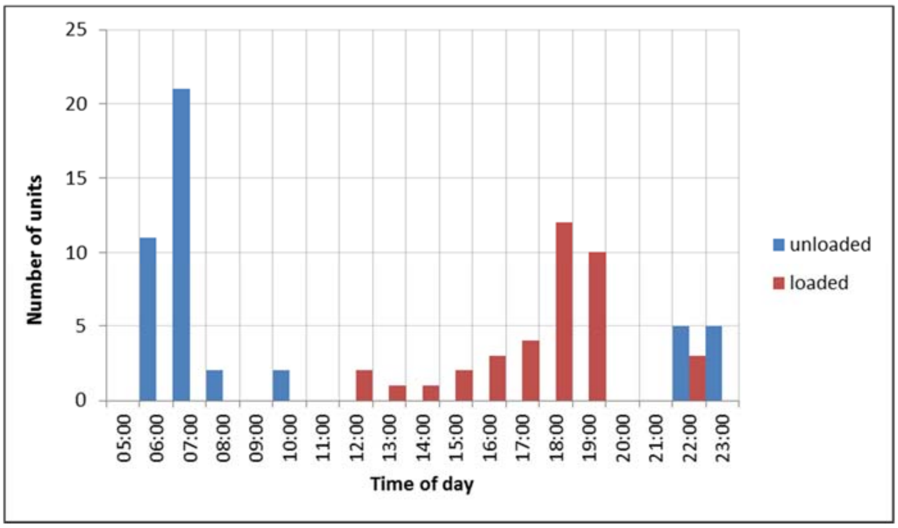
In the Gothenburg Combiterminal, units are said to be "active" in the terminal, when they have been unloaded from a train or truck (i.e. after the unloading work order is confirmed by the crane operator), and become inactive after they have been loaded onto a train or truck. In this way, the terminal office has control over each unit. Units which are physically in the terminal area, but are still on top of a train or truck, are not considered active in the terminal system, and are only present in the system as bookings. Furthermore, after a unit is loaded onto a train or truck, the unit becomes inactive when the completion of the work order is confirmed.

3.3.4 Lifting patterns

Graph 2 shows the distribution of train loading (red) and unloading (blue) operations on a given day of the Gothenburg Combiterminal, which give a clear picture of the variability of capacity utilization throughout a day. Peak times happen in the morning due to train arrivals, and in the evening due to departures, because of the night-leap operations as described in Section 2.3.

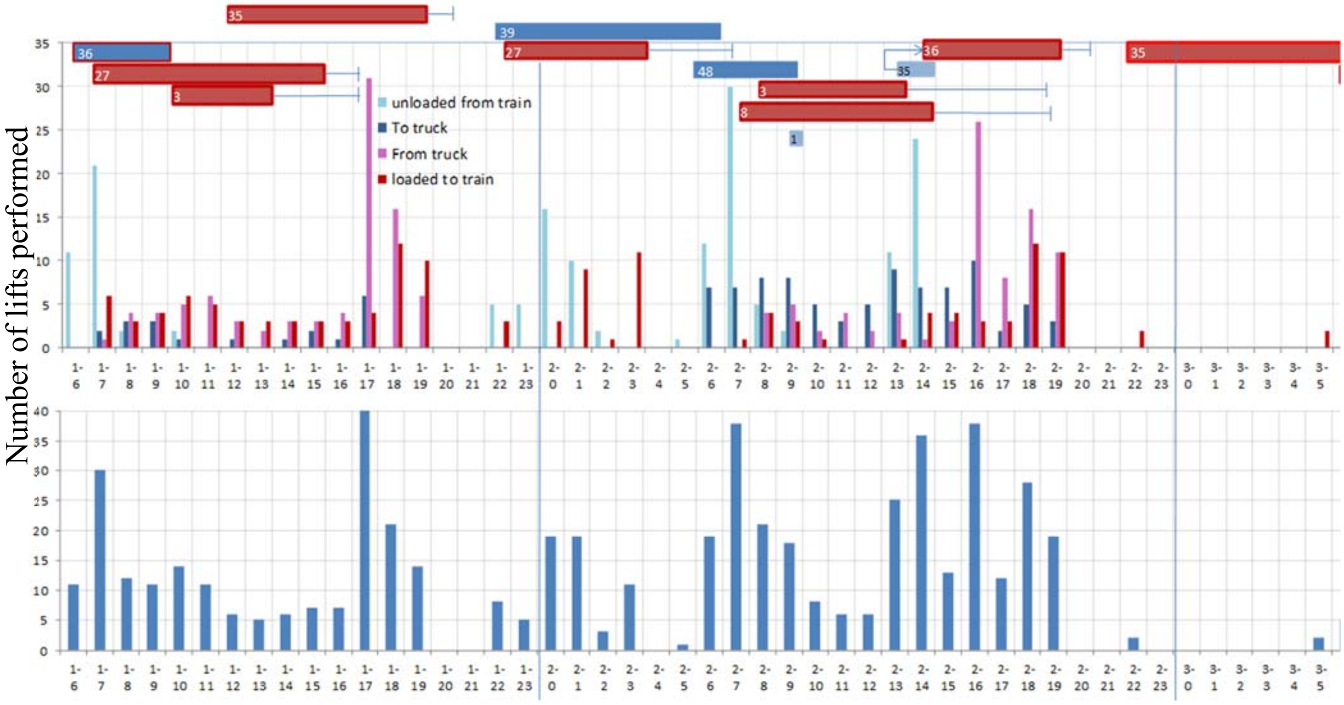
Graph 2: Units loaded and unloaded from/to trains in Gothenburg Intermodal Terminal during one day (1st June 2015)

(see Annex 10 for source of information)



Graph 3: Lifting and demand patterns in Gothenburg Intermodal terminal

(see Annex 10 for source of information)



Graphs 2 and 3 depict the number lifts performed per hour during three days in the Gothenburg Intermodal Terminal. The top bar graph shows the number of units handled according to type of activity, while the bottom graph shows the total number of lifts (therefore, the sum of the quantities in the top graph). The boxes drawn in the top represent incoming (blue) and outgoing (red) trains. The "interval" lines at the right of each outgoing train box represent the scheduled departure times. The numbers in each train box correspond to the number of units.

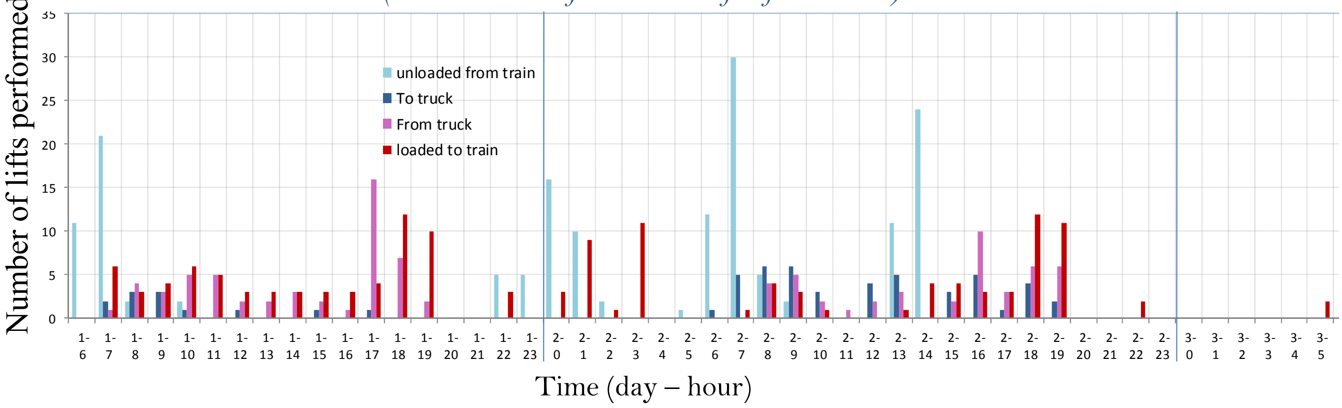
In Graph 3, some behaviors can be observed, such as:

- Arrival of load units by truck is rather constant throughout the day but peak greatly before train departure. Actually, this is the activity that presents the major difference between peak and low-demand times.
- Departure of load units by truck is stable throughout the day.
- Train loading operations are rather distributed through the day and peak just slightly in the evening/before train departure.
- Train unloading operations have large peaks at train arrival but are rather absent afterwards.
- During night hours, the terminal is closed for road vehicles but still performs train loading/unloading operations
- Trains do not necessarily present the late-night departure and early-morning arrival scheme.

In total, the main peaks observed in Graph 4 are provoked by train unloading moves at train arrival, and road vehicle unloading in the evenings. The road peak of road vehicles in the evenings happens because the terminal closes at a certain time for road transport, but trains still depart through the night.

Graph 4: Lifting and demand patterns in Gothenburg Intermodal terminal, 1th to 3rd June 2015, disregarding trailers

(see Annex 10 for source of information)



Comparatively, when trailers are omitted (Graph 4), the largest peaks are formed by train unloading, while train loading and truck unloading form slighter peaks. In those three days, a total of 386 units were transshipped, and 588 lifts were performed[9]. 26% of the units were containers, 32% swap bodies and 42% were trailers.

The terminal´s actual patterns vary from the four phases presented by Ballis and Golias (2002) in Section 2.3, which also fail to describe the evening movements. In this real-life case, it is true that units are unloaded quickly after train arrival (see also Graph 6), but road vehicles do not queue to pick up units, as they arrive evenly during the day. Therefore some import units are not transshipped directly to trucks but unloaded into the buffer lanes and storage yard. This behavior might be differ depending on whether the terminal has a static or dynamic capacity, i.e. the flow factor.

Graph 5 depicts a histogram representing the behavior of train unloading operations. The x-axis represents the difference between the time each unit was loaded and the scheduled arrival time of its corresponding train. For example, of all the units examined, 37 of them were unloaded one hour after their train had arrived. Graph 6 depicts the opposite situation, so to say, the time in advance that each unit was loaded before scheduled departure.

It can be observed that all loading stops around one hour before train departure, and that the time trains stay at the terminal is longer than turnaround time. This suggests that in this case train dwell time is determined by scheduled arrival and departure times rather than turnaround times. Therefore, if a train runs on a tight schedule, dwell time depends on turnaround time. Else, it is determined by departure and arrival times, which are settled on the train´s path in the timetable. In turn, this determines the quantity assigned to the flow factor in equation 5.

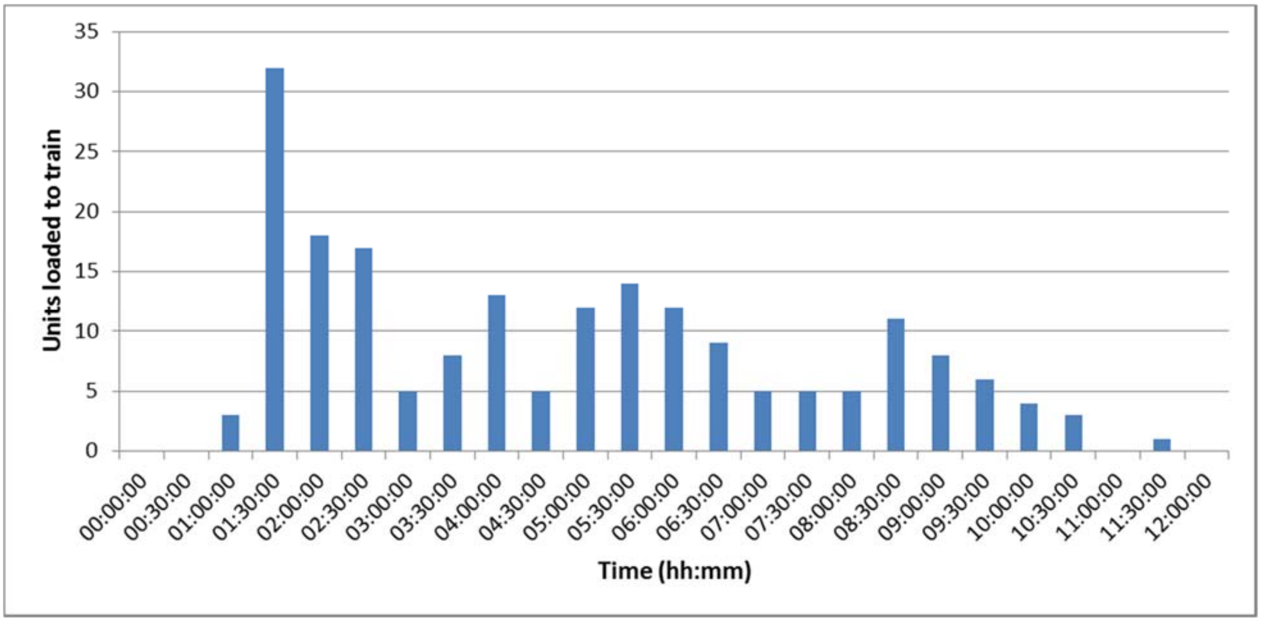
In the case that dwell time depends on the train’s timetable, the empty wagon train set is held until its outgoing time is due, either in the transshipment track (flow factor = 1) or in the side

[9] This number assumes that no trailer was loaded/unloaded from trucks by cranes, and considered two lifts per unit (i.e. one to hold up a unit, and one to load the unit)

yard (flow factor equal 2 or more). This depends on the terminal operator´s policies, the tightness of the schedule, and the availability and number of holding tracks. The train operator is benefited when it is allowed to keep wagons in the terminal if needed and may avoid the need for shunting. But this is not in the terminal´s best interests, since it occupies capacity which could be otherwise used to process another train. If one train is being handled (unloaded/loaded) continuously and stays through the day, then most probably the flow factor is one and one train stays throughout time in one transshipment track.

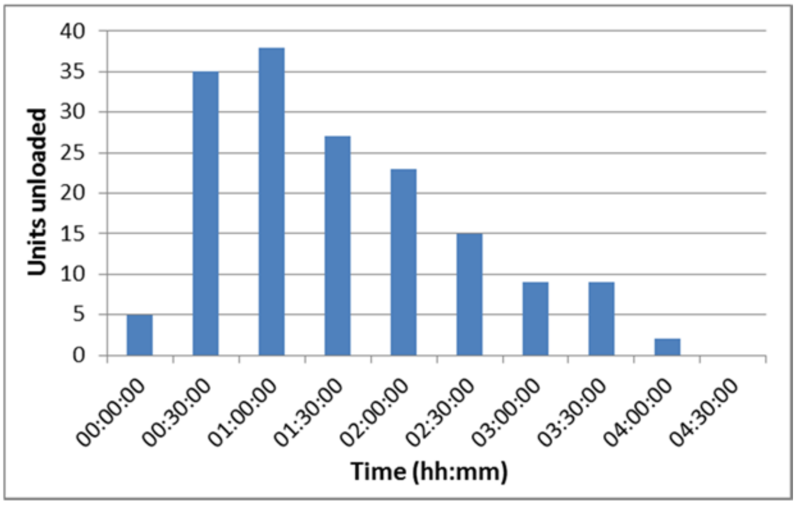
Graph 5: Number of units loaded vs time before train departure

(see Annex 10 for source of information)



Graph 6: Number of units unloaded vs time after train arrival

(see Annex 10 for source of information)



Consultant Fredrik Bärthel[10] is currently carrying out a study regarding the needs for transshipment capacity expansion at the terminal Alnabru, in Oslo, Norway. This information is located in this section, as the method does not count with a written report.

[10] Interview with Fredrik Bärthel, Consultant for society logistics WSP, May 2015.

The first step performed consisted on a market research to predict future volumes to and from the terminal, in terms of number of trains. Afterwards, the number of wagons was calculated according to incoming and outgoing train volumes, followed by an estimation of the number of empty units required to balance incoming and outgoing loaded unit flows. The arrival and departure times of trains are deterministic and based on rail schedules; delays are not considered. After the total number of incoming and outgoing units is determined, the necessary number of cranes and tracks is estimated.

Based on years of experience, he has developed a tool to determine loading/unloading times, for trains, and the utilization of cranes across a train´s turnaround time. For an incoming train, 50% of the units will be unloaded in the first hour, the next 20% are unloaded half an hour after that, and the next 10% another half an hour afterwards. Likewise, in the case of outgoing trains, 50% of units are loaded one hour before departure, while the next 20% are handled between 60 and 90 minutes before departure (see Figure 19).

In the case of train arrival, the Gothenburg terminal presents the same distribution as proposed by Bärthel (see Graph 6). Nevertheless, the same does not happen in the case of train loading (see Graph 5). This could be because in Alnabru, trains are "cleared out" to the side yard after unloading, while in the Gothenburg Combiterminal trains tend to stay in the transshipment tracks during the day.

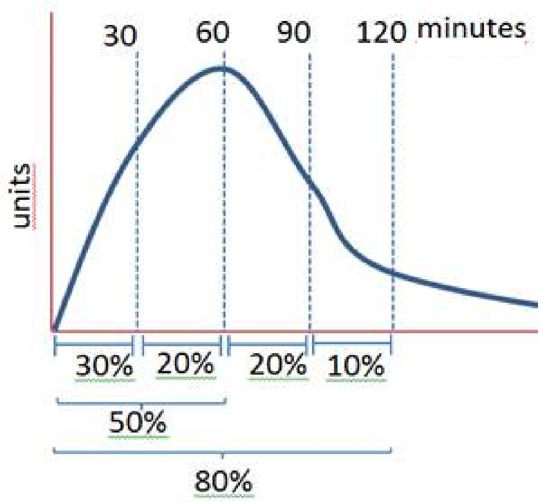


Figure 19: Distribution of units unloaded according to train arrival time
(adapted from method by Bärthel[11])

3.4 Interview findings

3.4.1 Components of capacity

The main influences for terminal capacity most mentioned by interviewees are:

- The number and length of tracks
- The number and type of cranes
- Mix of load units

[11] Interview with Fredrik Bärthel, Consultant for society logistics WSP, May 2015.

- Area of the terminal
- Operational policies

In some terminals, the length of tracks may be shorter than the maximum allowable train length[12]. This may happen when terminals were built before allowed length expanded, and/or because there is not enough area to allow for complete tracks. In these cases, train sets have to be split to enter the terminal. Furthermore, although the literature clearly classified between transshipment and holding tracks, it happens also that the terminal owner provides a certain number of tracks, which the terminal operator decides how to utilize[13].

