Scheduling rail freight node operations through a slot allocation approach

vorgelegt von
Dipl.-Verkehrswirtschaftler
René Schönemann
geb. in Halle (Saale)

von der Fakultät V – Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
– Dr.-Ing. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Kai Nagel
Gutachter:	Prof. Dr.-Ing. habil. Jürgen Siegmann
Gutachter:	Prof. Dr.-Ing. Thomas Siefer

Tag der wissenschaftlichen Aussprache: 31. Mai 2016

Berlin 2016
D 83

Abstract

The present thesis examines the performance of complex freight handling nodes with a focus on the transshipment from and to rail. Complex freight nodes can be found for example in major seaports or industrial railways of large industrial plants. When investigating the railway-specific processes of freight hubs, it has been particularly noticed that they practice an improvised rather than a scheduled operations policy, also known as timetable-free scheduling. That means that the necessary cargo handling process steps are performed rather spontaneous and without much planning efforts. The inadequate organisation of railway operations also leads to significant inefficiencies in the cargo handling capacity.

This thesis is motivated by the aspiration of increasing the productivity and thus the throughput capacity of goods, wagons and trains through complex freight nodes as an overall system. The existing shortcomings shall be resolved by means of a robust hub scheduling approach which synchronises the work processes in freight yards with the preceding and succeeding links in the transport chain. By replacing the timetable-free operations in the railway yard with a process-oriented coordination mechanism based on a time windows (slot management) strategy, the railway-specific requirements shall be integrated with customer's and logistics demands. Initially, the rail-specific production processes of freight nodes are elaborated in a process analysis. It examines the properties of various terminal types, the cargo handling techniques, and the process sequences that freight trains, which pass through the hubs, are involved in. In the process analysis it is pointed out that due to the large number of actors involved in complex hubs and due to the lack of a superior coordinator, it is expected that there is a significant potential for increasing the efficiency of such nodes due to new dispatching strategies.

A two-step simulation and optimisation model approach is developed to investigate the productivity of the cargo transshipment to and from rail. In the first stage of the modelling approach, the performance of freight nodes is analysed in a combined queueing and simulation environment. The model allows to examine the efficiency of individual system components and to identify bottlenecks. It is shown how the interrelationships between the various processes can have a negative influence on the terminal performance when process coordination is insufficient. In particular, it is illustrated, how the turnover rate of the hub decreases when overloaded. A central aspect of the simulation is the computation of distribution functions from observation data for determining stochastic process durations in the different areas of freight hubs. They are of essential significance for the development of the slot management system in the further course of this thesis.

In the second step, the optimisation phase, slot plan sequences for the processing of trains through the freight node are developed by means of a linear programming approach. These slot plans control the allocation of railway infrastructure in the formation tracks of the marshalling yard as well as of the loading tracks in the transshipment terminals. By gradually populating stochastic data from real-life observations, the slot system is tested for robustness under realistic conditions. A stability analysis examines in several scenarios, how slot plans get influenced by various optimisation rules and external factors. In particular, unpunctual incoming freight trains have been identified as having a significant impact on the slot stability. With the development of appropriate rescheduling mechanisms it is further shown that the stability of the slot plans can be ensured even with heavy arrival time deviations. With the rescheduling approach, the slot management has been extended to a powerful tool that keeps the hub production system in a steady state, providing the necessary transparency to other involved stakeholders, particularly to railway undertakings. It even allows reducing the dwell times of trains with arrival delays and thus reduces the likelihood that these delays affect the train's subsequent trips.

With the current production system, transshipment nodes were identified as the weakest link in the transport chain. The thesis shows that their position can be significantly strengthened through such a slot management in combination with a powerful rescheduling approach. The presented slot-based hub production approach optimises not only the processes of the hub itself, but also the processes for the benefit of other stakeholders in the transport chain, such as forwarders, shippers or railway undertakings.

Zusammenfassung

Die vorliegende Dissertation untersucht die Leistungsfähigkeit komplexer Güterumschlaganlagen mit Hauptaugenmerk auf den Umschlag von und zum Schienengüterverkehr. Komplexe Güterumschlaganlagen finden sich beispielsweise in großen Seehäfen oder bei Werkbahnen großer Industrieanlagen. Bei der Untersuchung der eisenbahnspezifischen Prozesse von Güterverkehrsknoten wird deutlich, dass diese zu einem großen Teil einer improvisierten statt einer geplanten Betriebsstrategie unterliegen, auch bekannt als ablaufplanfreie Disposition. Das heißt, die für den Güterumschlag erforderlichen Prozessschritte werden ohne lange Vorplanung operativ und spontan ausgeführt. Die unzulängliche Organisation des Bahnbetriebs führt auch zu wesentlichen Effizienzverlusten in der Güterumschlagsleistung.

Die vorliegende Dissertation ist motiviert durch das Bestreben, die Produktivität und damit die Durchsatzleistung von Gütern, Waggons und Güterzügen durch komplexe Knoten als Gesamtsystem zu erhöhen. Die bestehenden Unzulänglichkeiten sollen mit Hilfe eines robusten Hub-Scheduling-Ansatzes, welcher die Arbeitsprozesse in Güterbahnhöfen mit den vorausliegenden und den nachfolgenden Gliedern der Transportkette synchronisiert, gelöst werden. Durch das Ersetzen der planfreien Abläufe in den Eisenbahnknoten mit einem prozessorientierten Koordinierungsmechanismus, basierend auf Zeitfenster-Strategien (Slot-Management), sollen die eisenbahn-spezifischen Anforderungen mit denen der Logistik vereinbar gemacht werden.

In einer Prozessanalyse werden zunächst die Bahn-Produktionsprozesse von Umschlagknoten des kombinierten Verkehrs dargestellt. Dabei wird auf die Eigenschaften verschiedener Terminal-Typen, die Umschlagtechniken sowie die Prozessfolgen, die Güterzüge innerhalb der Hubs durchlaufen, eingegangen. In der Prozessanalyse wird herausgestellt, dass aufgrund der Vielzahl involvierter Akteure in komplexen Hubs und aufgrund des Fehlens eines übergeordneten Koordinators ein erhebliches Potential zur Erhöhung der Effizienz durch geeignete Dispositionsstrategien zu finden ist.

In einem zweistufigen Simulations- und Optimierungsmodell wird anschließend die Produktivität beim Güterumschlag von und zum Schienengüterverkehr untersucht. In der ersten Stufe des Modellansatzes wird die Leistungsfähigkeit von Umschlagknoten in einer kombinierten Warteschlangen- und Simulationsumgebung analysiert. Das Modell ermöglicht es, die Effizienz einzelner Anlagenteile zu untersuchen sowie Engpässe identifizieren zu können. Es zeigt sich, wie die Wechselbeziehungen zwischen den verschiedenen Prozessen bei unzureichender Prozesskoordination die Terminal-Performance negativ beeinflussen. Insbesondere kann dargestellt werden, wie die Umschlagleistung bei Überlastung abnimmt. Ein zentraler Aspekt der Simulation ist die Entwicklung von Verteilungsfunktionen aus Beobachtungsdaten für die Bestimmung stochastischer Prozessdauern in den verschiedenen Anlagenteilen. Diese werden für die Entwicklung des Slot-Management-Systems im weiteren Verlauf der Arbeit von wesentlicher Bedeutung sein.

Im zweiten Schritt, der Optimierungsphase, werden Slot-Pläne für die Führung von Güterzügen durch mehrere Anlagenteile mit Hilfe eines Verfahrens der linearen Optimierung entwickelt. Diese Slot-Pläne steuern die Belegung der Schieneninfrastruktur in den Vorstellgleisen wie auch in den Ladegleisen der Umschlagterminals. Durch das schrittweise Einpflegen stochastischer Daten aus Beobachtungen wird das Slot-System auf Robustheit unter realistischen Bedingungen getestet. Im Rahmen der Stabilitätsuntersuchung wird in verschiedenen Szenarien geprüft, welchen Einfluss diverse Optimierungsregeln sowie externe Faktoren auf die Slot-Pläne haben. Es zeigt sich, dass insbesondere unpünktlich ankommende Güterzüge einen wesentlichen Einfluss auf die Slot-Stabilität haben. Mit der Entwicklung geeigneter Rescheduling-Mechanismen wird das Slot-Management um ein leistungsfähiges Tool erweitert, das das Hub-Produktionssystem in einem stabilen Zustand hält. Es ermöglicht die Reduzierung der Verweilzeiten verspäteter Züge und verringert damit die Wahrscheinlichkeit, dass sich Verspätungen auf Folgefahrten auswirken.

Mit dem derzeitigen Produktionssystem wurden Umschlagknoten als schwächste Glieder in der Transportkette identifiziert. Durch ein Slot-Management in Kombination mit einem leistungsfähigen Rescheduling-Ansatz kann ihre Position deutlich gestärkt werden. Der vorgestellte Ansatz optimiert dabei nicht nur die Prozesse des Umschlagknotens selbst, sondern auch die Prozesse zugunsten der anderen Beteiligten in der Transportkette, wie z.B. Speditionen, Verlader oder Eisenbahnunternehmen.

Contents

Abstract I

Zusammenfassung III

List of figures VIII

List of tables XI

List of acronyms XV

List of symbols XVII

1 A slot management coordination mechanism for complex freight hubs 1
- 1.1 Background and motivation . . . 2
- 1.2 Recent advances in railway and freight terminal research . . . 4
  - 1.2.1 Freight terminal planning, simulation, and optimisation . . . 5
  - 1.2.2 Railway capacity, timetabling, and scheduling . . . 6
  - 1.2.3 Slot allocation . . . 8
- 1.3 Research objectives . . . 8
- 1.4 Research methodology and thesis structure . . . 11
- 1.5 Scientific and societal relevance . . . 14
- 1.6 Terminology . . . 15

2 Railway-based freight transport networks 17
- 2.1 Hub-and-spoke transport networks . . . 17
- 2.2 Rail freight production systems . . . 19
  - 2.2.1 Block train services . . . 19
  - 2.2.2 Single wagonload transport . . . 20
  - 2.2.3 Intermodal freight transport . . . 20
- 2.3 Intermodal network hub technologies . . . 22
  - 2.3.1 Inland road-rail terminal . . . 23
  - 2.3.2 Seaport container terminals . . . 26

2.4 Exchange operations in intermodal hubs 28
2.4.1 The functional structure of intermodal hubs 29
2.4.2 Stakeholder involvement in hub operations 31
2.5 Railway processes in intermodal terminals 33
2.5.1 Process sequence of incoming trains 33
2.5.2 Process sequence of outgoing trains 35
2.6 The quality of process coordination 37
2.6.1 Transportation planning process levels 38
2.6.2 Capacity planning of rail freight nodes 40
2.6.3 Growth in maritime traffic and its effects on hub development . 43
2.7 Chapter summary 45

3 Slot management as a planning mechanism **47**
3.1 A general definition 47
3.2 Slot management in a supply chain context 49
3.3 Successful implementations of slot management 50
3.3.1 Aviation 51
3.3.2 Parcel delivery services and freight distribution 52
3.3.3 Computer networking 54
3.4 Slot management design in railway operations 55
3.4.1 The general railway track allocation 55
3.4.2 Present state of slot allocation in rail freight yards 57
3.5 Chapter summary 59

4 Productivity and capacity of rail freight nodes **61**
4.1 General procedure on modelling 62
4.2 The hybrid simulation model 62
4.2.1 Queueing systems 64
4.2.2 Event-based simulation systems 65
4.2.3 Modelling approach 67
4.2.4 Expected simulation results 69
4.3 Initial experiment design 70
4.3.1 The train arrival process 70
4.3.2 The processes of incoming trains 72
4.3.3 The shunting process 75
4.3.4 The process of transshipment 76
4.3.5 The process of outgoing trains 78

4.4 Evaluation of terminal performance 79
4.4.1 Set of experiments 81
4.4.2 Simulation scenario 1: using deterministic times 82
4.4.3 Simulation scenario 2: stepwise increase of stochastic dwell times 85
4.4.4 Simulation scenario 3: stochastic arrival times 89
4.4.5 Conclusions on the productivity assessment 90
4.5 Chapter summary 93

5 Railway yard slot scheduling 95
5.1 Time slot development 96
5.1.1 Determination of the minimum required process times 97
5.1.2 Determination of the planned process times 102
5.1.3 Slot design 104
5.1.4 Robustness of time slots 105
5.2 Optimising rail freight yard slot plans 107
5.2.1 Principles of scheduling 108
5.2.2 Rail freight yard graph model 111
5.2.3 Hierarchical planning and reduction of scheduling space 112
5.3 Time slot scheduling in the non-delayed case 114
5.3.1 Prerequisites 114
5.3.2 The scheduling model 115
5.3.3 Input data and expected results 118
5.3.4 Computation results 119
5.4 Time slot scheduling with basic arrival time deviations 121
5.4.1 The arrival delay of a single train 121
5.4.2 The arrival delay of several trains 123
5.4.3 The early arrival of a single train 124
5.4.4 The early arrival of several trains 126
5.5 Time slot scheduling with arrival uncertainty 126
5.6 Rescheduling time slots 130
5.6.1 First come first served (FCFS rescheduling rule) 132
5.6.2 Minimisation of the maximum delays (MMD rescheduling rule) 133
5.6.3 Minimum slot swap (MSS rescheduling rule) 135
5.6.4 Comparison and evaluation of rescheduling measures 136
5.7 Chapter summary 138

6 Requirements and incentives to implement slot management **141**
6.1 Prediction of train arrival times 141
6.1.1 Recent advances in delay estimation 142
6.1.2 Independent methods for the estimation of arrival times 143
6.1.3 Summary on the prediction of train arrival times 151
6.2 Economic issues and pricing schemes 152
6.2.1 The yard manager's perspective 153
6.2.2 The railway undertaking's perspective 155
6.2.3 Slot management as an enabler to reduce uncertainties 157
6.2.4 Further pricing elements influencing the hub efficiency 158
6.2.5 Summary on economic matters and pricing schemes 160

7 Conclusions and closing remarks **163**
7.1 Resume of the main achievements 163
7.2 Concluding critics to the used model and methodologies 167
7.3 Future research potentials 168
7.3.1 New role of the hub manager in the transport chain 169
7.3.2 Simplification of processes 169
7.3.3 Effects of energy-efficient train driving on yard performance . . 170

A Appendix **171**
A.1 Initial slot plan 172
A.2 Planned time table 173
A.3 Slot plan with one technician team 174
A.4 Slot plan with two technician teams 175
A.5 Slot plan with delay of two trains 176
A.6 Slot plan with random delays 177
A.7 Rescheduled slot plan using the FCFS rule 178
A.8 Rescheduled slot plan Minimising the Maximum Delays 179
A.9 Rescheduled slot plan using the Minimum Slot Swap Rule 180

Bibliography **XXI**

List of figures

1.1 Classification of the research topic in scientific areas 4
1.2 Thesis structure 11

2.1 Hub-and-spoke transpot networks 19
2.2 Intermodal transport service with two hubs. 21
2.3 Design of an inland container terminal for road-rail exchange 24
2.4 Stacking ability of a reach stacker 25
2.5 Two types of gantry cranes 26
2.6 Scheme of an intermodal seaport terminal, view from above 27
2.7 Transshipment equipment for seaport terminals 29
2.8 Operation areas of a seaport container terminal 30
2.9 Generalised Transshipment Interface 31
2.10 Process sequence for handling incoming trains 33
2.11 Process sequence for outgoing trains 36
2.12 Decision levels and operations in freight node management 40
2.13 Input and output variables in the performance modelling 41
2.14 Waiting-time function 41
2.15 Growth in maritime container traffic 43

3.1 Job-oriented and machine-oriented Gantt chart 49
3.2 Blocking time diagram 55
3.3 Time-line of the capacity allocation cycle 56
3.4 A complex freight node in the supply chain 58

4.1 A simple queueing system 64
4.2 Queueing network with three serial service stations 65
4.3 Work centres in Simul8 66
4.4 Infrastructure layout in the simulation 67
4.5 A simple intermodal terminal in Simul8 68
4.6 Four-axle flat wagon Sgns of SBB Cargo 69

4.7 Histogram of freight train arrival deviations 71
4.8 Distribution function of arrival time delays 73
4.9 Dwell times in formation tracks for incoming trains 74
4.10 Fitted probability density curve for incoming trains on formation tracks 75
4.11 Uniform distribution of crane moves 77
4.12 Dwell times in formation tracks for outgoing trains 79
4.13 Fitted probability density curve of for outgoing trains 80
4.14 Overview on the basic set-up of the simulation runs carried out 81
4.15 Process plan for one train 82
4.16 Determination of the terminal's performance limit 83
4.17 Utilisation analysis of a transshipment track 84
4.18 Initial Track Allocation Schedule 85
4.19 Occupation plan in simulation (iii) of scenario 2a 87
4.20 Occupation plan in simulation (iii) of scenario 2b 89
4.21 Occupation plan in scenario 3b 91
4.22 Comparison of the average dwell time in the different scenarios 92
4.23 Average utilisation of tracks in the simulation 93

5.1 Decomposition of process delay effects 97
5.2 Tasks performed with trains arriving at freight nodes 98
5.3 Tasks performed with departing trains at freight nodes 99
5.4 Decomposition of the dwell times structure in formation tracks 103
5.5 Decomposition of the dwell times structure in loading tracks 103
5.6 A single time slot 104
5.7 Exemplary slot plan for one train 105
5.8 Integration over the Probability Density Function (pdf) 106
5.9 Scheduling Space 113
5.10 Slot plan pattern provided by the transshipment terminal 115
5.11 Slot plan with one technician crew handling the trains 119
5.12 Slot plan with two technician crews handling the trains 120
5.13 Schedule performance with increasing delay of train 15 122
5.14 Scheduling with the EDD rule when train 15 is 150 min late 123
5.15 Schedule performance with increasing delay of trains 15 and 16 124
5.16 Slot plan when train 21 arrives 120 min early 125
5.17 Priority rules with random train arrivals 128
5.18 Slot plan with randomly deviant arrivals 129

6.1 Estimated arrival time depending on the train origin 144
6.2 Histogram of delay probabilities 146
6.3 GPS records of three trains running on the same route 148
6.4 Path-time diagram of the trains travelling the considered route. 149
6.5 Regression and 95% confidence intervals of the GPS data points 150
6.6 Boxplots of the ANOVA 151
6.7 Pricing schemes for rail freight yards 154
6.8 Two-stage linear cost function $c(t)$. 155

7.1 Integration of the slot management in the annual planning process . . 166

List of tables

2.1 Comparison of hinterland transport modes 44

3.1 Airport coordination levels 51

4.2 RMG crane paramaters 77

5.1 Comparison of the yard productivity in different scheduling runs . . . 120
5.2 Generated random arrival times 127

A.1 Initial slot plan for the scheduling model 172
A.2 Initial timetable for the scheduling model 173
A.3 Rescheduled slot plan using the FCFS Rule 178
A.4 Rescheduled slot plan using the MMD Rule 179
A.5 Rescheduled slot plan using the MSS Rule 180

List of acronyms

AGV	Automatic Guided Vehicle
ANOVA	Analysis of Variance
CSV	comma-separated values
DES	Discrete Event Simulation
DNM	Directorate Network Management
ETA	estimated time of arrival
FCFS	First Come First Served
GPS	Global Positioning System
IATA	International Air Transport Association
JSSP	Job-shop Scheduling Problem
KPI	Key Performance Indicator
LTL	Less than Truck Load
OR	Operations Research
pdf	Probability Density Function
PERT	Program Evolution and Review Technique
RMG	Rail-Mounted Gantry Crane
RTG	rubber tired Gantry Crane
RU	Railway Undertaking
SC	Supply Chain
SCM	Supply Chain Management
SWL	Single Wagonload
TEU	Twenty-foot Equivalent Unit, Standard ISO Container
TOC	Train Operating Company
USA	United States of America

List of symbols

sets

F Set of formation tracks ($F \in I$)

G Graph, i.e. $G = (H, I)$

H Set of nodes of G

I Set of edges of G, Set of tracks (equates to stations $i \in I$)

I_n Set of stations train n is calling

L Set of loading tracks ($L \in I$)

N Set of trains, jobs $n = \{1, ..., N\}$

variables

C_{max} Maximum completion time (makespan) of the whole schedule

$\delta\bar{T}$ Relative alteration of the average time in system

$\Delta t^{arr}_{n,f}$ Difference between the scheduled and the actual arrival time of train n at station i

μ_i Service rate at station $i \in I$

ρ Degree of capacity utilisation

A Arrival process

LQ Queue length

NT Number of transshipments of the terminal

S Service process

c_i, c_j Service time for the coupling process at stations $i, j \in I$

$d_{i,j}$ Travel time between stations i and j

$t^{min,shunt}_{i,j}$ Minimum required shunting time between stations i and j

$t^{p,shunt}_{i,j}$ Planned shunting time between stations i and j

t_{dwell} Observed dwell time

t_{trans} Observed transshipment time

$t^{arr}_{n,i}$ Actual arrival time of train n at station i

$t^{buf}_{n,i}$ Planned buffer time for train n in station i

$t^{d}_{n,i}$ Delay time for train n in station i

$t^{del}_{n,i}$ Process delay for train n in station i

$t^{dwell}_{n,i}$ Actual measured dwell time of train n in station i

$t^{min}_{n,i}$ Minimum required process time for train n in station i

$t_{n,i}^{p}$ Planned process time for train n in station i

$t_{n,i}^{p,arr}$ Planned arrival time of train n at station i

$t_{n,l}^{buf,trans}$ Planned buffer time in the loading track

$t_{n,l}^{del,trans}$ Process delay in the loading track

$t_{n,l}^{min,trans}$ Minimum required process time in the loading track

$t_{n,l}^{p,trans}$ Planned process time in the loading track

$t_{n,f}^{arr}$ Actual arrival time of train n

$t_{n,f}^{buf,in}$ Planned buffer time in the incoming formation track

$t_{n,f}^{buf,out}$ Planned buffer time in the outgoing formation track

$t_{n,f}^{del,arr}$ Arrival delay of train n

$t_{n,f}^{del,dep}$ Departure delay of train n

$t_{n,f}^{del,in}$ Process delay in the incoming formation track

$t_{n,f}^{del,out}$ Process delay in the outgoing formation track

$t_{n,f}^{dep}$ Actual departure time of train n

$t_{n,f}^{dwell,in}$ Actual measured dwell time of incoming train n in formation track f

$t_{n,f}^{dwell,out}$ Actual measured dwell time of outgoing train n in formation track f

$t_{n,f}^{min,in}$ Minimum required process time in the incoming formation track

$t_{n,f}^{min,out}$ Minimum required process time in the outgoing formation track

$t_{n,f}^{p,arr}$ Planned arrival time of train n

$t_{n,f}^{p,dep}$ Planned departure time of train n

$t_{n,f}^{p,in}$ Planned process time in the incoming formation track

$t_{n,f}^{p,out}$ Planned process time in the outgoing formation track

$\bar{t}_{dwell}$ Average dwell time

$\bar{t}_{trans}$ Average transshipment time

$\bar{T}$ Average time of trains in the system

$\hat{T}_n$ Time of train n in the system, turnover time

parameters

λ Arrival rate

λ_d average arrival delay of the delayed trains

λ_e average earliness of the early arrivals

C_e Slot costs for early arriving train

C_l Slot costs for late arriving train

C_p Slot costs for punctual arriving train

C_{slot} Standardised costs for the slot usage

C_{slot}^{arr} Standardised costs for the arrival slot

C_{slot}^{dep} Standardised costs for the departure slot

$C_{terminal}$ Standardised costs for the terminal slot

d_i Due time of job i

l_n Train length of train $n \in N$: number of load units/rail wagons

lf_n Load factor (number of loading units per rail wagon) of train n

p_d Proportion of delayed trains

p_e Proportion of early trains

p_i Processing time of job i

r_i Release time of job i

$slot_n$ Slot number of train n

so, si Source and sink nodes

$t_{slot_n}^{start}$ Start time of slot n

$t_{slot_n}^{end}$ End time of slot n

functions

$C(t)$ Cost function, time-dependent

$E(A)$ Expected value (mean) of train arrivals

$E(S)$ Expected value (mean) of the operating time

$D(A)$ Variance of train arrivals

$D(S)$ Variance of the operating time

$F(t)$ Distribution function

$f(t)$ Density function

$r_{n,f}^{in}(t)$ Reliability of the planned dwell time of incoming train n in formation track f

$r_{n,f}^{out}(t)$ Reliability of the planned dwell time of outgoing train n in formation track f

$f_{n,i}(t)$ Probability Density Function of the dwell time of train n at station i

$r_{n,i}(t)$ Reliability of the planned dwell time of train n at station i

T_n Tardiness of train n (= $t_{n,i}^{del}$)

T_{max} Maximum tardiness

$\sum T_n$ Total tardiness (Sum of all delays, slot conflicts)

$\sum U_n$ Number of late / conflicting trains

1 A slot management coordination mechanism for complex freight hubs

Reliability and punctuality are critical quality measures of scheduled train operations and are important to both operators and customers. According to a market study (Hertenstein and Kaplan, 1991, cited in Hallowell and Harker, 1998), a 1% improvement in the reliability of cargo delivery time could yield as much as a 5% revenue increase in several markets. The high economic potential in the improvement of freight transport operations becomes consequently apparent and has not lost relevance until today.

Reliability and punctuality are also important parameters in the provision of rail freight transport services. A major performance deficiency is seen in the nodes, i.e. the intersections of the rail network where traffic is broken. Rail freight operators perceive railway nodes as "black holes" with a low productivity and as a source of delay and procedural insecurity (Gille and Schönemann, 2009). This is especially valid for complex freight hubs, e.g. those of seaports, or of complex industrial sites with widely branched railway networks for connecting factories and customers to the transport network. Those complex freight hubs consist of two organisational parts: a logistics terminal for the cargo handling and an associated rail yard. However, rail yards are prone to be slow and inefficient due to their complexity. Among others, Marinov and Viegas (2011a) state: "In practice it has been shown that the rail yards are the major sources of delay in the rail freight networks." The reasons therefore are found in the complexity of the handling processes and the little co-ordinated interrelations between the actors performing the hub operations.

With a view on the increasing challenges for the transport market as well as the increasing process orientation in logistics service provision, this drawback calls for new organisational methods in the freight hub management. This thesis aims at identifying the deficiencies in complex freight hubs more accurately and at developing an efficient process coordination mechanism which increases the service performance of the hub-internal operations. This new coordination mechanism is intended to enable the hub having more influence on the transport chain and possessing a better control over internal processes, e.g. fostering steady work load or the compensation of delays. The approach does not only take into account the railway-related processes, but also those of logistics matters and intends to integrate their interests into one planning conception.

1.1 Background and motivation

By studying the real-life operations in several European freight hubs, it could be confirmed what has been reported in various scientific publications – the processes between the intermediaries of a transport chain are poorly adjusted. Both the railway-specific and the logistics-specific processes of several complex freight hubs were examined in this context. When investigating the railway-specific processes, it has been particularly noticed that the rail yards practice an improvised rather than a scheduled operations policy, also known as undisciplined or timetable free dispatching policy (see also Marinov and Viegas, 2011a). Considering a transport chain, a rail yard is positioned between a preceeding and a succeeding facility: On the one side there is the national railway system which operates according to a timetable. On the other side are sidings and transshipment terminals which operate a time slot technique (Gille and Schönemann, 2009). Both facilities organise their processes individually and thereby influence the operation sequences of external parties.

With its improvised policy, the rail yard becomes the weakest member of the transport chain where it suffers the operational unsteadiness of the both adjacent chain links. Consequently, the productivity of the rail yard can be regarded as low which influences in the end the performance of the whole freight hub. The low productivity is characterised by long dwell times for trains,[1] infrastructures blocked by unproductive processes and non-optimal train scheduling. The poor performance is of course partly caused by disturbances and delays from the rail network. While it is argued that rail yards also act as buffers to absorb such disturbances to some extent, unproductive processes belong therefore to their tasks. However, this must not result in the impairment of the yard performance or the overall transport quality. Unproductive and buffering processes are hence required to be minimised.

When studying the operations in several European freight hubs, it has further been identified that capacity constraints (e.g. the number of trains processable per day) are either unknown, vague, or effectively not respected. An excessive capacity utilisa-

[1] The dwell time is the time a train spends in a yard or terminal.

tion, however, results in overloads and thus in a reduction of processing speed. With the aforementioned non-scheduled, timetable-free operations strategy practised at railway yards, it is impossible to determine their productivity properly. Where an efficient production planning is absent, the level of service can hardly be measured or improved.

The issue of hub operations optimisation is not restricted to railways only. In other transport modes similar challenges are evident. Neuman and Atkin (2013, p. 1) for example state for the organisation of airports: "Although simultaneously handling all of the problems may result in more effective resource utilisation, historically different types of airport resources have been handled independently. Despite introducing new support systems the historical separation has often remained. This may increase congestion, which has a negative impact on both the passenger's comfort and the environment." The authors therefore aim to match the relationships between the landing slots, the gates, and other elements of an airport.

This thesis is motivated by the aspiration of increasing the productivity and thus to raise the throughput of cargo, trains, and wagons through complex railway-based freight hubs such as seaports or large intermodal terminals. The identified current deficiencies shall be tackled by developing an integrated and robust hub scheduling method which is able to synchronise operations in the rail yard with preceding and succeeding links of a transport chain and therefore to integrate railway-specific requirements with customer's and logistics demands. The intended production method shall be achieved by substituting the conventional timetable-free operations strategy of the rail yards with a process-oriented coordination mechanism based on the time window (slot) strategy. Slot management is a technique to steer the hub-internal process organisation and to control the access to the infrastructure capacities. The individual processes that are carried out on freight wagons and trains are being organised in time windows which are synchronised with the processes of the cargo transshipment in the loading sites and also with the track access routes on the rail network.

With the introduction of such a capacity management method it is expected that the unproductive waiting times for wagons calling at a freight node can be eliminated which can lead to decreased dwell times and shorter vehicle turnaround cycles. Simultaneously, the existing infrastructure capacities shall be optimally utilised and overloading shall be avoided. Increasing the throughput rate denotes not only benefits for the hub productivity itself but is also favourable for rail freight operators, railway undertakings and for the competitiveness of rail freight services in general.

Finally, a robust production system within the hub contributes to more stable operations on the rail network.

The enthusiasm for this topic originated from working on various research projects focussing on hub management for rail freight services. It was confirmed from practitioners on the national and the European level, how important the issue is in the daily operation of complex freight nodes. This involved not only the optimal operation of terminal facilities, but also the linking in transport chains, the configuration of interfaces between transport modes and not least the connection of the last mile. The work is supported by the experience and knowledge gained during visits of a series of large and small intermodal terminals in Europe which is incorporated in a couple of passages. The aim is, however, to deliver a universally valid slot management approach without developing a proprietary solution for a specific hub.

1.2 Recent advances in railway and freight terminal research

The topic of this thesis touches several fields of research. The central matter is scheduling and optimisation, which is approached from three directions (Figure 1.1). Scheduling, i.e. the optimal spatio-temporal allocation of resources, is of essential

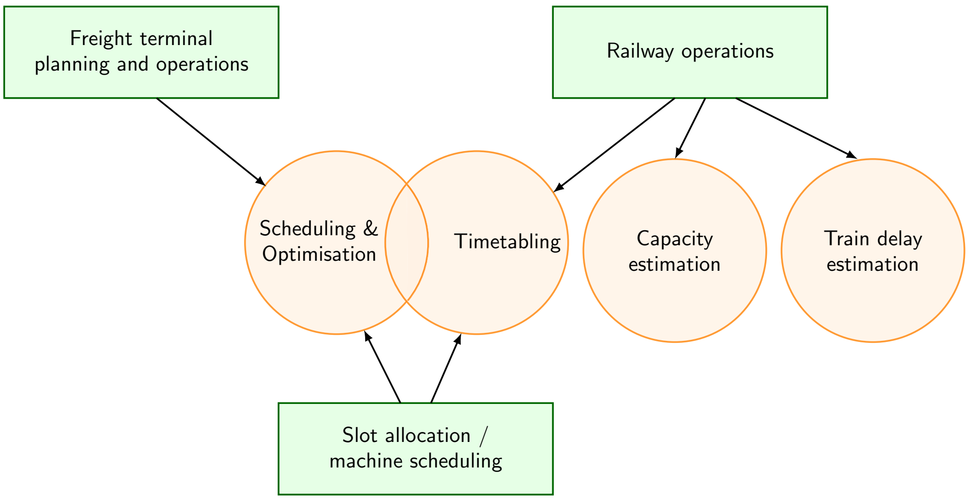


Figure 1.1: Classification of the research topic in scientific areas

relevance for efficient freight terminal operations. The methods in scheduling are strongly related to timetabling in railway operations science. Slot allocation builds thematically on optimisation approaches such as scheduling and timetabling as well.

In Operations Research, inter alia, techniques of machine scheduling are used. Furthermore, the determination of railway infrastructure capacities and the estimation of train delays are relevant fields for the development of slot plans in freight nodes. The current state of research in the respective areas is presented in the following sections.

1.2.1 Freight terminal planning, simulation, and optimisation

The capacity and productivity of intermodal terminals is usually measured separated from influences of the railway network using terminal simulation tools. A large number of publications deals with the research on freight terminals either by simulation or by optimisation approaches. Recent works that use simulation approaches are for example Henesey (2006); Carteni and De Luca (2009); Baldassarra et al. (2010). In operations research different optimisation problems with regard to freight terminals have been defined, in particular concerning large seaport terminals. The *berth allocation problem* (see e.g. Vacca et al., 2008) describes the optimal assignment of berthing times and berthing positions of container vessels calling at seaports. The *crane scheduling problem* concerns the optimal loading and unloading of containers in terminals for both, quay cranes (e.g. Zampelli et al., 2013), and yard cranes (e.g. Zeng and Yang, 2009; Boysen et al., 2010). The *storage space optimisation* characterises the organisation of container storage blocks to reduce the number of crane shifts and sorting procedures (Kozan and Preston, 2006; Wu et al., 2013). The *internal transport optimisation* (e.g. Huynh, 2008) aims to improve the traffic flows and vehicle utilisation and therefore to accelerate the throughput of cargo through the terminal. Recent publications are dealing with the efficiency of various handling equipment (Carteni and De Luca, 2010) or the energy-efficient improvement of the container terminal throughput (Xin et al., 2013).

Moreover, other optimisation problems emerged, as well as variations and combinations of the mentioned ones. Very comprehensive literature reviews covering the recent developments in terminal optimisation can also be found in Steenken et al. (2004) and Stahlbock and Voß (2007). From the literature reviews it becomes clear that the most research addresses primarily the improvement of sea-to-shore-transshipment. The ocean carrier industry has recognised early the great potential in reducing the demurrage[2] of sea ships. The ship operators determined the berthing time as the most important criterion in optimising the terminals since costs can be reduced significantly.

[2]Idle time periods for ships in ports.

Research on the land-side of freight terminals focusses for example on *network and location studies* (e.g. Kreutzberger, 1999; Arnold et al., 2004; Limbourg and Jourquin, 2009), on *innovative transshipment* (e.g. Bontekoning, 2006; Gasser, 2010), or on node-internal optimisation such as *crane scheduling* and *container storage optimisation* (e.g. Crainic and Kim, 2005, and the aforementioned). The integration of railway operations is considerably underrepresented in the aspect of whole terminal optimisation. The analysed terminal models are usually bordered at the transfer point to the rail system (see also later Figure 2.8, p. 30).

1.2.2 Railway capacity, timetabling, and scheduling

Research in railway capacity estimation and timetabling is usually carried out off a logistics context. Simulation and optimisation activities focus mostly on capacity estimation of either routes or nodes and on timetable generation.

Route scheduling: Most of the publications deal with problems like how to plan and operate railway lines and large railway networks. Railway timetabling is simulated using specific microscopic software tools like RailSys or OpenTrack (see Nash and Hürlimann, 2004; Radtke and Bendfeld, 2001). This has the disadvantage that the railway systems are largely analysed in isolation from logistics and other external influences. Some approaches (e.g. Rizzoli et al., 2002; Malavasi and Ricci, 2003; Hansen, 2004) aim to incorporate railways and logistics but neglect for example railway-specific demands and railway-operational dynamics.

Many, mainly American and Australian, papers investigate methods of scheduling trains on a single-line track (e.g. Chen and Harker, 1990; Higgins et al., 1997). This is due to the facts that the railway lines in these countries are often single-tracked and equipped with passing points. Contrarily, European lines are mainly double-tracked. A wide range of solution methods is found; for example branch & bound methods, heuristics, or integer-programming formulations. Timetabling methods and solutions for capacity estimation are described all-embracing for example in Hansen (2010) or Hansen and Pachl (2014).

Other publications deal with the problems in rescheduling and supporting human dispatchers (Jovanović and Harker, 1990; Kraay and Harker, 1995) or real-time scheduling (Rodriguez, 2007; Cacchiani et al., 2013). Some of these publications discuss the problem of re-scheduling trains when disruptions occur. Important in this field are the works of Törnquist (2006) and the alternative graph approach by

D'Ariano et al. (2007). More recent works are those of Kroon and Huisman (2011) or Chu and Oetting (2013).

Rail (freight) node scheduling: Some researchers recognised the particular importance of junction and node capacity in the train scheduling (Zwaneveld et al., 1996; Carey and Carville, 2000; Carey and Crawford, 2007; Chen et al., 2010). Martin et al. (2015) developed an assessment method for the microscopic performance and bottleneck detection in railway nodes. A fraction of researchers is engaged with the simulation and optimisation of railway marshalling yards. Petersen (1977a,b) developed an analytic model of rail yards to determine the train throughput time through marshalling yards as well as the effects of yard congestion. The approach was designed using queueing models of which probability distributions for the major operations in the yards were calculated. Also newer publications (e.g. Kavička and Klima, 1998; Adamko and Klima, 2008; Marinov and Viegas, 2009) focus on the infrastructure design and the operations in marshalling yards (e.g. train bundling, scheduling). Tomii et al. (2001) proposed an algorithm for managing shunting operations based on probabilistic local search and Program Evolution and Review Technique (PERT). However, transshipment, cargo loading and other logistics processes are insufficiently considered in these publications.

A more integrated approach is shown by Rizzoli et al. (2002) where a simulation tool for combined rail/road transport in intermodal terminals has been developed. Also Hartmann (2004) incorporated different transport modes in a container terminal simulation but rail freight is considered rudimentary and railway-specific requirements are not considered. Lee et al. (2006) developed a simulation study for the design of the railway facilities in a container terminal but again disregarded inferences with the adjacent railway network. As a preliminary study for this thesis, Baron (2011) developed a simulation of the rail yard in a container terminal that compared different loading strategies and their effects on the scheduling of freight trains.

Train delay propagation: The research on train delays is closely related to timetabling and scheduling. As a measure of quality, delays describe the situation if a designated schedule cannot be met. Consequently, there is a large interrelation between the two topics. As an early work, Chen and Harker (1990) shall be mentioned again. For the performance of freight yards especially important is the delay propagation (e.g. Siefer and Radtke, 2006; Yuan, 2006; Büker and Seybold, 2012). This is followed by the estimation of arrival times of trains at a freight node (Persson, 2005).

The works of Conte (2007) and Flier et al. (2009) focus on identifying dependencies among delays. With the more profound discussion on schedule quality, the term *service reliability* (e.g. Arcot, 2007; Landex, 2012) became apparent in timetabling.

1.2.3 Slot allocation

Slot management and allocation generally comes from the field of production planning, where the term describes, for example, the assignment of jobs to machines for a certain time period. Koolstra (2005) provided a first holistic aggregation of slot management approaches applied to transport problems.

The vast majority of scientific publications about slot management in transportation refers to aviation. With Neuman and Atkin (2013) and Smith et al. (2014) exist two recent works that investigate two fields of interest for this thesis – (i) the optimisation of node processes and (ii) the complex interplay among involved actors.

In the railway sector, slot management approaches are found in the timetabling process, but are often not explicitly mentioned as such. In rail freight transport, there are some works that deal with the optimisation of the transshipment in terminals. Boysen et al. (2011), for example, is concerned with the assignment of trains to terminal loading slots from a theoretic perspective. However, the relationship with the rail network and the marshalling processes is not within the study area. The question of the slot management in rail is dealt with in depth in section 3.4.

1.3 Research objectives

The objective of the present thesis is to integrate slot management techniques with the production processes of complex freight nodes in order to schedule the rail yard operations therein optimally under the consideration of external influences from the rail network and the adjacent freight terminals. An integrated analysis, using a combination of simulation and optimisation technologies, is envisaged in order to shape the production system of a complex freight hub as a whole. For the description of the main research objective, the following three major challenges were identified during on-site analyses in several complex freight hubs:

Finding 1: Each actor involved in the transshipment processes plans and optimises primarily its own work flow. There is no general coordination of the overall train and freight handling process.

Finding 2: The railway-specific and the logistics processes are poorly coordinated and often interfere with each other. Local process optimisation leads to non-optimal global solutions.

Finding 3: The medium term capacity planning in complex freight is not considered during real-time operations. Train movements in rail yards are usually not scheduled but are carried out on an operational basis which leads to large proportions of unproductive processes and idle times in the provision of services.

Based on these major drawbacks, the main research objective is stated as follows:

Main research objective: Develop a combined simulation and optimisation model that identifies sophisticated production methods for rail freight yards which lead to a better operational performance at complex logistics hubs and which have a positive effect on the reliability of the rail freight operations in such hubs.

More specific, the main objective can be better described when it is decomposed in its essential single components:

Combined simulation and optimisation model: A model, defined as a limited depiction of reality by Stachowiak (1973), has the capability to illustrate complex situations at logistics hubs in a transparent way. One or several models developed in a suitable simulation environment allow to understand hub operations and correlations between processes. The purpose of the modelling is to evaluate the performance of logistics hubs under various conditions. Model parameters are modified accordingly in several stages to identify favourable operational methods that support the main objective. The transparency of models makes it possible to detect effects or relationships when changing input parameters. The modelling approach is implemented in a dynamic simulation environment to gather information on how objects behave over time. At a later stage, adequate optimisation methods are introduced to improve procedures in the dynamic environment.

Production methods: Innovative production methods for rail freight yards differ fundamentally from the conventional non-scheduled operations strategy by relatively improved economic, technological (and environmental) advantages. New production methods often go together with higher planning efforts which are to overcome with the use of new planning tools based on innovative information technologies. The improved operational performance prevails regularly the increased planning efforts.

Operational performance: Operational performance signifies the speed and flexibility of hub operations. Whereas the cargo handling speed has a specific influence on the vehicle dwell time, the flexibility is regarded as the capability of the hub to react on changing circumstances such as train delays, etc. The benefits of a performance improvement become noticeable for the hub itself (e.g. reduced costs, higher productivity), for railway undertakings (shorter dwell time for wagons), and for the intermodal competitiveness of rail freight transport in general. Operational performance can be measured by studying several Key Performance Indicators (KPIs). Thus the operational performance is a measure to analyse and compare the efficiency of different production methods.

Reliability of the rail freight operations: Reliability can be understood as the ability of a failure-free performance over a specified time period. In hub management, the term does not refer to technical breakdowns but rather to conflict-free process flows that do not cause delays or infrastructure assignment collisions. Concerning the transport chain, reliability implies that the hub operations are synchronised with the preceding and succeeding members of the transport chain and that the hub itself is not considered as a risk factor for the transport chain or its customers.

To clarify the main research objective more precisely, it shall be supported by the following three secondary research questions which will be answered in the course of the thesis:

Q 1: How can an improved process synchronisation increase the yard performance and reduce unproductive times?

Q 2: How can freight nodes flexibly respond to stochastic (random) external conditions, e.g. arrival time deviations? What is the effect of deviations on the yard performance?

Q 3: What effects does a higher degree of organisation in freight hubs have on the external railway network? How can it contribute to a more reliable work flow?

The successful achievement of the main research objective and the three secondary research questions required the decision about the appropriate work instruments. The answer to the questions was therefore preceded by an analysis of considerable simulation and optimisation tools. After the definition of the research objectives, the following work structure could be developed.

1.4 Research methodology and thesis structure

For solving the research questions, a combined methodology consisting of simulation and optimisation was chosen. This is also reflected in the document structure, as presented in Figure 1.2. The research is structured in three parts:

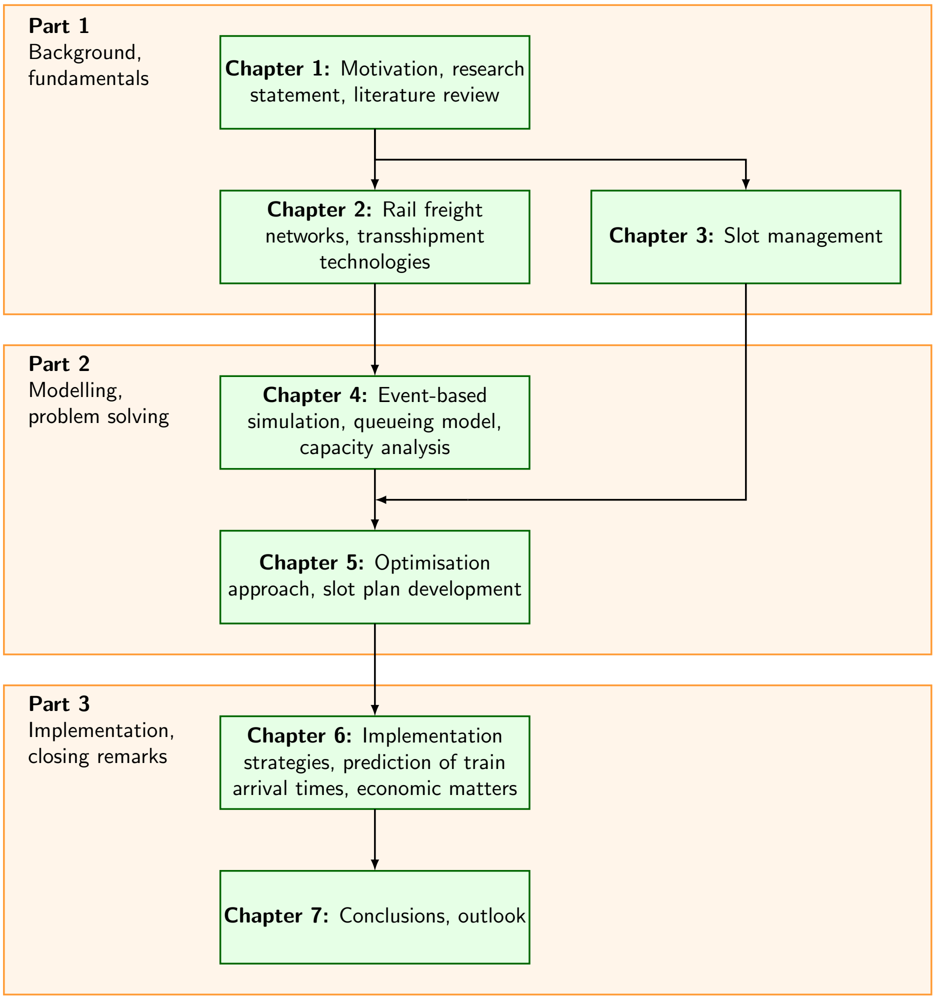


Figure 1.2: Thesis structure, order of chapters

Part 1: covers the identification of the research topic, the determination of the current state of research with literature review as well as introductions to the prevailing rail freight production systems and to slot management. This part comprises the chapters 1, 2, and 3.

Part 2: embraces the capacity estimation for complex freight nodes carried out in a simulation environment as well as the development of the slot management optimisation approach. These activities are elaborated in the chapters 4 and 5.

Part 3: investigates strategies and incentives on the implementation of the developed optimisation approach and summarises the results. The outcomes of this part are documented in chapters 6 and 7.

The following part describes the research methodology at the various parts:

Part 1: Background and fundamentals

Chapter 1 consists of introductory activities such as the definition of the research topic and the development of the thesis work plan. A general literature study on relevant subjects of railway operations, timetabling, and slot management allowed confining the research topic. It has been supported by a parallel fieldwork study of cargo transshipment operations at several intermodal terminals and rail freight yards and with the determination of the operational challenges in such facilities.

Chapter 2 outlines the character of railway-based freight transport networks and the prevailing production systems. It describes the possible technologies in freight transshipment yards and their performance. Likewise, the general yard design in intermodal terminals and large seaports is presented. A special focus of the chapter is the description of process sequences that take place in a complex freight hub. The involved stakeholders are identified and introduced. Ultimately, the stages of the hub capacity planning process are illustrated.

Chapter 3 gives an introduction to slot management as a planning mechanism. It describes the general approach and presents existing implementations in other fields of transport science. Finally, a notion of the implementation of slot management on complex freight hubs is given.

Part 2: Modelling, simulation, and problem solving

The joint approach using simulation and optimisation techniques has the purpose to evaluate the freight hub capacity and to develop reliable work schedules that utilise the given capacities efficiently. All modelling efforts were carried out on a microscopic infrastructure level, which is work-consuming but provides the most detailed results.

Chapter 4 covers the queueing model of a railway-based freight node incorporated in the event-based simulation environment Simul8 (Concannon, 2003). The model simulates railway operations in a freight terminal and it has the ability to vary logistics parameters, e.g. the change of the cargo handling equipment, and thus to examine their impact on the performance of the railway operations. It enables the identification of bottlenecks, the determination of the production quality and other key figures in terms of node performance. Several experiments are carried out on the simulation model in order to understand the functioning of the yard under different aspects.

The use of event-based simulation models for analysing rail freight yards was successfully shown by Marinov and Viegas (2011b). However, the present thesis goes further by performing an optimisation of the process sequences on the basis of the simulation model and thus putting the entire node in a coordinated operational structure.

Chapter 5 embraces the time slot development approach and its formalisation as an optimisation problem. It starts with the decomposition of slot components and the determination of minimum slot times based on the results of chapters 3 and 4. An optimisation model is then presented, which creates slot plans for the rail yard based on a machine scheduling approach. The model is examined experimentally by, for example, modifying the optimisation strategy or varying the available resources (shunting locomotives, staff). Main point of the investigation is the test of slot plans against operational robustness and the development of efficient rescheduling algorithms that modify the slot schedule in order to achieve a new adapted optimal solution if the process flow was disrupted. The examination is carried out by a structured set of experiments.

Part 3: Implementation strategies and closing remarks

After the development and experimental tests, the question arises, how can the slot management approach be introduced at existing logistics nodes and transport chains;

how can the method become accepted by the customers and other stakeholders? Further, this part comes back to the initial research questions and to the issue of how they have been answered in the course of the thesis.

Chapter 6 is twofold. It comprises approaches for the estimation of train arrival times which is found to be a key input parameter for the reliability of slot schedules. An approach based on independent positioning data from GPS devices will be developed to calculate forecasts for the arrival of trains at the considered hub. The second part of the chapter embraces economic concerns for the pricing of rail infrastructures that foster a short wagon dwell time. The chapter has a strongly normative character as it focusses on the question of how slot management implementations should be made.

Chapter 7 contains a critical discussion of the introduced slot management design, the main conclusions of this work and a brief outlook on further research needs.

The thesis pursues a normative character, i.e. it focusses on the question of how slot allocation decisions in complex freight hubs ideally should be made. Though, it is also relevant, how slot management strategies are currently implemented (see section 3.3), this is of secondary importance. The normative character is intended to give decision makers support and to show the feasibility and transferability of the concept to the operations of complex freight hubs.

1.5 Scientific and societal relevance

Scientifically, the thesis aims at contributing to the further development of intermodal transport and a better intermodal competitiveness of rail freight transportation. Its scientific contribution comprehends the improved integration and simplification of transshipment processes from and to rail and the improvement of the transshipment interfaces between transport modes. By investigating the interrelations between railways and logistics, the current drawbacks in freight train processing shall be identified, which outlines possibilities for stronger collaboration between the stakeholders involved in the transshipment operations.

The scientific contribution of this thesis concretely embraces a systematic analysis of railway-related process flows in complex freight hubs and an analysis of the operation methods of involved stakeholders and their interrelations. The normative approach and the conceptualisation of synchronised rail yard operations support the improved integration of railways and logistics demands. This is enforced by the two-step planning approach composed of simulation and optimisation.

The fast and efficient processing of cargo through hubs makes great demands on the hub-internal process organisation. The thesis develops appropriate methods to measure the productivity of logistics hubs.

This thesis has also a societal relevance: It gives insight into the overall logistics hub process handling, the measurement of productivity in hubs, and the depiction of interactions between rail freight services and logistics. The open depiction allows interested stakeholders to gain an understanding of the hub as a whole and thus to respect this in the planning of own processes. For railway operators, the thesis provides a demonstration of new operational methods for freight train handling in yards which can have positive effects on their own production planning. Finally, the thesis aims at strengthening rail transport as a whole and thus to support the greening of transport. The slot management approach can be regarded as an indirect transport greening measure.

1.6 Terminology

To avoid confusion about the terminology used in this thesis, the following note should be considered. The work focusses on the processes of so-called complex logistics hubs, which can be found for example at large seaports or big industrial railway areas. From an organisational point of view, such a hub is divided into a railway-related part and a logistics (transshipment) part. To distinguish them, the following terms are used:

Complex logistics *hub* or freight *node*:	refers to the logistics node as a whole
Marshalling or rail *yard*:	describes the railway operational part
Transshipment or freight *terminal*:	refers to the cargo handling facilities

2 Railway-based freight transport networks

As in other transportation systems, the railway system may be regarded as a network, consisting of links and nodes. Links of the network refer to lines of rail track on which the movements of traffic are performed. Nodes refer to stations where lines intersect. They serve as interfaces with other transport systems and as access points for customers to pick up or deliver traffic to. In the case of rail freight transport, the nodes represent classification yards, marshalling yards, or intermodal terminals. Junctions or passing points which are particularly relevant from a railway operational point of view, are also among the nodes of the transport network.

2.1 Hub-and-spoke transport networks

Modern railway-based freight transport concepts follow a hub-and-spoke network architecture in order to bundle flows and to achieve cost advantages. At the macroscopic level, the railway infrastructure network is composed of stations and routes. The network can be illustrated as a graph model with nodes and edges for transport infrastructures (cf. Kösters, 2010, p. 27 ff.). It is defined as an undirected graph $G = (H, I)$ with a node set $H = (h_1, h_2, \dots)$ and an edge set $I = (i_1, i_2, \dots)$.

The hub-and-spoke concept derives originally from aviation, where a system of hub-and-spoke networks emerged as airlines in the United States rationalised the efficiency of their services with the deregulation movement at the end of the 1970s (Rodrigue et al., 2006, p. 111). Other sectors, for example parcel delivery networks, adapted the concept at an early stage.

By steering traffic connections not directly but via one or several central hubs, the concept of hub-and-spoke systems consolidates the transport flows in hubs and redistributes them to the connecting destinations. Underused, non-efficient direct connections are thus avoided. Direct services between two points exist only if sufficient traffic is in demand. The greater flexibility can be observed in Figure 2.1, where a point-to-point network involves 15 direct connections (A), whereas a hub-and-spoke network requires only 6 connections (B). According to Rodrigue et al. (2013, p. 23) there are three main advantages of hub-and-spoke networks compared to point-to-point connections:

- Economies of scale on **connections** by offering a higher frequency of services: Several indirect connections substitute one direct connection between a pair of nodes.
- Economies of scale in **hubs**: potential of establishing an efficient distribution system since the hubs handles larger quantities of traffic.
- Economies of scope in the use of **shared transshipment facilities**: e.g. lower costs for customers as well as higher quality infrastructures.

Depending on the configuration of the transport network, the transports in hub-and-spoke networks are being interrupted by one or more sorting processes in hubs. Aviation networks are often star-shaped with one hub in the centre, forming a two-staged network like in Figure 2.1 (B). Rail freight networks, however, form rather multi-staged networks, involving two or more hubs in order to facilitate highly utilised trains during the long-distance run between two hubs like in Figure 2.1 (C). There, freight trains run between logistics hubs h_i, $i \in \{1,2\}$ on the main run where goods from various origins so_i, $i \in \{1,2,3\}$ and to various destinations si_i, $i \in \{1,2,3\}$ are being bundled. The pre- and onward carriage is carried out either by other modes of transport, e.g. by trucks or by local freight trains, depending on the production system (see section 2.2). Bundling allows to increase the efficiency of rail freight transportation. Thus, the utilisation of trains is increased and the use of larger transport units (longer trains) results in positive economies of scale. Due to a comparatively high energy efficiency and low personnel costs, rail-based logistics concepts can thus be provided more cost-effectively.

From hub-and-spoke networks also arise disadvantages. The sorting operations in hubs require a temporal synchronisation of incoming and outgoing trains. If, for example, delays occur during the pre-carriage of a transport, this can have far-reaching consequences to the entire transport chain. Connecting trains may not be granted and thus promised delivery times cannot be held. Hub-and-spoke networks are thus sensitive to disruptions and delays at hubs. This fact should be of further significance in chapters 4 and 5.

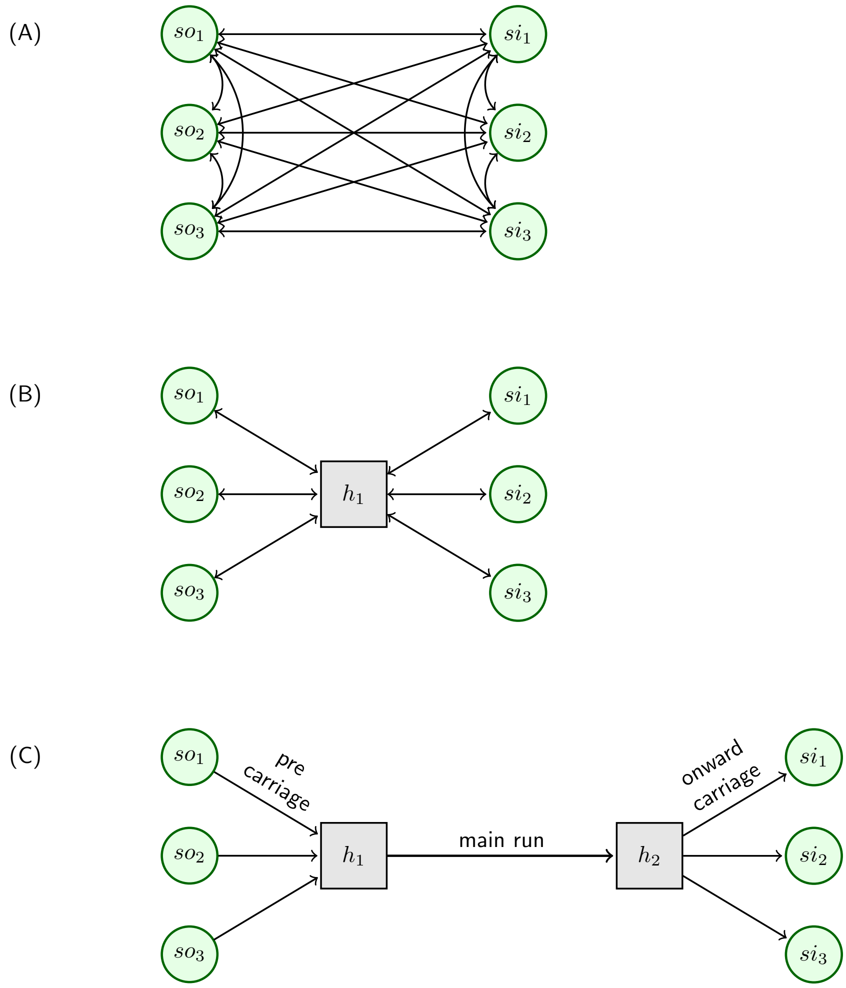


Figure 2.1: Point-to-point network (A), two-staged hub-and-spoke network (B) and a three-staged hub-and-spoke transport network (C).

2.2 Rail freight production systems

2.2.1 Block train services

The simplest form of rail freight transport, from an operational point of view, is the block train service in which all transported goods belong to one shipment. Block trains travel without intermediate treatment from sender to receiver. Ideally, they perform as shuttle trains with a fixed wagon set travelling between two nodes with regularly recurring operating times. Block train services are characterised by comparatively low transport times and simplified business processes, because complex shunting and sorting effort is avoided. Due to the high utilisation risk, block trains are popular only on selected routes, for example in chemistry transport or in seaport hinterland traffic. In a one-stage block train network, each node is connected to every other node as shown in Figure 2.1 (A). If each node is simultaneously source and sink of a main run, a network with N nodes would have maximum $N \times (N-1)$ direct routes available.

2.2.2 Single wagonload transport

The classic Single Wagonload (SWL) transport is used when the entire transport is carried out by rail but several shipments are concentrated on one train. SWL is an open production system to which the customer can bring in any amount of goods at any time. Wagons or wagon groups are usually loaded at the customer's sidings and fetched by the railway operator. All wagons of several customers in a specific geographic area are then collected in marshalling yards. There, wagons are sorted and those for the same direction are bundled to trains. After the transfer to to destination region, the trains are separated and re-distributed by running through one or more other marshalling yards. The collection and distribution processes are usually carried out over a hierarchical hub-and spoke network. Often, with SWL also non-economic relations with low transport volumes are served. The production system is typically used for the transportation of general cargo in non-standard vessels.

The European SWL is affected by a permanent decline of transported volumes since several decades. Some European countries have discontinued their SWL services at all. This is because the system is prone to be slow and expensive. Indeed, the intensive sorting and train making processes in the marshalling yards are time-consuming and labour intensive.

2.2.3 Intermodal freight transport

Conventional, pure rail transport is only partially suitable to satisfy the demands of the substantial transport market for high-valued goods. Not only that access to the rail network in many regions is insufficient. Also the costly and long-lasting shunting of a few wagons is not compatible with the requirements of the logistics. The preferred field of application for rail derives from its cost structures: The high fixed terminal costs and the low variable haulage costs make railways particularly suitable for transporting large amounts of heavy goods over long distances (cf. Woxenius, 1998, 1 ff.).