The mix of load units is determined by the market demand, and terminals have to buy equipment that meets the type of units as well as the expected weight. Terminals which serve markets more related to maritime transport, will have the container as the dominant unit, while terminals based on inland transport will have trailers as the main unit. Cranes may be either dedicated to one type of unit or flexible. Flexible cranes increase transshipment capacity as they can be assigned to any kind of unit, but they can be more expensive. Due to the extra equipment, they may also have less lifting capacity in terms of weight compared to an equivalent crane[14].

The area of the terminal determines how much units can be placed in the storage yard, and may also limit the possibility of adding new tracks if needed. Often, the chance of expanding a terminal is limited by the price of the surrounding land, and/or the fact that there is no more available land in the surroundings[11]. At the time an intermodal terminal is built, the infrastructure and superstructure capacity is low due to limited resources and initially low demand[12]. The objective is to attract as much market as possible so to cover costs. Some time later, if the terminal was successful, congestion problems may arise, while the surrounding land might have increased in price or is already occupied. This is related to the fact that offering buffer storage is not really profitable, especially for the terminal owner, since the price of land is much higher than the minimal revenue which can be collected from buffer storage.

The main operational policy mentioned, is related to whether outgoing trains wait for late units or not. In intermodal trains, only around 20% of the income is profit, while the majority represents path, fuel, staff and other costs. If a train is less than 80% full, it results in a loss[11]. Late units that miss the train represent lost income, as the operator charges nothing or very little when they are not loaded. In order to cover costs and have fuller trains, some terminals choose to wait for late units and run the train on the operator´s next scheduled path, which might be the next day or later. It is regarded as better practice not to wait for late units, in order not to affect other (final) customers served by the same train, not to delay the next trains to be served. In fact, losing a path that is already paid, may be more costly than running a partially full train[14].

The DIOMIS project (2007), which combines the UIC and industry actors with the aim of enhancing intermodal transport, presents many bonus-malus incentives for incentivizing customers and haulers to deliver units on time. These are regarded as innovative, and include

---

[12] Interview to Anna Elias, Green Cargo´s Terminal operations manager at Gothenburg terminal and Jonas Emanuelsson, Jernhusen Systems Manager, May 2015.

[13] Interview with Bosse Eriksson, Terminal Operations Manager at Eskilstuna intermodal terminal. May 2015

[14] Interview with Lennart Hammarbäck, Consultant for society logistics. *WSP Group*. May 2015.

serving punctual trains first. Therefore, apparently this measure is not widely implemented yet. According to the consortium's report, railway companies are likely to discourage this measure.

3.4.2 Reliability and capacity utilization

It was also pointed out by interviewees that while larger capacities allow for shorter loading and unloading times, the factor of time reliability is perhaps most important for the customers. This is because transport customers plan their operations counting on determined transport times and schedules. Total time might not be a problem, as long as long lead times are established in advance and respected. This allows for operations at the receiver´s location to be planned accordingly and to run smoothly. Late or early arrivals, however, represent a bigger problem. One single late arrival might even cause customers to choose another service if the load was an important one, while early arrivals might represent inventory costs for the receiver[12, 13].

Capacity utilization is a determinant of the reliability of the terminal, as well as a relevant economical factor since overcapacity hinders economic performance.

Most agree in that the leading factors affecting capacity utilization and the reliability of intermodal terminals are external. The utilization of capacity depends on train and truck arrival patterns, as well as the services demanded by each customer. On the other hand, time reliability is influenced by the railway system and the operation of haulers. Meanwhile, the nature of the demand in each terminal as well as the communication with customers and other intermodal companies greatly determine their operational and capacity needs.

The rail network is highly interdependent, and any problem happening along the network may affect other traffic directly or indirectly. When units are delayed for delivery to the hauler, it is in almost all cases due to late arrival of trains, and the final cause is to be found in the railway system[15,16,17]. Unscheduled delays occur mostly along the lines due to weather problems, maintenance activities, technical malfunctions, interaction with other delayed trains, etc[13,14]. In turn, most late train departures from the terminal may be attributed to late arrivals. An exception is, for example, one Dutch terminal which was mentioned to have periodically late departing trains due to a shortage of staff and a great amount of demand.

The key measure that would significantly impact capacity utilization and reliability, as well as the competitiveness of intermodal transport, is related to information and cooperation. This entails both coordination with rail operators and haulers.

Most decision-making in terminals, especially regarding loading plans, are carried out manually. No DSS are used since they are considered not to take into account all the possible situations that occur in real life. It was even mentioned that the assumptions used by some of these tools, simplify the system so much that make their use irrelevant, as decisions in similar environments could be easily carried out manually by the staff[15]. Therefore, the benefits are not considered worth the investment.

---

[15] Interview with Pär Svensson. Logistics development manager, *Eskilstuna Logistik*. May 2015 and

[16] Interview with Anna Elias, Gothenburg Combiterminal, and Jonas Emanuelsson, Jernhusen Systems Manager, May 2015.

[17] Interview with Fredrik Bärthel, Consultant for society logistics WSP, May 2015.

Some of the models reviewed, assume that all actors in the terminal know the location of each unit through the terminal management system, but this is not always the case. For example, in the Gothenburg Combiterminal units are arranged in the yard by unit type, and the crane operators must know the yard layout in order to find the units related to work orders.

Two major problems have been identified regarding information sharing from the IM and rail operator the terminal. Firstly, terminals are not informed by the rail operator or the origin terminal about the actual wagon and unit arrangement of an incoming train, so the loading plan is prepared at the terminal only once the train has arrived. Depending on the type of train, loading decisions may be very complex and therefore involve large amount of time, as explained in Section **3.3.1**. Second, intermodal terminals do not receive information from the IM nor the rail company about the incoming trains´ status in the line, and are normally not notified in case a train deviates from schedule. When a train has not arrived past its scheduled time, the terminal operators may manually communicate (i.e. phone call or email) with the other actors in order to know about the train status and adjust their own operations, but do not always get this information[13]. This problems is also reported are also reported by the terminals which are part of the DIOMIS project (2007).

The second problem is a crucial one for the competitiveness of intermodal transport, as no traceability is offered to customers, and when shipment delays occur, the customer cannot be informed. In this sense, traceability can be considered as a component of reliability. Because of the nature of the rail network and the relationship between the terminal and the line, any minor delay can be multiplied and may originate important deviations from schedule. The fact of not being able to inform this on-time hinders the trust towards intermodal transport.

Second, terminal operators normally have no means of communicating with road haulers, who are in turn subcontracted by the rail companies, fowarders, and/or the intermodal operator. Therefore, the terminal does not have knowledge of the exact times that units are to be delivered and picked up, and cannot inform truck drivers about any deviation from schedule of incoming trains. Having a means of communication with the of haulers and drivers would be beneficial for coordination and operational planning[18], a statement also mentioned in the DIOMIS project (2007). In this way, the arrival of trucks could be arranged to prevent congestion at the gates and in surrounding roads, and to achieve direct transshipment.

In general, there is consensus in the fact that a wide implementation of information technologies for internal and external information is very necessary, which also corresponds to the conclusions of several literature sources. For example, the results of the PLATFORM project point out that an adoption of new transshipment equipment has no real impact unless there is an implementation of information computer systems coordinating the terminal, rail and road activities (Rizzoli et al., 2002, p. 81). This suggests that a major factor to be taken into account into the determination of practical is the use of information systems. This is because even if a terminal has a huge infrastructure, what determines it real capabilities is how it is operated as well as how it coordinates with customers and other actors in the system.

[18] Interview with Jonas Emanuelsson, Jernhusen Systems Manager, June 2015.

The fact that neither the terminal nor the intermodal operator have an automatized mean of tracking units along the rail network hinders operational planning as well as customer service. Currently, in the Swedish intermodal transport sector there is a project proposal in order to solve this problem. It consists of installing RFID detectors at certain points along rail lines, which could communicate to the intermodal transport operator the status of each unit. Nevertheless, the technology is relatively expensive.

Curiously, the train dispatcher working for the IM does have the necessary information about train status in the line. From pairing train numbers with bookings, operators could know the status of units without the need of such costly technologies. Nevertheless, they do not receive this information.

The problem of market loss and financial struggle is also mentioned. Some terminals are publicly owned or have been subsidized, and were built expecting to create new traffic volumes and establish transport routes. In Sweden, there have been a great amount of terminals developed over the last few years. Nevertheless, as train operators could not gather the volumes to fill the required train load factor, terminals have great overcapacity and barely reach their break-even point, and as a result some have already closed down. Generally, terminal operators try to offer complimentary logistic services, or to operate own trains in order to make up for the low demand[19].

3.5 Information and coordination

It was recognized from the observations and interviews, that the qualitative factors regarding a terminal have a great effect on its practical capacity, since they shape how the infrastructure and equipment is used. This is also true for any kind of production system, but especially so in the case of intermodal transport where so much coordination of different organizations is needed, as the interdependence of operations is extraordinarily high.

A widespread standard for the definitions of the different types of capacity and their implications is urgently necessary, in order to best coordinate capacity planning efforts at all levels and avoid the waste of resources. The current structure of the intermodal transport market is related to poor coordination and information sharing, which are otherwise crucial for the optimal function of the chain.

At the operational level, the most important areas where coordination affects intermodal terminal´s capacity utilization and service quality are:

- Terminal operators have no knowledge about the status of incoming trains and deviations from schedule
- Terminal operators are not given information about wagon and unit arrangement of incoming trains.
- Terminal operators do not receive information about the delivery and pick-up times of units by road vehicles
- Internal coordination and automation.

[19] Interview with Lennart Hammarbäck, Consultant for society logistics. *WSP Group*. May 2015.

- Coordination of bookings throughout the chain, from the shipper, to carrier, to intermodal transport operator, to terminal operator. This means that bookings shall be done some time in advance, and unit characteristics correctly specified.

The type of information systems that a terminal would require to have seamless operations is a Terminal Management System (TMS) which automatically imports bookings from an online booking system. Furthermore, the next level of automation would be to utilize DSS for job assignments to cranes, and for the creation of the loading plan. In terms of external coordination with other actors in the system, an Electronic Data Interchange (EDI) would be used. Bookings from customers, unit status, train status, and train manifests would be automatically updated in the TMS and DSS.

At the tactical level, there is general lack of adoption of common practices, policies, and information technologies.

At the strategic level, the lack of coordination in terminals planning and capacity expansion in terminals within a line or corridor is pointed out as a great problem. According to the International Union of Railways (UIC, 2008) congestion in one terminal does not only affect the terminal itself, but may also produce overcapacity in connected terminals due to the inability of allocating new train services in the line. When strategies to enhance infrastructure in terminals are not coordinated, capacity in terminals along a corridor is unbalanced and congestion at the bottleneck terminal, constraints intermodal operators to plan new routes or increase frequencies according to market demand. This implies a huge amount of wasted resources in the capacity expansion efforts of the enlarged terminals. In turn, when the terminals causing the bottleneck are enlarged, the demand might be already absent (UIC, 2008).

The adoption of information technologies is not only necessary for seamless operations. Information technologies automatize record keeping, allowing for operational data to be translated to useful information for strategic planning.

3.6 Relation between line and node capacity

The capacity utilization of lines and nodes have a strong co-dependency. On the one hand, the way capacity is managed in lines affects the node capacity utilization, and on the other hand, nodes are frequently mentioned as bottlenecks for line capacity, mostly in the case of congestion and delay (e.g. Mussone & Calvo, 2013, p. 13).

Perhaps the main way in which line capacity influences nodes is through the timetable. Firstly, the fact that passenger trains occupy most of the daytime capacity creates peak demand periods in rail terminals (Section **3.3.4**), and since intermodal trains are normally operating in night-leaps.

Second, due to the stiffness of the timetable, time unreliability in nodes has a disproportional impact on a train´s total travel travel time and delay, since trains that miss their designated path due to delays in the node must wait until an ad-hoc path in the line is designated to them. Therefore, delayed trains must may have to wait from hours to days in the node, occupying its capacity. In the case that one train obtains a path to leave the terminal, this does not mean that it may travel the complete route at once. Instead, it may have to wait at sidings many times along the route for each ad-hoc path to be assigned[20]. This makes it crucial for trains to maintain their connections, a fact also expressed by Bohlin et.al. (2010, p. 11).

Third, if the timetable headways within the links connected to the terminal are larger, not only is line reliability larger and practical capacity lower, but the same would likely happen whithin the terminal. Naturally, this is also subject to the kind of node in which the terminal is located, and the amount of links with which it connects. As trains arrivals become more spaced, there ir a bigger buffer available to absorb delays, so operations would become more reliable. Nevertheless, this might not be preferred by the terminal operator and terminal owner since it implies lower demand and capacity utilization.

Meanwhile, the leading way in which nodes affect line capacity happens when nodes are too congested to receive incoming trains. As these trains approach the node, they must wait at sidings in the main line, where they consume capacity and also limit passing and meeting operations for other transport flows (Dirnberger & Barkan, 2007, p. 53). This may happen in intermodal terminals, in the case that the transhipment tracks or the side yard (if there is need to change locomotives) are full. It also may also occur in marshalling yards, if the arrival yard cannot receive more trains.

Said occurrence was observed in the visit to Hallsberg Mashalling yard, where queues formed in the main line as a result of congestion in the arrival group during midnight hours. Such queues intervened with otherwise unrelated traffic, such as shuttle intermodal trains and passenger traffic. In turn, the trains contained in the arrival tracks has to wait a long time to be marshalled, as the tracks in the marshalling group were full. However, it was observed that in average only three of eight tracks of the departure yard were occuppied across time. The conclusion of the study visit was that due to a shortage of staff, some of the outgoing trains were being prepared for departure within the classification group, which involves not only the wagon preparation but also the time for path clearance and the time waiting for the arrival of the locomotive. As a result, many outgoing trains were greatly delayed. In one case, it occurred that a train leaving the yard was indirectly bloqued by the queue caused by the arriving trains.

Furthermore, according to Khoshniyat (2012), who modelled the same yard and developed an optimization program for marshalling operations, it is not possible to avoid late departures when stochastic arrivals are modelled. In other words, when arrivals deviate from the schedule, delays would always happen in the simulation results, even with optimal operations. Deviation from schedule is a problem not only in the case of late arrivals, but also in the case of early arrivals since trains consume capacity which was not planned for them. This is likely also the case for intermodal terminals.

Another way in which nodes interact with line capacity, as well as with the capacity in other nodes, is when they "push" trains out. When a node is full, it might try to dispatch a train earlier than its departure time with ad-hoc paths. Sometimes it also happens when the terminal operator wants to make sure that the train arrives on time Nevertheless, the train will disrupt operative plans in other parts of the network by consuming capacity along the lines and in the

[20] Interview with Anna Elias. Operations Manager at Gothenburg Combiterminal, *GreenCargo*. June 2015.

final terminal. In this case, a local view of optimization generates a system-wide sub-optimization.

Although it is not mentioned in literature, it is worth noting that in some cases blocking distance is much larger than the blocking distance needed the minimum headway. This is because the construction of numerous signal boxes (the apparatus controlling signaling systems) along a line represents a large investment. Therefore, some IMs might choose to establish block distances that are much longer than required[21], with the consequence that the actual length of the block section is larger than the distance needed for safe operations, i.e. minimum block length and therefore capacity is further constrained.

3.7 Type of capacity and related assumptions

It is difficult to determine which kind of capacity the review analytical methods for terminal capacity aim to calculate. Most of the line and terminal capacity methods consider that the capacity utilization that gives a reasonable level of reliability corresponds to approximately 80% of the calculated capacity (this will be called the *80% limit*). The inclusion of the 80% limit suggests that the calculated capacity corresponds to the nominal or the theoretical capacity.

Nevertheless, let us consider the UIC 406 method (Section **2.4.3.5**) (UIC, 2004). When the timetable is compressed, the percentage of time occupied by the compressed paths represents the capacity utilization. If additional paths were added in order to reach 100% capacity utilization, this does not mean that theoretical capacity is reached. This is because theoretical capacity would only happen with an ideal traffic mix, but in this method real train paths are utilized. Therefore, the limit is rather a percentage of practical capacity which is established to avoid delay. For example, in the CUI report (Sinclair Knight Merz Group, 2012), it was mentioned that certain lines presented a 100% capacity utilization, which is impossible unless this makes reference to practical capacity. The following paragraphs will examine the parameters utilized in the analytical methods for the calculation of terminal capacity (sections **2.4.4.1** to **2.4.4.3**)

In the report by Kassel+Partner (Section **2.4.4.1**) (in UIC-GTC, 2004), it is indirectly implied that the calculated capacity corresponds to nominal capacity, since operating hours are inserted in **Equations 6** and **Equations 7**, and the term *nominal capacity* is used in other sections. However, some parameters which include realistic operational conditions are inserted in the formulas, such as the flow factor in tracks (**Equations 6**), and the inclusion of empty unit runs.

In ideal conditions, which are assumed by theoretical or nominal capacity, trains and trucks would be synchronized and no split lifts would be performed. Additionally, timetable schedules would not be considered, so the track flow factor would not be applicable. As the flow factor depends upon timetables and side yard capacity, it can be regarded as more approximate to practical capacity.

In all formulas for track capacity in the analytical methods, the number of units corresponds to the number of units of incoming trains PLUS the units of outgoing trains.

[21] Interview with Dag Hersle, WSP, May 2015.

Regarding the factor of mobile crane utilization (see , **Equations 10** and **:**

), unloaded crane trips may be regarded as unavoidable, therefore this factor shall be considered for the ideal conditions. Nevertheless, the load factor used for track capacity and the factors to account for empty unit lifts shall not be considered.

It is not clear from the original formulas, whether the parameter *equipment performance* makes reference only to vertical movement (i.e. the lift itself), or if it concerns also horizontal movement, in which case it would depend on the terminal's area and layout. It will be assumed that performance equals the reverse of cycle time, which includes vertical and horizontal movement. A determination of cycle time is described by Bontekoning (2006), see Section **2.4.4.6**.