By promoting the system advantages of rail, intermodal transport has become an important branch for the rail freight sector. It is understood as “The movement of goods in one and the same loading unit or road vehicle, which uses successively two or more modes of transport without handling the goods themselves in changing modes” (United Nations Economic Commission for Europe, 2001, p. 17). As a sub-category, the *combined* transport is defined as “intermodal transport where the major part of the European journey is by rail, inland waterways or sea and any initial and/or final legs carried out by road are as short as possible” (United Nations Economic Commission for Europe, 2001, p. 18). Combined transport came up in Europe in the 1960s following the developments in the United States of America (USA) (Seidelmann, 2010). The hierarchical system, carried out as a hub-and-spoke network architecture, is illustrated in Figure 2.2. The transport between the hubs h_i is called the *main haulage*. The transport from the origin to the first hub is called *pre-carriage* while the *onward carriage* indicates transportation from the second hub to the final destinations.

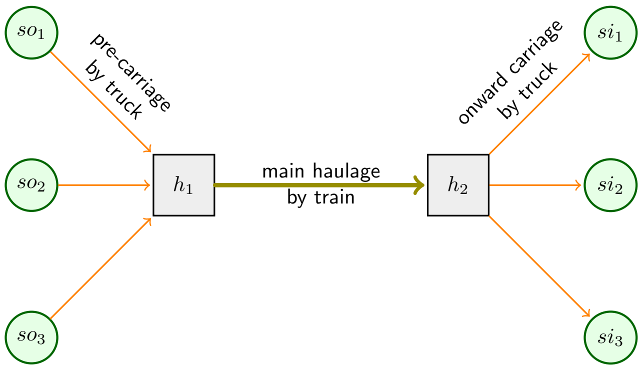


Figure 2.2: Intermodal transport service with two hubs.

Both terms, *intermodal* and *combined* transport, are often used synonymously among experts. It is, however, widely accepted that the share of road in intermodal transport is to be minimised due to economic and environmental reasons. In combined transport, a distinction is made between accompanied and unaccompanied traffic. In accompanied combined transport, full trucks are horizontally loaded onto rail cars and transported to the destination area by rail to continue their journey on the road to the final destination. During the rail journey the trucks are accompanied by their drivers.

Of greater interest within the scope of this thesis is the unaccompanied combined transport. Here, unified loading units such as ISO-Containers,[3] swap bodies,[4] or semi-trailers[5] that can be handled between different vehicles are used. For the transshipment of loading units, standardised handling equipment is required in transshipment hubs located at the system boundaries.

The advantage of intermodal compared to single wagonload transport is the abandonment of shunting operations as far as possible and thus the reduction of non-productive processes during the journey by rail. Direct transport from source to sink, as with block train services, are rarely possible and one or more transshipment operations from or to other modes of transport are required. Thus, intermodal transport is considered as the most economic production system for complex transport chains today.

2.3 Intermodal network hub technologies

It has already been indicated on the preceding pages that the hubs play an important role in intermodal transport networks. Hubs are the locations where traffic flows merge and disperse and are thus interfaces for collection and distribution processes. They connect transport chains and markets. Consequently, they can be points of interchange between the same and also between different modes of transport. Contrariwise, hubs are also the places where transport flows break. In railway-based intermodal nodes, the requirements of railway operations meet with those of logistics. While from a logistics point of view the fast transshipment and onward transport of goods is of the highest importance, the most efficient use of facility capacities is important from a railway operational point of view. Both objectives can be contradictory.

The core functions of logistics nodes are the traditional services handling, storage, and transport. Adapted to network nodes, they can be defined as follows (see Schönknecht, 2009, p. 23 ff.):

[3]ISO-Container: "A large standard size metal box into which cargo is packed for shipment aboard specially configured oceangoing containerships and designed to be moved with common handling equipment enabling high-speed intermodal transfers in economically large units between ships, railcars, truck chassis, and barges using a minimum of labor. The container, therefore, serves as the transfer unit rather than the cargo contained therein." (see United States Maritime Administration and National Association of Stevedores, 1993)

[4]Swap Body: Standardised, non-stackable loading unit equipped with support-legs, used for inner-European combined transport. For further definition see Deutsches Institut für Normung e.V. (2007).

[5]Semi-trailer: A lorry trailer without front axle to be mounted onto a semi-trailer tractor. For further definition see Deutsches Institut für Normung e.V. (2001).

The Handling Function involves the composition and dissolving of vehicle units or loading units at the interface between long and short distance traffic and the transshipment of goods between transport modes.

The Storage Function provides the temporal buffering of goods since the transshipment between many small and one large mode of transport can indeed take place in the same location, but not for all small modes at the same time. Other tasks of a storage centre, for example, packaging or container repair, might be offered by intermodal terminals as well.

The Transport Function embraces the movement of goods between modes of transport between the handling and storage facilities. At large hubs also internal transports can be necessary between individual operating sites.

In order to carry out the transfer and bundling of freight, intermodal network hubs are designed as transshipment terminals possessing the appropriate technologies to handle loading units. Layout, configuration and the transfer operations at hubs depend on the infrastructural organisation and the specific purpose of a hub. Specific loading and unloading equipment is required depending on the vehicles (trains, ships, trucks) which the terminal has to accommodate. Intermodal terminals can be differentiated according to their handling performance, i.e. the size of the terminal and the transport modes handled. Two types of terminals, relevant for this study, and their technical characteristics are shown in the following subsections: A domestic terminal for handling goods between road and rail and a large-scale seaport terminal for handling ship-to-rail transfers.

2.3.1 Inland road-rail terminal

Intermodal inland terminals generally connect up to three transport modes – rail, inland waterway, and road. They act as an interface between local and long distance traffic, whereas long distance transports are carried out by freight train or barge. Trucks are used for local and regional dispersion. Under certain circumstances, inland terminals also transship barge-to-train or rail-to-rail. This is especially true for dry-ports or seaport hinterland hubs which can take over certain sorting and handling operations from seaports in order to relieve them.

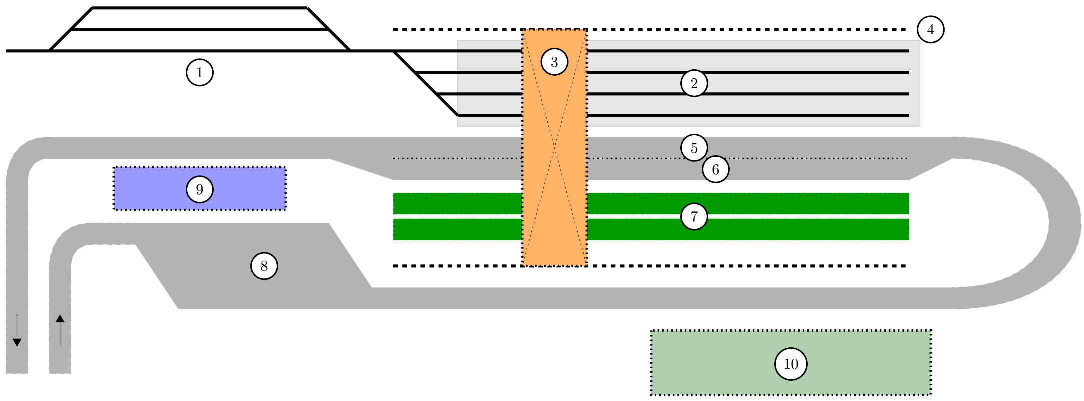


Figure 2.3: Design of an inland container terminal for road-rail exchange

Terminal structure: In the recent past, a specific standard design for intermodal terminals has prevailed. It enables a cost-effective construction and operation of the facilities. A schematic view of such a terminal is shown in Figure 2.3. It depicts the following elements:

1. **Holding tracks:** Storage area for the buffering of incoming and outgoing trains. The area connects the terminal to the rail network.
2. **Loading tracks (rail):** Hold wagons for loading and unloading activities.
3. **Gantry crane:** Required for the transshipment of cargo units.
4. **Crane rails:** Guiding rails where the gantry crane is mounted on.
5. **Loading lane (road):** Area where trucks stop to deliver or pick up cargo units.
6. **Passing lane:** Enables road vehicles to pass each other while staying in the crane area.
7. **Container storage area:** Space to store unloaded ISO containers for later reforwarding. In this example, containers can be stored in two lanes.
8. **Truck waiting area:** Location where arriving road vehicles park until they are allowed to move to the loading lane.
9. **Administration office:** Reporting point for truck drivers, yard administration, etc.
10. **Depot, repair centre:** Area to park mobile transshipment equipment, space for additional services, container repair, etc.

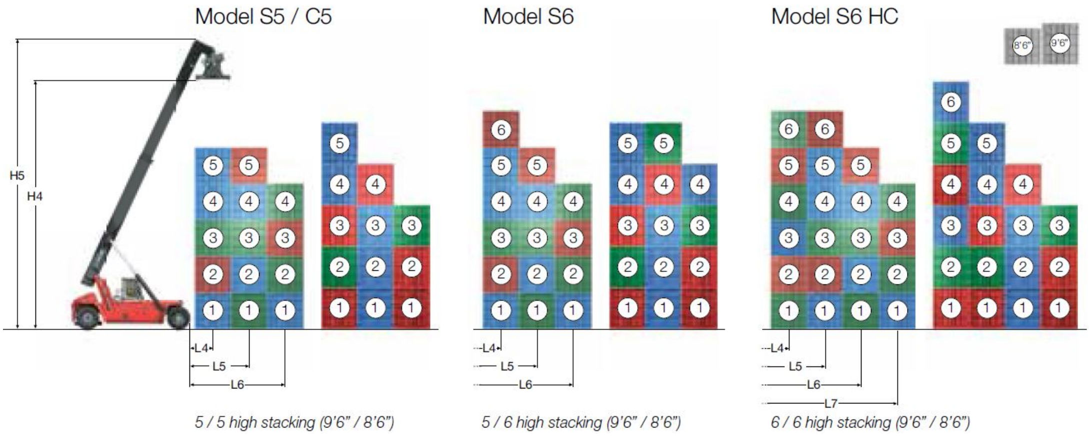


Figure 2.4: Exemplary illustration of the stacking ability of a reach stacker with a stack height of up to six ISO containers. (Cargotec Finland Oy, 2014)

The railway infrastructure equipment is rather simple. There are limited or no options for the sorting of wagons. Marshalling is reduced to a minimum, i.e. moving wagon groups between the holding tracks and the loading tracks. Switches are generally hand-operated or electrical locally operated. The train movements are generally performed as shunting trips which allows simplified signalling and train control.

Transshipment equipment: The productivity of terminals differs greatly by the type of transportation and handling equipment used. For smaller inland intermodal terminals, the most common types are Rail-Mounted Gantry Crane (RMG) and reach stackers. More rarely, also rubber tired Gantry Cranes (RTGs) are in use.

Reach stackers are floor-borne vehicles for smaller terminals. Similar to forklifts, they are mobile around the terminal area. Their extendable telescopic slide-out enables the loading of cargo units onto rail cars in the second row as well as the stacking of the container depot, see Figure 2.4. According to Günther and Kim (2006), reach stackers have a practical storage capacity of about 500 TEU/hectare.

The **Rail-Mounted Gantry Crane** (see Figure 2.5a) is spanning the loading zone of a terminal. He rests on two parallel rails which are placed along the loading zone. The steel construction is thus capable of moving independently along the loading tracks. Attached to the bridge is a movable trolley with the lifting unit. Thus, movements in three directions (horizontal, vertical, diagonal) is usually also simultaneously possible. Gantry cranes are characterised by a higher rate of turnover capacity compared with reach stackers. According to Günther and Kim (2006), RMGs have a practical

(a) Rail mounted gantry crane (R. Schönemann, June 2015)

(b) Rubber tyred gantry crane (R. Schönemann, November 2009)

Figure 2.5: Two types of gantry cranes

storage capacity of about 1,100 TEU/hectare. The technical performance of gantry cranes is approximately 20 moves per hour.

Rubber tyred gantry cranes (Figure 2.5b) are similarly constructed to RMGs. Instead of the guide rail they posses rubber tyres to move sideways along the loading tracks on a concrete surface. According to Günther and Kim (2006), RTGs have a practical storage capacity of about 1,000 TEU/hectare.

2.3.2 Seaport container terminals

The layout and design of intermodal seaport terminals has been widely studied in the recent decades. Good overviews on this topic can be found for example in Steenken et al. (2004) and Kim and Günther (2007). Modern seaport container terminals connect three transport modes: sea, rail, and road, whereas the main objective is always the transshipment from or to sea traffic. Container moves from road to rail or vice versa are usually not practised in such terminals.

Like for inland terminals, a certain ideal layout has been established also for seaport terminals. Differences exist due to local geographic conditions and the transshipment equipment employed. Seaport terminals can use in principle the same devices as inland terminals. To accommodate the distinctly high container turnover, some

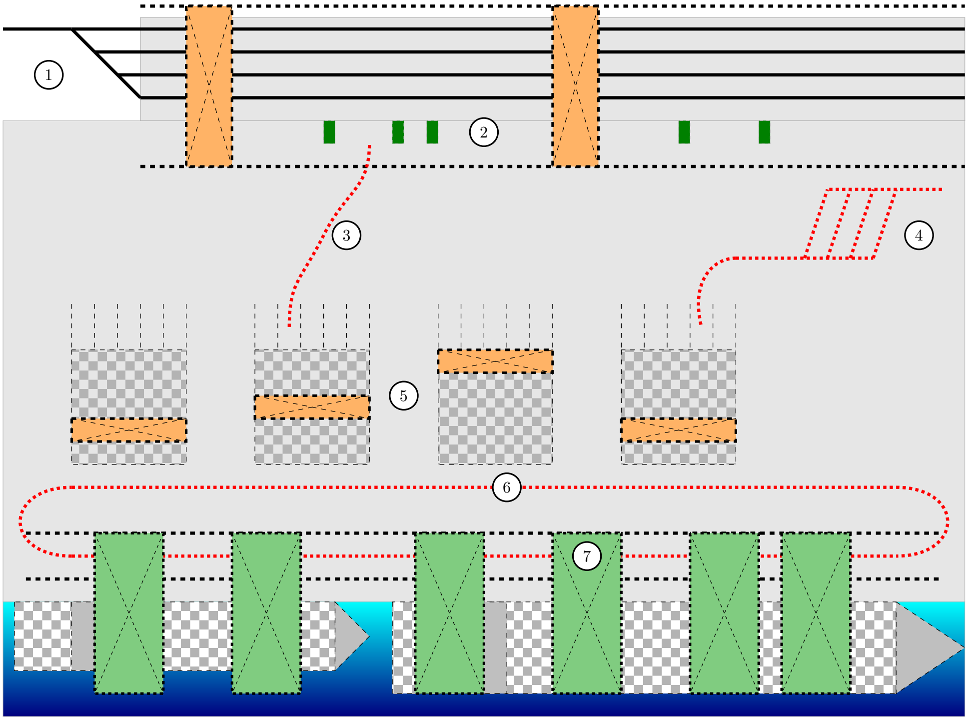


Figure 2.6: Scheme of an intermodal seaport terminal, view from above

other machines are applied as well. An exemplary outline of a modern terminal is illustrated in Figure 2.6. It depicts the following elements:

1. **Railway operation area:** A set of loading tracks covered by generally more than one RMG. The design of this area is similar to the road-rail terminal discussed in section 2.3.1.

2. **Container intermediate storage:** Short term stay area for containers to be loaded to / unloaded from rail.

3. **Container-Transfer:** Loading units from and to rail are not put directly to the container stack. Instead, an additional transfer using a terminal tractor (Figure 2.7d) is required. These are simplified, not road-allowed versions of semi-trailers.

4. **Truck waiting area:** Arriving road vehicles wait here until they are allowed to enter the area. They pick up from and deliver directly to the storage areas, unlike in shipment by rail.

5. **Container storage:** The container storage area is usually separated into different blocks, of which some are reserved for special containers (e.g. reefers which need electrical connection, or for dangerous goods). There are separate blocks for import, export, and empty containers. Storage blocks are operated by RMGs or by straddle carriers (see below).

6. **AGV driving surface:** Special vehicles move containers between the storage area and the quay. Either Automatic Guided Vehicles (AGVs) (Figure 2.7b) or straddle carries (Figure 2.7c) are applied here. As terminal tractors, AGVs require a crane to load and unload containers. The driver-less vehicles travel on a predefined circuit track and were a milestone in the development of terminal automation. Contrarily, straddle carriers pick up loading units by themselves. The tall vehicles can be applied universally for transport and stacking. Whether AGVs or straddle carriers are used is a question of principle but it has a crucial effect on the productivity of the cranes and the terminal itself: The ability of autonomously lifting containers allows for decoupling the work flow of transport and crane operations.

7. **Container gantry crane(s):** When arriving at a port, a container ship is assigned to a berth. Highly productive quay cranes (Figure 2.7a), located at the berth, load and unload containers. Typically, several cranes operate per ship. They are mounted on guiding rails allowing them to move along the berth. Modern triple spreaders allow to move three containers at once and such reaching a transshipment performance of about 40-60 boxes per hour (see Lind et al., 2007).

In addition to these, also further functions can belong to the scope of a container terminal. Typical services are for example the operation of empty container depots or supplemental logistical tasks, such as the packaging of goods, the filling of containers, etc.

2.4 Exchange operations in intermodal hubs

The previous section gave an overview on transshipment terminal freight hub and technology. This section will therefore explain the terminal functions from an operational perspective in order to understand the flow of goods and information through the hub. Figure 2.8 illustrates a seaport terminal exemplarily on a more functional level as analysed by Steenken et al. (2004, p. 6 ff.). In the grey-coloured operation

(a) Container gantry crane (R. Schönemann, June 2015)

(b) Automatic Guided Vehicle (source: porttechnology.org (2012))

(c) Straddle Carrier (R. Schönemann, June 2015)

(d) Terminal tractor (R. Schönemann, November 2009)

Figure 2.7: Transshipment equipment for seaport terminals

areas, cargo units are handled, stored, or transshipped (handling and storage function). The arrows, however, depict the movement of loading units between the areas (transport function). The figure clearly shows that the interdependencies between the logistics functions are crucial and that the performance of a freight terminal depends on the sound interactions between the single facilities. On the upper and the lower end, the operation areas of the adjacent transport modes (truck, train, ship) are outlined but are not further regarded in the model.

2.4.1 The functional structure of intermodal hubs

Steenken et al. (2004) and similar papers focussed on freight terminal operations. Typically, the interaction with other transport systems is poorly considered there, as stated in section 1.2.1. To enlighten the exchange operations between the transship-

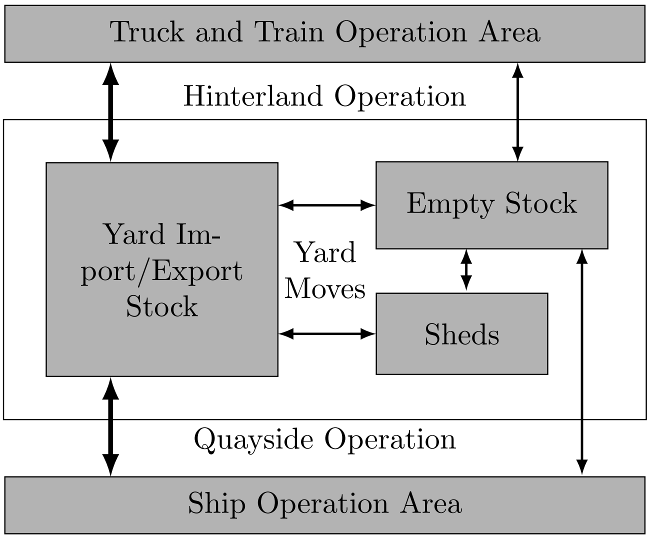


Figure 2.8: Operation areas of a seaport container terminal and flow of transports, (source: Steenken et al., 2004, p. 6)

ment terminal and the rail network, the *Train Operation Area* from Figure 2.8 is subjected to a more in-depth analysis. The truck operations are left out here since they are not further relevant for this analysis. A more detailed view on the train operation area given with Figure 2.9. Trains entering the area from the railway network pass through it in a logical circuit as illustrated by the arrows. Initially, the site is divided into two central logic units:

Marshalling Yard: Holds trains received from the external rail network and is used for train handling operations, such as arrival check, wagon sorting and marshalling before they are being transferred to the loading sites. Similarly, outgoing trains are handled in this area before they leave the freight hub. The green boxes depict operational areas where sequences of processes are being performed. These processes (here indicated as p_n, q_n) are in depth analysed in section 2.5.

Transshipment Yard: The interface to the underlying transshipment terminal as seen in Figure 2.8 and to other transport modes, accordingly. Here, the intersection between Figures 2.8 and 2.9 can be found. Larger freight nodes, such as seaports contain usually more than one transshipment yard or terminal.

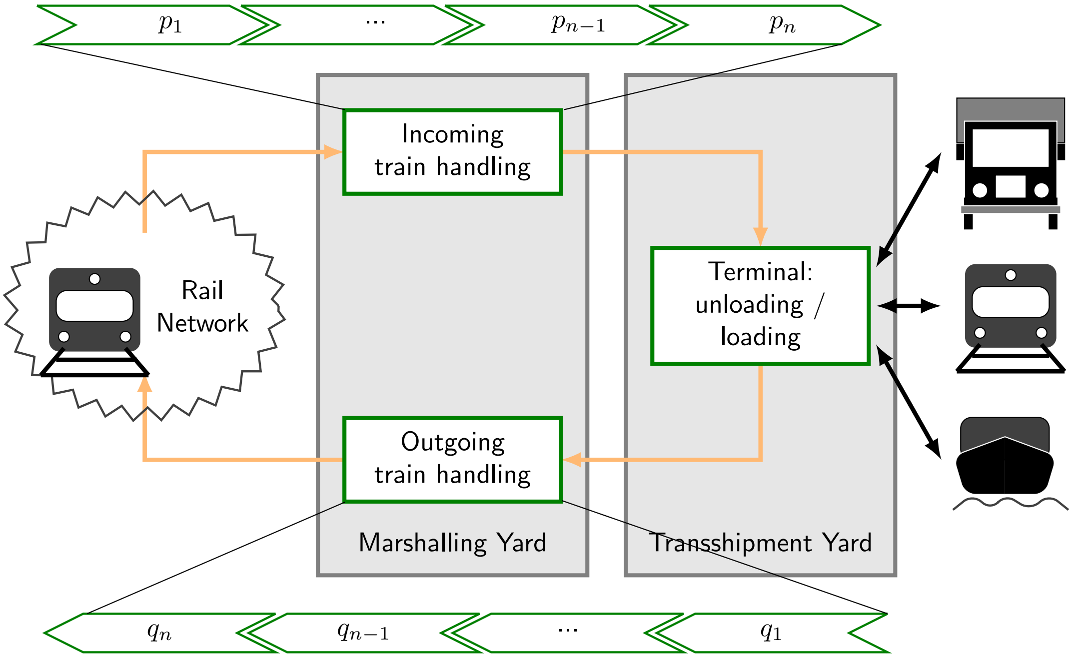


Figure 2.9: General Transshipment Interface: wagons pass through the facilities in a logical circuit on the orange arrows calling at several process stations (green boxes).

2.4.2 Stakeholder involvement in hub operations

Numerous companies and organisations are involved with various roles in the production systems of intermodal hubs. As a consequence of specialisation and division of labour, the various stakeholders pursue partially very different objectives during the service provision in freight forwarding. A terminal manager for example is interested in processing freight trains as fast as possible in order to maximise the number of wagons handled in a time period. A Train Operating Company (TOC), contrarily, might be interested in optimising its train loading schemes which can lead to a significantly longer dwell time for the wagons in the terminal.

Every role may be performed by a different actor, but it is also possible that a single actor is responsible for one or more service roles. The following listing provides a general overview on the most relevant stakeholders (roles) during the exchange operations:

Customers: Owners of the goods transported and clients of the enterprises operating in the yard.

Freight operators: Companies that organise transports for its customers or operate intermodal transport networks.

Forwarders, shippers: Companies that execute transports. Depending on the transport mode it is to differentiate between sea freight forwarders, inland water freight forwarders, train operating companies, road freight forwarders, or multimodal transport operators. These companies are entrusted with the transport of goods between the remote destination and the considered freight node.

Shunting operating companies: Railway undertakings offering transport services inside the hub. Their core business comprises mainly the shunting and marshalling of wagons and wagon groups, and related wagon handling services.

Terminal operators, loading sites: The operators of the transshipment points executing the cargo handling between two transport modes. This can also be a port operator, an intermodal inland terminal, or a warehouse. Larger hubs (e.g. sea ports) feature often more than one terminal or loading point.

Terminal infrastructure manager: Larger freight nodes with extensive internal road and rail networks hold their own companies that manage these infrastructures. Sea ports for example possess their own rail infrastructure managers.

National railway infrastructure manager: The infrastructure manager that operates the rail network outside the freight node. Incoming and outgoing train traffic must be co-ordinated between the two infrastructure managers.

Others: Stakeholders and actors not further relevant for this study such as customs authorities, container maintenance and repair services, special logistics operators, vehicle provider etc.

The presence of specific stakeholders depends on the size and purpose of the freight node. In small nodes, for instance, the functions of the terminal operator and the terminal infrastructure manager are bundled in one entity. Similarly, only at larger hubs specialised shunting operators will provide their services. With the size of the hub increases the specialisation and labour division and thus the number of involved stakeholders. The efforts for co-ordination and information exchange increase at the same time.

Which stakeholders are actually involved in the operations, depends on the individual production method of the hub but also on the specific transshipment task.

Each transshipment task consists of a plurality of successive process steps, whereas it is to distinguish between railway-specific and logistical processes. In addition to the physical processes, information exchange procedures between the stakeholders play an important role. A detailed analysis of the production processes and the necessary exchange of information is further discussed in section 2.5.

The requirements for information exchange and the sharing of work require complex contractual structures between the stakeholders. Contractual and regulatory aspects are not discussed further in this work. For advances on this issue it is referred to a recent paper on the policy and regulatory framework, carried out by Bergqvist and Monios (2014) at Swedish and British intermodal terminals.

2.5 Railway processes in intermodal terminals

While dwelling in a node, freight trains are involved in a series of linked processes as indicated in Figure 2.9. The single consecutive processes are now described in more detail. They form a logic process sequence whereas the processes themselves can be carried out at different locations within the yard. Besides, for better illustration, a distinction is made between process sequences for incoming (p_n) and outgoing trains (q_n).

The following process sequences represent a generalised interpretation. The detailed arrangement of the processes and the consideration of additional, optional processes depends on the specific hub and the specific job (e.g. the customers requirements, order situation for return freight, etc.) and may differ from the general case.

2.5.1 Process sequence of incoming trains

The first process sequence, *incoming train handling*, describes individual processes that are carried out on trains entering a freight hub before they can be unloaded. The

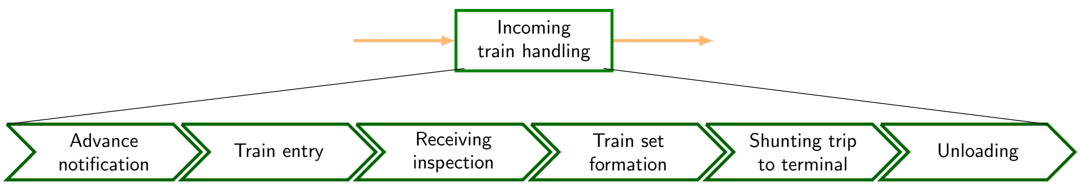


Figure 2.10: Process sequence for handling incoming trains in a yard

sequence is depicted in Figure 2.10, following up as a detail of Figure 2.9. Behind the individual processes the following details embody:

Advance notification: The train operating company (TOC), the customer, or the forwarder generates a transport order and informs the receiving yard about the upcoming train arrival. The advance notification contains information such as the scheduled arrival time, the wagon order, the number of axles, weight of the wagons, information on oversized loads, dangerous goods, target loading site, etc. and thus all information required to handle the train in the receiving freight node. The notification data are ideally transmitted electronically to the node via proprietary or web interfaces. However, still today numerous TOC send advance notifications by fax. This implies an increased workload at the receiving yard since all data have to be typed into the local IT system. The advance notification is sent at the latest with the departure of the train from the originating node and thus at relatively short notice. With it, the operational yard schedule is updated. The effective receiving track in the marshalling yard, however, is generally not appointed until the yard's station manager gets informed about the train arrival from the neighbouring station.

Train entry: The national infrastructure manager (or the neighbouring station manager respectively) reports the upcoming train arrival to the yard's station manager who determines a suitable inbound track within the marshalling yard. The corresponding track is reported to the train operating company. After the train arrived, its completeness is reported back to the national infrastructure manager. The long-haul locomotive is being detached from the train and parked or assigned to the next job by the TOC.

Receiving inspection: Irrespective of the train data reported with the advance notification, several checks are undertaken after the train arrived at the marshalling yard. This includes comparing the real wagon order to the reported wagon order as well as a check for possible damages. The loaded freight is compared with the shipping documents. If not done before, the proper loading site, siding or terminal is determined for each wagon. The loading sites will be informed about the train arrival by the yard manager. At border stations customs treatment might be carried out[6].

[6]Customs treatment is not considered further in this thesis.

Train set formation: If the train consists of wagons destined for several loading sites, the wagons are decoupled and sorted, respectively. The process is carried out by a shunting operator, authorised by the forwarder or the TOC. Depending on the number of sorting events, the process can be time-consuming until conditions as in SWL occur. Though, intermodal trains are usually block trains or maximally split into two to three sections.

Shunting trip to terminal: After the marshalling operations have ended, the train set can be moved to the terminal or loading site. The process is carried out by a shunting operator, authorised by the forwarder or the TOC. After the shunting movement, the track at the marshalling yard is free for other purposes. Time of shunting, travel route, and target tracks are negotiated operationally by radio or telephone between the infrastructure manager, shunting operator and the terminal operator at most freight nodes.

Unloading: Train sets arrived on loading tracks can be unloaded and loaded by the terminal operator according to the shipping documents.

2.5.2 Process sequence of outgoing trains

The following process sequence illustrates the processes that trains are involved in when they leave the hub. Not necessarily all trains or wagon groups that have been unloaded will immediately be used afterwards again and will participate in both, incoming and outgoing process sequences. In practice it happens for example that wagons are unloaded and then returned to the marshalling yard. For a new transport task they are then moved in the terminal a second time for loading. Consequently, this sequence contains some optional processes or those which depend on other prerequisites. Hence, some variants of the sequence are possible. The generalised process sequence is depicted in Figure 2.11 as a detail of Figure 2.9. Behind the individual processes the following details embody:

Transport order planning: The TOC or the freight forwarder places a transport order. This allows the freight terminal to prepare the shipment and to allocate the requested loading units or goods from the storage areas. Also, the loading site can request the required number and type of wagons for the transport. The wagon list of the envisaged train set is then handed in later by the TOC.

Train set formation: If the wagons are not already standing on a terminal track (from a former incoming train), empty wagons have to be delivered. Based

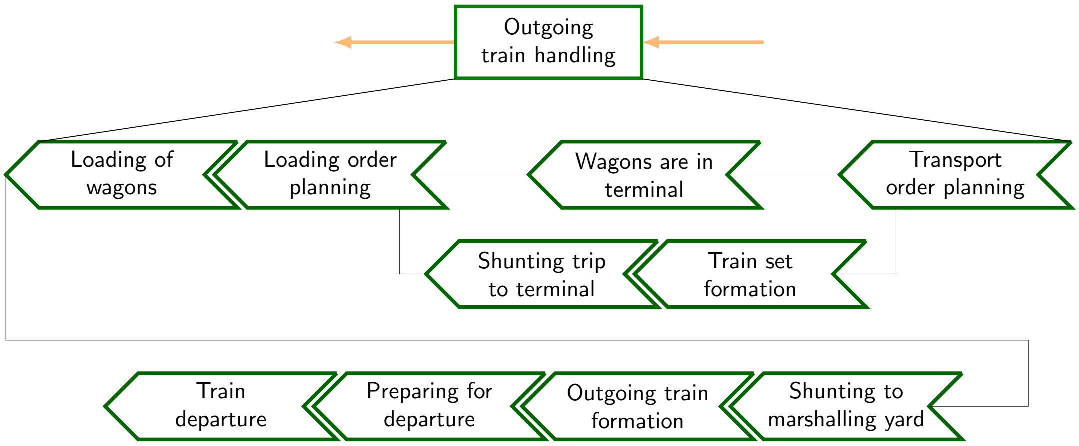


Figure 2.11: Process sequence for the handling of outgoing intermodal trains

on the requirements provided by the loading site, empty wagons are arranged and formed to a train set. This is usually done either outside the hub area or in the hub's marshalling yard. The wagon list is then provided to the loading site.

Shunting trip to terminal: The empty train set is shunted to the transshipment yard either alone or joined with a set of wagons that will be unloaded. The process is carried out by a shunting operator, authorised by the forwarder or the TOC.

Loading order planning: The TOC assigns the available loading units, containers etc. to the storing positions on the freight wagons according to its customer's requirements. Safety regulations, weights, clearance outlines, etc. are to be respected here. When done, the loading order is sent to the transshipment yard.

Loading of wagons: With the receipt of the loading order, the loading site or terminal can start the physical transshipment of loading units onto the freight wagons. (Optionally customs controls can be carried out again at border stations.)

Shunting trip to marshalling yard: The train set is moved to the marshalling yard by a shunting locomotive. The hub infrastructure operator assigns the route and the final formation track at the marshalling yard, accordingly. The loading track at the transshipment terminal is freed and can be prepared for the next loading operation.

Outgoing train formation: If the loaded wagon groups are no block trains, they are combined with other wagon groups. A shunting operator takes care of the marshalling and places the final train set on the assigned outgoing track. The wagon list of the new train is created and sent to the TOC.

Preparing for departure: Vehicle inspection, brake test at all wagons, and check of transport documents is carried out by technical inspectors of the TOC or the shunting operator. The train is reported as ready to the yard manager. The long-distance locomotive is attached and a simplified brake test is executed.

Train departure: The train will leave the yard at the scheduled departure time. The actual departure time is negotiated between the station manager and the network infrastructure operator. As soon the train leaves the hub, the formation track is clear for the next use.

Again, it shall be mentioned that the detailed arrangement of the processes depends on the specific hub and the specific job. It may differ from the general case.

2.6 The quality of process coordination

In transportation systems, supply and demand for transport services are bound to factors which are limiting the productive efficiency. A freight node is capable of processing only a certain amount of vehicles per day, for example. If demand exceeds the capacity of a transport system, supply and demand are becoming imbalanced and require a coordination (cf. Windt, 2001, p. 49). To avoid overload, access to freight nodes is regulated by the administrative authorities.

In economic theory, the balancing of demand and supply is controlled over the market price. Infrastructures such as complex freight nodes have, however, rather characteristics of natural monopolies which do not function as such (e.g. Varian, 2010, p. 439 ff.). Their, supply and demand can be balanced in other ways, for instance by the allocation of time periods to users or by auctions.

Core of the supply and demand coordination is the optimal design of individual processes as well as the ideal interaction between related processes. The planning of node capacities is therefore carried out in different levels. The current as well as the forecasted demand is taken into account in order to identify the optimum utilisation and possible bottlenecks.

2.6.1 Transportation planning process levels

Transportation management influences directly the overall logistics performance of a transport chain and respectively its network nodes. A vast amount of academic literature is concerned with the definition of production management in general and transportation management in particular. In the commonly accepted multi-stage planning approach, a distinction is made in three levels of decision making:

Decision level	**Strategic**	**Tactical**	**Operational**
Perspective	Long-term	intermediate	short-term
Time range	more than 1 year	1 year to 1 month	weeks, days, hours

The strict separation between the three levels, however, is often not clearly transferable into practice since methods and procedures may overlap. Crainic and Laporte (1997) adapted the three-stage concept to intermodal transportation planning. The following classification of the three planning horizons relates mostly to the author's description:

Strategic planning involves typically long-term decisions such as the determination of the physical service network layout, the planning of facilities (e.g. construction of transshipment terminals), or the determination of transport demands. Furthermore, the determination of the ideal hub size is discussed at this level. Many other contributions focus on the impact on transport chain performance as well as on transportation costs. A good overview can be found in Crainic (2000). Strategic planning includes also the decision on investments in vehicle fleets. With the recent emergence of locomotive and wagon rental corporations however, vehicles can be available to freight forwarders on relatively short notice. This issue can therefore rather be seen in tactical planning.

The planning horizon often exceeds 5 years (Crainic, 2000). Because of the long planning horizon, strategic plans have to rely usually on aggregated input data, where for example customers, products or locations are merged to groups. Strategic decisions take place on a rather macroscopic level. Policy makers, consultants, international shippers, etc. engage in this type of activity to determine the general development and broadly shape the operating strategies of the system.

Tactical planning comprises decisions over a medium term horizon of several months up to one year. The efficient and rational allocation of existing resources is at the core in order to improve the performance of the whole system (cf. Crainic and Laporte, 1997). Tactical planning decisions concern particularly the design of the service network. Therefore the transport system is for example divided into long distance distribution and regional traffic (cf. Wieberneit, 2008). This includes the consolidation of transport flows and therefore the definition of transport routes (route choices), timetable construction, vehicle and crew scheduling, or general operating rules for terminals etc. Transportation lot sizes and service frequencies are determined.

For intermodal freight terminals, the tactical planning phase denotes for example the planning of track utilisation, the scheduling of cranes and other machines, and the scheduling of staff. Data are still aggregated at that stage. Crew scheduling, for example, is carried out to determine the number and type of required employees without assigning concrete names. The decisions are sensitive only to broad variations, e.g. the seasonal changes.

Operational planning involves the day-to-day and real-time decisions that are performed by the local management, e.g. by yard managers or dispatchers. At this level, decisions in rail freight operations are based in large part on "rules of thumb" and require planners with extensive experience. It comprises the implementation of detailed schedules for staff, machines, vehicles and other activities and their adjustment. In intermodal terminals, the actual routing of trains through the hub or the assignment of wagon groups to terminal tracks are essential operational tasks. Other decisions to be made concern for example the load order of trains, the redistribution of empty wagons, or the scheduling of shunting locomotives, as it was discussed more deeply in the process analysis in section 2.5. Data are required on a microscopic level and not any more aggregated. Thus, individual freight wagons or loading units are considered, for example. The decisions made may have a direct impact on other facilities or network nodes but also on other stakeholders and their internal short-term production plans.

Bontekoning (2006) pointed out that the scientific challenges in operational management are threefold: "First, to develop techniques to deal with the immediate planning problem. Second, to develop heuristics to optimize as much as possible for broad planning problems. Third, to develop fast heuristics for real-time application." In particular, the topics of real-time control and rescheduling are current challenges in operational planning to replace the widely prevailing rules of thumb.

A summary of the planning levels is depicted in Figure 2.12. The slot management system to be developed in this thesis is to be classified in the tactical and the operational level.

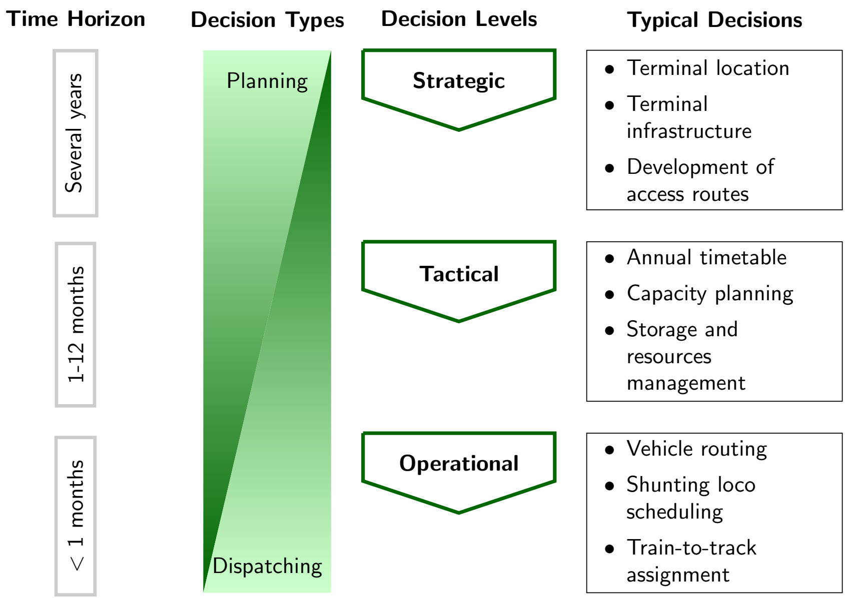


Figure 2.12: Decision levels and operations in freight node management (extension of Vacca (2011, p. 22))

2.6.2 Capacity planning of rail freight nodes

Capacity or performance of railway infrastructures can be defined as the "Maximum number of trains N that may be operated using a defined part of the infrastructure at the same time during a defined time period $[1/T]$, T: Time period, e.g. 24 h" (Hansen, 2013). In terms of freight transport nodes, the capacity assesses the fulfilment of service requests, i.e. the number of transport units (trains, ships, trucks) whose requests for a service (handling of cargo) can be satisfied. Capacity or performance assessment defines thus the modelling of the operations as service processes with its corresponding input and output variables. Accordingly, the prevailing modelling approaches are based on the concepts of service systems.

The corresponding input variable for the service process, and thus the load on the system is described with the *transport demand*. The amount of effectively served requests forms the output of the service process that is described as the *transport volume* (see Figure 2.13). The transport demand represents the amount of incoming transport units with demand for service. Indicator for the transport demand is the traffic intensity, defined as the number of incoming traffic units per time (e.g. trains/h). In addition, the arrival rate λ can be introduced as a measure of the intensity for transport demand. It is defined as the reciprocal of the time intervals between successively arriving elements, also referred to as inter-arrival times. The arrival rate is taken up

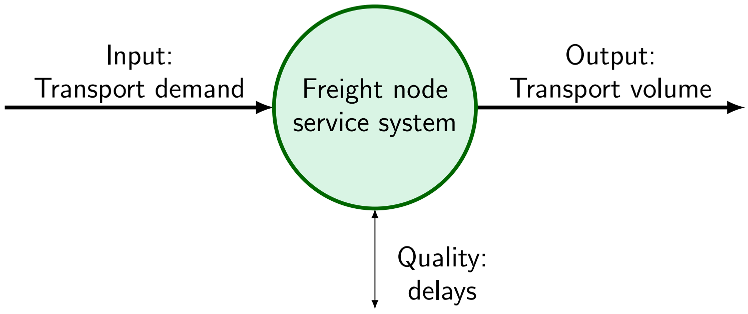


Figure 2.13: Fundamental input and output variables in the performance modelling of freight nodes

again in chapter 4 where it plays an important role in the capacity analysis of freight nodes.

In case of excessive demand, not all service requests can be served immediately. Waiting times arise and operations are getting delayed. Waiting times are a measure of operational quality of the system since they indicate delays and have thus a negative influence on the transport operation. Waiting times can be described as a function of load as depicted in Figure 2.14. The waiting-time function has a progres-

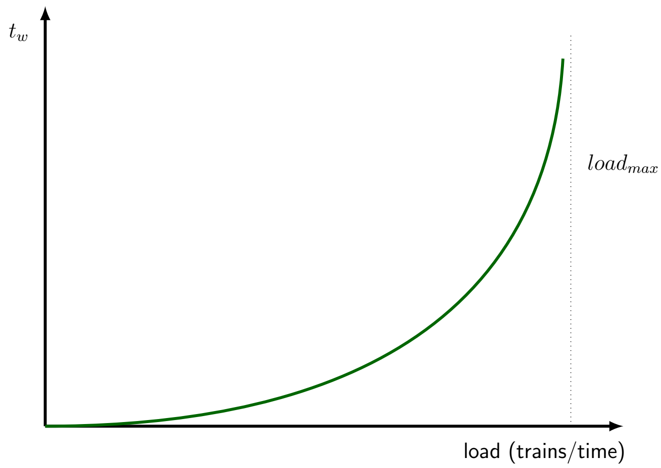


Figure 2.14: Waiting-time function

sively growing course and converges asymptotically to a load value $load_{max}$, which is referred to as the maximum performance. At this value, the waiting time t_w would theoretically be infinite. This does not mean that the train movements in the system come to a complete standstill. Rather, a growing queue builds up which can not be reduced by the system with the maximum possible throughput (cf. also Pachl, 2013, p. 129 ff.).

Of particular relevance to the performance is the temporal distribution of the traffic demand, and thus the regularity of vehicle arrivals. A distinction can be made between cyclical and stochastic deviations. Fluctuations in the transport demand are called cyclically if they depend, for example, on the day, year, or week. Stochastic fluctuations relate to the distribution of the number of arriving trains in a time period or to the inter-arrival times (cf. Potthoff, 1975). Stochastic deviations are often identified in capacity analyses as the source for congestion.

In complex rail freight hubs, the performance is planned on the tactical level. Railway undertakings interested in calling at a freight node are usually requested to report their demands until an appointed date (deadline). The node operator develops an annual timetable from the obtained transport demand and can plan track occupations in detail for formation tracks and terminal tracks. The goal is the most efficient allocation of tracks to trains. For example, trains are classified according to their length to an appropriate track so that infrastructure usage is optimised. Possible occupation conflicts are resolved in accordance with the transport companies. In addition to the annual planning, the allocation of occasional train services and short-term traffic to the remaining gaps is worked out during the year.

Notably, complex freight nodes do not pursue a holistic capacity planning. In many systems, individual system components such as the marshalling yard and the freight terminal itself are being scheduled independently because they belong to different owners. This leads, depending on the ownership of the components, to increased coordination efforts. In contrast, the Bremen port railways for example, pursue in this regard already an innovative approach by scheduling both, the formation yard and the terminal tracks in one step (Schönemann, 2015) despite various ownerships. The method presupposes an intensive collaboration between the port railway and the terminal operator Eurogate but it reduces the efforts for the railway undertakings enormously and increases the performance of the facilities towards a global optimum.

Another phenomenon in practice is that the long-term capacity plan is of no importance any more in the implementation phase at the operational level (as well Schönemann, 2015; Gille and Schönemann, 2009). Dispatchers on duty book the incoming trains with an advance of 2-3 hours onto available tracks. Here, the spontaneous, improvised operation, already mentioned in Marinov and Viegas (2011a) comes to the fore. Reasons for the lack of coordination are diverse and are often the result of the difficult coordination between infrastructure operators, railway undertakings and terminal operators. The quite spontaneous train handling creates increased disposition efforts, which lowers the total performance of the node. In chapter 4, the effects will be studied in detail with the aid of an example freight node.

Further processes such as the coordination of shunting locomotives are usually not coordinated by a superior organiser. The shunting operator is appointed by the railway undertaking that booked the time slot in the freight node and has no direct contractual relation to the infrastructure manager who is scheduling the train movements. Accordingly, the shunting operator optimises its own processes depending on the jobs assigned to him. Necessary train movements are coordinated spontaneously with the dispatcher on duty.

Operators have limited control over system constraints so that the quality of process coordination depends accordingly on the optimisation strategies of the individual actors. It does not necessarily lead to an overall optimal solution. With regard to the growth in freight transport, as it will be described in the next section, it becomes apparent that the quality of process coordination requires a better organisation in order to absorb the expected increase of traffic volumes in freight hubs.

2.6.3 Growth in maritime traffic and its effects on hub development

The continuous increase of transport volumes in the recent decades becomes noticeable in the growth of port container flows, as evident in Figure 2.15. With it, container

Figure 2.15: Flow of containers from land to sea transport modes, and vice versa, in twenty-foot equivalent units (TEUs), development between 2000 and 2012 (The World Bank, 2014)

ships are getting bigger and in relation to their size and growth apparently only a few unsolvable technical or technological limits remain (see Schönknecht, 2009, p. 3).

Currently the largest container ship in the world is the Maersk Mc-Kinney Møller with space for more than 18,000 containers (TEU) (Maersk Group, 2013). Container ships with a capacity of 10,000 TEU and more are not uncommon in the major ports of the world. This development has far-reaching effects on the utilisation of port hinterland connections, because the large number of containers must be transported and distributed with the help of numerous individual shipments on the land side. Land side modes of transport (rail and road) could not keep up with the growth of the sea ships. Due to legal and infrastructural restrictions, as stated by Schönknecht (2009, p. 4), none of these modes may be able to grow in its dimensions or weights notably. It can therefore be assumed that the differences in the transshipment performance between sea ships and other modes will further increase. Table 2.1 illustrates the maximum dimension of hinterland transport modes today. The table contains also

Mode of transport	Dimension	Max. container capacity
Truck	40 t	2 TEU
Train 700 m	700 m	100 TEU
Large river barge	110 x 11.45 m, 3 layers	208 TEU
Pushing unit	185 x 11.45 m, 4 layers	320 TEU

Table 2.1: Comparison of the transport performance of different hinterland transport modes (cf. Schönknecht, 2009)

data for barge and feeder ship transport. However, river navigation is not further considered in the scope of this thesis. Depending on the quantity of cargo that is to be transferred between the sea ship and the shore-based transport modes, Schönknecht (2009) determines the following effects that may occur with increasing vessel size:

- Increase of the land-side transport operations
- Increase of the storage consumption and the storage density
- Prolongation of the ship dwell time in the port

The growth in the shipbuilding industry of the past 10-15 years gives an impression what challenges are to come to the networks of land-side transport modes in the next years and decades.

2.7 Chapter summary

Railway networks are organised in a hub-and-spoke architecture from the infrastructural point of view but also from the perspective of the production systems. The network nodes are of varying type and importance (hierarchical grading), depending on different criteria such as their geographic location, traffic volumes or the types of handled goods.

Of the most interest for this thesis are complex freight hubs for intermodal transport with heavy traffic. Such nodes are for example seaports or industrial railway hubs. The process sequences taking place in those hubs require a temporal and spatial coordination of infrastructure resources by different users (trains). A microscopic examination of the rail-based processes in hubs revealed the complex relationships between them and the multitude actors clearly.

In complex railway networks, timetables are used for the coordination of vehicle movements. These timetables are updated on a regular basis and in multiple time horizons. For freight hubs and complex transshipment terminals that applies only inadequately. For the determination of the actual conditions in rail freight hubs, observations and interviews with rail freight experts were carried out at several places in Europe. It became known that tactical infrastructure utilisation planning in such nodes is currently poorly coordinated and rather spontaneous. Further, the quality of process coordination depends on the optimisation strategies of the individual involved actors. Due to the absence of a superior coordinator it is expected that there is a significant potential in the short-term planning and in the dispatching strategies for increasing the efficiency of such nodes.

3 Slot management as a planning mechanism

With manifold operations, carried out by numerous actors, complex freight hub infrastructures, as they have been described so far, require a certain degree of planning and coordination to perform properly. Slot management is a traffic control measure suitable to raise punctuality and productivity in network nodes by applying structured planning mechanisms.

This chapter is engaged with the definition of slot management techniques. However, as already stated by (Koolstra, 2005, p. 5), "there is no well-established state-of-the-art of slot allocation research, because it is mainly confined to a few specific topics." The work by Koolstra is one of the very few that analyse slot management applied to transport scientifically. Building on it, this chapter attempts to develop a valid definition of slots and slot management related to complex freight hubs. After the scientific classification of the concept, existing implementations of slot management in transportion are presented. Theron, the focus is again directed to the railway sector and first existent implementations in the field of rail freight are illustrated.

3.1 A general definition

The term *slot management*, requires the definition of the two individual words slot and management. In terms of process management a slot can be paraphrased as a *time window*. The Cambridge University Press (Cambridge Dictionary of British English 2015) defines a slot as "an amount of time that is officially allowed for a single event in a planned order of activities or events". This clearly shows the temporal character of slots. In terms of transportation, this definition should be extended insofar as that it typically means assigning elements to space *and* time.

The most clear definition of the term with regard to transportation derives from research in aviation. The European Union (1993, Article 2) defines a slot as "the scheduled time of arrival or departure available or allocated to an aircraft movement on a specific date at an airport". With regard to rail freight operations, the term shall be defined in this thesis:

> A slot is a spatio-temporal scheduled assignment of an element of scarce infrastructure to a certain occupant.

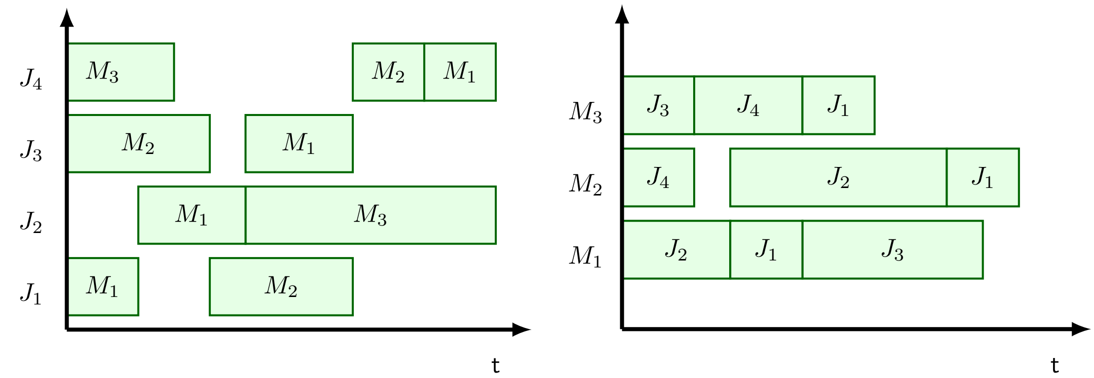
This definition specifies an elementary slot, meaning the spatio-temporal assignment to a single piece of infrastructure. However, a train is running on several infrastructure elements over time and thus, the determination of elementary slots is not sufficient for slot management. For the processing of trains through different facilities of complex freight nodes (preceding and succeeding track occupations), slot paths (sequences of single slots) are needed. A major challenge tackled in the further course of this thesis is precisely the development of slot sequences satisfying a determined service quality of freight nodes.

Slot *management* is the procedure of the allocation, administration and control of slots. The allocation of slots outlines that capacity constraints are obeyed when slots are assigned but also that available capacities are optimally exploited. Slot management is a cross-divisional coordination mechanism which touches several fields of research:

- Business economics,
- Mechanical engineering,
- Computer science,
- Information science.

As a management mechanism, it is a discipline of business economics. Particularly when applied to transportation, mechanical engineering influences slot management due to the necessary consideration of technical and transport-related peculiarities. Also, the development and implementation of slot plans requires today the utilisation of appropriate IT services and it relies on relevant input data. It is therefore home in the domain of computer and information sciences.

Slot management is scientifically related to production process planning or machine deployment planning where a set of jobs is assigned to be executed on a set of machines (resources). Thus, it specifies the sequence of tasks on one or more regarded resources. The ordered sequence of jobs forms a production schedule containing the planning intervals in which each job can be executed on the required



(a) Job-oriented Gantt chart **(b)** Machine-oriented Gantt chart

Figure 3.1: Two variants of Gantt charts. (a) job-oriented, showing the sequence of machines a job is executed on and (b) machine-oriented, showing the sequence of jobs being executed on a machine.

resources. A production schedule may be visualised by a Gantt chart that may be either job-oriented or machine-oriented (Figure 3.1).

Returning to railway operations, production planning methods can be found amongst others in the track allocation process. The classic timetabling procedure, however, differs from industrial or logistics production planning (see later section 3.4.1). Moreover, slot management for freight nodes considers further scientific approaches, such as supply chain management (SCM) and operations research (OR). The support provided by methods of SCM is necessary because slot planning in complex freight hubs is a multi-dimensional optimisation approach that requires the consideration of transport-related as well as logistics-related matters. In freight nodes, the slot sequences of different facilities are to be coordinated in time so that all resources can be optimally utilised. To achieve this objective, methods of OR are used for scheduling freight yards of a certain complexity.

3.2 Slot management in a supply chain context

The term Supply Chain Management (SCM) has been mentioned at first by Keith Oliver in 1982 (Oliver and Webber, 1992). Since then, many definitions have been created; however, not one is universally valid today. A network-oriented, more recent definition of SCM is expressed by (Stadtler and Kilger, 2005), aiming to describe the essence of existing proposals: "Supply chain management (SCM) is the task of integrating organizational units along a SC [Supply Chain] and coordinating materials, information and financial flows in order to fulfil (ultimate) customer demands with the aim of improving competitiveness of the SC as a whole."

SCM is a strategic concept that describes the design of object and process flows in order to raise customer satisfaction and to increase the efficiency in the various stages of supply chains. It aims to coordinate flows across organisation's borders and thus comprises the fields of logistics, production planning, marketing, and controlling (see Günther, 2005). Being a management tool, SCM raises the question who is *managing* a supply chain? Two kinds of SCM strategies can be characterised: centralised and decentralised. The question of management goes along with the term *coordination*. Centralised, or focal, chains are coordinated by a powerful leader which decides on actions with impact to the whole supply chain but considers his own margins only (cf. Albrecht, 2009, p. 21). Typical examples for coordinated chains are component supplier chains in car manufacturing industries which are controlled by the big car manufacturers.

In intermodal transport, process chains, as they have been examined in section 2.4, can be perceived as decentralised, or polycentric. A leading coordinator is often not present and SCM, as a strategic concept, is poorly implemented, accordingly. The existing links of processes can rather be considered as pure transport chains which fulfil the logistics needs for forwarding goods. As a management tool, SCM does not only comprise the establishment of transport chains but supports also the integration of information flows and coordination which leads to a higher overall transport efficiency.

Slot management has the ability to accomplish SCM approaches in those transport chains and to enable efficient production process control at scarce infrastructure facilities. Consequently, the node in which slot management will be implemented, can experience a higher level of coordination. Moreover, the actor that introduces slot management at its own facilities will gain a higher power in leading the chain. Accordingly, the success of slot management depends on the node operator's ability of enforcing own objectives among other actors of the chain.

3.3 Successful implementations of slot management

This section examines implementations in transportation where slot management is already introduced as a coordination measure. However there are more, three applications have been chosen to illustrate the variety of possible applications: First, the slot allocation in aviation is considered as one of the most established implementations. Further, parcel delivery services are examined as an application in logistics, and at the end, the slot-based intercommunication concepts in IT networks.

3.3.1 Aviation

At airports, slots are assigned to scarce infrastructure resources, e.g. to runways for planes landing and taking off. The runway is usually the bottleneck of an airport system. However, time windows on big airports are often assigned to continuative resources like gates or parking lots, too. Thus, complex systems of slot sequences, similar to production schemes in SCM, are designed: "A slot covers the full range of airport infrastructure" (Wye River and de Wit, 2007, p. 3).

Not all airports control flight movements by slots. The International Air Transport Association (IATA) distinguishes airports by three levels of slot coordination as shown in table 3.1. The classification of airports to these levels depends on the excess of the demand for aircraft movements.

Level	Definition
1	**Data collection (monitoring):** capacity of the airport infrastructure is generally adequate to meet the demands of airport users at all times
2	**Schedules Facilitated:** potential for congestion during some periods of the day, week or season, which can be resolved by voluntary cooperation between airlines.
3	**Fully coordinated:** demand for airport infrastructure significantly exceeds the airport's capacity during the relevant period [...]

Table 3.1: Airport coordination levels (cf. IATA, 2012)

Slot allocation at coordinated airports is arranged by a coordinator institution that is generally appointed by the country in which the airport is located. The allocation of slots to customers (i.e. airlines) follows certain priority rules, e.g. the historical precedence (grandfather rights), retimings, or local guidelines. For more details on airport slot allocation it may be referred to relevant literature, e.g. Wye River and de Wit (2007) or European Union (2010).

Such procurement rules do not exist in rail freight node so far. But, when awarding slots in complex freight hubs, another important aspect is to be regarded in contrast to aviation: The slot assignment time on a runway is very short. Hence, many planes can be handled within a short period. In contrast, in rail freight transport it is to assume that the occupancy times of the infrastructure elements (e.g. during loading and unloading of trains) will be much longer. Accordingly, only few slots per infrastructure element can be scheduled over a period. In the dynamics of slot plans and their rescheduling (chapters 4 to 6), this effect will therefore be taken into account.

The slots in aviation discussed so far are more precisely landing or departure slots. Their purpose is the coordination of ground-based facilities. Slots can be assigned to airways[7] too. Airway slots are being assigned in Europe by the Directorate Network Management (DNM)[8] when airspace is scarce, e.g. in bad weather conditions in order to shift capacity allocation conflicts to the ground. Especially the flying of holding patterns can thus be avoided. If airway slots are assigned, the originally advised planned landing slot plan gets disarranged, consequently. The airlines and the airports have no influence on that.

Airway slots indicate that there are capacity constraints on spokes of aviation networks and thus, there are certain similarities to railway networks from the organisational point of view. Congested airways can result in delays or the usage of alternative routes. Analogies to rescheduling actions in railways are noticeable. However, the operational processes differ significantly. An air plane will, for example, not start if it is known that his destination airport has no free arrival slot. The operational rail freight scheduling does not work that far. Trains leave their origin yards and run piece by piece up to the next route conflict where they wait at a siding or a passing loop until the track is clear again.

When regarding the trip announcement process, significant differences between rail and aviation appear, too. The submission of a flight plan (announcement of a flight) is carried out on a short-term basis, generally at least 60 minutes before departure, (cf. for example Civil Aviation Authority, 2009). The announcement of a train trip at short notice takes longer, for example 48 hours in the German rail network (cf. DB Netze AG, 2014). This is again more closely discussed in section 3.4.1.