In general, the assumptions for intermodal terminal theoretical capacity shall englobe:

a) Dwell time is equal to minimum turnaround time. Therefore, a train set is never idle at the terminal; arrival, unloading, loading and departure operations instantly follow one another.
b) Transshipment moves are direct, as the arrival of trucks is coordinated with the movement of trains
c) Trains succeed each other continuously in the transshipment tracks, so that train arrivals are not subject to the timetable but to availability of tracks.
d) Each train leaves just after it is finished to be loaded and prepared for departure. In this way, train departures are not subject to the timetable.
e) All units are TEU containers of equal type and length.
f) All wagons are flat wagons designed to carry twenty-foot containers, with equal size and length.
g) Trains are of the maximum size allowable in the network and/or the transshipment tracks.
h) Optimal crane movements
i) The use of load units is optimal, and therefore there are no empty unit flows; trains are optimally full.

In this way, a theoretical capacity which assumes all ideal parameters, is useful to indicate the maximum capability of terminals. Although unreal, it is very useful for the exploration of improvement measures. Theoretical capacity shall also consider whether or not trains need to go to a side yard to change traction. Therefore, theoretical capacity shall be indicative of the capabilities of the infrastructure and superstructure.

Nominal capacity, in turn, is equivalent to the theoretical capacity adjusted to actual operating hours. It is difficult to determine whether nominal capacity should take into account train timetables, or if it should assume continuous occupation (i.e. an ideally congested terminal). To correspond to Kim et.al. and Nocera's methods, nominal capacity shall not take into account train arrival schedules according to the timetable.

Practical capacity shall include all the real-life operational and demand factors which cause the achievable output to be lower. Therefore, the following assumptions are proposed:

a) Real-time timetables, and stochastic deviations from schedule for trains.
b) The arrival and departure patters of trains and trucks are not perfectly coordinated.
c) Representative combination of units of each kind, weight and size.
d) Capability of cranes to handle units of each kind, weight and size.
e) Availability of staff
f) Must reflect the real percentage of trains that work as shuttle trains, and trains which are divided in wagon groups and need to be shunted
g) The influence of internal and external communication and coordination systems.

In general, it appears that the analytical methods combined parameters which in theory correspond to different types of capacity. In Table 3 the input factors from the method´s formulas are classified according to the type of capacity to which they correspond.

Table 3: Elements of capacity in reviewed analytical formulas

	Nominal capacity	**Practical capacity**
Global parameters	Terminal operating times. [hr/day], [day/yr]	Train arrival and departure schedules
	Crane operating hours per day and year	Train waiting time at side yard
	Train traction	Peak factor
Capacity of transshipment tracks	Length of transshipment tracks	Flow factor
	Number of transshipment tracks	Load factor [TEU/wagon]
	Length of train	
	Length of average wagon	Coefficient for empty units
	Capacity of train [TEU/train]	
	Time to receive train	
	Time to prepare train for departure	
	Time to clear track	
Equipment capacity	Performance of gantry crane	Real operating times of each crane. [hr/day], [day/yr]
	Performance of mobile crane	
	Utilization factor (factor for empty runs)	

All the analytical methods observed are based on yearly aggregated capacity. As mentioned in section 2.4.1, an aggregated capacity analysis is only appropriate in the case of products which can be stored for long periods of time, a situation much unlike the transshipment service. Due to the nature of the demand patterns in intermodal terminals, a yearly or daily average of 80% of capacity utilization—which is recommended by the analytical tools—does not necessarily avoid the occurrence of delay due to congestion. To illustrate this,

Figure 20 shows the variation in the utilization of capacity along a day (left graph). At certain points in the morning and evening, the required capacity is larger than the available practical capacity (undercapacity), which generates congestion. During the day however, the required capacity is lower than the practical capacity (overcapacity), which results in economic loss. Meanwhile, the right graph depicts the capacity utilization distributed throughout the day (supposing that area of the gray space under the curve is equal for both graphs). For this day, the average capacity utilization would be 50%—therefore, lower than the required 80%—and nevertheless, congestion could not be avoided due to the nature of train arrival and departure schedules.

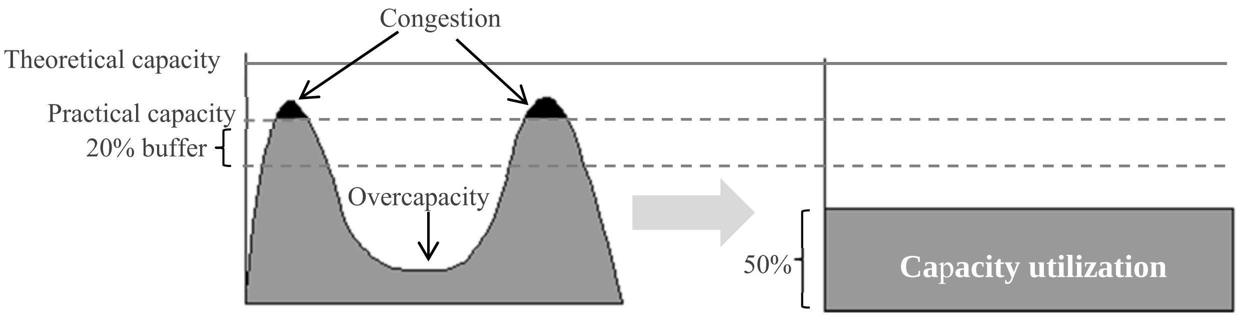


Figure 20: Real use of capacity (left) and average capacity utilization (right) along a day in an intermodal terminal

(own creation)

The UIC 406 method (see section 3.4.1.4) advises for line capacity utilization to be based on the two peak hours of a busy day, therefore capacity utilization in terminals shall also be measured in the two peak hours of a busy day of the week, or so to say, ***peak capacity utilization***.

3.8 Comparison and applicability of reviewed methods

3.8.1 Analytical methods

Analytical methods may be used whenever a quick, simple determination of capacity is needed. They are used to compare demand and capacity in a very aggregate basis, mostly using yearly volumes. Therefore, they are not appropriate when analyzing capacity at specific times of day, such as peak capacity. They shall also not be used to direct investment in terminals, except when utilized previous to simulation and optimization methods when planning new terminals, in order to have a first idea of the number of basic resources needed. On the other hand, are very useful for the understanding of capacity factors and the relationship between them.

Analytical methods have the advantage to be cheap and need little information on a terminal; therefore they could be used by third parties other than the terminal operator and owner. They could be applied, for example, by international organization or authorities within preliminary studies of large-scale transport network infrastructure capacity expansions (such as in the study by UTC-GCG, see section **2.4.4.1**).

It is tricky to compare the adequateness of each of the reviewed methods, as it would depend on the type of capacity to be analyzed, as well as the scope of the analysis. Furthermore, none of the methods could be tested by applying them to a number of terminals, as the required information was not available.

The method for track capacity (**Equations 6**) by Kassel + Partner may be most appropriate for the calculation of practical capacity in tracks, as it includes the flow factor (i.e. whether the terminal has a static or dynamic capacity). In turn, the flow factor depends on train schedules, unloading/loading times, number of transshipment tracks, and number of holding tracks. The flow factor shall not be used for the calculation of theoretical and nominal capacity, as these assume that dwell time is equal to minimum turnaround time. Furthermore, this makes the method more applicable in the case of renovation of existing terminals where the flow factor is more easily determined. On the other hand, the formulas for handling capacity ( ) may be used for theoretical, nominal or practical capacity by changing the value of adjustment factors (e.g. utilization factor and the factor for management handlings) .

The methods by Nocera and Lee et. al. for track capacity (**Equations 9, Equations 13**), seem equally fit for the calculation of theoretical or nominal capacity of tracks, as they assume that dwell time is solely dependent upon times for shunting, loading and unloading, i.e. excluding waiting and holding times. The main difference between these two is that Lee et.al give a tool intended for planning terminal capacity requirements according to demand, while Nocera's tool is intended for measuring capacity of an already existing terminal, which corresponds more to the approach of this thesis (although the difference is just based on different disposition of the variables within the equations). Furthermore, Lee et.al.'s tool has the advantage of linking loading and unloading times to the train handling times in the formulas for track capacity. The method for handling equipment capacity by Nocera (**Equations 10**) is rather difficult to apply to all terminals, as not all of them have separate equipment for internal and external operations.

Therefore, it is concluded that the formulas for equipment capacity given by Kassel+Partner ( ) are the most appropriate for equipment capacity calculation (both practical and theoretical/nominal). Regarding track capacity, Kassel+Partner's formulas for track capacity are most appropriate for the calculation of practical capacity of tracks, while Nocera's formulas for track capacity are mostly suitable for theoretical/nominal capacity of tracks, once the train handling times are linked to the times provided by equipment capacity calculation, as carried out by Lee et.al.

Of the reviewed analytical methods, the Envelope Analysis carried out by the University of Texas (Section **2.4.4.4**) is different from the others in the group, as it is an analytical empirical method. It is useful for determining the coefficients for a readily applicable capacity formula according to the amount of each type of resource, and when studying a very large group of terminals.

3.8.2 Simulation methods

The simulation methods are more flexible and may be used for the analysis of theoretical, nominal and/or practical capacity depending on the input parameters. As they are more complicated to implement, they are more suitable when infrastructure investments in a new or existing terminal or a small group of terminals is to be carried out, involving large capital investment. Therefore, the main users would be terminal owners and operators. They are more appropriate than analytical methods for the determination of practical capacity, since they are more capable of modelling realistic system behaviors, as well as capacity utilization throughout a day. According to Abril et.al., (2008) the practical capacity calculation must always be performed with simulation methods when an in-depth analysis is needed.

Of the simulation methods found (Bontekoning (2006), PLATFORM project (IT Ingegneria dei Trasporti, 1999), and SimContT (Gronalt et al., 2006)), only the last two are applicable to a road-rail terminal or group of terminals, as Bontekoning's method is more directed towards new hub terminals. Between these two tools, SimConT was not reviewed as the report for this method does not describe the way in which the intermodal terminal was abstracted as thoroughly as the PLATFORM report. Therefore, an adequate comparison of the two tools could not be carried out.

3.9 Proposed methods

There is not one unique method to be proposed, but rather a set of measures to be applied in order to calculate capacity at different levels and with different objectives.

3.9.1 Analytical methods

The scope of the proposed analytical methods includes only transshipment capacity, and not storage or gate capacity. This is because transshipment capacity is the primary factor interacting with the railway network, and analytical formulas do not relate yard and gate capacity to transshipment capacity. In reality, perhaps congestion in the driving lanes could slow down crane movement, and/or if the yard is full, the cranes have to wait for units to be picked up in order to empty a train. Nevertheless, this will not be considered as it would add too much complexity. In all cases, the bottleneck approach may be used, but not without disregarding its applicability in each individual case.

The spatial scope includes the side tracks, the transshipment tracks, the parking lanes and the storage yard. The railway network itself shall not be considered, nor the gate and the road network surrounding the terminal. The interface with the road and rail networks happens through the vehicles' arrival schedules and delay.

3.9.1.1 Theoretical and nominal capacity

Theoretical capacity and nominal capacity are unreal since ideal operational conditions are assumed. Nevertheless, for the terminal owner, it is useful to know the maximum capabilities for the infrastructure and superstructure, in which a certain amount of money was invested. This type is actually paid for, even when not fully utilized. For the terminal operator, it is useful in order to know to what extent operations could be optimized.

For an external observer, for example a train operator interested in opening new train services along a route, this type of capacity may be useful to know. If they cannot appoint a new service because the terminal is "congested", they might work together with the terminal operator in order to enhance the use of capacity. In other words, to raise the practical capacity of the terminal closer to theoretical capacity.

The following formula, adapted from (Nocera, 2008) (Section **2.4.4.2**), may be used to calculate nominal and theoretical capacity of the tracks. The same formula may be used for theoretical and nominal capacity, by assigning different values to the operating times. So to say, when calculating theoretical capacity, 24 hours per day and 365 days per year shall be considered, while nominal capacity shall use actual opening times. The number of units makes reference to the amount of incoming AND outgoing units by rail.

$$C_{rail} = n_{track} \cdot c_{train} \cdot n_{train} \cdot Ot_{\frac{day}{yr}} \qquad \left[\frac{TEU}{yr}\right] \tag{17}$$

$$n_{train} = \frac{Ot_{hr/day}}{t_a + t_b + t_d + t_e} \qquad \left[\frac{train}{track \cdot day}\right]$$

$$c_{train} = \frac{L_{train}}{L_{TEU}} \qquad \left[\frac{TEU}{train}\right]$$

$$L_{train} = \min(L_{track}, L_{max}) \qquad [\mathrm{m}]$$

$$t_b = \frac{c_{train}}{C_{equipment,hr}} \qquad \left[\frac{h}{train}\right]$$

Where:

$n_{track} = number\ of\ tracks$

$c_{train} = capacity\ of\ train\ [TEU/train]$

$n_{train} = number\ of\ trains\ per\ track\ per\ day$

$Ot_{day/year} = operating\ days\ per\ year$

$Ot_{hr/day} = operating\ hours\ per\ day$

$t_a = time\ between\ arrival\ and\ unloading\ [hr]$

$t_b = unloading\ time\ [hr]$

$t_d = time\ for\ departure\ [hr]$

$t_e = min.time\ between\ trains\ [hr]$

$L_{train} = lenght\ of\ train$

$L_{track} = lenght\ of\ transshipment\ track$

$L_{max} = maximum\ allowed\ train\ lenght\ in\ network$

The factor for empty units, the consideration of different types of wagons and unit lengths, and the time waiting for departure are removed from Nocera's formula (2008). t_a and t_e include the time to shunt and change locomotive into the formula. If there is need to get a path from the train dispatcher in order to shunt the train set from one yard to another, this waiting time is not taken into account. The time for receiving a train and shunting the train set, if needed, is to be considered in the time between arrival and unloading t_a. The time for preparing a train for departure and shunting the train set, if needed, is considered in the time for departure t_d. Furthermore, the minimum time between trains t_e, is necessary for changing of switches, preparing the track, and for safety reasons.

The following formula, adapted from (UIC-GTC, 2004) (see section **2.4.4.1**) is proposed to calculate nominal capacity of the superstructure (equipment):

$$C_{equipment,hr} = C_{gantry} + C_{mobile} \qquad \left[\frac{TEU}{hr}\right] \tag{18}$$

$$C_{equipment,yr} = \left(C_{gantry} + C_{mobile}\right) \cdot Ot_{\frac{hr}{day}} \cdot Ot_{\frac{day}{yr}} \qquad \left[\frac{TEU}{yr}\right]$$

$$C_{gantry} = N_{gantry} \cdot P_{gantry} \cdot U_{gantry}$$

$$C_{mobile} = N_{mobile} \cdot P_{mobile} \cdot U_{mobile}$$

Where:

$$U = utilization\ factor\ to\ account\ for\ unloaded\ movements$$
$$N = number\ of\ \ cranes$$
$$P = equipment\ performance\ \ [TEU/hr]$$

As stated in Section 3.7 the utilization factor is taken into account since it is assumed that even in ideal conditions, unloaded movement of cranes cannot be avoided.

The performance of the equipment shall correspond to the reverse of the average cycle time. The cycle time has been studied by Bontekoning (2006), as described in section 3.4.4.6, and includes the vertical and horizontal movement needed to perform a lift upon one load unit. In the equation, it is also assumed that the cranes operate the whole time the terminal is open.

3.9.1.2 Practical capacity

An analytical method for practical capacity can be determined by including certain factors to the formulas assigned for theoretical and nominal capacity. The formula from (UIC-GTC, 2004) (see section 2.4.4.1), is proposed for the calculation of practical capacity of tracks:

$$\boldsymbol{C_{rail} = c_{train} \cdot FF \cdot 2 \cdot n_{track} \cdot Ot_{year} \cdot f_e} \qquad \left[\frac{TEU}{yr}\right] \tag{19}$$

$$\boldsymbol{c_{train} = \frac{L_{train}}{L_{wagon}} \cdot EF \cdot LF \cdot \left(\frac{L_{unit}}{L_{TEU}}\right)} \qquad \left[\frac{TEU}{train}\right]$$

$$\boldsymbol{L_{train} = \min(L_{track}, L_{max})}$$
$$EF \leq 1, \quad FF \leq 1$$

Where:

$$L_{track} = length\ of\ transshipment\ track\ [m]$$
$$L_{wagon} = average\ length\ of\ wagon\ [m/wagon]$$
$$L_{unit} = lenght\ of\ average\ unit\ [m/unit]$$
$$LF = load\ factor\ \ [unit/wagon]$$
$$EF = empty\ factor\ \ [empty\ units/total\ units]$$
$$FF = flow\ factor, use\ of\ a\ track\ during\ the\ day\ \ [train/day]$$
$$t_{year} = terminal\ operating\ days\ per\ year\ \ [day/yr]$$
$$n_{track} = number\ of\ transshipment\ tracks$$

The flow factor shall be calculated for a busy day. It inclusion is considered as representative for practical capacity, since it reflects whether there is space in the side tracks for "clearing" trains after unloading, whether trains are shuttle trains with tight schedules, the average train turnaround times, and the average dwell times.

The load factor *LF* accounts for the fact that the size of load units is not always the same as the length of the wagons, and therefore some parts of the wagons are left empty, and/or some wagons may carry more than one unit. The *LF* shall not consider empty wagons, because that would correspond to capacity utilization, and not to practical capacity. The empty factor *EF* reflects the fact that in real life, empty units must be sent back to other terminals because of unbalanced flows.

It is assumed that the train operator sends the correct wagons for the terminal to be able to send all the booked units.

The following formulas, adapted from (UIC-GTC, 2004) are proposed for the calculation of the practical capacity of cranes:

$$C_{equipment,hr} = C_{gantry} + C_{mobile} \qquad \left[\frac{TEU}{hr}\right] \tag{20}$$

$$C_{equipment,yr} = C_{gantry} \cdot Ot_{day,g} \cdot Ot_{yr,g} + C_{mobile} \cdot Ot_{day,m} \cdot Ot_{yr,m} \qquad \left[\frac{TEU}{yr}\right]$$

$$C_{gantry} = N_{gantry} \cdot P_{gantry} \cdot U_{gantry} \cdot EF \cdot K_1 \cdot K_2 \cdot K_3$$

$$C_{mobile} = N_{mobile} \cdot P_{mobile} \cdot U_{mobile} \cdot EF \cdot K_1 \cdot K_2 \cdot K_3$$

$$K_1, K_2 \leq 1$$

Where:

$Ot_{gantry} = operating\ hours\ per\ day\ of\ gantry\ crane$

$K_1 = coefficient\ for\ internal\ coordination$

$K_2 = coefficient\ for\ direct/indirect\ transshipments$

$K_3 = coefficient\ for\ load\ unit\ mix$

$EF = empty\ factor\ \ [empty\ units/total\ units]$

The coefficient for internal coordination depends on how quickly and precisely work orders are communicated, the ability for each crane to find the place of load units in the yard, how the loading plan is performed. This depends on the information systems in use in the terminal, and on the experience of operators. The coefficient for external coordination accounts for the number of indirect and direct transshipment operations, which are mostly dependent on how coordinated the arrival of trains and trucks is. It shall also include to some extent, the number of trailers which do not need lifts from/to trucks, since their loading and unloading from trains may be considered as direct transshipment.