3.3.2 Parcel delivery services and freight distribution

The adoption of time slot management mechanisms in logistics did, unlike in aviation, primarily not arise for infrastructural capacity regulation but to increase customer satisfaction to form supply chains. “The growth of e-commerce and postal

[7] airway: pre-defined flight path in the air

[8] The Directorate Network Management is a division of EUROCONTROL, the European Organisation for the Safety of Air Navigation, (www.nm.eurocontrol.int).

shopping [..] have reinforced the importance of "just in time" policies in freight distribution.” (Deflorio et al., 2012). The development was supported by recent advances in new technologies (tracking and tracing, information technologies, smart phones, etc.) which created new logistics products. Besides the service quality improvements, also cost reduction can be identified as a motivation for using slot management in parcel delivery and freight distribution. Similar organisation structures can be found in Less than Truck Load (LTL) transport. Basically, slot management is applied in the following fields:

Home delivery and parcel services: Postal service operators provide time frames for delivering shipments in order to raise the likelihood that the customer is available for receiving the consignment. Going along with an increased customer satisfaction, this action allows to reduce delivery costs by avoiding empty runs and unsuccessful delivery attempts. Allowing the customer to choose pick-up or delivery times enables the operator to design vehicle routes “to satisfy a set of transportation requests, each involving a pickup and delivery location, under capacity, time window, and precedence constraints” (Ropke and Cordeau, 2009).

Distribution centre door allocation: “The total travel time of a vehicle trip [of freight distribution services] depends on travelled distance, its typical speed and the time spent on each pickup and/or delivery stop [...], but it is also affected by waiting and access time, congestion, deadlines or service features, etc.” (Deflorio et al., 2012, p. 2). To optimise operations at distribution centres, trucks are assigned to loading doors and to suitable time slots for (un)loading. According to Chmielewski et al. (2009, p. 199), arriving tours will be assigned on an operational level, hence on relative short notice. The first “objective of the planning task is to find an optimal allocation that leads to minimal total distances [of the distribution centre’s inner flows] and a minimal number of resources needed in operations. The truck which shall be (un)loaded is thus assigned to a door that is close to the storage area that it needs to use. A second objective is the minimisation of truck waiting times.” However, the second objective is rarely integrated, as Chmielewski et al. (2009, p. 199) further states. Significant to note is that the resource scheduling (door assignment) is already carried out by automated IT mechanisms in modern warehouses.

Since slot management approaches in delivery and parcel services are not primarily used to administer scarce infrastructure facilities, no central coordinator institution is appointed. The organisation of time slot mechanisms is operator-based and thus, heterogenic on a competitive market. Though, in the case of logistics facilities, e.g. warehouses, a central operator (and owner) manages the scarce infrastructure.

For the sake of completeness it shall be mentioned that in general road transport, slot management can also be applied to infrastructural bottlenecks such as motorway mergers. This topic is not further pursued because the organisation of road traffic (simultaneous, competing use without timetable) is different from railway operations which is not beneficial for the purpose of this thesis. From this section can be taken that slot management of road infrastructure is possible but established systems like the approach in parcel delivery services differ significantly from those in the air or rail transport sectors.

3.3.3 Computer networking

In CSMA/CD[9] computer networks such as Ethernet[10] (Ethernet Working Group, 2015), a *slot time* is used to determine the network device that can use the communication link next. The slot time is defined as "the amount of time a device waits after a collision before retransmitting [a data package]" (Rouse, 2007). The slot time is particularly important for half-duplex transmissions, hence those where communication in both directions is possible, but only in one direction at a time.

Slot time in computer networks is consequently used to coordinate traffic on scarce infrastructure facilities, too. However, no central coordinator exists - all partners of the network (computers, router, etc.) have equal priorities. In contrast to transportation networks, the slot time in computer networks is not used to avoid conflicts (collisions) but to detect communications errors and to initiate a second transmission. This approach differs significantly from the other examples. The consequences of conflicts in computer network transmissions are considerably lower than for example on an airport runway where comprehensible security issues arise.

The lack of ex-ante coordination allows an increased amount of data that can be transmitted in a given time interval. But, with the number of network members also the number of collisions rises and with it the slot times required for collision dissolving. High network traffic leads to higher ex-post coordination and thus to significant speed reductions. In summary, computer networks pursue a completely different slot strategy, which can not be transferred to transport networks, primarily due to security issues.

[9]CSMA/CD: Carrier Sense Multiple Access/Collision Detect
[10]Ethernet: Technology for Local Area Networks, specified in the standard IEEE 802.3

3.4 Slot management design in railway operations

Slot management in general is not new to railways. All railway traffic needs to be coordinated, at the latest on the operational planning level. Slot coordination is however not established in all relevant stages of the production process. This section describes the general track allocation methods in the European railway systems briefly. Then, the focus goes back to rail freight transport in which the main challenges for slot management can be found.

3.4.1 The general railway track allocation

The efficient coordination of train movements on the network is a central task in the production planning proces of railway infrastructure operators. Routes of all train runs of a given period are to be registered in a timetable. The central planners assign the necessary infrastructure resources to each train in the timetable in terms of a spatio-temporal allocation and thus ensure the conflict-free routing under the consideration of capacity constraints.

The planning is carried out in great detail. The network arcs are cut into block sections on which slot paths will be generated. A microscopic division of a track between two interlocking stations into block sections is illustrated in Figure 3.2. The green blocks represent figuratively time windows on a route assigned to a certain train run. In each block, only one train is allowed at a time[11].

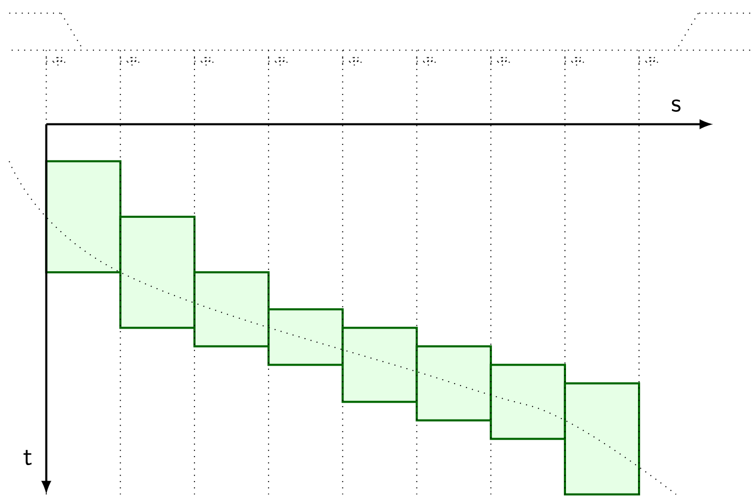


Figure 3.2: Blocking time diagram (cf. Pachl, 2008)

The compilation and feasibility check of timetables is today carried out simulation-based using software packages such as OpenTrack (Nash and Hürlimann, 2004) or RailSys (Radtke and Bendfeld, 2001). Since this is a very complex undertaking, new optimisation-based methods are continuously emerging (see e.g. Lusby et al., 2011, for an overview), though they have not been able to replace the simulation systems, yet.

The whole track allocation process is carried out on the tactical planning level and requires a certain time, usually several months. Since freight forwarders do not estimate transport demand so far in advance, the process keeps going until into the operational level and daily scheduling. The work-flow of the annual capacity allocation process is very comprehensively described, for example, by Klabes (2010, p. 33 ff.) so that here only the main core is depicted. Figure 3.3 compares the annual planning processes of European railways and the aviation branch conveniently. It can be seen

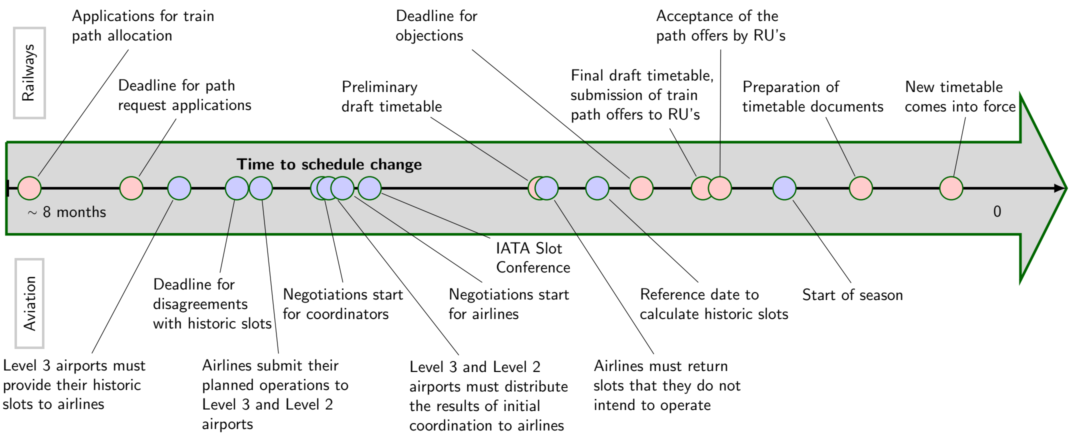


Figure 3.3: Time-line of the capacity allocation cycle in aviation (IATA, 2014) and European railways (DB Netze AG, 2015; European Parliament, 2001)

that the process in aviation is carried out shorter. Whereas aviation has two negotiation rounds per year (summer and winter schedule), railways update the timetables mostly annually. For rail freight transport, the annual timetable is increasingly insignificant, as the number of short-term route requests rises continuously as a result from flexible and more spontaneous logistics demands (e.g. Braun et al., 2005). Short-term route requests are not planned in the annual schedule and are being booked in both traffic systems in the free remaining slots on a First Come First Served (FCFS) basis. Understandably, not always ideal routes can be assigned thereby.

[11] The topic of headway control techniques in rail transport shall not be discussed further in this work. Reference is made to the well-established literature (e.g. Pachl, 2008).

3.4.2 Present state of slot allocation in rail freight yards

In complex freight hubs, slots are provided to shippers and railway undertakings (RU) respectively for the usage of loading tracks (under the crane). In these time windows, an assigned track is available to the RU exclusively. Within the reserved slot time, the desired cargo handling must be carried out. Following from capacity constraints, the loading slots are limited, as only a certain number of slots can be awarded per day. The shorter the dwell time of a train can be, the more slots may be provided per day. Based on a series of interviews with experts of different rail freight nodes (here and in the following Gille and Schönemann, 2009; Schönemann, 2011, 2014, 2015), the current state of process planning can be summarised as follows:

The allocation of loading slots takes place on the tactical planning level, especially for the regularly recurring traffic (for planning stages see again section 2.6.1). Additional traffic demand can be assigned to the remaining slots on short notice. Most of the complex nodes that were analysed, conduct consequently an annual capacity plan that follows from the preparation of the annual railways timetable which was explained in the previous section. The annual slot plan is being updated then monthly or weekly.

The slot allocation in loading tracks is mostly carried out by the terminal owner or terminal operator without the consideration of the railway operations of other stakeholders in the transport chain. Some complex hubs pursue already little integrated slot management approaches. The Bremen port railway, for example, assigns already loading time windows according to the planned arrival times of incoming trains in the **tactical planning phase**. However, the mapping is not as detailed and does not consider all existing processes. The port railway plans dwell times in the formation yard and in the terminals accurately for each track. The scheduled time window between the formation yard and terminal overlap by one hour which is available for the shunting moves between the two locations. Though, no exact shunting routes are planned in this phase.

Generally, no comprehensive software is in use for slot planning purposes. Most freight nodes use proprietary developments or own spreadsheet-based applications. A problem from praxis is the data storage – the amount of data for a whole year can be managed with the existing solutions only with great synchronisation effort. Considering the fact that several planners are working on the same data simultaneously, would make a database solution favourable.

In the **operational planning phase** and during production, the national railway infrastructure manager books incoming trains directly into the formation tracks and controls thus parts of the yard-internal infrastructures. This can be a problem in real-time dispatching situations. It is important to know that the initially developed tactical schedules are of no consideration any more in the effective operational scheduling. Dispatchers observe only the train arrivals in the next 2-3 hours and schedule them individually. Thus, the efforts of generating the tactical capacity plan are being questioned. The logical break between the tactical plan and the actual dispatching methods becomes apparent.

The further development of the existing slot management approaches is desired but also faces difficulties. A crucial nuisance are the manifold tasks and responsibilities which are scattered among a multiplicity of actors with some competing interests (cf. section 2.4.2). The successful establishment of slot management in rail freight yards depends on whether the hub manager is able to enforce its slot rules and thus to take a leading position in the supply chain.

General supply chain (SC) leaders in complex freight nodes are the globally operating sea-shipping companies. Large container terminals align their process organisation on the fastest possible handling of the large ocean-going vessels in order to minimise their berth time. The freight node operator could, however, enforce specific regulations on the land side. Though, the national railway infrastructure manager to whom it is physically connected to, is also a strong leader that aims in prevailing its working schemes as mentioned in the case above. Consequently, the freight node is organisationally trapped between two large Supply Chain (SC)-conductors which potentially weaken its own position and its operational flexibility (Figure 3.4).

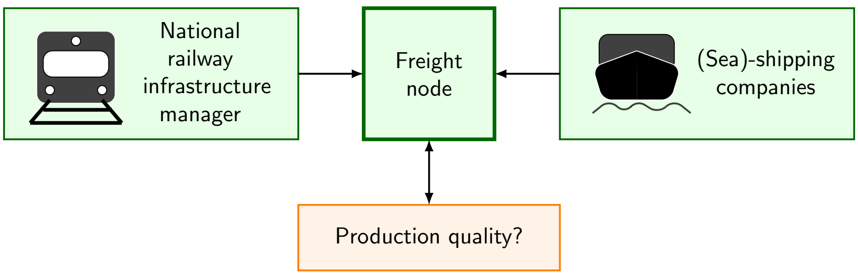


Figure 3.4: A complex freight node in the supply chain. The operational quality is suffering under the influence of the adjacent competing stakeholders.

Being located between two strong stakeholders, how can a freight node enforce a slot management system and become a leader in the (sub-) chain? This can only be achieved in a careful approach, with the cooperativeness of all stakeholders and the development of a quality assurance which ensures the customers a continuing safe train and freight handling. The remaining chapters of this thesis aim at tackling this issue by developing a two-level planning approach that integrates initial tactical capacity planning and operational scheduling. It thus aims at producing reliable slot schedules which are valid on all planning levels and support the affected stakeholders in overcoming the currently existing logical break in the present rail yard scheduling policy.

3.5 Chapter summary

The chapter defined the term *slot management* in a general way and applied it to production systems in rail freight yards. Thereon, the meaning of the term was examined in a supply chain context and three examples of existing slot allocation systems were presented. From the considered approaches, airport scheduling was the closest similar to rail yards from an organisational point of view. However, it can be expected that slot management for rail yards is in conclusion harder to plan than for airports since freight trains interrupt their journeys in case of route conflicts, which increases their uncertainty of arrival time. Air planes, in contrast, fly usually uninterrupted from origin to destination.

Further, the general slot allocation process in railway timetabling and the current situation of train handling in complex rail freight nodes were discussed. It became evident that the rail freight nodes suffer from certain drawbacks in process flows. Especially notable was the fact that the initial tactical capacity plans are not being observed in day-to-day scheduling which results in efficiency losses. Thus, a systematic gap between the tactical capacity plan and the operational yard scheduling was identified which shall also be tackled by the slot management approach.

It was noted that the rail freight node is organisationally trapped between two strong stakeholders of the supply chain and it is a challenge to enforce slot management production strategies on these premises. The following chapters will develop a solution on how this central challenge of this thesis can be tackled by a cooperative two-step scheduling approach that incorporates tactical slot planning and operational (re-)scheduling in order to satisfy the requirements of all involved stakeholders on the freight handling process in complex hubs.

4 Productivity and capacity of rail freight nodes

In order to define the structures of a possible slot management approach for rail freight nodes, the process analysis from section 2.5 will be further exploited in a yard simulation model in which the transport flows of intermodal transshipment terminals are investigated. The simulation model provides information on the yard productivity and allows determining the yard service quality based on selected indicators. This productivity analysis thus presents the basis for the development of slot plans for complex rail freight nodes.

The maximum productivity of an intermodal freight node is usually specified with the annual turnover (the number of loading units handled per year). The figure is usually based on the calculated optimal usage of the transshipment equipment and the facilities in the hub. Numerous factors from inside and outside are limiting the theoretical capacity of intermodal transshipment terminals when loading to and from rail. A hub's real productivity, depends on external influences such as the train timetable or the service quality of the railway network it is connected to, but also on the organisation of the own process flows.

For the development of a slot management approach, the analysis in this chapter enables identifying the durations of individual processes, potential bottlenecks and other capacity limiting factors from observation data. An event-based simulation system is used to set up a network-oriented reproduction of rail freight node operations on a microscopic level. The precise mapping of processes and interrelations among processes allows the consideration of all details. In the following two sections, the developed simulation model is presented. Section 4.3 depicts relevant input data. Subsequently, the evaluation and analysis of yard productivity follows in section 4.4.

4.1 General procedure on modelling

In order to ensure a structured modelling approach, the simulation will be performed in these steps with reference to the Association of German Engineers (Verein Deutscher Ingenieure, 2014):

1. Task definition
2. System analysis
3. Model formalisation
4. Implementation
5. Experiments and analysis

Steps 1 and 2 have already been discussed in the previous chapters. The steps 3 through 5 are explained in the following sections. The model formalisation (step 2) is described in section 4.2. Parallel to the five steps, the phases *data collection* and *data processing* are carried out. The necessary input data and assumptions are particularly relevant in step 4 (section 4.3). In order to obtain clear-cut results and to simplify the interpretation, the data are being processed through different simulation experiments in step 5 (section 4.4).

Before the simulation experiments can start, a model, based on the real processes must be created. The modelling is carried out in two sub-steps: First, the development of a "symbolic model" and second the transfer into the resulting "simulation model". The symbolic model is an abstract image of the original system in which the processes and parameters can be reduced and idealised. In the second step, this abstract model is transferred into an actual simulation model using a simulator environment.

4.2 The hybrid simulation model

Different investigation methods fall within the framework of analysing rail freight nodes. Microscopic railway simulation software (e.g. OpenTrack or RailSys (cf. Nash and Hürlimann (2004); Radtke and Bendfeld (2001)) provides good help when analysing the capacity of railway infrastructures and the performance of train operations. Logistics processes, such as loading of wagons however, cannot be modelled by them. Contrarily, terminal simulation models which map logistics processes do not consider railway-specific prerequisites (cf. section 1.2.1).

A practical approach for the integrated analysis of railways and logistics operations can be achieved in the combination of queueing systems and event based simulation methods, initially proposed by van Dijk (2000): *Queueing* is used to compute the performance of any kind of facilities. However, the application of queueing theory for daily-life problems has remained rather restricted. *Simulation*, contrarily, is found to model and evaluate realistic situations in both daily-life and industrial environments. The hybrid approach makes use of the advantages of both methods. By *queueing*, process sequences of freight hubs can be illustrated and assessed. General rules and insights might thus be developed. *Simulation* enables the evaluation of these findings and the comparison of variants. For a summary on the advantages and disadvantages of queueing and simulation, reference is made to table 4.1.

	Queueing	**Simulation**
Advantages	• Insights • 100% exact • Generic components • Few detailed data necessary	• Allows real-life complexity • Allows real-life uncertainties
Disadvantages	• Too restricted for reallife modelling • Strongly simplified uncertainty assumptions (exponentiality)	• Numbers • Confidence? • Too much complexity • Detailed data required

Table 4.1: Disadvantages and advantages of simulation and queueing analysis (cf. van Dijk, 2000)

Accordingly, a hybrid approach enhances the complexity of queueing, while it produces more confident simulation results. The concept is, with adaptations, transferable to the analysis of rail freight nodes. In a first stage, typical process sequences as those identified in section 2.5, are mapped in an analytical queueing model. A second stage uses simulation to extent the obtained results with a certain real-life complexity. This two-stage approach enables the implementation of more precise experiments than those who rely on simulations only.

4.2.1 Queueing systems

The capacity assessment using service or queuing systems refers to the following background. Queuing theory is concerned with the investigation of statistical processes in manufacturing systems (e.g. Ferschl, 1964; Gudehus, 1976; Hall, 1991; Gudehus, 2012). A queueing system consists of a service station, which is fed by an inbound stream and out of which comes an outbound stream[12]. Logistics objects (loading units, transport units, goods etc.) and information objects (information and data) run through the service stations. The objects are therein modified, processed or differently handled. A service station, as illustrated in Figure 4.1, is characterised by several parameters which determine the timespan between arrival and completion of an process:

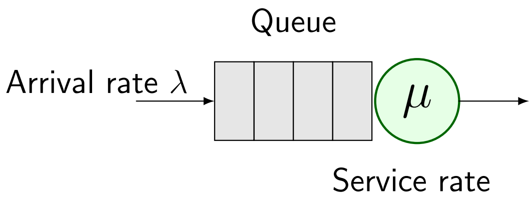


Figure 4.1: A queueing system with one queue and one service station (server). The arrival rate λ describes the process of incoming items. The service rate μ describes the processing speed of the service station.

New requests (incoming logistics objects, information objects) are depicted by the arrival process A. The time interval between the arrival of two objects is called inter-arrival time. Of the objects A_n, $n \in \mathbb{N}$, is assumed that they arrive randomly, that they are stochastically independent and equally distributed with the distribution function $F_a(n)$, the mean time between two arrivals $E(A)$, and the variance $D(A)$. The reciprocal value λ is called arrival rate and indicates how many objects arrive in average per time unit in the system:

$$\lambda = \frac{1}{E(A)} \tag{4.1}$$

The service process depicts the time taken to serve objects. Successive objects S_n, $n \in \mathbb{N}$, must also be stochastically independent and equally distributed random variables with the distribution function $F_s(n)$ and their corresponding mean operating time value $E(S)$ and the variance $D(S)$. The reciprocal value μ is called service

[12]For freight nodes, these streaming variables were already introduced with Figure 2.13, page 41.

rate and indicates how many objects can be handled per time unit at the station on average:

$$\mu = \frac{1}{E(S)} \tag{4.2}$$

In practical studies, the expected values $E(A)$ and $E(S)$ are typically expressed as the arithmetic means. For the characterisation of operating systems, Kendall (1953) introduced the notation $A/S/c$ where the letters A and S mark the type of distribution of the inter-arrival times and service times. Among others, the following abbreviations are the typically used to denote the distribution types A and S:

- D: deterministic distribution
- M: exponential (markovian) distribution
- G: general (independent) distribution
- others ...

The letter c defines the number of parallel servers. The denomination $M/M/1$ thus stands for a queueing system with exponentially distributed inter-arrival times, exponentially distributed service times and one server.

Transshipment nodes can be considered as networks of service stations that are connected by transport links. Thus, it is possible to map the processes and the interrelations between them to complex queueing network graphs. $M/M/1$ queues allow to interconnect individual queues to such queuing networks. Figure 4.2 illustrates the interconnection of three queues represented in a row.

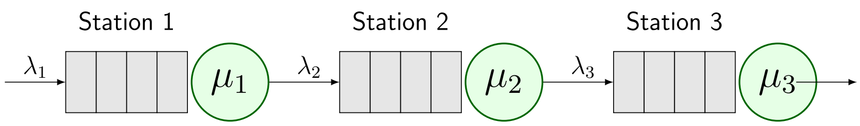


Figure 4.2: Notation of a queueing network with three serial service stations

Queues can be set in series but also in parallel with several service stations having the same function. In the following, the development of such a queueing network for freight nodes in an suitable simulation environment is described.

4.2.2 Event-based simulation systems

A simulation model of a freight node has been implemented in the event-based simulation environment Simul8 (Concannon, 2003). The software system is a universal tool for planning and optimisation of process flows, for example in production sites, manufacturing systems, or logistics. Simul8 is a discrete event simulation (DES) system. Discrete simulation is a method, which assumes that changes of the model state occur by leaps and bounds at discrete points in time. Event-based simulation is a discrete simulation wherein the time of events is set by the events themselves (cf. Verein Deutscher Ingenieure, 2013). It is assumed that no changes occur between two consecutive events and thus the simulation can directly jump in time from one event to the next. In contrast to event-based simulation, in *continuous or activity-based simulation* the system state is updated according to a set of activities happening in predefined time slices. Discrete-event simulation does not have to simulate every time slice and can thus typically run faster than a continuous simulation.

Simul8 is a microscopic simulation system which produces realistic results on a very detailed level. A Simul8 simulation model can be considered as a graph of queuing systems through which so called *work items* (freight trains in the sense of the thesis) flow. Work items can be broken down into smaller units so that, for example, individual wagons or loading units may be considered. Queueing systems are represented by so called *work centres* (service or queueing stations, cf. Figure 4.3). Work centres represent activities through which the work items go through. In the sense of the terminal simulation, different rail tracks are considered as work centres where activities on trains or wagons are carried out.

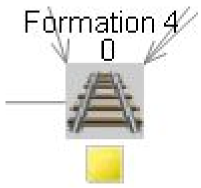


Figure 4.3: Work centres representing queueing stations in Simul8

In terms of the simulation, a work centre can have several states:

Waiting	Track is empty and waiting for new tasks.
Working	Track is occupied and an active process is carried out, e.g. a train is being loaded/unloaded.
Blocked	Track is occupied and processing is completed but train cannot leave due to an infrastructure conflict.
Resource starved	Track is occupied but resources are unavailable, e.g. crane not ready or staff is not available.

Trains pass through the model addressing different stations in a specific sequence. The work centres are therefore connected to a queueing network and a routing logic is implemented to ensure the succession of the train flow. For specific processes, staff and machinery is employed as so called *resources*. The inspection of incoming cars for example, requires the employment of wagon technicians. Loading and unloading require the provision of cargo handling equipment and the appropriate operators. Resources are implemented with a limited capacity (e.g. working hours, number). The productivity of the loading tracks, for example, is limited by the number and type of cranes.

4.2.3 Modelling approach

The simulation model represents an inland transshipment terminal based on the yard design depicted in section 2.3.1. First, the symbolic model was developed. Its infrastructure layout is represented in Figure 4.4. It consists of a marshalling yard with four

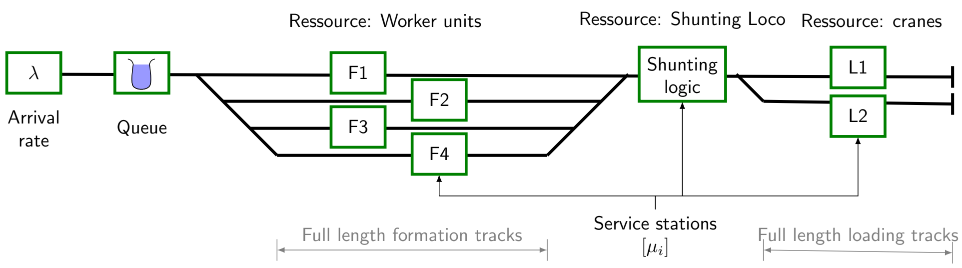


Figure 4.4: Infrastructure layout of the transshipment terminal (symbolic model)

formation tracks (F1 ... F4) and two loading tracks (L1, L2). The loading tracks are served by two gantry cranes for cargo transshipment. The design is inspired by real existing terminals, but does not emulate a specific one. For the purpose of clarity the depicted model is kept small. It is, however, possible to transfer the concept to larger simulations.

The symbolic model was then implemented in a Simul8 simulation model. A screen shot is presented in Figure 4.5. It represents the same terminal but from a

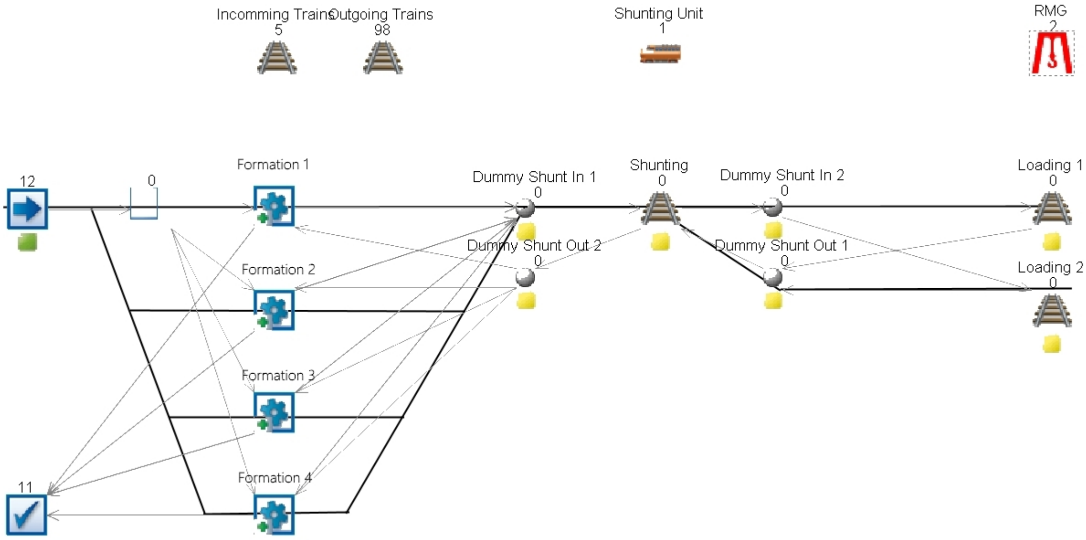


Figure 4.5: A simple intermodal terminal in Simul8 with two loading tracks and a small marshalling yard of four formation tracks. The physical rail network (black lines) is in the background layer for comprehension.

process-related perspective. Each formation and loading track is represented by a work centre (queueing station) in the sense of the queuing system. Furthermore, the resources such as cranes and shunting locomotives are visible in the upper part of the picture.

There is one entry point (top left corner) that describes the train arrival process λ. Wagon sets can be stored on four formation tracks where the freight inspection, brake test etc. are performed. The formation tracks are used concurrently by incoming and outgoing trains. A scalable shunting process takes care of the wagon transport to and from the two loading tracks where the freight handling is executed. Different loading strategies can be applied and also the handling equipment (crane, reach stacker) can be modified. It is assumed that loading units are transshipped to a storage area and not being directly transferred to other transport systems. Trains leave the model via the logical exit station (lower left corner).

A notable feature is that only one queue is used. In fact, a model should have no queues, as this would mean, the wagons or wagon groups could disappear off the tracks and could be parked between imaginary. Consequently, the service stations are directly linked. The queue inserted after the train arrival process is for informational purposes only and may indicate congestion effects.

All processes which are required for the service provision in freight nodes have been implemented in the model with suitable data. This includes transport processes such as the shunting of trains but also logistics processes such as train making or cargo loading/unloading. For the simulation and for further analyses in this thesis, a unified model train with universal properties was defined. The model train is defined as a full-length intermodal block train with the following properties:

- Train length: 30 four-axle wagons (e.g. wagon type Sgns, see Figure 4.6)
- Load factor: 1.6 (regarding Bontekoning (2006, p. 107) the most commonly applied load factor for ISO containers). This equals a maximum number of 30 wagons $\cdot$ 1.6 $=$ 48 cargo units per train set.

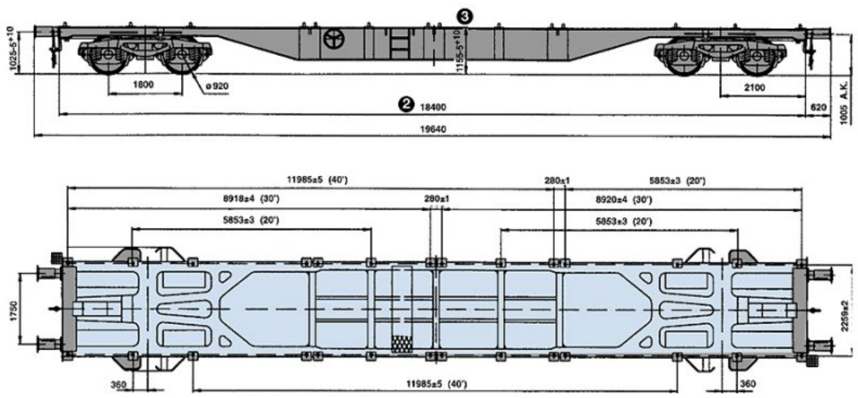


Figure 4.6: Four-axle flat wagon Sgns of SBB Cargo (source: SBB Cargo AG, 2015)

4.2.4 Expected simulation results

The simulation experiments allow the determination of various key performance indicators (KPI) for the terminal performance. A first key figure is the capacity utilisation ρ of the whole system and ρ_i for individual stations. In $M/M/1$ queueing models, the degree of capacity utilisation is calculated as the ratio between the arrival rate and the service rate:

$$\rho = \frac{\lambda}{\mu} \tag{4.3}$$

In Simul8, the corresponding values are generated accordingly after every simulation run. Also of fundamental interest is the average time a train spends in the queuing system:

$$\bar{T} = \frac{\rho}{\lambda(1-\rho)} \tag{4.4}$$

The average dwell time over all trains is calculated by Simul8 automatically. Also worth observing is the queue length LQ. It describes the number of trains in the queue before they enter the terminal. As already mentioned above, a value $LQ > 0$ indicated congestion and must be avoided.

A last key figure considered is the transshipment rate NT of the terminal with:

$$NT = \frac{T \cdot \rho_i}{t_{trans}} \tag{4.5}$$

whereas T denotes the time period (usually one year), ρ_i is the productivity of the considered loading track and t_{trans} is the average time per transshipment. Despite of these key figures, Simul8 produces a set of other details that can be evaluated accordingly.

4.3 Initial experiment design

For the most realistic simulations, data have been collected extensively at several intermodal inland terminals and sea port terminals in Europe. The data contain type and order of typical work processes, their time durations and the resources required (e.g. machines, workers, etc.). Vehicle dynamics of train movements were obtained from a preceding microscopic railway simulation of the marshalling yard. The service rates μ_i for each station i have been determined by empirical studies and are expressed as distribution functions. The train arrival rate λ was also empirically determined. Data preparation for the different stations μ_i is very individual and is presented in the following sections.

4.3.1 The train arrival process

Trains run usually by a known timetable which could be directly implemented in the simulation as input data. To describe the arrival process, the inter-arrival times can be calculated from such a given freight train timetable. If the values corresponded to a poisson-distributed arrival process, it is possible to calculate the equivalent distribution and thus the parameter λ.

As it often comes to train arrival deviations in rail freight transport, the timetables can not be simply transferred to the model assumptions. Train arrival uncertainty is one of the major constraints in productivity planning for container terminals, especially in operational (short-term) planning. In mixed railway networks, where freight trains have a lower priority than passenger trains, deviation from the timetable are prevalent. Late as well as early arrivals up to several hours are possible. Positive timetable deviation (delays) can disrupt the intended allocation of transshipment facilities (e.g. terminal slots) or lead to extended infrastructure occupations. Negative timetable deviations can cause (minor) issues, e.g. early departures from hubs may disrupt the planned track allocations on rail lines which can lead to delays of other trains in other sections of the rail network. Early arrivals may lead to exceeded capacity usages in hubs, especially during peak hours.

Figure 4.7 shows exemplarily a distribution function of timetable deviations of arriving freight train at a freight node with a mean value of $\mu = 57.24$ min and variance of $\sigma = 184.25$ min. Hence, each train arrives too late for about one hour on average.

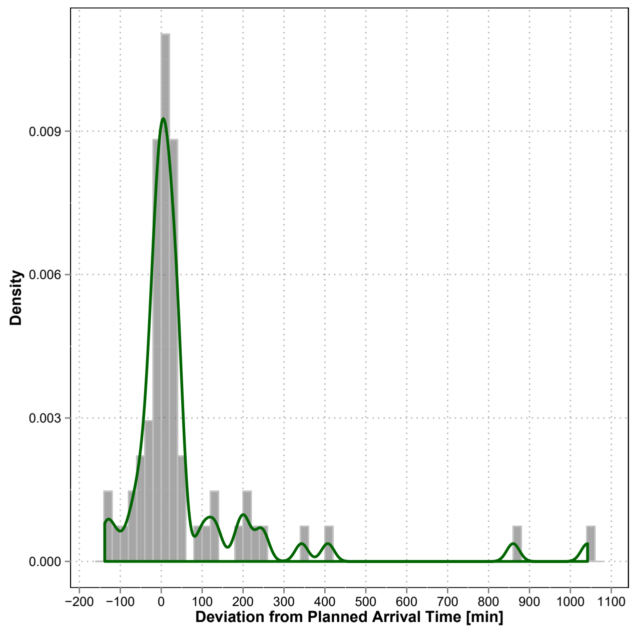


Figure 4.7: Sample of freight train arrival deviations on a network cross-section (histogram and density).

However, with over three hours, the variance is huge and can barely draw conclusions for arrival predictions.

The randomness of the inter-arrival times of trains suggests to avert the timetable as far as possible and describe the train arrivals with the mathematical scheme of probability calculus. Timetable deviations of arriving trains should be regarded as stochastic effects as proposed by Wendler and Naehrig (2004), where the authors describe a method for the statistical analysis of train delay data. Accordingly, it is possible to map the train arrivals in a distribution function.

To generate the distribution function of delays, it is necessary to measure the time difference $\Delta t_{n,i}^{arr}$ between the planned $t_{n,i}^{p,arr}$ and the actual arrival time $t_{n,i}^{arr}$ of a sufficient number of trains n at a certain cross-section i of the route:

$$\Delta t_{n,i}^{arr} = t_{n,i}^{arr} - t_{n,i}^{p,arr} \tag{4.6}$$

These data can ideally be gained from a train describer system[13] or from observations. It can be shown that the train delays follow a negative exponential distribution, initially introduced by Schwanhäußer (1974). The also occurring early arrivals in rail freight transport can be regarded as a positive exponential distribution as shown by Heister (1978). Thus, a case distinction is to be made when formulating the distribution function

$$F(t) = \begin{cases} 1 - p_d \cdot e^{-\lambda_d \cdot t}, & t \geq 0 \\ p_e \cdot e^{\lambda_e \cdot t}, & t < 0 \end{cases} \tag{4.7}$$

where p_d describes the proportion of delayed trains while λ_d describes the average delay of the delayed trains. Accordingly, p_e describes the proportion of early trains with λ_e being the average earliness of the early arrivals (cf. Wendler and Naehrig, 2004). To achieve proper results, it is useful to evaluate data over a certain time period, for example over 24 hours. Accordingly, the distribution function shown in Figure 4.8 has been calculated and could then be used to describe the train arrival process with a certain randomness in the simulation model. The randomness in train arrivals makes the further planning of yard activities more complicated but reflects a realistic behaviour during the simulations. As initially mentioned, the reliable estimation of train arrival times is the key for successful scheduling of intermodal terminals. As it will be shown later in section 5.6, an increase in terminal productivity can be achieved by using methods of rescheduling at the operational level when the estimated time of arrival (ETA) for freight trains is sufficiently predictable.

4.3.2 The processes of incoming trains

Trains arriving in a freight terminal will initially be parked on a formation track. Dwelling there, several tasks, e.g. the incoming check, are being performed (see sec-

[13] A system identifying trains along the journey: part of the train protection system.

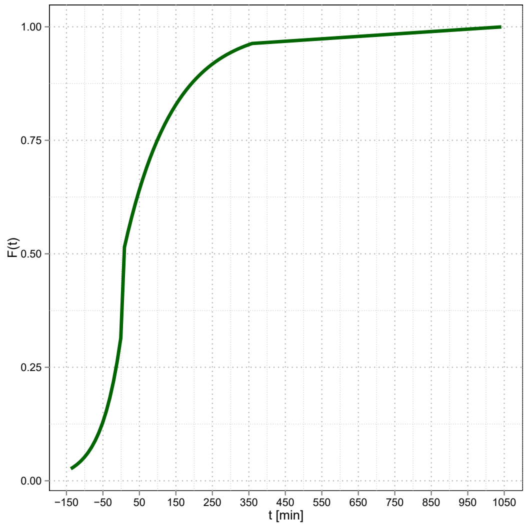


Figure 4.8: Distribution function of arrival time delays based on sample data

tion 2.5.1). The processes are carried out by several actors and partly at the same time. The technical wagon check can for instance be carried out by terminal staff whereas the check of the freight documents is performed by staff of the TOC. An eventual split of the train set into wagon groups might be carried out by a shunting operator. Thus, it is quite complicated to determine the overall dwell time of trains on formation tracks in a general way.

For the assignment of time slots to railway infrastructures, the duration of the individual processes is minor relevant. Up to now, of interest is rather the maximum necessary dwell time which a train set requires before it can be moved to a loading track. It is possible to determine realistic dwell times from databases containing wagon parking information as they are held in existing intermodal terminals. Another possibility is to reproduce dwell times from GPS-based wagon positioning data. From these information it is however not possible to determine the real productive time which is required for the works to be carried out on the wagons. It cannot be determined whether the works have been finished or the wagons are just waiting to be moved further.

In a data analysis at a larger railway node, amongst others the gross dwell time of freight wagons in formation tracks has been analysed. The observations are illustrated in the histogram in Figure 4.9. The observed dwell time averages $\bar{t}_{dwell} = 273$ min whereas the distribution has a strongly positive skew, so it has some outliers to the right. Contrarily to the observations, an approximation of the process dura-

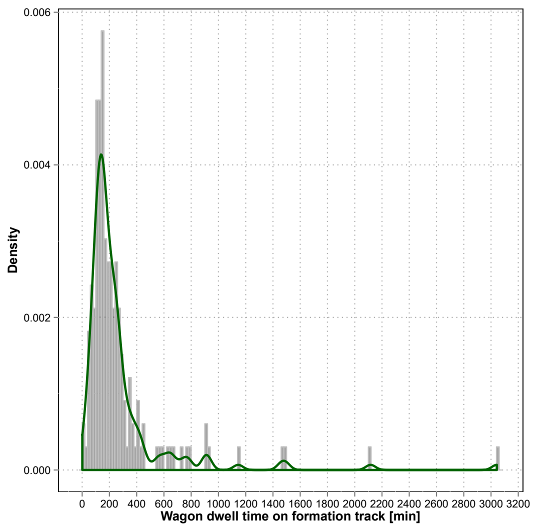


Figure 4.9: Histogram and density function of dwell times in formation tracks for incoming trains

tions of incoming trains showed that the maximum required dwell time in formation tracks ranges usually between $60 \leq t_{dwell} \leq 180$ min for a full length train set. The big range results in the fact that the actual dwell time depends on the on-site circumstances and the process cycle in a specific rail yard. In any case it appears that the observed wagons dwell in the formation tracks much longer than necessary. It is to assume that a large portion of the dwell time is non-productive waiting time. It is further to assume that the formation tracks are widely used to buffer trains which arrived too early or too late and wait for further processing.

In order to obtain a realistic mapping of the formation track dwell time for the simulation model, a distribution function was created from the histogram of Figure 4.9. The extreme outliers ($t_{dwell} > 1,000$ min) where truncated because they do not represent regular train operations. A curve fitting describes the data best as a Pearson type V distribution

$$f_{min,\alpha,\beta}(t_{dwell}) = \frac{\beta^{\alpha}}{\Gamma(\alpha)(t_{dwell} - min)^{\alpha+1}} \cdot exp\left(- \frac{\beta}{(t_{dwell} - min)} \right) \tag{4.8}$$

with the parameters

$$\begin{aligned} min &= -38.1 && \text{minimum } t_{dwell}, \\ \alpha &= 3.98 && \text{shape parameter} > 0, \\ \beta &= 796 && \text{scale parameter} > 0. \end{aligned}$$

The fitted function (Figure 4.10) can directly be incorporated in the simulation model.

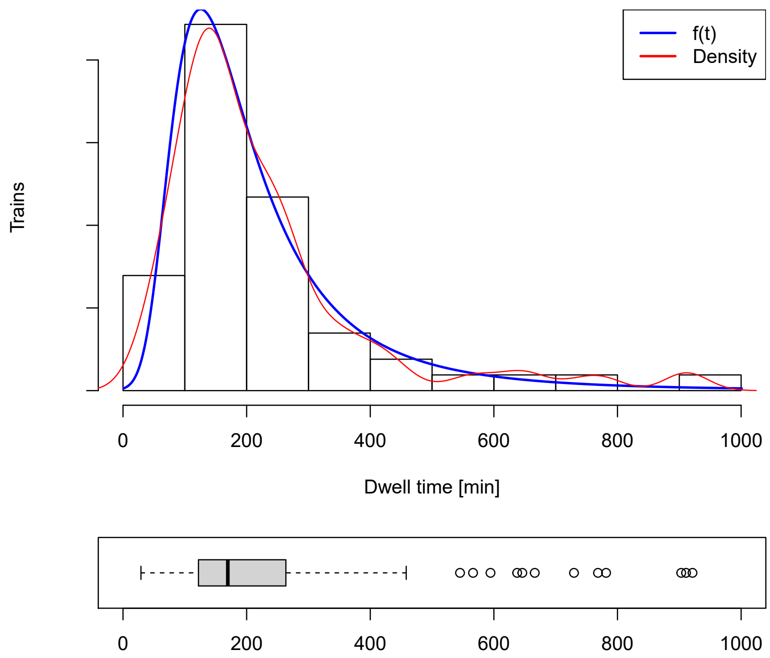


Figure 4.10: Fitted probability density curve for incoming trains on formation tracks based on historic dwell time data. The data could be fitted to a Pearson type V distribution (blue).

4.3.3 The shunting process

Shunting is needed to move wagons from a formation track to a loading track. Therefore, one shunting locomotive is available in the simulation. The process is kept simple: Travel times between all tracks have been calculated using the microscopic railway simulation OpenTrack. These travel times are directly taken over to the simulation as fixed values. The approach was considered as sufficient since shunting speed is low ($v_{max} = 25$ km/h) and travel times do not vary substantially.

The locomotive operational plans derive indirectly from the track occupations and the production plan of the terminal. At first sight it is prepared on an operational level (as usual in existing production schemes) but, as it will be seen later in this chapter, the schedule follows a projectable structure that has a tactical character under slot management conditions. For optimality however, attention is to be paid, that empty runs are avoided.

The process does not consider driver's working times or breaks. As it will be shown later, this is not necessary in this model since the utilisation of the shunting locomotive is rather low and thus enough time for breaks or driver changes is assured. Additional restrictions may be placed. These are, for the fundamental analysis of slot management however, not relevant.

4.3.4 The process of transshipment

The turnover rate of loading units depends on various infrastructural and operational characteristics. Most influencing are the type and number of handling equipment (e.g. reach stacker, RMG, etc.), the number of loading tracks, the order of trains being served, and the loading strategy in the terminal.

The simulation model uses two gantry cranes spanning two full-length loading tracks. Thus, the infrastructural conditions are determined. Hereupon, a couple of operational settings are to be defined. First, the order in which trains are processed must be determined. Larger transshipment terminals operate trains usually in so-called bundles or batches which suggests already a slot system (see section 3.4.2). This ensures that the transshipment processes in the terminal (timely provision of loading units, reduction of search efforts in the storage area, etc.) are executed optimally. The simulation model uses initially such a fixed slot plan. With the optimisation in chapter 5, further priority rules will be considered.

When the terminal is highly utilised, the loading slots are scarce, as only a certain number per day can be assigned. The shorter the trains dwell under the crane, the more slots can be provided per day. The reduction of the service time is directly related to an increase of handling capacity. The operational challenge is to determine the loading schemes of the trains and thus the therefore necessary lateral crane movements. The loading plans of the trains, as noted by Baron (2011), affect the handling capacity of the gantry cranes: The containers can be loaded either optimal to the terminal (reducing crane movements) or optimal to the train operator (reducing shunting and sorting efforts on the train runs). In existing terminals, the train operators usually provide loading lists from which it is apparent which container is to be placed on which wagon and position. From it, the terminal operator can calculate the optimum loading order strategy.

The gantry cranes themselves are described by certain parameters. Baron (2011) describes parameters of RMG cranes universally as shown in table 4.2. In the simulation model, the value *crane moves per hour* is the most relevant. Accordingly, the transshipment time t_{trans} of a loading unit has been implemented as a continuous uniform distribution

	Minimum value	Maximum value
Boom length	12 m	-
Lift height	7 m	-
Transshipment capacity (moves per hour)	*16*	*20*
Lift speed (m/min)	16	100
Crab speed (m/min)	100	150
Crane lateral speed (m/min)	160	240
Load pick-up	40 s	50 s

Table 4.2: RMG crane paramaters (cf. Baron, 2011)

$$f(t_{trans}) = \frac{1}{t_{trans}^{max} - t_{trans}^{min}} \tag{4.9}$$

with $t_{trans}^{min} = \frac{16}{60 \text{ min}}$ and $t_{trans}^{max} = \frac{20}{60 \text{ min}}$. Hence the transshipment time per loading unit ranges in the interval 3 min $\leq t_{trans} \leq$ 3.75 min. The corresponding uniform distribution function is depicted in Figure 4.11. The average transshipment time is expressed as $\bar{t}_{trans} = \frac{3+3.75}{2} = 3.375$ min per cargo unit.

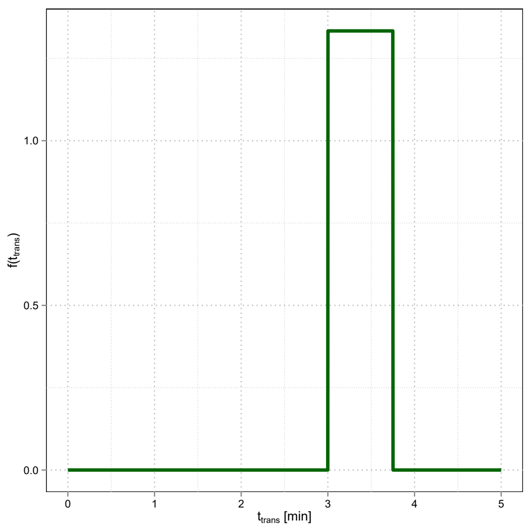


Figure 4.11: Uniform distribution for describing the duration of gantry crane moves

The distribution function describes the time needed for transshipping a single loading unit. To emulate the transshipment process of a wagon group or a whole train, some more parameters need to be considered. For this purpose, the previously defined model train (section 4.2.3) is used. Furthermore, the following parameters are used to characterise the whole transshipment process and thus the dwell time of a train in a loading track:

- Trains arrive completely loaded.
- All cargo units will be unloaded and the trains will be completely loaded again.
- All cargo goes through the storage area, no direct transshipment from/to other vehicles.

4.3.5 The process of outgoing trains

Wagon groups, being unloaded and loaded, return to a formation track where they will be prepared for departure. The processes taking place were described in section 2.5.2. For the simulation model it is important to note that here the same infrastructure as for incoming trains is used. Thus, the two processes compete on resources. A special control in the simulation prioritises outgoing trains so that they can exit the yard and no deadlocks occur.

The process sequence of outgoing trains is modelled similar to the sequence of incoming trains (section 4.3.2): the dwell times can be mapped from observations or GPS data. As for the incoming trains, also here it becomes apparent that the dwell times are significantly longer than necessary. As pointed out in Figure 4.12, the observed dwell times have a strongly positive skew and several extreme outliers to the right. This indicates some intense exceeding of the usual dwell time which can not be explained by operational necessity.

These extreme values ($t_{dwell} > 1{,}000$ min) where truncated again here. A curve fitting describes the dwell time data best as a Weibull distribution:

$$f_{min,\alpha,\beta}(t_{dwell}) = \frac{\alpha}{\beta}\left(\frac{t_{dwell} - min}{\beta}\right)^{\alpha-1} exp\left(-\left(\frac{t_{dwell} - min}{\beta}\right)^{\alpha}\right) \tag{4.10}$$

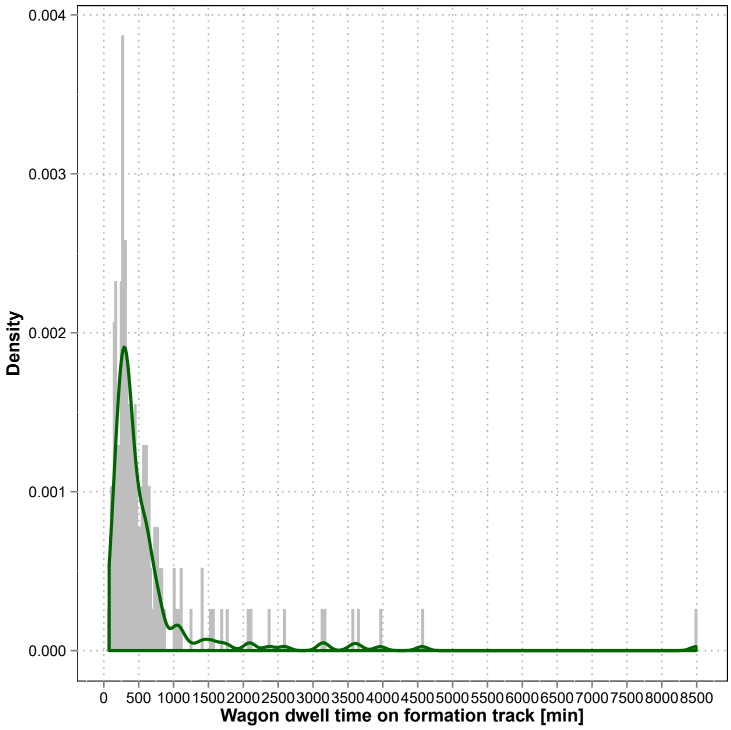


Figure 4.12: Histogram and density distribution of dwell times in formation tracks for outgoing trains

with the parameters:

$$min = 76 \qquad \text{minimum } t_{dwell},$$
$$\alpha = 1.71 \qquad \text{shape parameter} > 0,$$
$$\beta = 347 \qquad \text{scale parameter} > 0.$$

The fitted distribution function (Figure 4.13) was directly incorporated in the simulation model. It will be of further use later in section 5.1, too.

4.4 Evaluation of terminal performance

The previous sections described the generation of important input data and the definition of relevant variables for the simulation model. In this section, a set of experiments is developed. Their purpose is to obtain general insight into the process flows of intermodal terminals and to identify the influences of the variable changes on the yard performance. The experiments were carried out in a controlled and structured

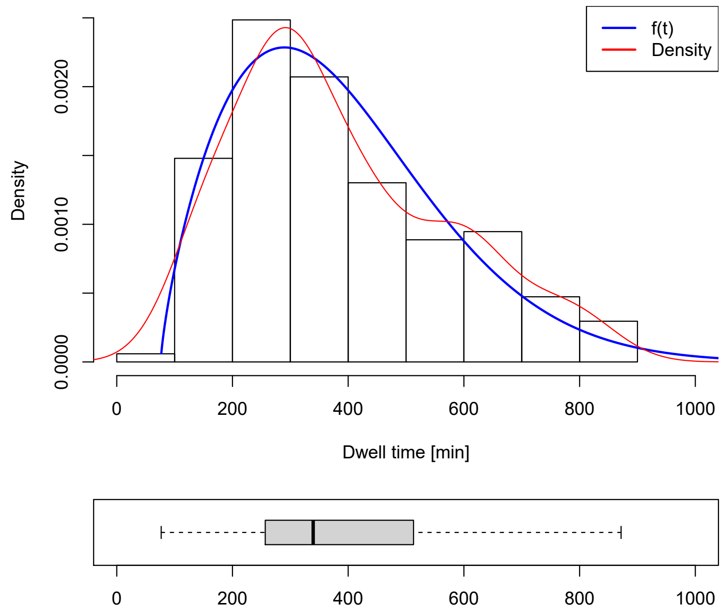


Figure 4.13: Fitted probability density curve of for outgoing trains on formation tracks based on historic dwell time data. The data could be fitted to a Weibull distribution (blue).

manner in order to set up the most constructive framework for slot management strategies. Controlled experiments are those that study the effects of changing a single variable on certain performance indicators. Structured experiments denote that the studies of a single variable were performed in a specific order. From the data analysis in sections 4.3.1 to 4.3.5 it becomes apparent that the following variables have considerable influences on the performance of a rail freight node:

- Arrival time of trains and inter-arrival time
- Arrival time delay of trains
- Dwell time of trains on the different tracks
- Train lengths and number of loading units

The load factor and the load order are further variables but have been identified of having only a minor influence on the slot management design.

4.4.1 Set of experiments

The variables that have the highest influence on the slot management performance are the dwell times on the different stations. Also, the train arrival time and train delays influence the performance significantly. Based on this fact, three scenarios were carried in this order:

Scenario 1: Deterministic process times: Track occupation times and the arrival processes are described as standardised deterministic values. This introductory scenario produces several basic key figures for the understanding of a freight terminal production system.

Scenario 2: Stepwise increase of stochastic dwell times: In two intermediate steps, the distribution functions for the train dwell times elaborated in sections 4.3.2 to 4.3.5 are introduced at the corresponding service stations.

Scenario 3: Stochastic arrival time: The findings on the arrival process elaborated in section 4.3.1 are introduced. Similarly, the data are introduced in two simulation steps.

The overview on all simulation runs is depicted in Figure 4.14. Apart from the de-

Simulation Run		Arrival Process	Formation Tracks	Loading Tracks
Scenario 1		deterministic	deterministic	deterministic
Scenario 2	Simulation 2a	deterministic	stochastic	deterministic
	Simulation 2b	deterministic	stochastic	stochastic
Scenario 3	Simulation 3a	stochastic	deterministic	deterministic
	Simulation 3b	stochastic	stochastic	stochastic

Figure 4.14: Overview on the basic set-up of the simulation runs carried out

scribed differences, all simulations have the following identical properties: All simulation runs have a start time $t_0 = 0$ min and run maximally until $t_{end} = 7,200$ min (5 days). A set of $N = 12$ trains is scheduled to be processed in the yard within the given simulation time. Of course, the arrival times vary depending on the simulation scenario, but the first train arrives always with $t_1^{arr} = 0$ min. The chosen set-up allows to regard a long-running simulation without that the warming-up period[14] has a too strong effect on the yard performance.

All simulated trains match with the model train, defined in section 4.2.3, having a length of 30 wagons and are fully loaded. All cargo units will be unloaded and the trains will be completely loaded again. The chosen setting describes thus the worst case generating the longest possible transshipment time. Any changes would reduce the transshipment time and thus shorten the train dwell time.

4.4.2 Simulation scenario 1: using deterministic times

A first simulation was run to determine the node's overall maximum capacity and the effects of overload. It is characterised by deterministic process durations, hence with constant service rates μ_i on all stations i. A deterministic system contains no randomness and thus always produces the same result from a given starting condition. The distribution functions for the process durations, which were produced in section 4.3, however, are not used directly in this first case. Instead, the deterministic process durations are composed of the arithmetic mean values of these functions. The Gantt chart in Figure 4.15 shows how a single train passes through the different locations (stations) of the model with these characteristics. The corresponding mean

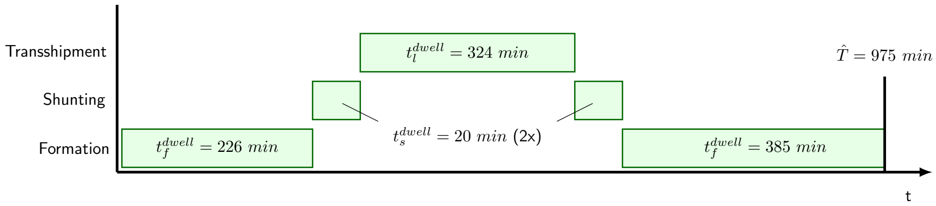


Figure 4.15: Machine-oriented Gantt chart of a single train journey through the model with deterministic process times

values of the single process durations can be taken from the figure. As a measure of quality, the average time in the system $\bar{T}$ was monitored. If regarded isolated, i.e., no interferences or allocation conflicts occur, the overall dwell time of a train is $\bar{T} = \hat{T}_n = 975$ min, or 16.25 hours, respectively. By altering variable values, other interesting key performance indicators can be obtained:

Theoretic capacity and marginal efficiency: The inter-arrival rate was initially set to $\lambda = \frac{1}{180\ min}$ As the only variable, λ was subsequently stepwise increased during

[14]warming-up period: phase until the simulation system comes to full operation

several simulation runs in the interval $\frac{1}{180\ min} \leq \lambda \leq \frac{1}{60\ min}$. Figure 4.16 illustrates the results. When the arrival rate λ is low, the average time of a train in the system $\bar{T}$ is

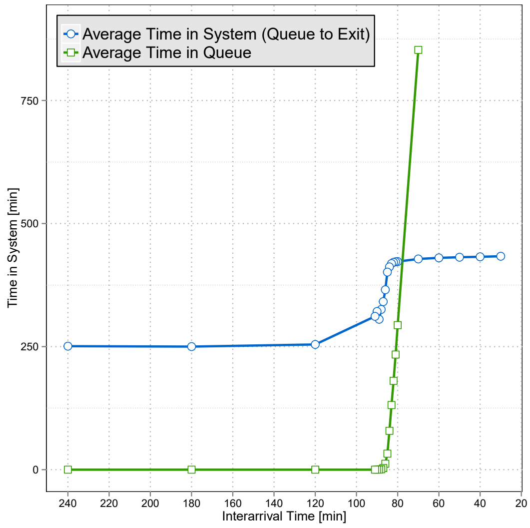


Figure 4.16: Determination of the terminal's performance limit: Congestion appears at the terminal entrance when the arrival rate is high (green line). Overload results in an overall slower processing and has a negative effect on the train handling time.

stable on a low level. With an increasing λ, more trains arrive at the hub per time period. Now, $\bar{T}$ increases at a certain point which is caused by hub-internal congestion effects. Due to infrastructure allocation conflicts, trains have to wait before they can proceed to the next station in the queueing system. As long as the arrival rate is lower than the outward flow, the congestion effects can still be caught and do not cause congestion phenomena outside the terminal. When further increasing λ, so that the arrival rate is higher than the outward flow, trains pile up in front of the hub (ca. near $\lambda = \frac{1}{90\ min}$). The queue length is now $LQ > 0$. If this is the case, arriving trains would not be able to enter the hub. This can have outreaching effects on the railway network and other railway nodes.