The coefficient for the mix of load units represents the fact that not all cranes are adapted for all weights, load unit types, and sizes.

When decisions are to be taken with the results from the last mentioned method, the 80% limit should be respected. In this respect, peak capacity utilization, should be no more than 80% of the practical capacity (corresponding to two hours). This gives a buffer to allow for an acceptable *level of reliability*.

3.9.2 Analytical empirical methods

Analytical empirical methods demand for a more specific study of each terminal, according to details of operational conditions. Their results may be used to calibrated modelling tools.

3.9.2.1 Empirical practical capacity

Perhaps the most exact method for determining the practical capacity of an existing terminal is to determine the maximum number of moves in two hours without delay. This can be determined by observing several congested two-hour intervals across several days, and selecting the period with the maximum amount of moves. The observation should include import and export units. Each indirect transshipment shall count for one move only, and each move from storage to train or vice versa, shall be included as a half move.

$$practical\ capacity = \frac{n_{moves}}{2hrs} \tag{21}$$
$$n_{moves} = n_{import} + n_{export}$$

Given total delay = 0

For units which need to be unloaded/loaded to trucks, the number of moves are:

Number of import units:

- Train-to-truck = 1 move
- Train-to-storage = 0,5 moves
- Train-to-storage-to-truck = 1 move

Number of export units:

- Truck-to-train = 1 move
- Truck-to-storage = 0,5 moves
- Storage-to-train = 0,5 moves

In the case of load units which do not need to be unloaded/to trucks (i.e. trailers), the number of moves are:

Number of import units:

- Train-to-storage = 1 move

Number of export units:

- Storage-to-train = 1 move

The conversion of the different indirect moves into direct transshipments is due to the fact that by definition, only value-added output shall be included in capacity calculations.

Ideally, the input for this method shall come from data collected by the terminal management system (TMS). With the resulting information regarding practical capacity, it would be possible to calibrate a factor *K*, which could englobe the factors $K_1 \cdot K_2 \cdot K_3$ from equations 20.

3.9.2.2 General study of capacity factors

A frontier analysis method as applied by (Anderson & Walton, 1998) (section **2.4.4.4**) is perhaps the most generic and simple to be applied (once it is actually developed), as its application would require a small amount of information. Nevertheless, its development would need a large pool of terminals in order to retrieve correct coefficients, which might prove complicated.

In this method, the practical capacity (calculated as in previous section) of the studied group of terminals would be plotted against one type of production input (side yard, number of cranes, etc.). The line corresponding to the upper envelope would determine the upper limit of practical capacity of the resource.

3.9.3 Simulation method

In general, the terminal module of the PLATFORM model seems the most appropriate of the simulation methods, despite being older. It is possible that the SimContT method is very similar, although it was not reviewed in more detail in the thesis as explained in section **3.8.2**. In the following paragraphs, several comments are made regarding the PLATFORM method and he adjustments that can be made.

The positive aspects of the model are:

- It takes into account status of outgoing train vs truck arrival, and the possibility to calibrate truck-train arrival coordination.
- The arrival by truck is realistic in the sense that it considers the status of the outgoing train, and the decisions taken on this basis. Nevertheless, the time limit for situating a unit in the parking lane or in storage is not specified.
- It takes into account the fact that tracks must be cleared before train arrival
- It gives the option of changing unit´s priority

Factors that could be modified:

- The platform planner indicates a train to leave the terminal as soon as it has been unloaded, which disregards timetables. This should be adjusted to allow for considering timetables, if it is to be used for the calculation of practical capacity.
- The model requires that at least one gantry crane should be active at each time because there might be storage areas which other cranes might not reach. However, this is only true whenever there are "trapped" rail transshipment tracks where mobile cranes cannot access.
- It is assumed that terminals do not wait for late unit arrivals, which does not always hold true in reality.
- Stochastic delays and early arrivals are not considered, and should be included in the arrival and departure patterns of vehicles.
- Does not seem to take into account the type of traction and if the yard is electrified or not; it assumes trains can enter directly as soon as a track is available.
- Seems to assume that all units are equal, so that all units can be lifted by all cranes and loaded in any wagon. Input parameters should include the type of load unit, maximum weight, and unit length that each crane can carry. I turn, the simulator should be able to assign different load units to different cranes.
- It should take into account that some trailers do not need to be loaded/unloaded from road vehicles.
- In the case of the arrival by train, the platform planner in the model must ask the yard planner where the unit may be stored. However, in reality the crane operator might take this decision himself. For example, in GCT, units are arranged in the yard according to unit type. Therefore, the crane operator already knows where a unit must be placed, without the need to ask the terminal office

Other considerations:

- If the terminal is located in a same (macroscopic) node as a complicated marshalling yard, this yard may be considered as an external source of delay, which may represented through deviations inserted in the arrival schedule of trains.
- The model should consider also trains with wagon groups that need shunting. Furthermore, the side yard should not be seen only as a place for arrival/dispatch of trains, but also
- It is necessary for the side tracks to be considered not only as a place to receive and dispatch trains, but also , since side track capacity gives the possibility of “clearing” the unloaded wagons and therefore have a larger load factor.

The road gate is not considered in the scope, as the service time of trucks is not calculated because the model shall focus more on transshipment capacity.

It is important to remark that the objective of the used method will NOT be to improve capacity/performance by optimizing procedures (operations, administration etc.), infrastructure or both. It will solely provide information about the practical capacity of a terminal as dependent of the relevant factors. For example, if an optimization exercise was to be done, the used method shall provide the data needed for setting the constraints (or a part thereof), but not to feed the objective function itself. In other words, it will provide valuable data for further use in decision making.

A method for the calculation of theoretical capacity is expected to provide numerical results about a terminal´s capacity, which have intrinsically no specific user or scope of application. However, this numerical result is meant to be further utilized as entry data on subsequent analysis and decision-making processes. These analyses, depending on the actor by which they are executed, will have a strategical, tactical, or operational scope.

3.9.4 Summary of capacity factors

Table 4 provides a short summary of the factors which have been mentioned in the Results and analysis section, grouped by categories. Note that it is not within the scope to specifically suggest best practices, but solely to indicate the different characteristics that influence terminal capacity.

Table 4: Summary of capacity factors

Factor type	*Factor*	*Variations*
Demand factors (external)	Mix of load units	Mostly containers Mixed Mostly trailers and swap bodies
	Train timetables	
	Wagons with dangerous goods	
	Maximum allowed train length in network	
	% of shuttle trains vs. trains with wagon cuts	
Length Infrastructure factors	Number of transshipment tracks	
	Number of side tracks	
	Length of transshipment tracks	
	Length of side tracks	

	Electrical access to terminal	
	Connectivity of side yard to the terminal	
	Yard size (affects size of storage yard and number of parking lanes)	
Superstructure factors	"Mix" of cranes (type of load units they serve)	
	Number of gantry cranes	
	Number of mobile cranes	
	Performance (speed) of cranes	
	Lifting (weight) capacity of cranes	
Policies	Prioritization of incoming trains	FIFO On-time trains first
	In-store empty wagon keeping (dynamic vs. static capacity)	
	Allowance for late units	
	Physical organization of units within yard	
Communication technologies	Internal coordination	a. Radio/walkie-talkie b. TMS c. TMS with DSS d. TMS, DSS, automatic cranes
	External coordination with transport chain actors (rail operator, intermodal operator, hauler)	a. Mail/email b. Online platform c. Online platform + TMS d. EDI + TMS e. EDI + TMS + DSS

3.10 Conclusions and further research

In this thesis, the internal and external factors affecting road-rail intermodal terminal capacity were studied with the objective to propose a method which could be used to calculate terminal capacity. For this purpose, a framework was developed in order to describe the road-rail intermodal system, terminal functioning, and capacity concepts. Furthermore, existing models for the determination of rail line and intermodal terminal capacity were reviewed, with the purpose of studying their abstraction of the real system, and to determine the relation between line and node capacity. The focus on external factors was given to railway transport, and the effect of the road transport network was only taken into account in the nature of road vehicle arrivals. Next, seven interviews and two study visits were performed, in order to better structure the framework, and to provide empirical input for analysis.

During the analysis, the real system is compared and contrasted with the reviewed models. A differentiation between theoretical, nominal and practical capacity was performed, as well as an explicit definition of the peak factor and capacity utilization concepts. Based on the methods reviewed, two analytical methods were proposed which can be used to calculate theoretical, nominal and practical capacities at a very rough level. The thesis also proposes two analytical empirical methods which can be used on pair. One tool can be used measure the real practical capacity in an existing terminal, and another method to compare the practical capacity across many terminals. Such a cross-study of a number of terminals would allow determining more accurately the effect of each resource on capacity, including information technologies.

Finally, one of the simulation methods previously reviewed is proposed as the tool to use whenever a detailed analysis is needed, i.e. when the output is to be used for big scale investments. Several recommendations are given for an increased applicability of the method.

A relevant finding of this thesis is the big influence that information technologies and coordination have on a terminal´s practical capacity. It is considered that fixed resources (i.e. infrastructure and superstructure), determine theoretical capacity. Nevertheless, theoretical capacity is only useful to determine the maximum, ideal output that a system could achieve. What represents its actual capabilities, in turn, is practical capacity, which is shaped according to how fixed resources are utilized. Furthermore, the way how resources are used is mostly determinant on operational policies, train and road schedules, and coordination with other actors, which is enhanced by information technologies.

In general, the lack of standardization in terms of terminology and operational concepts used by authors and industry actors across different backgrounds, affects information and knowledge transmission. Furthermore, if intermodal transport is to be established as a mature research area, there is a need for common fundamental concepts, and to translate a lot of empirical knowledge into explicitly described systems.

Currently, the actors involved in the intermodal system regard each other as competitors and not as business partners. This affects the strategic developments of a network with balanced capacity, as well as the flow of information needed for streamlined operations. Nominal capacity is a wasted asset unless coordination measures are applied to raise practical capacity. In this sense, the reports by the UIC “Best practices for the management of combined transport terminals” (in DIOMIS, 2007) and “International co-ordination of combined transport terminal development” (UIC, 2008) provide examples and suggestions on how cross-European coordination efforts could be conducted.

The issues of standardization, coordination and knowledge transmission provide for a vast source of research yet to be carried out. This is not only through the point of view of information technologies and business, but also of the social sciences. Further research could focus on problems such as the need to reconcile private and public interests in different planning levels, as well as the provision of incentives for cooperation between actors. Another line of research could be focused on the potential of market share rise due to the establishment of alternative bundling schemes made possible with the implementation of information technologies. This is because the concept of new hub terminals is only applicable when operations are completely coordinated.

Finally, this thesis has a limited scope in that it does not thoroughly explore storage yard and gate capacity, as well as the operations on the road network side. Furthermore, it is possible that certain capacity factors in terminals were not explored due to time and resource limitations. A more thorough study could explore the road, rail and terminal systems altogether, and could also include an in-depth exploration of the tradeoff and optimization of cost versus capacity factors in the system.

4 References

Abril, M., Barber, F., Ingolotti, L., Salido, M. a., Tormos, P., & Lova, a. (2008). An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review*, *44*, 774–806. doi:10.1016/j.tre.2007.04.001

Ahrens, G., Dellmann, T., Gies, S., Hecht, M., Hefazi, H., Henke, R., … Tegel, O. (2009). Railway Systems – Railway Engineering. In K.-H. Grote & E. K. Antonsson (Eds.), *Springer Handbook of Mechanical Engineering* (pp. 1011–1147). California, Magdeburg: Springer. doi:9783540307389

Anderson, K. M., & Walton, C. M. (1998). *Evaluating Intermodal Freight Terminals: a Framework for Government Participation*. *Research Report SWUTC/98/467505-1*. Retrieved from http://library.ctr.utexas.edu/digitized/SWUTC/467505-1.pdf

Arcot, V. C. (2007). *Modeling uncertainty in rail freight operations: Implications for service reliability*. *University of Maryland*. Maryland: ProQuest, UMI Dissertations Publishing. doi:1450224

Armstrong, J., Preston, J., Khadem Sameni, M., Potts, C., Khosrav, B., Bektas, T., & Bennell, J. (2010). Overcoming the Capacity Constraints Imposed by Nodes on Railway Networks. *University of Southampton*. Retrieved from http://www.academia.edu/2824404/Overcoming_the_Capacity_Constraints_Imposed_by_Nodes_on_Railway_Networks

ARTC. (2015). Proposed Ambleside Rail Crossing Loop. *Australian Rail Track Corporation Ltd.* Retrieved from http://www.artc.com.au/library/Ambleside Loop - FAQs.pdf

Ballis, A., & Golias, J. (2002). Comparative evaluation of existing and innovative rail-road freight transport terminals. *Transportation Research Part A: Policy and Practice*, *36*, 593–611. doi:10.1016/S0965-8564(01)00024-6

Ballis, A., & Golias, J. (2004). Towards the improvement of a combined transport chain performance. *European Journal of Operational Research*, *152*, 420–436. doi:10.1016/S0377-2217(03)00034-1

Barber, F., Ingolotti, L. P., Tormos, P., Lova, A., Management, R., & Analysis, C. (2007). *Survey of automated Systems for Railway Management: technical Report*. Retrieved from http://www.dsic.upv.es/docs/bib-dig/informes/etd-01152007-140458/AutomatedSystems.pdf

Bärthel, F., Östlund, B., & Flodén, J. (2011). MINT Deliverable 1. State-of-the-Art. *MINT – Model and Decision Support Systems for Intermodal Terminal Networks*. Retrieved from http://www.sir-c.se/getfile.ashx?cid=257578&cc=3&refid=17

Beck, A., Bente, H., & Schilling, M. (2013). *Railway Efficiency*. *International Transport Forum, OECD*. Retrieved from http://www.internationaltransportforum.org/jtrc/DiscussionPapers/DP201312.pdf

Behrends, S. (2006). *The modal shift potential of intermodal line-trains from a haulier ' s perspective*. Gothenburg.

Behrends, S., & Flodén, J. (2012). The effect of transhipment costs on the performance of intermodal line-trains. *Logistics Research*, *4*, 127–136. doi:10.1007/s12159-012-0066-0

Behrends, S., Liljestrand, K., Bäckström, S., & Pahlén, P.-O. (2011). Intermodal transport from a haulier's perspective. *SIR-C – Swedish Intermodal Transport Research Centre*.

Bergqvist, R. (2008). Evaluating road-rail intermodal transport services - A heuristic approach. *International Journal of Logistics*, *11*(781071018), 179–199. doi:10.1080/13675560701633273

Bergqvist, R. (2014). The role of contracts in achieving effective governance of intermodal terminals Jason Monios. *World Review of Intermodal Transportation Research*, *5*(1), 18–38.

Birdsell, L. B. (1985). *Evaluating intermodal truck/rail terminal operations*. University of Waterloo, Canada.

Bohlin, M., Dahms, F., Flier, H., & Gestrelius, S. (2010). Optimal Freight Train Classification using Column, 10–22.

Bontekoning, Y. M. (2006). *Hub exchange operations in intermodal hub-and-spoke networks*. TRAIL Research School; TU Delft.

Bontekoning, Y. M., Macharis, C., & Trip, J. J. (2004a). Is a new applied transportation research field emerging?—A review of intermodal rail–truck freight transport literature. *Transportation Research Part A: Policy and Practice*, *38*(1), 1–34. doi:10.1016/j.tra.2003.06.001

Bontekoning, Y. M., Macharis, C., & Trip, J. J. (2004b). Is a new applied transportation research field emerging?—A review of intermodal rail–truck freight transport literature. *Transportation Research Part A: Policy and Practice*, *38*(1), 19. doi:10.1016/j.tra.2003.06.001

Bontekoning, Y. M., & Trip, J. J. (2004). Rail-rail Hub-terminals as an Alternative for Shunting: An Explorative Comparative Case Study. *Transport Developments and Innovations in an Evolving World Advances in Spatial Science*, 235–251.

Boysen, H. E. (2013). General model of railway transportation capacity. In A. Brebbia, N. Tomii, P. Tzieropoulos, & B. Ning (Eds.), *Computers in railways XIII*. Southampton: WIT Press. Retrieved from https://books.google.se/books?id=f7eE3ZaTdMcC&pg=PA337&lpg=PA337&dq=Capacity+is+a+measure+of+the+ability+to+move+a+specific+amount+of+traffic+over+a+defined+rail+line+with+a+given+set+of+resources+under+a+specific+service+plan.+(&source=bl&ots=a6vwKm4nao&

Boysen, N., Fliedner, M., Jaehn, F., & Pesch, E. (2012a). A Survey on Container Processing in Railway Yards. *Transportation Science*, (March 2015). doi:10.1287/trsc.1120.0415

Boysen, N., Fliedner, M., Jaehn, F., & Pesch, E. (2012b). Shunting yard operations: Theoretical aspects and applications. *European Journal of Operational Research*, *220*(1), 1–14. doi:10.1016/j.ejor.2012.01.043

Bromma. (n.d.). *RSX40C combi spreader*. Retrieved from http://www.bromma.com/show.php?id=1138811

Brown, S., Bessant, J., & Lamming, R. (2013). *Strategic Operations Management* (3rd ed.). New York: Routledge.

Burdett, R. L., & Kozan, E. (2006). Techniques for absolute capacity determination in railways. *Transportation Research Part B: Methodological*, *40*, 616–632. doi:10.1016/j.trb.2005.09.004

CargoNet. (2012). *Requirements for containers, swap bodies and semitrailers for transport by CargoNet*. Retrieved from http://www.cargonet.no/Global/dokumenter/dok ny web/Publikasjoner/engelsk/June 2012_Requirements for containers swab bodies and semitrailers for transport by CargoNet.pdf

Chandra, S., & Agarwal, M. M. (2001). *Railway Engineering*. New Delhi: Oxford University Press.