When the node operates above its maximum capacity, $\bar{T}$ rises to a higher level. Since all process times are deterministic, the new level (ca. near $\lambda = \frac{1}{80\ min}$) is constant, too. Consequently, internal occupation conflicts are a first sign for overloads. They result in slower overall processing of trains and lower the productivity. The throughput time of trains or wagons (and cargo) through the hub increases but does not necessarily lead to a complete standstill or deadlock. Hence, the capacity limit is in practice difficult to recognise but can be elaborated in such a simulation model.

Detection of bottlenecks: While the previous paragraph investigated the hub as a whole, it is convenient to examine the performance of single stations in detail to discover bottlenecks. Figure 4.17 illustrates for this purpose the utilisation of a loading track subject to the arrival rate λ. The productive usage increases with λ (green line).

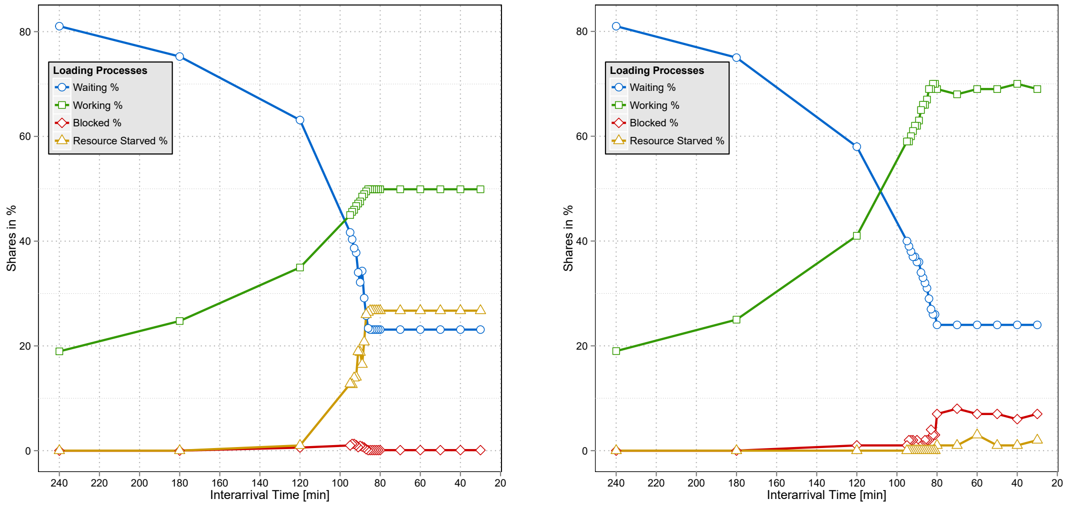


Figure 4.17: Utilisation analysis of a transshipment (loading) track. Left: Resource starvations occur because transshipment facilities are not ready. Productivity is low. Right: Increasing transshipment capacity allows higher productivity but causes another bottleneck elsewhere.

Logically, the productivity with larger traffic volumes initially increases. Simultaneously, the idle time decreases (blue line). In the left graph, the utilisation rate remains at about 45% but resource starvations arise at a certain point (yellow line). In these cases it was identified that no gantry crane was available while trains were waiting to be served which caused congestion. By employing an additional crane, it was possible to raise the utilisation rate of the loading track and reducing thus the resource starvations (right graph of Figure 4.17). After that, however, blockades (red line) occur which indicate that trains are ready to leave the station but can not do so due to a bottleneck on a subsequent infrastructure element. It must be solved by analysing the other relevant hub elements.

Initial job schedule: An initial job schedule with twelve trains calling at the hub was created. The arrival times are chosen in such a way that the utilisation of the terminal is maximised and no occupancy conflicts or waiting times occur. For that purpose, an alternating arrival rate of $\lambda \approx \frac{1}{20\text{ min}}$ and $\lambda \approx \frac{1}{385\text{ min}}$ interacted perfectly with the service times of the stations in the queueing model so that the maximum perfor-

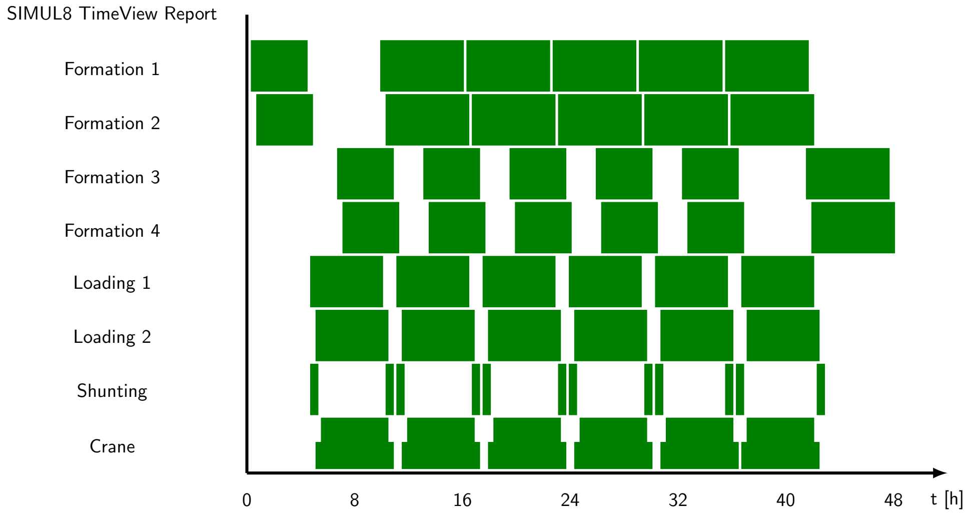


Figure 4.18: Occupation plan for all tracks and resources in the first simulation scenario. Train arrival $\lambda \approx 0,32$ per hour and average time for trains in the system $\bar{T} = 16.31$ hours.

mance could be estimated. In this idealistic process flow scenario (see Gantt chart in Figure 4.18), the idle times of all stations could be minimised. Since the processing times are still deterministic, the process flow generates a periodically recurring pattern. It shows that the best utilisation of a terminal can be achieved when train arrivals are uniformly distributed and adapted to the operating speed of the terminal. The result illustrates the best (theoretic) job schedule for the simulation model. The overall handling of the 12 trains takes about 50 hours, whereas the average time of a train in the system is $\bar{T} = 978,85$ min (16.31 hours) and the average arrival rate is $\lambda \approx 0,32$ trains per hour which equals to an inter-arrival time of approximately 185 min between two trains.

4.4.3 Simulation scenario 2: stepwise increase of stochastic dwell times

The initial work plan of scenario 1 is not very realistic because it does not consider variations that can occur in real freight hubs. It was, however, useful to get an overview on the process flows and the interaction between the different queueing stations of the simulation model. Now, stochastic impacts are introduced to determine the impact of disturbances on the hub productivity. Therefore, the distribution functions which were computed in section 4.3 are stepwise introduced to replace the uniform, deterministic process durations from the first scenario. Two simulations

were run: **Scenario 2a** uses stochastic process durations in the formation tracks. **Scenario 2b** considers additionally stochastic process times in the loading tracks.
The arrival rate in both scenarios 2a and 2b varied in three simulation runs:

(i) $\lambda = \frac{1}{360 \text{ min}}$: equates to 4 trains in 24 hours

(ii) $\lambda = \frac{1}{240 \text{ min}}$: equates to 6 trains in 24 hours

(iii) $\lambda = \frac{1}{185 \text{ min}}$: equates to ca 8 trains in 24 hours (same arrival rate as in scenario 1)

The simulation runs (i) and (ii) have been defined because the handling of 4 or 6 trains correspondents to a typical output of an intermodal terminal of this size. The simulation run (iii) allows the comparison of the results with scenario 1.

Results of scenario 2a: After introducing stochastic dwell times in the formation yard occur already assignment conflicts at a small arrival rate of $\lambda = \frac{1}{360 \text{ min}}$. Logically, the conflicts increase with the arrival rate. The following table outlines the resulting conflict minutes[15] in the three simulations. The four formation as well as the two loading track values were grouped:

	Conflict minutes			Average time
Simulation	**Formation**	**Loading**	$\sum$	**in system [min]**
2a (i)	90.53	23.42	113.95	966.76
2a (ii)	209.05	30.59	239.64	947.82
2a (iii)	1,852.06	637.09	2,489.15	1,141.61

Although the number of simulated train arrivals remains constant at 12, it was observed that the average processing time of trains in the system increases. The resulting conflicts rise with the arrival rate. As the following table shows, the conflicts have relative small percentages on the total processing time in the simulations (i) and (ii) but increase significantly in (iii). For clarity, parallel tracks of the categories Loading and Formation were grouped here as well. The table shows the average proportions of blockades in the respective groups:

[15]Conflict minutes: The cumulated number of minutes where conflicts have occurred.

	Averaged blockades		Total process
Simulation	**Formation**	**Loading**	**time [min]**
2a (i)	0.46%	0.24%	4,916.57
2a (ii)	1.47%	0.43%	3,561.10
2a (iii)	14.55%	10.01%	3,182.38

They vary between 0.24% and 1.47%. Such conflicts are resolvable by operational dispatching and have no far-reaching impacts, for instance, on the rail network or other freight nodes. In simulation (iii), however, the allocation conflicts have proportions of 10.01% and 14.55% on the total process time. Such conflicts are no longer operationally manageable and can have already an impact on the rail network. The table informs also about the total process time, i.e. the overall time required for all 12 trains to pass through the modelled freight node. The total process time decreases with an increasing λ because the idle times in the course of the simulation are being reduced.

A machine-oriented Gantt chart in Figure 4.19 informs about the assignment of trains to tracks and the occurring allocation conflicts from simulation (iii). Appar-

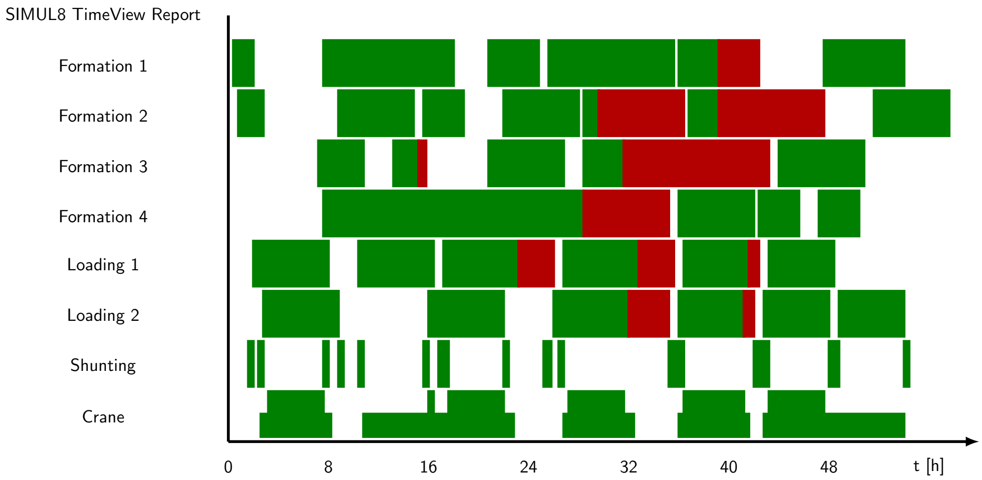


Figure 4.19: Occupation plan for all tracks and resources in simulation (iii) of scenario 2a. Train arrival $\lambda \approx 0,32$ per hour and average time for trains in the system $\bar{T} = 19.03$ hours.

ently, numerous track allocation conflicts (red blocks) occur. Due to the stochastic effects, the average train dwell time raised from $\bar{T} = 978,85$ min (16.31 hours, scenario 1) to $\bar{T} = 1,141.61$ min (19.03 hours) while the arrival rate was the same. The results of this simulation show already a fairly realistic behaviour. Hence, it gets clearer, what challenges have to be considered in the development of the slot management strategies.

Results of scenario 2b: In addition to the settings in scenario 2a, now the loading tracks have variable process durations using the distribution function from chapter 4.3.4. Since this is a uniform function with a relative small interval, no significant changes were determined. Small assignment conflicts occur already in simulations (i) and (ii) and increase significantly in simulation (iii). The calculated conflict minutes are presented in the following table[16]:

	Conflict minutes			Average time
Simulation	**Formation**	**Loading**	$\sum$	**in system [min]**
2b (i)	86.26	2.28	88.55	956.50
2b (ii)	241.81	23.85	265.65	918.06
2b (iii)	1,807.86	695.60	2,503.46	1,132.59

As in simulation 2a, the average time of trains in the system increases with the congestions in the same way. It is slightly lower compared to simulation 2a. The relative proportions of conflicts on the total process time is comparable, too:

	Averaged blockades		Total process
Simulation	**Formation**	**Loading**	**time [min]**
2b (i)	0.44%	0.02%	4,889.96
2b (ii)	1.77%	0.35%	3,408.19
2b (iii)	14.16%	10.89%	3,192.67

The blockades in simulations (i) and (ii) are again few and thus solvable within the operational yard dispatching. The conflicts of simulation (iii), however, are considerably higher and hardly solvable in a day-to-day business. It can be noted that there are slightly but not significantly higher amounts of conflicts in the loading tracks compared to simulation 2a (10.89% vs. 10.01%). A graphical illustration of the assignment of trains to tracks and the occurring allocation conflicts from simulation (iii) is provided in Figure 4.20. The average time in the system lowered slightly from $\bar{T} = 1,141.61$ min (19.03 hours, scenario 2a, Figure 4.19), to $\bar{T} = 1,132.59$ min (18.9 hours).

[16] Again, the parallel tracks were grouped for a better clarity.

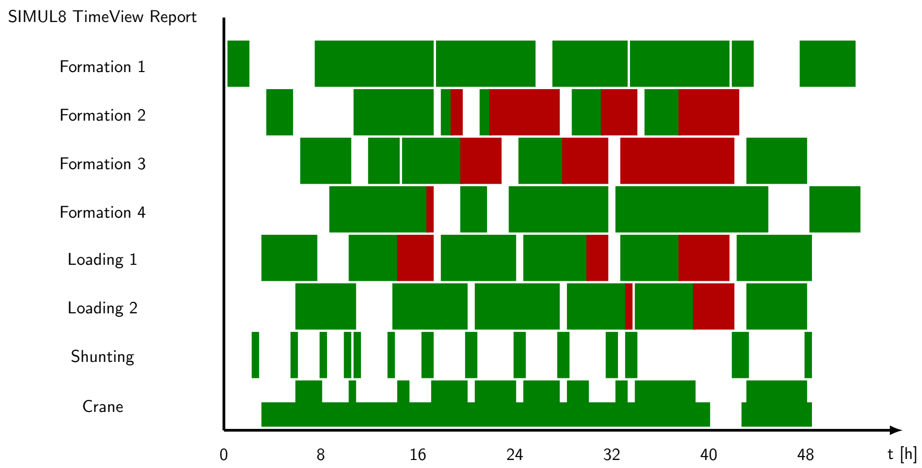


Figure 4.20: Occupation plan for all tracks and resources in simulation (iii) of scenario 2b. Train arrival $\lambda \approx 0,32$ per hour and average time for trains in the system $\bar{T} = 18.9$ hours.

In summary, the consideration of stochastic loading times does not have a significant influence on the yard performance. This can be explained by the relative stable uniform distribution function (Figure 4.11) used for the variable loading times. Under the stochastic conditions of both scenarios, 2a and 2b, it is in conclusion difficult to map the track allocations into a schedule structure that satisfies a regular slot management plan.

4.4.4 Simulation scenario 3: stochastic arrival times

In the previous scenarios, the train arrivals were either modelled using a well defined schedule which was harmonised to the yard operation speed (scenario 1) or a constant arrival rate (scenario 2). Now it is to consider that the train arrivals depend in reality on different bounds such as the time of the day, the day of the week or the origin of the trains.

Accordingly, the arrival process was redefined following the beta distribution calculated in section 4.3.1. The original distribution function was scaled so that the mean time between arrivals corresponds to $\frac{1}{\lambda} = E(A) = 185$ min which matches the performance of the former scenarios for comparability. Accordingly, inter-arrival times for 12 trains were calculated and incorporated in the simulation model. Two simulation runs were conducted: (3a) using deterministic processing times in the formation and loading tracks (as in scenario 1), and (3b) using variable processing times (as in scenario 2b).

At first sight, it becomes clear from the following table that the average dwell time per train in the system is in both simulations higher than it was in scenario 2:

	Conflict minutes			Average time
Simulation	**Formation**	**Loading**	$\sum$	**in system [min]**
3a	1,462.90	461.20	1,924.10	1,191.52
3b	3,072.78	1,304.93	4,377.72	1,595.70

The stochastic arrival times have thus a strong negative influence on the throughput time of individual trains. Furthermore, the total process time is significantly larger (3,728 min in simulation 3b vs. 3,192.67 min in simulation 2b (iii)) and the average blockages have increased, in both absolute and relative:

	Averaged Blockades		Total process
Simulation	**Formation**	**Loading**	**time [min]**
3a	10.81%	6.82%	3,382.30
3b	20.61%	17.50%	3,728.00

This leads to the fact that the stochastic train arrivals disarrange the terminal-internal processes in such a way that the productivity decreases far more than in the preceding scenarios. The deriving work plan does not lead to an acceptable slot schedule. A structured arrangement of time slots is hardly observable as it can be seen in Figure 4.21. Trains dwell significantly longer (average time in system $\bar{T} = 26.6$ hours) in the terminal whereas the time-share where they are actively involved in logistics processes, declined.

4.4.5 Conclusions on the productivity assessment

The previous sections showed how various stochastic effects can influence the performance of the railway infrastructure of a freight node. Scenario 1 illustrated the most idealistic variant which produces smooth time slot structures in all terminal tracks and thus created a track utilisation pattern that slot management strives for. The stepwise increase of stochastic effects resulted in a decrease of the terminal performance as it becomes particularly evident when regarding the average dwell time of the trains in the yard (Figure 4.22). The average time per train in the yard increases in scenarios 2a and 2b compared to scenario 1 by about 15%. The simultaneous occurrence of all stochastic effects in scenario 3b reflects the most realistic situation of

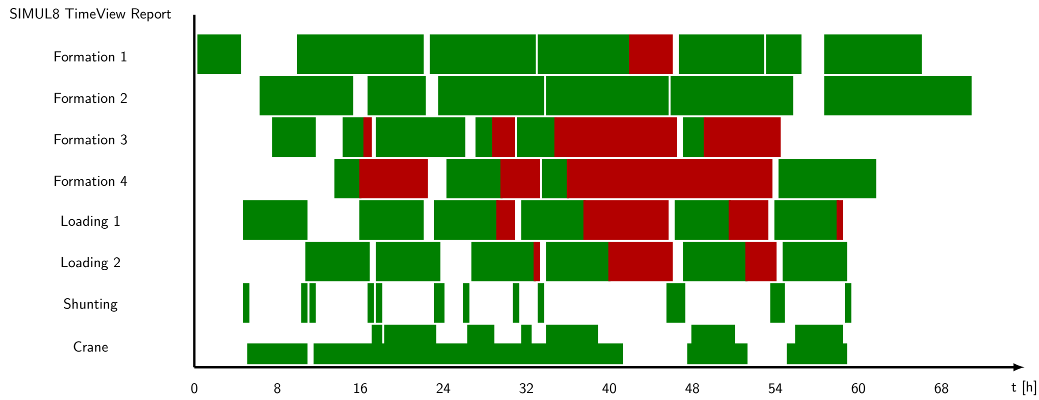


Figure 4.21: Occupation plan for all tracks and resources in scenario 3b. Train arrival $\lambda \approx 0,32$ per hour and average time for trains in the system $\bar{T} = 26.6$ hours.

the process sequences in intermodal terminals. It ultimately prolonged the average time in system per train by about 63% compared to scenario 1:

$$\delta\bar{T} = \frac{\bar{T}_{\text{Scenario 3b}}}{\bar{T}_{\text{Scenario 1}}} = \frac{1,595.7 \text{ min}}{978.8 \text{ min}} = 1.63 \tag{4.11}$$

Allocation conflicts of infrastructures in the depicted occupation plans (e.g. Figure 4.21) denote that trains dwell long time periods in non-productive processes.

Investigating the utilisation of the different queueing stations across all simulation runs, makes clear that the loading tracks have a higher workload compared to the formation tracks, but also a higher share of idle times (Figure 4.23). The formation tracks, however suffer more allocation conflicts and thus are identified as the bottleneck of the system under the given process times. Comparing these findings with air transport, one can find analogies: The runway system at airports has a similar function as the formation area at railway yards – it serves as the entry and starting point for vehicles. On large highly frequented airports, the runway system is often identified as the critical bottleneck element (e.g. de Neufville and Odoni, 2003, p. 367).

Particular attention is paid to the loading tracks as they contribute significantly to the creation of value at intermodal terminals. Compared to the others, scenario 1 achieves the highest utilisation rate regarding the loading tracks. As illustrated in Figure 4.23 (b), the utilisation rate of the loading tracks decreases from 64.53% in scenario 1 to 52.76% in scenario 3b. With it, the utilisation of the cargo handling equipment (cranes) decreases accordingly.

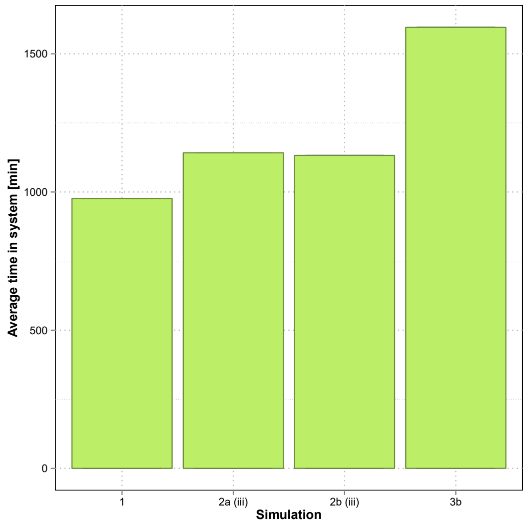


Figure 4.22: Comparison of the average dwell time per train in the terminal model: Stochastic effects prolong the dwell time in the conducted simulation scenarios by about 63% (scenario 1 compared to scenario 3b).

The overall terminal performance is expressed by the annual transshipment rate NT using equation 4.5. Assuming for simplicity a 24/7 operation over one year (= 525,600 min) and an average transshipment speed of 3.375 min per unit, the productivity for a loading track is

$$NT = \frac{\text{525,600 min per year} \cdot 64.53\%}{\text{3.375 min}} = 100,498 \text{ LU / year} \tag{4.12}$$

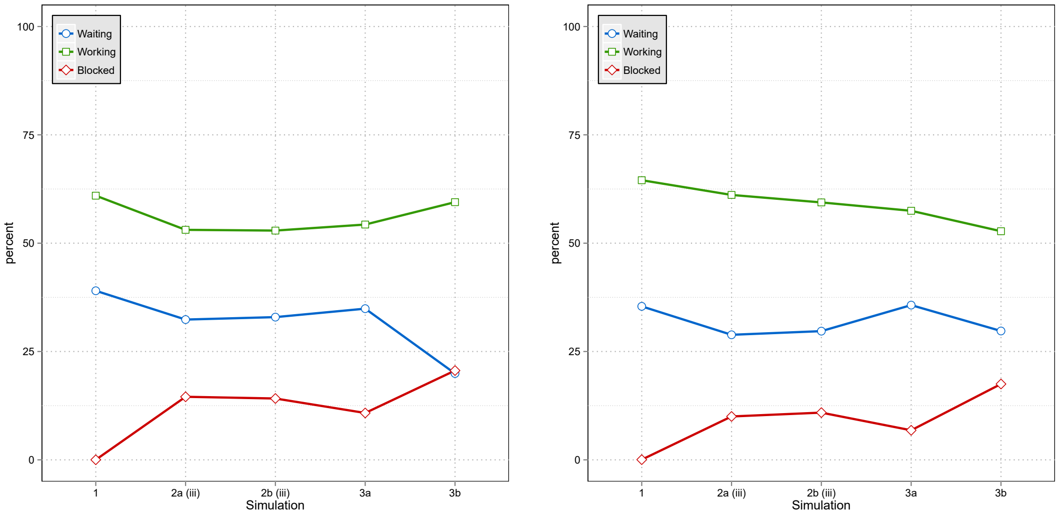
in scenario 1 and

$$NT = \frac{\text{525,600 min per year} \cdot 52.76\%}{\text{3.375 min}} = 82,165 \text{ LU / year} \tag{4.13}$$

in scenario 3b. The realistic transshipment capacity amounts accordingly about

$$\frac{82,165 \text{ LU}}{100,498 \text{ LU}} = 81.76\% \tag{4.14}$$

of the theoretical maximum capacity. It is easy to realise that a certain potential exists to improve the yard performance through higher-level planning of the processes.



(a) Formation Tracks **(b)** Loading Tracks

Figure 4.23: Average utilisation of (a) formation and (b) loading tracks in the simulation scenarios 1 to 3.

4.5 Chapter summary

The chapter assessed the productivity of intermodal freight nodes. The process flows were illustrated in a combined queueing and simulation model. On the basis of an exemplary terminal design, the basic processes of train movements and cargo transshipment have been investigated and explained.

A key aspect of the chapter was the determination of input data for the simulation model and further analyses. Based on observation data, distribution functions for the process durations in the different yard facilities were computed. These will serve also in the further course of this thesis as a basis for the development of the slot management system.

A number of scenarios was developed and applied to the simulation model to determine the performance of the individual facilities of such freight hubs and to identify bottlenecks. It could be illustrated, how the interactions between the various processes might affect the terminal performance when the process coordination is insufficient. In particular, it became clear how the performance of the entire terminals decreases in the event of overload. Further, the utilisation of single stations of the freight node was investigated to obtain information about bottlenecks in the transshipment yard.

Stochastic influences were stepwise applied in a set of various simulation scenarios. The non-correlated time behaviour between the numerous actors (infrastructure mangers, train operators, freight forwarders, terminal operators) causes stochastic, random process durations across the freight node which have a significant impact on the slot management design and the productivity of the node. The increase of stochastic influences has a negative effect on the yard performance: more stochastic disturbances in the production processes increase the overall dwell time of trains significantly. With it, the stochastic effects also lower the productivity of the terminal as a whole since less cargo units can be handled per time unit.

With the occurring stochastic process durations, it is difficult to recognise a structured work pattern in the occupancy of the terminal tracks. Process durations are hardly predictable. It is thus comprehensible that freight terminals and railway freight yards in general operate largely with improvised scheduling decisions on an operative planning level as quoted in section 1.1.

The challenge in the next chapter is to determine, how a slot management approach can be developed under these conditions. From this chapter can be taken that the assessment of the process reliability plays an essential role in the determination of time slots. It is required to stabilise the scheduling processes and to increase the system stability against externally caused influences.

5 Railway yard slot scheduling

In chapter 4 it has been demonstrated how trains pass through the freight node and how they utilise the railway facilities. The focus was on the simulation of process sequences, the dependencies between processes and on the determination of the yard productivity. In that first part of the two-level planning hierarchy, no schedules were considered. Process operations started spontaneous just when trains arrived.

This chapter aims at integrating the railway-specific processes with the logistics processes by establishing time windows for the stopovers of the trains in the different yard facilities. Thus, the spontaneous operations methods shall become obsolete. The time window strategy structures the work flows and provides a certain robustness to the terminal procedures against disruptions, such as train delays. It will denote the second part of the two-level planning hierarchy.

The time window or slot management strategy was explained in chapter 3. Applied to rail freight yards, time periods in which trains or train sets dwell on specific tracks, are to be determined. These time periods, must be long enough for performing all the necessary tasks but at the same time as short as possible in order to keep the yard productivity at a high level. This optimisation problem is solved using methods of Operations Research (OR). The chapter starts with the determination and the robustness of single timeslots in the different yard sections. Subsequently, single timeslots are merged into slot plans using appropriate scheduling methods. Finally, the slot plans are tested for stability. Rescheduling algorithms will be introduced to ensure the optimal flexibility of the freight yard due to late and early arriving trains.

In order to demonstrate the findings well, a certain traffic is required. The time slot analysis is therefore carried out on a larger yard model than the one used in chapter 4. The larger infrastructure model allows more parallel traffic and has thus more capabilities to demonstrate the slot management performance. It possesses the following characteristics:

- Formation yard with 16 tracks (instead of 4)
- Freight terminal with 6 loading tracks and four RMGs (instead of 2 tracks and 2 cranes)

- The characteristics of the model train remain as described in section 4.2.3 (30 four-axle wagons, load factor 1.6, etc.)

5.1 Time slot development

After arrival at the destination freight node, a train is only ready for a following journey after it has been unloaded and, if necessary, loaded again. The critical path of the sequentially activities determines the minimum turnaround time. In simulation scenario 1 (section 4.4.2, page 82), the process path of a train through the terminal was already shown by using a Gantt chart in Figure 4.15. This representation, in which time windows of the infrastructure assignment already can be seen, corresponds to an envisaged slot system.

The first step towards a yard-wide slot plan is the determination of the required individual process durations and their merging to a process sequence so that a train set can proceed through all stations with minimal conflicts. When developing terminal schedules, it needs to be ensured that the processing times are assumed reliable. Reliability according to Törnquist (2004, p. 5) is defined as:

> ... the ability to perform a service according to an implicit or explicit agreement, and in this case the service is to deliver on time or within a feasible time interval. In this context, reliability is a measure of likelihood of succeeding to reach the specified goal, and can be seen as equivalent to a frequency of having a varying degree of punctuality.

For determining the reliability of potential freight node schedules, the already known service time μ_i of a machine i is decomposed into individual components. Unlike in chapter 4, where the dwell time which a train set spends in a station i was rather taken unspecified as a whole, it is now analysed in detail. Taking into account the wide variance of the service time distributions identified in chapter 4, the pure process times must be supplemented by a time buffer to compensate eventual deviations. The dwell time thus consists of several active and passive components, which need to be identified and separated. Active components are required for executing essential processes. In passive components, e.g. time buffers, no economic value is being added. The introduction of time buffers is always at the expense of the maximum possible productivity. The longer a time slot is, the higher is its reliability but on the expenses of the station productivity. Accordingly, a compromise must be found between robustness and productivity.