Christoph, S. (2010). *40 years of Road-Rail Combined Transport in Europe*. Frankfurt: International Union of Combined Roa-Rail transport companies.

Clarke, D. B. (1995). *An examination of railroad capacity and its implications for rail-highway intermodal transportation*. *UMI Dissertations Publishing*. The University of Tennessee.

COSMOS. (2015). *Intermodal Terminals*. *Cooperative Solutions for Managing Optimized Services*. Retrieved from http://www.intermodal-cosmos.eu/content/intermodal-transport-in-south-east-europe/intermodal-basics/intermodal-terminals/index_eng.html

Crainic, T. G., & Bektas, T. (2007). A Brief Overview of Intermodal Transportation. *Logistics Engineering Handbook*, (January), 1–16. doi:doi:10.1201/9781420004588.ch28

Crainic, T. G., & Kim, K. H. (2007). Chapter 8 Intermodal Transportation. *Handbooks in Operations Research and Management Science*, *14*(06), 467–537. doi:10.1016/S0927-0507(06)14008-6

Damij, N., & Damij, T. (2014). *Process Management: A Multi-disciplinary Guide to Theory, Modeling, and Methodology*. Springer.

Department of Transport. (2013). *Private sidings*. *Western Australia Department of Transport*. Retrieved from http://www.transport.wa.gov.au/freightrail/private-sidings.asp

Dettoni, J. (2014). Mongolia's railway to China turns into national security threat. *BNE Intellinews*. Retrieved from http://www.bne.eu/content/story/mongolias-railway-china-turns-national-security-threat

Devore, J. (2014). *Probability and Statistics for Engineering and the Sciences* (9th ed.). Cengage learning.

DIOMIS. (2007). *Best practices for the management of combined transport terminals*. International Union of Railways. Retrieved from http://www.uic.org/diomis/IMG/pdf/DIOMIS4_final.pdf

Dirnberger, J., & Barkan, C. (2007). Lean Railroading for Improving Railroad Classification Terminal Performance: Bottleneck Management Methods. *Transportation Research Record*, *1995*(1995), 52–61. doi:10.3141/1995-07

EC. (2001). WHITE PAPER: European transport policy for 2010: time to decide. *Commission of the European Communities*, (2001), 124. doi:9289403411

EC. (2011). *White paper. Roadmap to a Single European Transport Area–Towards a competitive and resource efficient transport system*. *Office for Official Publications of the European ….* Brussels. doi:10.2832/30955

EC. (2013). *Mobility and transport: rail*. *European comission*. Retrieved from http://ec.europa.eu/transport/modes/rail/market/index_en.htm

EC. (2015). *Road transport: Reducing CO2 emissions from vehicles*. *Climate Action*. Retrieved February 20, 2015, from http://ec.europa.eu/clima/policies/transport/vehicles/index_en.htm

EEA. (2014). *Focusing on environmental pressures from long distance transport*. Luxembourg: European Environmental Agency. doi:10.2800/857401

EUROSTAT. (2009). *Panorama of Transport 2009*. doi:10.2785/28475

Ferreira, L., & Kozan, E. (1992). Intermodal terminals. *Australian Transportation Research Forum*.

Flodén, J. (2007). *Modelling Intermodal Freight Transport - The Potential of Combined Transport in Sweden*.

Flodén, J. (2011). *MINT Deliverable 2. Framework for Evaluation*.

Gatto, M., Maue, J., Mihalák, M., & Widmayer, P. (2009). Shunting for dummies: An introductory algorithmic survey. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5868 LNCS*, 310–337. doi:10.1007/978-3-642-05465-5_13

Gibson, S. (2003). Allocation of capacity in the rail industry. *Utilities Policy*, *11*, 39–42. doi:10.1016/S0957-1787(02)00055-3

Golinska, P., & Hajdul, M. (Eds.). (2012). *Sustainable Transport: New Trends and Business Practices*. Berlin: Springer Science & Business Media. Retrieved from https://books.google.se/books?id=h5z8V8WzmioC&pg=PA277&lpg=PA277&dq=shunting+yard+operations+sustainability&source=bl&ots=DvvNQkAESr&sig=Ho8gDPuzwPl8ReK

LPM_vGeNs4wc&hl=es-419&sa=X&ei=fvTlVLuqGIukygO6sIDQBA&ved=0CC0Q6AEwAg#v=onepage&q=rail&f=false

Goverde, R., Corman, F., & D'Ariano, A. (2013). Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions. *Journal of Rail Transport Planning and Management*, *3*, 78–94. doi:10.1016/j.jrtpm.2013.12.001

Greasley, A. (1999). *Operations Management in Business*. United Kingdom: Stanley Thornes.

Gronalt, M., Benna, T., & Posset, M. (2006). SimConT Simulation of hinterland Container Terminal operations. *Complexity Management in Supply Chain Concept*, (August), 227–246. Retrieved from http://books.google.com/books?hl=en&lr=&id=5Z82-IVi8GAC&oi=fnd&pg=PA227&dq=SimConT+Simulation+of+hinterland+Container+Terminal+operations&ots=02WQkp5Xxy&sig=Zf7SpcsASodI1_n3WL7Huw74gOM

Gudehus, T., & Kotzab, H. (2012). *Comprehensive logistics* (2nd ed.). Hamburg: Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-642-24367-7

Gütle, H., & Drewello, S. (2013). The need for investment on the "Rheintalbahn" in the Upper Rhine Valley – Results of a railway capacity analysis near the town of Lahr. *University of Public Administration in Kehl*, (August).

Hilmola, O. P., Ujvari, S., & Szekely, B. (2007). Deregulation of railroads and future development scenarios in Europe: analysis of the privatisation process taken place in the USA, the UK and Sweden. *World Review of Intermodal Transportation Research*, *1*(2), 146. doi:10.1504/WRITR.2007.013948

Indian Railway Employee. (2015). *Line capacity*. Retrieved from http://www.indianrailwayemployee.com/content/line-capacity

IT Ingegneria dei Trasporti. (1999). *PLATFORM: final report for publication*. COMMISSION OF THE EU ROPEAN COMMUNITIES; Partners: DFKI, ETRA, FIT, IDSIA, ADL, CEMAT, CONTSHIP, DSP, ESL, NOVATRANS. Retrieved from http://www.transport-research.info/Upload/Documents/200310/platform.pdf

James A. Tompkins, John A. White, Yavuz A. Bozer, J. M. . A. T. (2010). *Facilities planning*. John Wiley & Sons. Retrieved from https://books.google.se/books?id=-xBIq6Qm2SQC&dq=tanchoco+et+al.+recognizing+the+relationship+that+apple+alluded+to,+defined+a+unit+load+as&source=gbs_navlinks_s

Jonsson, P. (2008). *Logistics and Supply Chain management*. McGraw-Hill.

Katsarova, I. (2014). *Fourth Railway Package Still Divides Member States*. *European Parliamentary Research Service*. Retrieved from http://epthinktank.eu/2014/12/03/fourth-railway-package-still-divides-member-states/

Khadem Sameni, M., Landex, A., & Preston, J. (2005). *Developing the UIC 406 Method for Capacity Analysis*.

Khoshniyat, F. (2012). *Simulation of Planning Strategies for Track Allocation at Marshalling Yards*. KTH.

Koenig, R. (2001). GERMAN RAILROAD MAY SCALE BACK INTERMODAL TERMINALS. *The Journal of Commerce*. Retrieved from http://www.joc.com/economy-watch/german-railroad-may-scale-back-intermodal-terminals_20010107.html

Kreutzberger, E. (2010). Lowest cost intermodal rail freight transport bundling networks: Conceptual structuring and identification. *European Journal of Transport and Infrastructure Research*, *10*(10), 158–180.

Lai, Y.-C., Huang, Y.-A., & Chu, H.-Y. (2014). Estimation of rail capacity using regression and neural network. *Neural Computing and Applications*, *25*(7-8), 2067–2077. doi:10.1007/s00521-014-1694-x

Landex, A., & Kaas, A. H. (2006). Evaluation of railway capacity. *Annual Transport Conference at Aalborg University 2006*, 1–22.

Law, J. (2009). Effective Capacity. In *A Dictionary of Business and Management*. Oxford University Press. doi:9780191726545

Lee, B. K., Jung, B. J., Kim, K. H., Park, S. O., & Seo, J. H. (2006). A simulation study for designing a rail terminal in a container port. *Integration The Vlsi Journal*, 2180–2187.

León, M. (2000). *Diccionario de tecnología ferroviaria*. Madrid: Babel. Retrieved from https://books.google.se/books?id=1CSdAAUJlBgC&pg=PA574&lpg=PA574&dq="direct-access+siding"+rail&source=bl&ots=ApW8Xxw_gz&sig=hxfK1ze0sTs79qPcpLGyI2ZRYNg&hl=es-419&sa=X&ei=WZojVb-yFYmVsgHQ-4GIDA&ved=0CCMQ6AEwAA#v=onepage&q=direct access siding&f=false

Liilennevirasto. (2012). *Organizing the rail-related services in deregulated railway markets – Learning from Swedish, Danish and British experiences*. Retrieved from http://www2.liikennevirasto.fi/julkaisut/pdf3/lts_2012-26_organizing_the_rail-related_web.pdf

Lindner, T. (2011). Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity. *Journal of Rail Transport Planning and Management*. doi:10.1016/j.jrtpm.2011.09.002

Lowe, D. (2005). *Intermodal freight transport* (ebook.). Amsterdam; Boston: Elsevier Butterworth-Heinemann.

Lumsden, K. (2007). *Fundamentals of logisics*. Gothenburg: Chalmers University of technology.

Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, *4*(3). doi:http://dx.doi.org.proxy.lib.chalmers.se/10.1057/jos.2010.3

Macharis, C., & Bontekoning, Y. M. (2004). Opportunities for OR in intermodal freight transport research: A review. *European Journal of Operational Research*, *153*, 400–416. doi:10.1016/S0377-2217(03)00161-9

Marinov, M., Giovanni, L. Di, Bellisai, G., Clevermann, J., Mastellou, A., Victória, D., … Deleva, L. (2014). Analysis of rail yard and terminal performances. *Journal of Transport Literature*, *8*(2), 178–200. doi:ISSN 2238-1031

Marinov, M., Sahin, I., Ricci, S., & Vasic-Franklin, G. (2013). Railway operations, time-tabling and control. *Research in Transportation Economics*, *41*, 59–75. doi:10.1016/j.retrec.2012.10.003

Marinov, M., & Viegas, J. (2011). Tactical management of rail freight transportation services: evaluation of yard performance. *Transportation Planning and Technology*, *34*(May 2015), 363–387. doi:10.1080/03081060.2011.577155

Mattsson, L.-G. (2007). Railway Capacity and Train Delay Relationships. *Critical Infrastructure: Advances in Spaital Science*. Retrieved from http://www.springerlink.com/index/L4455467T0278714.pdf

Mayhew, S. (2009). Network. In *A Dictionary of Geography*. Oxford University Press. doi:10.1093/acref/9780199231805.001.0001

Mc.Nair, C., & Richard Vangermeersch. (1998). *Total Capacity Management: Optimizing at the Operational, Tactical, and Strategic Levels*. IMA Foundation for Applied Research; CRC Press LLC.

Morant, a, & Westerberg, M. (2014). *Knowledge Management in a Railway Network : The Case of Signalling Systems*. Luleå. Retrieved from Lule?

Mussone, L., & Calvo, R. W. (2013). An analytical approach to calculate the capacity of a railway system. *European Journal of Operational Research*, *228*(1), 11–23. doi:10.1016/j.ejor.2012.12.027

NetworkRail. (2010). *Route Plan 2010*. Retrieved from http://www.networkrail.co.uk/RoutePlans/PDF/RouteA-Kent.pdf

Nilsson, J. (2011). *Ownership of Real Estate and Terminals in a Railway Industry with Open Entry*. *VTI*.

Nocera, S. (2008). Terminalkapazitäat im kombinierten verkehr. *Internationales Verkehrwesen*, *60*(11).

OECD. (2010a). *Globalisation, Transport and the Environment*. Brussels: Organization for Economic Cooperation and Development. doi:978-92-64-07291-6

OECD. (2010b). *Globalisation, Transport and the Environment*. Brussels: Organization for Economic Cooperation and Development. doi:978-92-64-07291-6

Poole, E. (1962). *Costs- a tool for railroad management*. New York: Simmons-Broadman Publishing Corporation.

Poot, J. (2004). *On the Edge of the Global Economy*. MPG Books.

Pouryousef, H., Lautala, P., & White, T. (2015). Railroad capacity tools and methodologies in the U.S. and Europe. *Journal of Modern Transportation*, *23*(February), 30–42. doi:10.1007/s40534-015-0069-z

PPIAF. (2015). *Technology*. *Railway reform: Toolkit for improving rail sector performance*. Retrieved from http://www.ppiaf.org/sites/ppiaf.org/files/documents/toolkits/railways_toolkit/ch1_3.html

Prentice, B. E. (2003). Importance of Intermodal Connectivity and Bottleneck Elimination. *Transport Institute, University of Manitoba*, 1–10. Retrieved from https://umanitoba.ca/faculties/management/ti/media/docs/CTRF03_Bottleneck.pdf

Public Transport Corporation. (1999). *Track Technical Definitions*. Victoria. Retrieved from http://ptv.vic.gov.au/assets/PTV/PTV docs/VRIOGS/Track-Technical-Definitions.pdf

Railway Technical Web Pages. (2014). *The development and principles of UK signalling*. *RAILWAY SYSTEMS, TECHNOLOGIES AND OPERATIONS ACROSS THE WORLD*. Retrieved from http://www.railway-technical.com/sigtxt1.shtml

Reinach, S., & Gertler, J. (1998). *Training Requirements for Railroad Dispatchers: Objectives, Syllabi and Test Designs*. Washington. Retrieved from http://ntl.bts.gov/lib/10000/10300/10376/ord9808.pdf

RFF. (2009). *Who's who in the railway sector*. Retrieved from http://www.rff.fr/en/our-company/who-s-who-in-the-railway-sector

Rizzoli, A. E., Fornara, N., & Gambardella, L. M. (2002). A simulation tool for combined rail/road transport in intermodal terminals. *Mathematics and Computers in Simulation*, *59*, 57–71. doi:10.1016/S0378-4754(01)00393-7

Rodrigue, J.-P., Comtois, C., & Slack, B. (2013). *The Geography of Transport Systems* (3rd ed.). Routledge.

Rowangould, G. (2013). Public financing of private freight rail infrastructure to reduce highway congestion: A case study of public policy and decision making in the United States. *Transportation Research Part A: Policy and Practice*, *57*, 25–36. doi:10.1016/j.tra.2013.09.007

Rudjanakanoknad, J., Deethanasuwan, P., Wongchaivit, P., & Engineering, C. (2013). Time-Space Diagram Models for Single- and Double-track Railway System Capacities in Thailand, *9*.

Sherman, D., & Zhu, J. (2006). DATA ENVELOPMENT ANALYSIS EXPLAINED. In *Service Productivity Management*. Springer. doi:10.1007/0-387-33231-6

Sinclair Knight Merz Group. (2012). *Assessment of capacity allocation and utilisation on capacity constrained parts of the GB rail network*. London. Retrieved from http://orr.gov.uk/__data/assets/pdf_file/0017/467/skm-capacity-utilisation-aug-2012.pdf

Slack, N., Chambers, S., & Johnston, R. (2013). *Operations management* (7th ed.). Pearson.

Smith, R. a. (2003). Railways: How they may contribute to a sustainable future. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *217*, 243–248. doi:10.1243/095440903322712847

SNCF. (2007). Appendix 2- Glossary. In *National Rail Network Statement- 2008 Timetable*. Paris. Retrieved from file:///C:/Users/SEAH16292/Downloads/uk_docref_0_anx2 (1).pdf

Sommar, R., & Woxenius, J. (2007). Time perspectives on intermodal transport of consolidated cargo. *European Journal of Transport and Infrastructure Research*, *7*, 163–182.

Steadieseifi, M., Dellaert, N. P., Nuijten, W., Van Woensel, T., & Raoufi, R. (2014). Multimodal freight transportation planning: A literature review. *European Journal of Operational Research*, *233*(1), 1–15. doi:10.1016/j.ejor.2013.06.055

Stenström, C. (2012). *Performance Measurement of Railway Infrastructure with Focus on the Swedish Network*. *Program*. Retrieved from https://pure.ltu.se/ws/files/36836951/MPM_of_Railway_Infrastructure.pdf

Taha, H. A. (2007). *Introduction to Operations Research*. (M. J. Horton, H. Stark, V. O'Brien, & D. A. George, Eds.) (8th ed.). London, Sydney: Pearson Prentice Hall. Retrieved from http://moodle.sumdu.edu.ua/taha_h_a_operations_research_an_introduction.pdf

Trafikverket. (2012). *Network Statement 2014*. Stockholm. Retrieved from http://www.trafikverket.se/contentassets/1fea27da4b5f4f3cb809ce7c82293d22/3_infrastructure_2014_03_24.pdf

TRB. (1998). *Multimodal Corridor and Capacity Analysis Manual, Issue 399*. Cambridge: Transportation Research Board. Retrieved from https://books.google.se/books?id=0WH_-RXiweUC&pg=PA67&lpg=PA67&dq=relationship+line+node+rail+capacity&source=bl&ots=Ws7aVS2K6J&sig=vKxE_FDjOqwJqJ0vF55iGoXx4YQ&hl=es-419&sa=X&ei=P8ECVbDYD4nXyQPSkYDQCg&ved=0CCAQ6AEwAA#v=onepage&q=relationship line node rai

TRB. (2003). Rail Transit Capacity. In *Transit Capacity and Quality of Service Manual* (2nd ed.). Washington, D.C.: Transportation Research Board.

TRB. (2013a). Operations concepts. In *TCRP Transit Capacity and Quality of Service Manual* (3rd ed.). Washington, D.C.: Transit Cooperative Highway Research Program (TCRP) Report. Retrieved from http://www.trb.org/main/blurbs/169437.aspx

TRB. (2013b). Rail transit capacity. In *Transit Capacity and Quality of Service Manual—2nd* (3rd ed.). Washington D.C.: Transportation Research Board. Retrieved from http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_165ch-08.pdf

Tstita, K., & Pilavachi, M. (2013). Evaluation of next generation biomass derived fuels for the transport sector. *Energy Policy*, *62*, 443–455.

U.S. Congress. (1984). THE PROBLEM OF CAPACITY AND DELAY. In *Airport system development*. Washington: U.S. Government Printing Office. Retrieved from https://www.princeton.edu/~ota/disk3/1984/8403/840305.PDF

UIC. (n.d.). *Block Train / Full Train*. *UIC Rail Freight Portal*. Retrieved from http://www.railfreightportal.com/Block-Train-Full-Train

UIC. (2004). *Uic code 406*. *International Union of Railways*. Retrieved from http://banportalen.banverket.se/Banportalen/upload/1753/HandbokUIC406.pdf

UIC. (2008). *International co-ordination of combined transport terminal development*. Retrieved from http://www.uic.org/diomis/IMG/pdf/Livre_DIOMIS_A8.pdf

UIC. (2012). *Combined Transport in Europe*. Paris. Retrieved from http://www.uic.org/IMG/pdf/2012_report_on_combined_transport_in_europe.pdf

UIC-GTC. (2004). *Infrastructure Capacity Reserves For Combined Transport By 2015*. Freiburg / Frankfurt am Main / Paris.

UNECE. (2001). *Terminology on combined transport*. Geneva. Retrieved from http://www.unece.org/fileadmin/DAM/trans/wp24/documents/term.pdf

UNECE. (2010). *Illustrated Glossary for Transport Statistics*. Geneva. Retrieved from file:///E:/00 Rail/Literature/Illustrated glossary for transport statistics.pdf

Vanichkobchinda, P. (2007). The service expansion potential of railway system for freight transportation in thailand, 305–310.

VECTOR. (2015). *Freight Analysis Network – Terminal Capacity Estimation*. *Vanderbilt Center for Transportation Research (VECTOR)*. Retrieved from http://www.vanderbilt.edu/vector/currentresearch/research-archives/freight-analysis-network-terminal-capacity-estimation/

Vrenken, H., Macharis, C., & Wolters., P. (2005). *Intermodal Transport in Europe*. Brussels: European Intermodal Association.

Wezel, W. Van, & Riezebos, J. (2011). Case Study: Advanced Decision Support for Train Shunting Scheduling. In J. C. Fransoo, T. Waefler, & J. R. Wilson (Eds.), *Behavioral Operations in Planning and Scheduling*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-13382-4

Wichser, J., Weidmann, U., Fries, N., & Nash, A. (2007). Strategies for increasing intermodal freight transport between Eastern and Western Europe. In *PROCEEDINGS OF THE EUROPEAN TRANSPORT CONFERENCE*. Leiden: Institute of Transportation Planning and Systems ETH Zurich. Retrieved from file:///E:/00 Rail/Literature/strategies-for-increasing-intermodal-transport-between-eastern-and-western-eur.pdf

Woodburn, A. (2014). Development of rail freight in Europe : What regulation can and cannot do United Kingdom Case Study. *CERRE: Center on Regulation in Europe*, (December), 1–43. Retrieved from http://www.cerre.eu/sites/cerre/files/141211_CERRE_RailFreight_UK_Case_Study_Final.pdf

Woodburn, A. G. (2008). Intermodal rail freight in Britain: a terminal problem?, (August 2015). doi:10.1080/02697450802423708

World Shipping Council. (2015). *History of containerization*. *About the Industry*. Retrieved from http://www.worldshipping.org/about-the-industry/history-of-containerization

Woxenius, J. (1994). *Modelling european combined transport as an industrial system*. Chalmers University of Technology.

Woxenius, J. (2002). The Organisation of the European Intermodal Road / Rail Freight Transport Industry, (May).

Woxenius, J. (2007). Alternative transport network designs and their implications for intermodal transhipment technologies. *European Transport*, *35*, 27–45.

Woxenius, J., Roso, V., & Lumsden, K. (2004). The dry port concept–connecting seaports with their hinterland by rail, (September), 22–26.

WSP. (2014). Kombiterminal Gullbergsvass. *WSP Analys & Strategi*. Retrieved from http://www5.goteborg.se/prod/fastighetskontoret/etjanst/planobygg.nsf/vyFiler/V%C3%A4stl%C3%A4nken - station Centralen-Plan - samr%C3%A5d-Kombiterminal Gullbergsvass/$File/18_Kombiterminal_Gullbergsvass_20140903.pdf?OpenElement

Annex I: questionnaire

This annex contains the questions that were loosely followed in the exploratory interviews with terminal operators and managers.

Questionnaire/ semi-structured interview

Punctuality and time reliability in road-rail intermodal terminals

Thank you for taking your time in responding to this interview. The purpose of the following questions is to explore the factors affecting capacity and time reliability in road-rail intermodal terminals, both external and internal. This will be part of a Master Thesis performed at Chalmers University in Collaboration with WSP Sverige, which has the aim of analyzing the theoretical and practical capacity in combi terminals. The overall aim is to provide information that will help enhance the competitiveness of intermodal transport.

Please do not hesitate in bringing up examples and anecdotes from your personal experience. This is a semi-structured interview, which means that the questions provide a guide to follow, but are open for you to come up with additional issues.

Motivation of the study

In order to increase the market share of road-rail intermodal transport, it is necessary to solve one of the main challenges it is facing nowadays: time reliability or punctuality. In other words, it is necessary to minimize time variability in order to provide an optimal service quality and to prove reliable towards customers. It is considered that one step in the intermodal transport chain which presents considerable time variability takes place in the intermodal road-rail terminals.

Part I: Description of the operations within the terminal

Part II: External factors that affect time reliability

1. From your point of view, how is the performance of an intermodal terminal affected by the activities of railway operators?

2. How do operations related to traditional carload trains (as opposed to intermodal load) affect the time reliability of intermodal transport?

3. How do you consider that the time reliability of road-rail intermodal transport is affected by the national rail administration, its activities and the infrastructure it provides?

4. In which ways is the capacity utilization of intermodal terminals affected by short-term decisions from customers?

5. For rail-carriers offering transport services both for wagonload and intermodal traffic, how frequently are intermodal railcars mixed with bulk cargo (wagonload) in the same train?
   a. When it occurs, are intermodal cars marshalled with the wagonload cars?

6. In the case of late arrival and/or departure of trains, which party is affected and how? How does it affect your main customers?
7. If the terminal is very full during peak hours and there are new trains arriving, are there times when they have to wait along the main line?

Part III: Internal factors

8. How is the current capacity utilization of the terminal you are working with? Are delays frequent?

9. When you have encountered internal problems causing delay, how do they occur? And which of the following do you consider to be the major cause?
   Possible categories:
   a. Operational policies and administration. *For example:*
      i. *Training of personnel*
      ii. *Information sharing*
      iii. *Use of IT systems.*
   b. Terminal infrastructure. *For example:*
      i. *Number and length of rail parking tracks*
      ii. *Number and length of rail transshipment tracks*
      iii. *Warehouse capacity*
      iv. *Intermediate storage capacity*
      v. *Critical joints and switches*
   c. Terminal equipment. *For example:*
      i. *Number of transshipment cranes.*
      ii. *Capacity of transshipment cranes*
      iii. *Number of land transport vehicles*

Part III: Closing questions

10. In general, do you consider internal factors or external factors to be the main cause in the cases when time delays occur?

11. In the day to day work, it is normal that one encounters aspects that could be optimized, and which may cause a problem to appear repeatedly. Sometimes such situations might require a great influence within the company or a big investment to be solved. Other times, they are easy to optimize but there is not enough awareness of the problem. Could you bring up from your own experience, an example of sub-optimal situations that you have identified, which influenced time reliability? *This could refer to terminal operations, administration and/or to terminal layout.*

Annex 2: Railway infrastructure

*This annex serves as a complement to section **2.1.1**. It is relevant to clearly define infrastructure components, in order to determine the meaning of relevant terms. This was necessary for the project since the use of words is different and may overlap between sources.*

Marinov et.al. (2013, p. 69) mention "tracks, lines, signals, platforms, buildings, sidings, catenary, junctions, switches, bridges and interchanges" as the components of the railway infrastructure. All infrastructural elements, except the terminals, are owned by IMs, while infrastructure present in terminals and yards belong to terminal owners.

Tracks, lines and sidings

Railway tracks perform the function of guiding the direction of the locomotive and wagons through a fixed route by means of the parallel metal rails running along the track. Rails are laid on sleepers and ballasts which have the functions of supporting the rails and providing load distribution, damping, and noise absorption (Ahrens et al., 2009, p. 1070).

Tracks can be either *running tracks* or *sidings* depending on their function. ***Running tracks*** are those tracks meant for traffic circulation, while sidings, also called service tracks, are any type of track which is not intended for traffic flow but for support operations. A running track is sometimes called ***main track*** so as to distinguish it from service tracks.

Railway lines consist on running tracks that join two consecutive points, i.e. the links between nodes (when seem from a macroscopic and mesoscopic viewpoint). They can be formed by one single track or by several parallel tracks, and they are usually many kilometers long (Mussone & Calvo, 2013, p. 13).

The definition of a railway corridor varies across literature. The SNCF defines a ***railway corridor*** as a set of heavily transited train lines that connect two or more terminals within a main route, which links main industrial areas or large urban centers (SNCF, 2007). Meanwhile, Burdett & Kozan define a it as a "single serial line (track) that is made up of one or more sections of specific length that are sequentially traversed" (Burdett & Kozan, 2006, p. 618). This definition can be considered incorrect as it seems to combine the concepts of tracks, lines and corridors.

The definition which seems best suited to the usage of the term in literature is the one provided by the UIC, which defines a corridor as "all possible journey routes (main route or alternative routes), according to market needs, between a defined source and target"(2004, p. 6). The term ***route*** is similar *to rail corridor*, but disregarding market needs and actual traffic conditions: "consecutive lines and nodes as a whole, between a defined source and target" (UIC, 2004, p. 6).

Railway tracks can have a wide variety of technical characteristics. As a consequence, not all types of wagons and locomotives can travel along all tracks. These variances in track configuration occur frequently between different countries due to the national nature of the system, and also within one country´s railway system, since lines might be intended for different types of traffic. The *track gauge*, or distance between the two parallel steel rails, and the *loading gauge*, or the rolling stock´s maximum height and length, restrict a train´s physical dimensions, while its weight is limited to the *axle load*, or the maximum weight that the tracks can support.

Likewise, tracks have different speed designs; higher speed tracks require a larger curve radius and gradient to counteract centrifugal force, can support less axle weight, and must have smaller slopes. Therefore a heavy goods train cannot circulate high speed tracks since it could exceed the axle load, while a high speed train might not be able to circulate in normal tracks due to the lack of appropriate curve layout and slope. Other distinctions include the maximum train lengths allowed in a line (important for the safety and braking systems), as well as the power supply, i.e. whether there is need for diesel locomotives, or electric locomotives coupled to overhead contact wires (Ahrens et al., 2009, p. 1070). Additionally, in the case of electrified lines, the power supply must also have a specific voltage. Together with hangers, poles and other electric systems, overhead contact wires form part of the *catenary* (Trafikverket, 2012, p. 6).

A ***line section*** is a "section of railway track between two locations specified for operating purposes" (Public Transport Corporation, 1999), therefore it can adopt different meanings depending on the purpose of analysis. It may be described as a part of a line in which the mix and amount of traffic, the signaling systems, and the infrastructure do not present relevant changes (UIC, 2004, p. 6).

As such, railway lines are usually composed by two or more line sections (UIC, 2004, p. 6). Railway administrators classify lines and line sections according to various factors such as amount of traffic (e.g. tons per day), technical characteristics of tracks (e.g. speed design, gage, axle weight), number of tracks (single-track, double-track or multiple-track), and line´s relevance (primary, secondary, rural, freight only, and city commuter trains) (NetworkRail, 2010, Stenström, 2012, p. 67, Trafikverket, 2012, p. 12) .

As stated before, sidings are tracks intended for support functions. Some sidings, such as passing and holding sidings, are located along the main tracks and house "on the way" or running trains for a short period of time. ***Passing sidings*** are used for train passes and meets. ***Holding sidings*** are the tracks assigned to set aside or park freight trains on the way (i.e. along main lines) in order to conduct operations such as a change of locomotive and/or a change of driver (León, 2000, p. 223; SNCF, 2007). It is relevant to point out that this term is also used to refer to *side tracks/parking tracks* in terminals.. A ***private siding*** is any kind of siding that is owned and managed by an organization different for the IM (Department of Transport, 2013). The term *siding* is also used refer to tracks within rail terminals.

Railway nodes

A railway node is generally defined as any point in a network in which two or more lines meet, and englobes junctions, passenger stations and rail freight yards (UIC, 2004, p. 6). When railway networks are modelled from a macroscopic level, stations and terminals constitute the nodes, while junctions represent the nodes in a microscopic level (Marinov et al., 2013, p. 68).

A track *junction* has the objective of directing vehicles from one track to another, or to aid them cross a perpendicular track. It is formed by a layout of crossings and points. A set of points form the switches or turnouts, which are the parallel rails that are moved sideways to lead a train into the desired tracks. Junctions are present both in main tracks and sidings (Chandra & Agarwal, 2001, Chapter 15).

The effect of the physical layout of junctions might affect wagon flow in rail yard operations, since turnout layouts determine how restricted is the movement of trains between tracks. For example, in the junctions called diamond crossings, tracks cross each other but the trains are not able to transfer between tracks. In other layouts, the train must occupy other tracks before reaching the destination track, and thus takes more time and occupies extra capacity (Armstrong et al., 2010, p. 3; Chandra & Agarwal, 2001, Chapter 14,15).

All of the junctions in main tracks must be supplemented with signaling systems in order to control the circulation according to their priority, which depends upon their schedule, direction, and whether it transports goods or passengers. Ultimately, these decision factors depend on each IM and are reflected in the timetable. Junctions sometimes intensify the capacity constraints caused by signaling systems and the interference between trains, except in the case of double level junctions. As an example, a train crossing a perpendicular track though a diamond junction will occupy capacity in both tracks, even if the two routes are independent from one another.

A *passing siding* (also called or *bypass track* or *loop line*) laid parallel to another track in a main line forms a *passing loop.* This allows for trains with different speeds, to *pass* (or *overtake*) each other safely (León, 2000, p. 387). Similarly, a *crossing loop* allows two trains with opposite directions to *meet* (or *cross*) each other. It is important to note that passing sidings must be longer than trains, since the braking distance, as well as the distance needed for the train to obtain cruising speed after a complete stop must be included (PPIAF, 2015). Passing and crossing loops are may also have the name of *virtual nodes* (2007, p. 106).

Passenger stations have the primary function of embarking, disembarking and exchanging passengers in a safely manner, and may have the complimentary functions of performing the meeting and passing operations of trains, in the case that they are located at a junction or crossing between lines. In some cases shunting operations to change train composition, are also performed (Marinov et al., 2014).

Passenger stations are divided in transit stations or terminal stations. *Transit stations* are located along a main line and are rather small. Trains usually stop for a very short period, as necessary to embark and disembark passengers, in order to occupy the least possible capacity in the main line. Not all passenger trains stop at these stations. Terminal stations are located at the end of a line. They are normally larger, and have additional infrastructure for servicing passengers as well as trains. They have shunting tracks and parking sidings where trains can be held for cleaning, maintenance, and other operations (Marinov et al., 2014).

A *rail yard*, in the most general sense, is a system of several sidings laid out to perform a certain function. Rail yards might also receive the name of rail terminals or freight stations, in which case they include buildings and platforms. According to their function, they may be classified as goods yards, marshalling yards, and transshipment yards (intermodal terminals). Goods yards are described below, while marshalling and shunting yards are described in section **2.2** and intermodal terminals in section **2.3.**

Goods yards are meant for the reception, loading and unloading of goods wagons, and are thus intended for wagonload/carload (Chandra et. al., 2007, p.466). A type of goods yards is formed by *industry tracks*, which are *private sidings* located in industrial sites, connected to the national rail infrastructure. Wagons are loaded and unloaded at these sidings, and then circulated between several sites of the same industry. Other goods yard include *free loading areas* (see Bärthel et al., 2011, Chapter 7).

Signalling systems

Railway signaling systems are composed by infrastructural and software elements that interact with each other in order to supervise, regulate, and safeguard rail traffic along lines and junctions. These systems are crucial to assure traffic safety and avoiding train collisions, by maintaining a safe separation between trains, as drivers are given the instructions of how to make use of the line ahead (Morant & Westerberg, 2014, p. 2; TRB, 2003, p. 2). They are normally owned and managed by the IM.

Since several types of signaling systems must be present and homologized both in the tracks and in the locomotives (including the training of driver and staff), each rail system requires a specific type of appropriately equipped locomotives, as well as aptly trained drivers. A more thorough explanation of the functioning of signaling systems and technical concepts may be found in Annex 5.

Additionally to signaling systems, ***Automatic Train Protection*** Systems (ATPs) were developed to avoid human error during the train operation, such as a train overpassing a stop signal. These have the function to detect driver´s response to signaling, and may even automatically brake a train if the velocity was not lowered in time (Goverde et al., 2013).

Annex 3: bundling networks

*This annex serves as a complement for section **2.1.1** and describes the bundling network typologies used by rail and IRTT transport operators.*

Direct links

In direct links (a), trains travel directly from a beginning to an end terminal without intermediate stops and marshalling operations.

Shuttle trains

Shuttle trains (e) are a type of direct links in which trains travel back and forth between two terminals with fixed *wagon sets* or *train sets* (i.e. type, order and number of wagons in a train set) (Bärthel et al., 2011, p. 84).

Block trains

Block trains or *full train* is the name commonly given to direct links in the full-train-load wagonload system. They are used whenever one transport customer has enough volumes to fill up a train, such as a large industry. Therefore, only one shipper owns the load of a complete train (UIC, n.d.).

Transport corridor

In a transport corridor, also called *liner trains*, the route between the O/D terminals contains several intermediate feeder points, making it able to serve intermediate markets. This configuration has been promoted by intermodal transport researchers in order to compete with road over short distances (Behrends, 2006, p. 2). However, providing a good service quality calls for the need of very quick unloading and loading operations in intermediate terminals, which up to now has been not economically and/or technically feasible. This scheme has not been widely implemented, due to low economic performance and the fact that many intermodal terminals do not have the required direct connection to the main line (COSMOS, 2015).

Hub-and-spoke network

In a general sense, a transportation hub is any type of facility where persons or goods are collected, consolidated and transferred between transport modes, or between vehicles of the same transport mode. A hub can be an airport, a rail passenger station, a cross-docking yard, etc. Nodes may be classified as O/D nodes or as hubs depending on the scope, i.e. the extent and the type of network(s) covered. For example, when looking at the complete IRRT chain, which involves the road and rail networks, Crainic and Kim (Crainic & Kim, 2007, p. 472) classify intermodal terminals as hubs, since road shipments are collected, consolidated and changed between modes. In this sense, the O/D nodes are located in the shippers´ and receivers´ doors. However, when regarding solely the rail network within IRRT, intermodal terminals constitute O/D nodes, which is the way they are classified in most of the IRRT literature with a rail focus.

In a railway hub-and-spoke layout (a), all transported flows coming from several feeder terminals pass through a centralized node, called *hub*, in which they are classified, consolidated, and sent out to their respective destination. Incoming flows from different regions are sorted and grouped together with shipments with the same destination (Woxenius, 2007, p. 30). In the strict sense, hubs are not feeder nodes (i.e. the nodes where loads enter and exit the rail network.), but bundling points.

Within the hub-and spoke networks, the EU COSMOS project (2015) further differentiates between different systems with the hub-and-spoke network:

- Group train. In the departure terminals, outgoing trains are divided in *wagon groups*, which is a group of wagons with the same trip type (i.e. same origin and destination). In the node, trains quickly exchange wagon sets while keeping their respective locomotive, and their train number remains unchanged. This was the dominant scheme for combined transport in the EU until the 90´s, when the market allowed for direct shuttle services.
- Turntable traffic. At the departure terminals, wagons are normally unsorted, i.e. wagon groups are rarely formed. Trains meet at a hub called marshalling or shunting yard, in which their respective locomotives are decoupled and single wagons or wagon groups are interchanged. Operatively, the hub represents an O/D terminal in the sense that incoming trains are disassembled and new outgoing trains are formed (i.e. the incoming trains don´t have the same number as the outgoing trains). Nevertheless, the transport service towards the shipper is usually direct, i.e. it comprises from beginning to final terminal destination. This system is related mostly to wagonload traffic.
- Megahub production. This system is similar to turntable traffic but dedicated only for intermodal services. In the hub, the traction unit is ideally not decoupled, and wagon sets remain unbroken. Load units are interchanged vertically between trains through train-train transshipment, using equipment similar to that found in intermodal terminals.

Hierarchic network

A hierarchic network (b), also called *connected hubs,* is similar to a hub-and-spoke network. Local flows are consolidated in regional hubs which are in turn connected to other hubs (Woxenius, 2007, p. 30). Related systems, called *trunk feeder networks* and *trunk collection and distribution systems* are presented by Kreutzberger (2010, p. 163).

Annex 4: operations flowchart of marshaling yard

*Figure 21 depicts the operations performed in marshalling yards, serving as a complement to section **2.2**.*

The boxes represent static operations while box arrows represent transport or movement operations. Dotted arrows superposed by D-boxes represent variable waiting times.

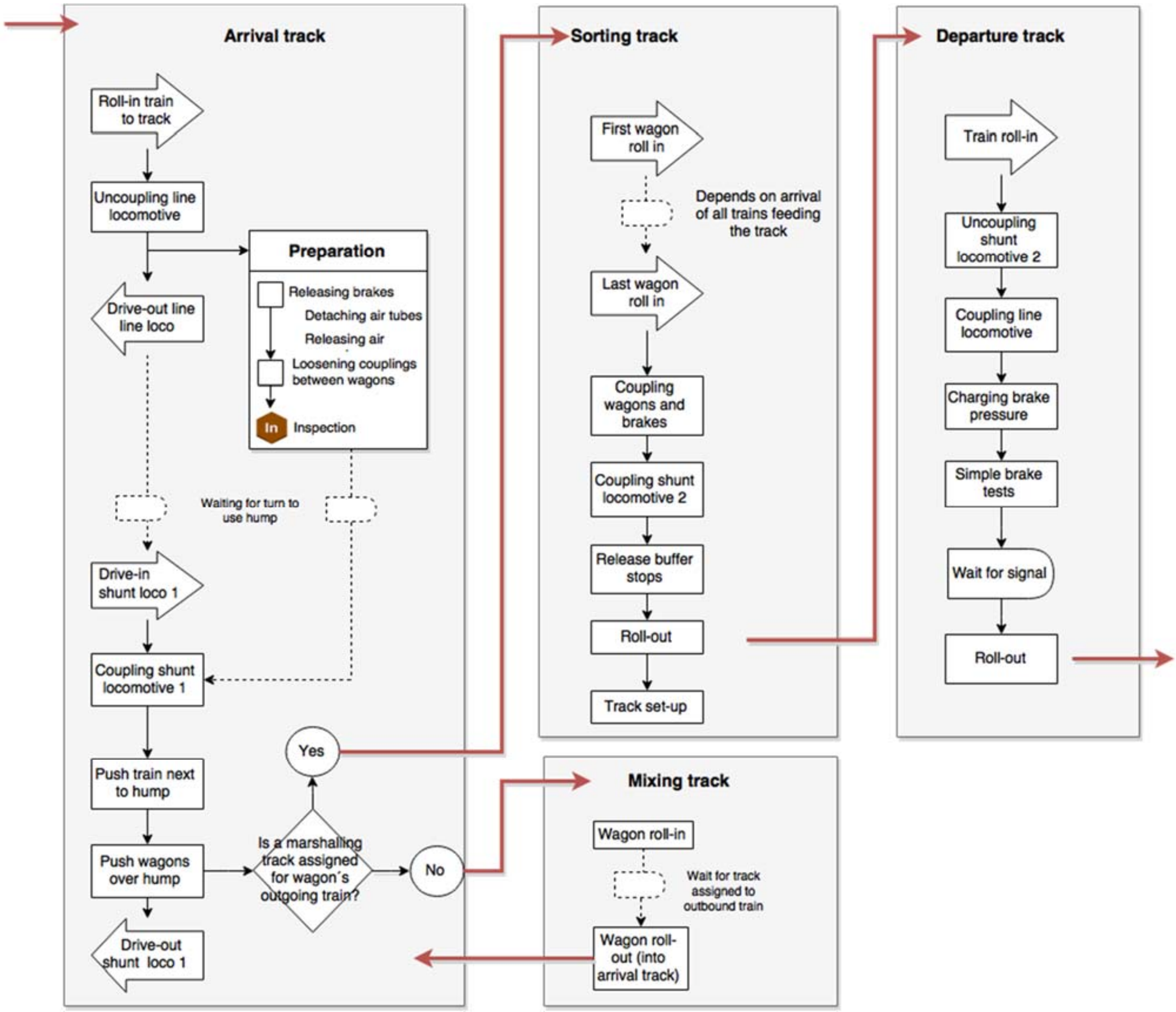


Figure 21: Flowchart of the operations performed in marshalling yards

Annex 5: signaling and safety systems

This annex has the objective of presenting the functioning and concepts of railway signaling and safety systems.

In railway systems, the major challenge for achieving safe operations comes with the interplay of two factors: speed/direction variability and braking distance. The first comes from the fact that one single track might be shared between trains travelling with different speeds and directions, which origin passing and meeting operations. Second, trains take an enormous amount of distance to brake, both due to the low amount of friction between the wheels and rails, and because of the heavy weight which characterizes them, making it easy for accidents to happen without an appropriate control system.

The most basic illustration of this problem is in the hypothetical case that there are two consecutive trains travelling through a single track, without passing or meeting loops. If the second train has a much greater speed than the first (Figure 22 a), the driver´s line of sight in the second train is much shorter than the distance needed to stop or slow the train down. By the time the first train ahead would be visible, there would be not enough distance to brake in order to avoid a collision. In the case of two trains travelling in opposite directions on a same track, such as in Figure 22 (b), the problem is amplified.

Unlike the case of road transport, where the experience and eyesight of truck drivers is enough to avoid collisions, in rail transport is necessary to provide train drivers with the necessary information them to know how to operate further (Railway Technical Web Pages, 2014).

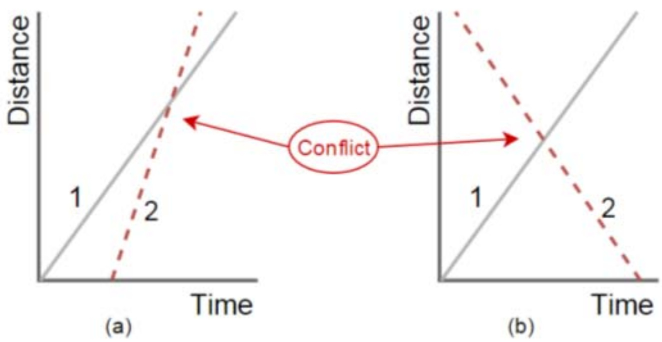


Figure 22: Time-distance graphs showing the movement of two trains with same direction but different speeds (a), and different directions but same speed (b).

Passing loops make it possible for trains running with different speeds and/or directions to overpass or cross each other safely. In these operations, the train with the lowest priority is signaled to enter the siding and wait until the other train has passed and the block section ahead is clear (ARTC, 2015; Gütle & Drewello, 2013, p. 7). Passing and crossing operations might also be performed at transit passenger stations, but in the case of freight trains, they are commonly carried out in loops.

Signaling systems in different tracks and track sections range from very automated and centrally controlled, to ones that are rather manually operated. In any case most of them have evolved from the traditional ***Fixed Block Signaling*** method. This method is based on dividing tracks in consecutive *block sections* (also called *signal blocks* and *signal sections*), with the principle that each block section can be occupied by just one train at the same time (Abril et al., 2008, p. 777; TRB, 2003). The following lines will explain the principles of the basic type of block signaling system, called two- aspect signaling.

The process of dividing a line in block sections is called ***line blocking***. Block distance is intrinsically related to headway and is calculated according to various factors, but by large the most important one comes from the safety requirement: "any train entering a block at its maximum permitted speed must be able to stop before the end of the block, thus maintaining a safe separation between trains". Thus, the minimum block length is proportional to the minimum safe headway. Blocks in one same line do not always have the same length; they are shorter where speeds must be lowered (e.g. when approaching stations and junctions), and longer in the areas between stations where speeds are higher (TRB, 2013b, p. 6).

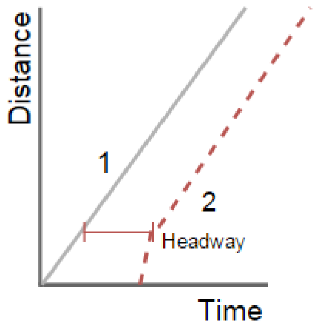
Blocking time refers to the time interval "during which a block section is allocated exclusively to a specific train… and is thus is thus blocked for other trains" (Goverde et al., 2013, p. 6). This depends on running time, as well as the headway (Mattsson, 2007, p. 131).

In two-aspect signaling, each block section is equipped with a track circuit, a simple electric system which detects if it is being occupied by a train. This implies that the exact position of the train within the block section is not known. If the section is clear or unoccupied, a signal at the end (or both ends) of the section will display a green light, indicating that the section is open and an incoming train can proceed. When it is being occupied by a train, the signal displays a "stop" (red light), in order to prevent the second train to enter, which is also called "protecting" the block. In this case, the second train must wait at the stop signal. The signal will again turn green once the first train has leaved the section, indicating that the section is clear again and the second train may go forward (TRB, 2013b, p. 8).

Additionally, it is necessary to ensure that the second train does start braking at a certain distance before it actually reaches the stop sign at the end of its section, or else it would violate the block principle by overpassing the beginning of the next section. Hence, one or more distant signals are placed ahead of the "main", constituting three-aspect signaling. In this scheme, two signal blocks may be occupied by only one train. A distance signal is green if the "main" signal is clear, or yellow if the "main" signal is red, which indicates a train to start braking. Such signals are also used to protect junctions, avoiding two trains from passing through a junction at the same time. Signaling systems which involve more distance signals are called Multi-Aspect Signaling, and involve several kinds of signal displays and blocking methods (Railway Technical Web Pages, 2014).

The headway is an important parameter for the development of the timetable (see section **2.4.2**). To illustrate the concept, Figure 23 shows a train graph, in which time is represented in the x-axis, while the y-axis shows the distance travelled.

Figure 23: Train graph depicting headway distance



In order to overcome certain constraints caused by fixed signaling, further signaling systems have been developed, such as ***Cab Signaling*** and ***Moving Block Signaling***. In cab signaling, the fixed block concept is also used, with the difference that the exact position of the train is known and a continuous signal is sent to the driver which updates the authorized speed. This varies according to the part of the section that the train is occupying, as well as the state of the next section, so that the train can run at an optimal velocity. The moving block signaling system further eliminates the constraints posed by fixed blocs, and consists on a continuous calculation of minimum train headway according to real-time traffic conditions (TRB, 2013a, p. 12).

Annex 6: capacity definitions

This annex presents the general definitions of capacity encountered in literature, serving as a complement for section 2.4.1.

Theoretical capacity	(Slack et al., 2013, p. 324)	*"The capacity which the technical designers had in mind when commissioning the operation"*
Theoretically maximum capacity	(Jonsson, 2008, p. 309)	*"Capacity that could be achieved if production could continue round-the-clock every year"*
Theoretical capacity	(Damij & Damij, 2014, p. 79)	*A resource's "maximum sustainable flow rate if it were fully utilized (without interruptions, downtime, times wasted to setups, idle periods, and so on)"*
Design capacity	(Slack et al., 2013, p. 324)	*"The capacity of a process as it is designed to be, according to actual operating times"*
Design capacity	(Brown et al., 2013, p. 398)	*"the maximum output of a process under ideal conditions"*
Nominal capacity	(Jonsson, 2008, p. 309)	*"The capacity that can be normally used, (according to) number of production units, number of shifts per day, number of hours per shift, and number of days in each period"*
Net capacity	(Jonsson, 2008, p. 309)	*"The capacity estimated to be at the company's disposal to carry out planned production activities" (including indirect unavoidable downtime and non-planned manufacturing such as repairs, urgent orders)*
Effective capacity	(Slack et al., 2013, p. 324)	*The maximum capacity that can be achieved once time losses are accounted for (only time losses due to routine and scheduled procedures necessary for the process, such as maintenance, staff breaks, changeover between batches, and general technical difficulties intrinsic to the process)*
Effective capacity	(Brown et al., 2013, p. 398)	*"Maximum output that can be realistically expected under normal conditions" (takes into account setup times, breakdowns, stoppages, maintenance, etc.)*
Effective capacity	(Law, 2009)	*"The capacity of a system in units per time, taking into account such factors as staffing decisions (e.g. whether to run 24 hours a day, 7 days a week, or to operate one 8-hour shift, 5 days a week), set-up times, maintenance, and an allowance for unforeseeable failures"*
Practical capacity	(Mc.Nair & Richard Vangermeersch, 1998, p. 28)	*"The level of output generally attainable by a process; it is theoretical capacity adjusted downwards for unavoidable nonproductive time, such as setups, maintenance and breakdowns"*
Practical capacity	(U.S. Congress, 1984, p. 47)	*"Practical capacity is the number of operations that can be accommodated with no more than a given amount of delay, usually expressed in terms of maximum acceptable average delay"*

Annex 7: railway capacity definitions

This annex presents the different definitions of railway capacity encountered in literature, classified according to their similarity with general (Table 5), theoretic (Table 6) and effective & practical (Table 7) capacities. Some works have been repeated since they give first a general impression of the definition and then provide one for the particular paper, which are both relevant.

Table 5: Definitions of rail capacity

System/ subsystem	Definition	Source
Rail system/ critical line section	*The general idea is to measure the maximum amount of traffic that a certain railway system, or a certain critical rail section, can accommodate in a given period of time.*	(Mattsson, 2007, p.132)
Rail system/ critical line section	*In this paper capacity is defined as the maximum number of trains that can traverse the entire railway or certain critical (bottleneck) section(s) in a given duration of time.*	(Mattsson, 2007, p.132)
Node	*Freight nodal capacity is the maximum processing rate through the node expressed in units of weight, cubic volume, dollar value, or equivalent equipment movement such as truckloads or containers. The maximum rate of flow is limited by that which is safe as well as technologically feasible.*	(TRB, 1998, p. 13)
Line	*Line capacity is the maximum number of trains that can be operated over a section of track in a given period of time, typically 1 hour.*	(TRB , 2003) [1]

[1] As cited by (Khadem Sameni et al., 2005, p. 3)

Table 6: Definitions of rail capacity related to theoretical or design capacity

System/ subsystem	Definition	Source
Corridor	*The capacity of a rail corridor is defined as the number of trains that can safely pass a given segment within a period of time.*	(Pouryousef, et.al., 2015, p.30)
Rail system	*Engineering capacity is the ultimate capacity of a railway system component under ideal conditions.*	(TRB, 1998, p. 69)
Rail system	*Absolute capacity is defined… (as) a theoretical value (overestimation) of capacity that is realized when only critical section(s) are saturated .*	(Burdett & Kozan, 2006, p. 617)
Rail system	*The simplest definition and the most prevalent encountered in the literature is that the capacity of a single line is the total number of standard train paths that can be accommodated across a critical section in a given time period.*	(Burdett & Kozan, 2006, p. 617)
Line	*It is the number of trains that could run over a route, during a specific time interval, in a strictly perfect, mathematically generated environment, with the trains running permanently and ideally at minimum headway consecutive trains).*	(Abril et al., 2008, p. 776)
Generic	*A theoretical maximum capacity expressed in terms of the maximum number of trains can be calculated by defining ideal circumstances.*	(UIC, 2004, p. 4)
Generic	*Capacity is measured as the count of valid train paths over a fixed time horizon within an optimal master schedule.*	(Harrod,2009)[1]

Table 7: Definitions of rail capacity related to practical and effective capacity

System/ subsystem	Definition	Source
Line	*Capacity is a measure of the ability to move a specific amount of traffic over a defined rail line with a given set of resources under a specific service plan.*	(Krueger, 1999)[2]
Rail system	*Railway capacity is generally de- fined as the maximum flow rate of trains or goods over a segment of track given a specific service plan that affects the scheduling and speed of trains.*	(Rowangould, 2013, p. 29)
Line section	*Capacity is the highest volume (trains per day) that can be moved over a subdivision under a specified schedule and operating plan while not exceeding a defined threshold.*	(Krueger, 1999)[1]
Generic	*The maximum number of trains that may be operated using a defined part of the infrastructure at the same time as a theoretical limiting value is not reached in practice.*	(Hansen et. al., 2008) [1]
Rail system	*In a number of European sources railway capacity is commonly defined as the number of trains in a given period and as a function of various traffic parameters*	(H. E. Boysen, 2013, p. 337)
Rail system	*Actual (sustainable) capacity is the amount that occurs when interference delays are incorporated on the critical section(s).*	(Burdett & Kozan, 2006, p.617)
Line	*Is the practical limit of ''representative'' traffic volume that can be moved on a line at a reasonable level of reliability.*	(Abril et al., 2008, p.776)
Line	*Capacity is the level of traffic (i.e. number of trains per day) that a rail line can accept without exceeding a specified limit of queuing time.*	(Marwich and Partners, 1995)[1]
Transport network (carrier)	*The ability of the carrier to supply as required the necessary services within acceptable service levels and costs to meet the present and projected demand.*	(Kahan, 1989)[1]
Line section	*Capacity is the highest volume (trains per day) that can be moved over a subdivision under a specified schedule and operating plan while not exceeding a defined threshold.*	(Krueger, 1999)[1]

[1] As cited by (Khadem Sameni et al., 2005, p. 3)
[2] As cited by (Abril et al., 2008, p. 776)

Annex 8: frameworks of reviewed terminal capacity methods

This Annex describes the framework in which the methods for capacity calculation described in section 3.3.1 and 3.3.2 were developed and utilized.

Background of the method used by Kessel+Partner and CombiConsult (UIC-GTC, 2004).

The report "Infrastructure Capacity Reserves For Combined Transport By 2015" was prepared in 2004 by CombiConsult and Kassel+Partner for the UIC, with the objective of answering the question: *"Will sufficient infrastructure capacity in terms of rail network and terminal transshipment capacity be available to meet an increased demand for international combined transport, and if not, which investments or other actions are required to overcome infrastructure capacity bottlenecks*?". Special attention is put to the challenge for the rail to absorb extra traffic as a result of the EU's strategy of modal shift to rail.

In order to provide an answer, the first step was to elaborate a comprehensive database of all actual transport flows consisting of a matrix of all O/D relations in 18 trans-European transport corridors, measured in terms of TEUs. The usage of infrastructure capacity by passenger and other freight trains was also considered. Next, the prognosis for traffic and infrastructure development for 2015 was developed (the two steps form the network model). Then, the results were evaluated in order to identify lines and intermodal terminals with potential capacity deficiencies, and based on that to provide capacity improvement recommendations. This was carried out with and without considering scheduled capacity enlargements.

General assumptions:

- "Persistence of the current international transport structure" (p.86) and logistic structure (p.117). So to say, for the prognosis of 2015 transport, the only changes taken into account are those of transport flows, assuming O/D relations and transport areas are constant.
- The study's prognosis for 2015 takes into account only the increase in international transport flows, assuming constant national transport volumes (p. 102).
- A macroscopic view was taken when terminal capacity was evaluated; terminals at similar O/D regions were grouped together into a "transport area", with a total capacity equal to the sum of their individual capacities. The capacity of each transport area was then analyzed according to the flows determined by the network model, as all transport flows served at each of the terminals were aggregated to the level of the transport area. It is however mentioned that this approach might not be the most realistic due to different customer relationships, network access and operational policies of each terminal.

After performing the calculation of terminal capacity, the authors requested actual transshipment information from the terminals, broken down into national and international flows. The capacity provided by the method was compared to the numbers provided by the terminals and normalized.

The authors mention the problem of the lack of accessibility and quality of the data needed for determine terminal transshipment capacity, since this information is often deemed as confidential (p. 89). Furthermore, the lack of intra and international coordination for the long-term planning of terminals in the EU was of concern to the authors and may avoid the development of combined transport. For example, an enlargement of a terminal in one country would have no positive effect if a related terminal, even if located in another country, has under capacity (p.117).

Background of the method developed by Silvio Nocera (Nocera, 2008)

According to the author, the quantity of terminal capacity gives insight into the performance and efficiency of the terminal, as well as the productivity of the facilities, which depend on the infrastructure features, conditions of the facility, and productivity features. Capacity also depends of complex interaction between operational decisions, demand and supply conditions, and market equilibrium, which is considered as rarely addressed in literature.

The general assumptions of the method as descried by the author are:

- Observing a large time period allows for using average values; therefore fluctuations, occasional peak values, and stochastic components can be neglected. Nevertheless, this must be confirmed according to the actual values provided by terminal companies.
- In an idealized scenario, the following steps take place: (1) containers arrive in a train or lorry; (2) containers are unloaded with cranes from the train or lorry; (3) the containers are brought into a warehouse (where they might undergo other logistic operation), or directly loaded into the new vehicle; (4) in the case that the containers were stored, they are picked up again by a crane after a determined time; (5) the containers are loaded into an incoming vehicle and leave the terminal.

The paper can be broadly divided in three sections. In the first section, the problem is stated as a linear programming (LP) optimization problem with the objective function of minimizing container dwell time, with restrictions according to demand, sub-system inter-relations, and sub-system capacity. Dwell time is defined as the sum of unloading, storage and unloading times. Secondly, the method for calculating terminal capacity of the sub-systems according to infrastructure and superstructure characteristics is introduced (which is the method of relevance for this thesis). In the last section, this calculation method is applied to two intermodal terminals.

The method for calculating terminal capacity is supposed to be the solution methodology for the LP problem. Nevertheless, the mathematical relation between both is not clearly described, and the LP formulas are not utilized in the third section.

Annex 9: List of observed capacity factors

This annex provides a list of factors which were observed in the site visit, to have an influence on terminal capacity.

Factors related to the terminal´s infrastructure and superstructure:

- Number, length, and layout of tracks
- Number and type of transshipment devices
- Weight lifting capacity of transshipment devices
- Flexibility of transshipment devices
- Whether the terminal has an overhead contact wire

Factors related to the terminal´s operational procedures and policies:

- Whether trains are served either simultaneously or one after another.
- The prioritization of incoming trains
- The allocation of cranes, forklifts and reach stackers to the loading and unloading jobs.
- Whether train operators are allowed to park unloaded trains within the terminal (in the case of not-constrained arrival and departure schedules).
- The time considerations to determine whether a unit will be put in storage or in the parking lanes
- The organization and physical allocation of units in the storage yard
- Whether outgoing trains wait for late load unit arrivals
- How road vehicles are directed inside the terminal.
- Whether the terminal operates extra hours and night shifts
- Whether the terminal offers shunting services
- The usage given to each track

Factors related to internal communications and IT:

The information and work order management within a terminal may be classified in three levels:

- Level 1: radio or walkie/talkie communication with operators of transshipment devices
- Level 2: display of work order on transshipment devices
- Level 3: automatized gantry cranes, and use of software decision-support systems for the loading plan.

Factor related to external communications and IT:

- Whether the terminal automatically receives information about line status and expected train delays from the infrastructure manager and/or rail operator
- Suitability of booking system
- Provision of automatic track-and-trace system to cutomers

Factor related to the intermodal terminal staff:

- Familiarity with operational protocols and yard layout
- Manual ability to handle cranes
- Knowledge of load units and wagons

Factor related to external staff (specifically road and rail vehicle drivers):

- Training on handling load unit and experience in communicating with reach stacker or forklift operators.
- Familiarity with the layout of the specific yard.

Demand- related factors:

- Whether trains are shuttle trains or formed by wagon groups
- In the latter case, number of cuts of outgoing trains
- Train length
- Number of units in each train
- Mix of load units
- Dangerous goods
- Weight of load units
- Train traction

Annex 10: Gotenburg Combiterminal loading patterns

Graph 7: Lifting and demand patterns in Gothenburg Intermodal terminal, 1th to 3rd June 2015depicts the lifts performed during three days in the Gothenburg Intermodal Terminal. The top bar graph shows the number of units handled according to type of activity, while the bottom graph shows the total number of lifts (therefore, the sum of the quantities in the top graph). The boxes drawn in the top represent incoming (blue) and outgoing (red) trains. The "interval" lines at the right of each outgoing train box represent the scheduled departure times. The numbers in each train box correspond to the number of units.

The graph is based on data from past booking provided by Jonas Emanuelsson, consisting of a spreadsheet with the information of the trains arriving the 1st and 2nd of June, and the trains departing the 2nd and 3rd of June, containing:

- Train trip (train number, date, time).
- Direction (i.e. train arrival or departure)
- For each unit within the train:
    - Unit ID
    - Unit type
    - Weight
    - Start time
    - End time
    - Incoming transport vehicle (i.e. train/truck)
    - Outgoing transport vehicle (i.e. train/truck)

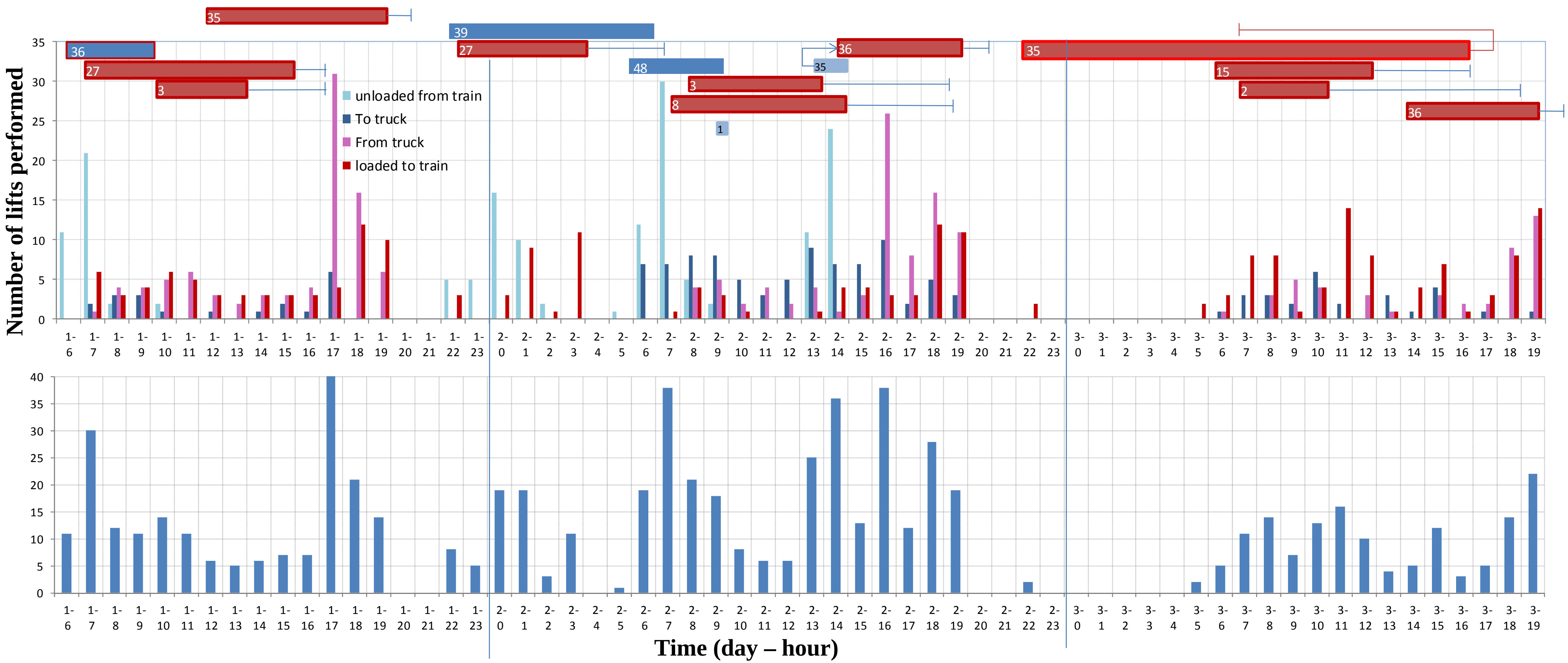
The start time refers to the time the unit entered the terminal, i.e. the time a work order to unload a train or a truck was performed. Similarly, the end time refers to the time the work order to load a train or truck was confirmed by the train operator.

Assumptions:

- In the case of trailers, the unit is registered as active in the system at the moment of truck arrival and not at the moment of truck unloading.
- The arrival time of a train to a transshipment track, corresponds to the arrival time of the unit with the earliest unloading time.

- A train´s departure time is equal to the scheduled time, unless the last unit´s loading time overpasses the scheduled departure time.

Graph 7: Lifting and demand patterns in Gothenburg Intermodal terminal, 1th to 3rd June 2015



Glossary

Transport schemes

Multimodal transport: "the carriage of goods by two or more modes of transport" (UNECE, 2001, p. 16,17). Multimodal freight transport is a broader term which does not necessarily imply the use of standardized load units. Therefore, it may also include chains in which goods are managed in boxes, pallets or individually, such as in express delivery, parcel, and postal services (Crainic & Bektas, 2007, p. 1; Steadieseifi et.al., 2014, p. 2). In comparison, in an intermodal transport chain goods are loaded and unloaded exclusively in the beginning and the end of the transport chain, and are never handled individually along the chain (Lowe, 2005, Chapter 1).

Unimodal transport: a transport chain which utilizes only one transport mode, such as pure road transport.

Combined transport: a multimodal transport scheme in which the main trip is conducted by vessel or rail, while the beginning and end parts of the route are made by road. Goods may be loaded either in intermodal load units (i.e. intermodal transport) or complete vehicles. It is relevant to point out that the definitions for intermodal and combined transport vary across literature, as mentioned by Bontekoning (2004a, p. 9), Woxenius (1994, pp. 4–8), and Crainic & Kim (2007, p. 467).

Trains and wagons

Wagon cut / wagon group (may also be called *blocks*): set of wagons which share the same path and normally remain attached to each other from the origin to the destination terminal. Sometimes they are also called blocks

Wagon set / train set: the set of wagons with a certain arrangement which are hauled by a railway locomotive. A train is formed by a wagon set and a locomotive. It is not clear whether the two terms have an exactly similar meaning or represent different slightly concepts.

Shuttle train: a train running in a direct link between two terminals (i.e. is not marshalled) whose wagon set remains constant along the route. This means that all wagons and load units share the same route

Flat wagons: railway wagons which carry containers and swap bodies.

Recess wagons: railway wagons used for carrying semi-trailers, which have a space for containing the semi-trailer's rear axis.

Traction units: network and shunting locomotives.

Network components

Network: A system of interconnected nodes and links (Mayhew, 2009).

Hub-and-spoke network: A type of network where the lines, or *spokes*, connect to a central node, denominated *hub*.

Node: a railway node is generally defined as any point in a network in which two or more lines meet, and englobes junctions, passenger stations and rail freight yards (UIC, 2004, p. 6). When railway networks are modelled from a macroscopic level, stations and terminals constitute the nodes; on the other hand, when modelled from a microscopic level, railway junctions represent the nodes (Marinov et al., 2013, p. 68).

Corridor: “all possible journey routes (main route or alternative routes), according to market needs, between a defined source and target”(UIC, 2004, p. 6).

Line (also called *main line*): consist on running tracks that join two consecutive points, i.e. the links between nodes. They can be formed by one single track or by several parallel tracks, and they are usually many kilometers long (Mussone & Calvo, 2013, p. 13). Railway lines are normally composed by two or more line sections (UIC, 2004, p. 6).

Corridor: “all possible journey routes (main route or alternative routes), according to market needs, between a defined source and target” (UIC, 2004).

Route: any possible journey route made of “consecutive lines and nodes as a whole, between a defined source and target” (UIC, 2004, p. 6).

Line section: is a “section of railway track between two locations specified for operating purposes” (Public Transport Corporation, 1999).

Siding: (also called service track), is any type of track which is not intended for traffic flow but for support operations. Can be present in rail yard or along main lines.

Railway tracks: perform the function of guiding the direction of the locomotive and wagons through a fixed route by means of the parallel metal rails running along the track (Ahrens et al., 2009, p. 1070).

Running tracks: are all tracks meant for traffic circulation, which form a **main line**.

Holding sidings: are the tracks assigned to set aside or park freight trains on the way (i.e. along main lines) in order to conduct operations such as a change of locomotive and/or a change of driver (León, 2000, p. 223; SNCF, 2007). The term is also used to name the sidings in an intermodal terminal´s *side yard*.

Private siding: is any kind of siding that is owned and managed by an organization different for the IM (Department of Transport, 2013).

Junction: an infrastructural element in railway networks which has the objective of directing vehicles from one track to another, or to aid them cross a perpendicular track. It is formed by a layout of crossings and points. A set of points form the switches or turnouts, which are the parallel rails that are moved sideways to lead a train into the desired tracks (Chandra & Agarwal, 2001, Chapter 15).

Passing siding (also called or *bypass track*, *loop line, passing loop* or *virtual node*): is a siding laid along a single track in a main line which allows for trains with different speeds, to pass (or *overtake*) each other safely (León, 2000, p. 387; Lumsden, 2007, p. 106).

Crossing siding (also called *crossing loop* or *virtual node*): is a siding laid along a main line which allows two trains with opposite directions to *meet* (or *cross*) each other (León, 2000, p. 387; Lumsden, 2007, p. 106).

Rail yard: in the most general sense, a rail yard is a system of several sidings laid out to perform a certain function. Rail yards might also receive the name of rail terminals or freight stations. According to their function, they may be classified as goods yards, marshalling yards, and transshipment yards (intermodal terminals).

Goods yards: are rail yards meant for the reception, loading and unloading of goods wagons, and are thus intended for wagonload/carload (Chandra et. al., 2007, p.466).

Industry tracks: private sidings located in industrial sites, connected to the national rail infrastructure. Wagons are loaded and unloaded at these sidings, and then circulated between several sites of the same industry (see Bärthel et al., 2011, Chapter 7).