It is assumed that the necessary tasks in a station take a minimum time to be executed. This denotes the minimum required process time $t_{n,i}^{min}$ of the respective train n on track i. The introduction of delay variables adds aspects of robustness and quality. The planned time buffer $t_{n,i}^{b}$ and the unanticipated process delay $t_{n,i}^{d}$ together form the delay time relative to a train's ideal process time. Figure 5.1 depicts a decomposition of process delay effects. The planned process time is the period that describes the

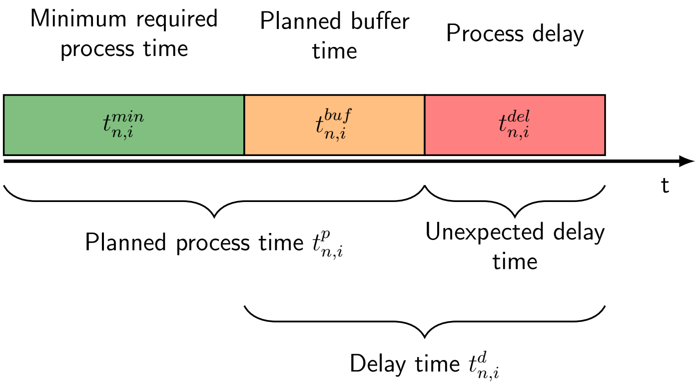


Figure 5.1: Decomposition of process delay effects (on the basis of Ryerson et al., 2014, p. 288)

time slot and is thus the period in which a track is reserved for a train. It is composed of the minimum required dwell time and the planned buffer time:

$$t_{n,i}^{p} = t_{n,i}^{min} + t_{n,i}^{buf} \tag{5.1}$$

The delay time, respectively, comprises the planned buffer time and the unexpected delay:

$$t_{n,i}^{d} = t_{n,i}^{buf} + t_{n,i}^{del} \tag{5.2}$$

5.1.1 Determination of the minimum required process times

The determination of the minimum required process times $t_{n,i}^{min}$, $\forall i, n$ is a challenge in assessing the service quality of freight hubs. It is not possible to determine them from the previously used sources such as dwell time distributions developed in chapter 4.3 because observation data are too vague. These observation data could serve at most for determining the planned process time $t_{n,i}^{p}$. Rather, the minimum required process times can be determined analytically by investigating the single tasks of the dwelling processes in the different stations. In the following, this is carried out for all different facilities (formation tracks, loading tracks, etc.) of the simulation model case. Hence, the task sequences are not universally valid and must be analysed in each real case individually.

Minimum required process times of incoming trains: The formation tracks receive the incoming trains. Several specific tasks are performed there (see also section 2.5.1). Assumptions about the durations of the tasks to be performed were evaluated in numerous observational studies (cf. e.g. Potthoff, 1977, p. 53 ff.). The tasks of train making have not changed significantly until today (cf. e.g. Homeyer, 2000) and can thus be expected to be of equal duration[17]. An overview on the composition of the most relevant tasks on arriving trains, distinguished between technical, operational and transport-related functions, is shown in Figure 5.2. As it can be seen, some tasks

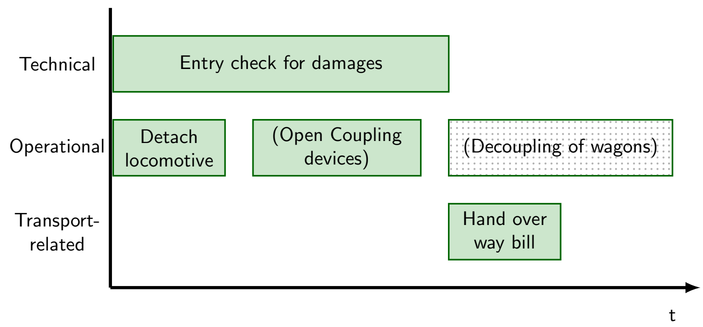


Figure 5.2: Composition and course of tasks performed with trains arriving at freight nodes

can be carried out simultaneously.

The most important and longest task is the receiving inspection and check for possible damages. Following Potthoff (1977, p. 55), a working speed of $v = 0.22$ m/s of the wagon technicians walking along the train is assumed (with two men working as a technician team). Applied to the model train, consisting of 30 four-axle wagons, the technicians require

$$30 \text{ wagons} \cdot \frac{19.74 \text{ m per wagon}}{0.22 \text{ m/s}} = 44.9 \text{ min}$$

for the check of a train. Other sources report significantly shorter time requirements for this task. To be on the safe side, the value following Potthoff is used.

[17]Recent improvements for rail freight yards concern technologies for yard automation and the introduction of automated, electronic data exchange. The developments apply mostly for the train handling of single wagonload services. The train handling tasks in intermodal transport did not change significantly thereby.

The detachment of the locomotive can be carried out in parallel but requires usually only 60 to 80 seconds. Hence, it is not time critical. Also parallel to the receiving inspection is the eventual splitting of the train set into groups of wagons[18]. If the split into wagon groups is desired, the coupling devices must be opened and the air pipes be separated between the relevant wagons. The separation of wagons is carried out either using a marshalling hump or a shunting locomotive, depending on the facilities in place and the principles of the shunting operator.

The check and handover of freight documents (way bill) is the final task. It takes usually only several minutes and is widely supported by electronic data exchange. Thus, generously 10 minutes are assumed here. Summing up, the critical path of the minimum required process times for the incoming trains can be determined with:

Technical	44.9 min
Operational	0 min
Transport-related	10 min
$\sum$	54.9 min

Minimum required process times of outgoing trains: The formation tracks hold the outgoing wagon sets where they are being prepared for the departure (see also section 2.5.2). Also here, a distinction is made between technical, operational and transport-related tasks. A general overview of the necessary tasks is shown in Figure 5.3. Again,

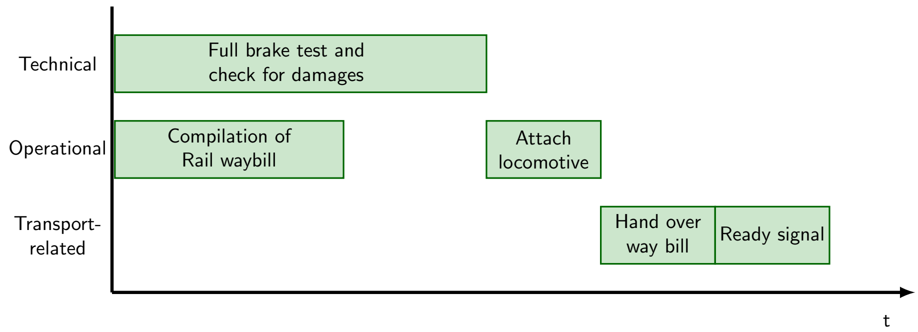


Figure 5.3: Composition and course of tasks performed with trains departing from freight nodes

some tasks can be carried out simultaneously.

The technical tasks contain the check of wagons and freight on damages as well as the filling of the air pipes. For the filling of the brake system Oberländer (1959)

[18] In intermodal transport, the split of trains in groups is usually avoided. At freight nodes serving several loading sites (e.g. seaports), a split in two or three wagon groups may occur.

assumes that this happens for freight trains with 1.5 s/m train length. After that, a wagon technician walks with $v = 50$ m/min along the train and checks the proper function of the brakes. A second walking cycle for examining the proper release of brakes is assessed with the same time consumption (Potthoff, 1977, p. 129). Again, assuming that a train set consists of 30 four-axle wagons (e.g. wagon type Sgns), the technical tasks require:

Air pipe filling	$30 \text{ Wagons} \cdot 19.74 \text{ m} \cdot 1.5 \text{ s/m} =$	14.8 min
Full brake test	$((30 \text{ Wagons} \cdot 19.74 \text{ m})/50 \text{ m/s}) \cdot 2 \text{ ways} =$	23.7 min
	$\sum$	38.5 min

If the full brake test is not required, the task does not apply and is shortened accordingly[19]. Subsequently, two operational tasks are considered: The *compilation of the freight documents* (rail waybill) is carried out mostly automatised using IT systems today. It is not a time-critical task. The second task is the *attachment of the locomotive* together with the simplified brake test. Following Potthoff (1977, p. 129), this takes no longer than 10 minutes. Also the two transport-related tasks, the *handing over of the freight documents* to the train driver and the *notification of readiness for departure* are considered generously with 10 minutes, each. In summary, the minimum required process times of outgoing trains can be determined with:

Technical	38.5 min
Operational	10 min
Transport-related	20 min
$\sum$	68.5 min

The actual durations of the tasks for incoming and outgoing trains are very individual and depend generally on the number of wagons, the working conditions in the yard in terms of weather, the technical equipment and the human-related differences in the operation and thus the speed of operation. The calculated figures provide, however, a fairly realistic indication.

Minimum required process time at loading sites: The loading and unloading processes are not as complex as the ones in the formation tracks. For determining the

[19] At many European railways, only a simplified brake test is required if the train set has not been decoupled since its last trip.

minimum required process time, it is again assumed the worst-case scenario as described in section 4.3.4: complete unloading and loading a train of 30 wagons with all other conditions as stated there. How long the handling of one train lasts at least, depends on the number of trains simultaneously handled and the number of operating RMGs. Without information on a specific loading strategy, general assumptions can be made. Assuming a terminal as described in this chapter's introduction with six parallel loading tracks which are served by four RMGs, the average minimum required service time per train is:

$$\begin{aligned} t_{n,l}^{min,load} &= \text{number of LU} \cdot \text{transshipment rate per LU} \cdot \text{crane availability } (\forall\, l \in L) \\ &= 48 \text{ LU} \cdot \frac{60 \text{ min}}{18 \text{ LU}} \cdot \frac{4 \text{ cranes}}{6 \text{ trains}} \\ &= 106.67 \text{ min} \end{aligned} \tag{5.3}$$

The minimum required process time at loading sites will have a minor impact on the slot management design since slots for the loading tracks will be already provided by the terminal operator. They can be considered as predefined and will be taken into account as the planned slot times $t_{n,l}^{p,load}$ in the next section.

Minimum required shunting times: The shunting machine slots are considered as train rides between two holding tracks. The minimum process times are accordingly represented by the respective journey times between two tracks i and j and the times for the two coupling operations:

$$t_{i,j}^{min,shunt} = d_{i,j} + c_i + c_j \tag{5.4}$$

The journey times $d_{i,j}$ between two stations i and j can be computed using a microscopic railway simulation tool and can be directly incorporated into the analysis. In the considered infrastructure model, the shunting trips are fairly short and shunting locomotives have a stress-free schedule. The coupling processes c_i and c_j are considered as constants since the coupling with a shunting locomotive is carried out using a shunting coupling device that allows easy and quick coupling and decoupling operations of only several seconds. When the shunting locomotive is running empty, the corresponding coupling times are 0 s. Since no particular disturbances are expected during shunting, the minimum required shunting times correspond to the planned shunting times and thus holds:

$$t_{i,j}^{p,shunt} = t_{i,j}^{min,shunt}. \tag{5.5}$$

5.1.2 Determination of the planned process times

The sum of the minimum processing time and a buffer supplement determines the planned process time. The determination of the buffer supplement connotes a criterion of quality, as the track dwell time must not exceed the planned process time value. Otherwise delays occur with negative influences on other processes. For the analysis, the process times in the formation and loading tracks are of special importance.

Planned process times in the formation tracks: A planned dwell time of $t_{n,f}^{p,in} = t_{n,f}^{p,out} = 300$ min ($\forall\ f \in F$) in the formation tracks is considered as a reasonable empirical value, sufficiently for the preparation of intermodal freight trains (Kreft, 2015). The planned dwell time is thus defined as a fixed time period equal at all formation tracks and independent of incoming or outgoing trains. A dwell time $t_{n,f}^{p} > 300$ min represents thus a delayed train n with a delay time $t_{n,f}^{del} > 0$. Based on the determined minimum process times from section 5.1.1, the planned buffer time $t_{n,f}^{buf}$ can be calculated. For the incoming trains it results in:

$$\begin{aligned} t_{n,f}^{buf,in} &= t_{n,f}^{p,in} - t_{n,f}^{min,in} \\ &= 300 \text{ min} - 54.9 \text{ min} \\ &= 245.1 \text{ min} \end{aligned} \tag{5.6}$$

Equivalent for the outgoing trains applies:

$$\begin{aligned} t_{n,f}^{buf,out} &= t_{n,f}^{p,out} - t_{n,f}^{min,out} \\ &= 300 \text{ min} - 68.5 \text{ min} \\ &= 231.5 \text{ min} \end{aligned} \tag{5.7}$$

The different time components can now be assembled into a slot concept for a formation track. The composition of the calculated time shares is illustrated in Figure 5.4. At first sight it becomes evident that the minimum required processing time is quite

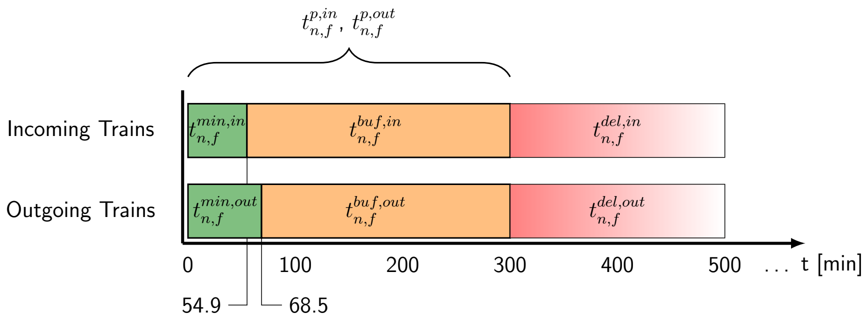


Figure 5.4: Decomposition of the dwell times structure in formation tracks showing the minimum dwell time $t^{min}_{n,f}$, the time buffer $t^{buf}_{n,f}$ and a possible delay $t^{del}_{n,f}$.

small in relation to the planned process time. Contrariwise it is known from the empirical distributions from section 4.3 that dwell times < 300 min are hard to realise under the prevailing operational circumstances. The planned process times are considerably below the empirically identified dwell time values. It is thus to determine whether the planned buffer times are robust enough and do not lead very likely to infrastructure occupation conflicts. This question will be investigated in section 5.1.4 in detail.

Planned process times in the loading tracks: Time slots in large intermodal terminals are often predefined by the operators. They are mostly between 4 and 6 hours long for a full length container train (Gille and Schönemann, 2009). For simplicity, the loading track dwell time in the model case is assumed to be $t^{p,trans}_{n,l} = 300$ min $(\forall\, l \in L)$ and thus adopts a well-recognised average. Apparently, a model train used in this analysis can be fully unloaded and loaded in the given space of time as illustrated in Figure 5.5.

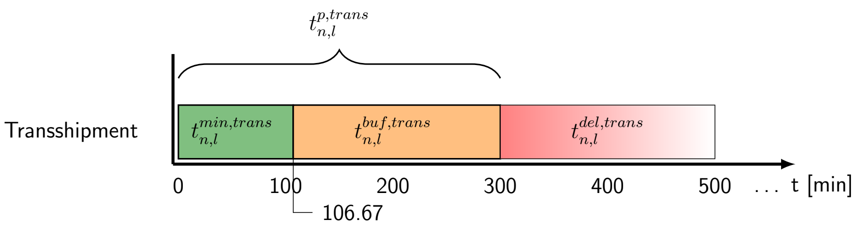


Figure 5.5: Decomposition of the dwell times structure in loading tracks showing the minimum dwell time $t^{min,trans}_{n,l}$, the time buffer $t^{buf,trans}_{n,l}$ and a possible delay $t^{del,trans}_{n,l}$.

Moreover, also the loading processes contain a remarkable time buffer. This can be explained due to the fact that the terminal handles up to six trains in parallel and does not focus work load on only one train at a time. In summary, at each processing station (formation and loading track) in which a train set is dwelling, 300 minutes are available for carrying out all necessary handling tasks.

5.1.3 Slot design

A general, initial slot design, valid for all processing stations, has been defined by determining the minimum and the planned process times. The time slot and thus the allowed dwell time for a train n in station i is defined by the planned process time $t^p_{n,i}$. The real, actual processing of the train is much shorter. It is described as the period $t^{min}_{n,i}$ which must be completely placed within $t^p_{n,i}$ but is flexibly movable therein. The time buffer $t^{buf}_{n,i}$ encloses $t^{min}_{n,i}$. Figure 5.6 illustrates the internals of an isolated time slot. The location of $t^{min}_{n,i}$ within the slot depends on the temporal availability of the

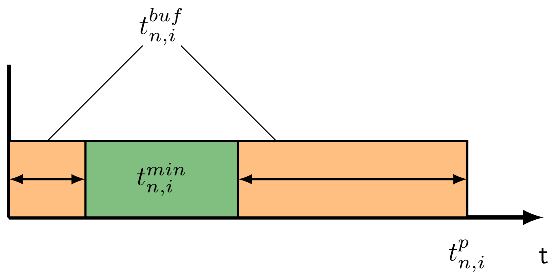


Figure 5.6: A single time slot: The minimum processing time $t^{min}_{n,i}$ lies within the time space of the planned process time $t^p_{n,i}$ and can be flexibly moved according to the scheduling requirements.

necessary resources (e.g. wagon technician, cranes). The resources must be scheduled accordingly. A train's journey through a freight hub consists of the concatenation of a couple of such time slots. By joining a train's single slots in the different yard facilities, its slot plan or chronological slot sequence will be generated as illustrated in Figure 5.7. Subsequently, the slot sequences of multiple trains are coordinated. This requires a certain co-ordination effort which will be discussed in section 5.2. Before that, the following subsection will briefly investigate the question of robustness of the formulated time slots.

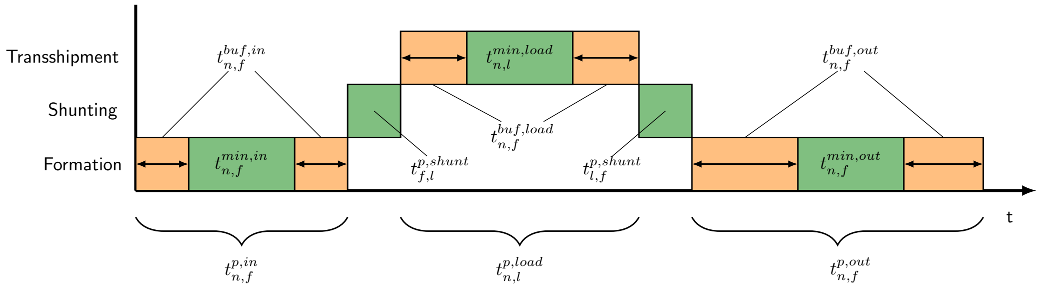


Figure 5.7: Exemplary slot plan for one train illustrating the minimum and the planned process times.

5.1.4 Robustness of time slots

A crucial goal of the slot management approach is to introduce a certain quality in the yard production processes and to specify appropriate target process times. A first step was the determination of the planned process time in section 5.1.2. Compared to the observed dwell times of the real-life data, the planned slot times are significantly lower and thus more efficient. To achieve the desired production quality, the consideration of the target slot times needs to be enforced also towards the terminal customers (railway undertakings). To what extent and how the planned slot times are to prevail over the decision making of the operating railway undertakings is discussed later in section 6.2.

At the present stage of the thesis, the planned slot times can be compared to the empirically determined dwell times from the real-life observations in order to estimate their robustness when applied to real freight yards. To estimate the robustness of a time slot itself, another stochastic simulation could now take place. Instead, the schedule reliability can also be measured using heuristic methods as proposed by Carey (1999). The author mentions that "the advantage of using heuristic measures is [...] that they can be computed more easily than detailed simulations, and they require less data, which may not be available." The approach was considered as useful for a range of transport scheduling problems, particularly those where knock-on delays have a large impact on the reliability, i.e. that delays propagate easily (cf. Törnquist, 2004, p. 75). This applies thoroughly to the scheduling of freight nodes. Following Carey's proposal, the robustness probabilities of the dwell times may now be calculated. For this purpose, the probability density function (pdf) for the dwell times in the formation tracks from chapters 4.3.2 and 4.3.5 are taken up again.

Let $f_{n,i}(t)$ be the pdf of the dwell time of train n at station i. The reliability $r_{n,i}(t)$ is calculated by integrating the pdf over a given interval:

$$r_{n,i}(t) = P(\text{train n does not suffer or produce an own-delay})$$
$$r_{n,i}(t) = P(t_{n,i}^{dwell} \leq t_{n,i}^{p}) = \int_{t=0}^{t_{n,i}^{p}} f_{n,i}(t)dt \quad (5.8)$$

The upper bound must be set to the planned dwell time $t_{n,i}^{p}$. The example in Figure 5.8 illustrates the situation. The whole area under the curve has the value 1. The

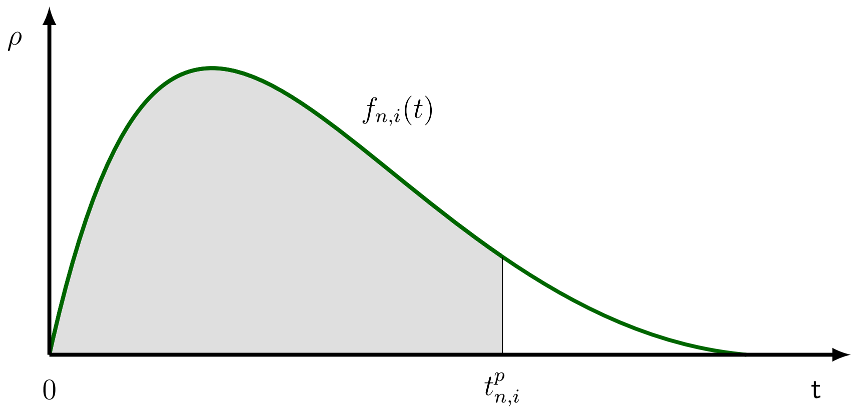


Figure 5.8: Integration over the probability distribution function to determine the reliability of the planned dwell time $t_{n,i}^{p}$.

grey area is a fraction of 1 and determines the space of a robust dwell time. A train n with a dwell time $t_{n,i}^{dwell} < t_{n,i}^{p}$ is thus within the reliability range. To calculate the reliability of dwell times for incoming trains at formation tracks, the pdf from section 4.3.2 is applied and the upper bound was set to the planned dwell time $t_{n,i}^{p} =$ 300 min accordingly. The reliability $r_{n,f}^{in}$ is thus:

$$\begin{aligned} r_{n,f}^{in}(t) &= P(t_{n,f}^{dwell,in} \leq 300 \text{ min}) \\ &= \int_{t=0}^{t_{n,i}^{p,in}=300} f_{min,\alpha,\beta}(t_{dwell})dt \\ &= 81.82\% \end{aligned} \quad (5.9)$$

For the outgoing trains at formation tracks respectively, using the pdf calculated in section 4.3.5, the reliability $r^{out}_{n,f}$ is:

$$\begin{aligned} r^{out}_{n,f}(t) &= P(t^{dwell,out}_{n,f} \leq 300 \text{ min}) \\ &= 40.35\% \end{aligned} \tag{5.10}$$

It can be seen that a conflict-free dwell time is not given safely in every case. A probability value of 81.82% denotes that about 18% of the incoming trains will exceed their planned dwell time which is likely to cause further conflicts on the subsequent processes. Much worse is the extremely low value of 40.35% for the outgoing trains. Almost 60% would exceed the planned formation track dwell time. The development of slot plans thus faces two challenges:

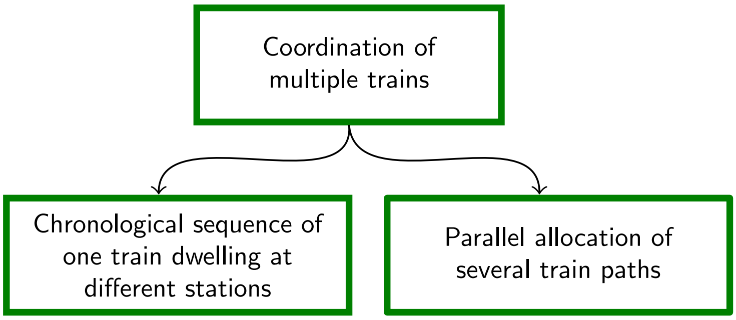
(a) The determined planned process times (section 5.1.2) already contain a considerable time buffer when compared with the minimum process times and thus lower the yard productivity.

(b) The determined planned process times are being extensively exceeded when applying the real-life observations from section 4.3. During the subsequent optimisation process it has to be checked whether the planned buffer times are sufficiently calculated.

Reasons for the dwell time exceedings are manifold but did not become evident from the observation data from chapter 4 since it is not possible to determine the real productive process times thereof. They might be found in the complex interaction of actors and interest groups during the train handling in the node. For outgoing trains it might also be thinkable that the locomotive is not available in time or that the planned departure is at a later time and trains thus wait to leave the node. In the context of a time slot optimisation it is aimed in the next section, to organise the process sequences of a multitude of freight trains in such a way that the time slot exceedings are minimised or even avoided.

5.2 Optimising rail freight yard slot plans

The next step after the planning of process durations is now, to determine how the numerous slot plans of a multitude of trains can be coordinated so that an optimal slot plan is created. The coordination process takes place in two dimensions:



For the purpose of solving this multidimensional optimisation problem, a scheduling model is developed. This section begins with a short introduction to Operations Research (OR) and scheduling before a rail yard scheduling model is presented. The model is thereupon applied to realistic scheduling situation in order to analyse the stability and feasibility of slot plans.

5.2.1 Principles of scheduling

The basic concept of scheduling is the allocation of resources to tasks over given time periods (cf. Pinedo, 2008, p. 1). Speaking in a modelling language, scheduling denotes the processing of N jobs on I machines in accordance with a specific order restriction. The processing time of job n on machine i is $p_{n,i}(n = 1, 2, \ldots, N;\ i = 1, 2, \ldots, I)$. In the scope of this thesis, freight trains and wagons are regarded as jobs and tracks are considered as machines. With scheduling, it is aimed to find a suitable slot plan (schedule or process sequence) in which these N jobs can be processed on these I machines so that a pre-defined optimisation objective (e.g. the minimisation of number of late jobs) is met.

Scheduling problems can be described by three characteristics: machine characteristics α, job characteristics β and the optimisation objective γ. The properties are usually depicted using the three-field notation developed by Graham et al. (1979):

$$\alpha/\beta/\gamma$$

The **machine characteristics** α describe the kind, order and number of machines of a production system. In the simplest case, with $i = 1$, a scheduling model is described as the single machine scheduling problem. From a macroscopic point of view, a railway node can be considered as a single machine problem. Its internal mechanisms, which shall be examined in this thesis, are, however, not observable in that case. A better insight in the production system is possible by investigating its vertical range.

If exist several machines at the same level, the optimisation problem is defined as a parallel machine problem. Three kinds of possible types of parallel machines exist: identical, uniform and heterogeneous. Identical machines ($\alpha := IP$) have the same production time and are applicable at the same time. Uniform machines ($\alpha := UP$) have machine-dependent production times in contrast and heterogeneous machines ($\alpha := HP$) have machine-dependent and job-dependent production times.

In contrast to the parallel arrangement, machines can also be ordered sequentially. This is referred to as the machine sequence in literature. Basically, each stage possesses only one machine. If the machine sequence for all jobs are equal, the optimisation problem is then called a *flow shop problem* ($\alpha := F$). If the machine sequences differ for the jobs, it is called a *job shop problem* ($\alpha := J$). If parallel machines exist in the respective stages, a general description of the problem is difficult and should be done in a qualitative way.

Different **job characteristics** or constraints must be observed when solving scheduling problems. These are represented by the identifier β of the three-field notation, where β may contain none, one, or more parameters. Possible values for β are manifold. Jaehn and Pesch (2014) describe the most relevant values as follows:

p_n — Processing time of job n. For $p_n = p$, all jobs have an equal processing time.

r_n — Release time (arrival time) of train n. If provided, jobs are not available at the beginning but arrive during the scheduling run.

d_n — Deadline (departure time) of train n. Defines the time when the job must be finished. For $d_n = d$ all jobs have the same due time.

pmtn — An interruption (preemption) of the processing is allowed. The processing of a job on a machine may be interrupted at any time and may be continued (possibly on a different machine) at a later time. If not stated, no preemtion is allowed (not suitable for modelling rail freight nodes in this context).

prec — Defines the sequence, priorities or precedence relationships between jobs. If stated, a job can be started only after its predecessor is finished. In this case order relations between jobs exist.

s_{nk} — Sequence-dependent set-up times between job n and job k apply. The machine must remain unoccupied for the specified time between the two jobs.

In addition, various other restrictions, such as permutation and wait restrictions, are possible. In this regard, reference is made to the relevant literature (e.g. Jaehn and Pesch, 2014; Pinedo, 2008).

The identifier γ describes the **optimisation objective** of the model whereas the objective value is to minimise. Most objective functions minimise the deadline d_n or the completion time C_n of job n. Again referring to Jaehn and Pesch (2014), the following optimisation objectives can be significant:

C_{max} The maximum completion time (makespan) over all jobs shall be minimised. The total time of a schedule corresponds to the completion time of the job that will be completed as the last one: $C_{max} := max\{C_1, \dots, C_N\}$.

L_{max} The lateness of job n is defined as $L_n := C_n - d_n$. The lateness of the job with the maximum delay $L_{max} := max\{L_1, \dots, L_N\}$ is to minimise.

$\sum C_n$ The sum of the completion times is particularly suited as the target criterion to minimise the inventory costs. The minimisation of $\sum C_n$ is equivalent to minimising the average throughput time $\sum C_n / N$.

$\sum w_n C_n$ As a generalisation, also the sum of the weighted completion times can be minimised. Thus, e.g. different cost rates or times can be considered.

The values described herein for the triple $\alpha/\beta/\gamma$ are far from complete. The numerous possible combinations results in a huge number of possible scheduling problems. The terminology used for describing scheduling problems is as in the two following examples:

$1||\sum w_n C_n$ A one-machine problem with no special job characteristics minimising the the weighted completion times.

$P_i|r_n|C_{max}$ A parallel machines problem with release times for the jobs minimising the maximum completion time.

Also relevant for the slot allocation are the so-called **priority rules**. Competing jobs which are available for a machine at the same moment can be interpreted as a system of queues. A priority rule now permits, according to the assigned numerical values, a selection from the conflict set. Accordingly, they determine the criteria by which the jobs are scheduled on the next available machine. Important priority rules are:

MST Minimum slack time: jobs are ordered by their slack (buffer) time $slack_n = d_n - p_n$. The job with the lowest buffer time starts first.

FCFS First come first served: jobs are processed in the order in which they arrive.

EDD Earliest due date: the job that has the nearest due date is served first.

LPT Longest processing time: jobs are scheduled on a machine with the descending order of their processing time.

SPT Shortest processing time: jobs are scheduled on a machine with the ascending order of their processing time.

WSPT Weighted shortest processing time: jobs are arranged on a machine in an ascending order considering their individual priority.

Priority rules can sometimes provide results with large deviations from the optimal solution. In practice, however, they prove to be – regardless of the difficulty of a scheduling problem – simply applicable. For difficult problems, they serve as a heuristic solution method for the determination of a suboptimal start solutions.

5.2.2 Rail freight yard graph model

In the case of the freight node model, the numerous rail tracks form not only parallel but also successive machines (stations). The optimisation problem is thus to be defined as a so-called flexible job-shop scheduling problem. More precise it is to speak about a *job-shop scheduling problem with zero buffers* since in between any two successive machines are no buffers implemented to queue trains. In the process sequence of the present simulation model, the machines *Shunting* and *Formation* are visited twice by the trains. An optimisation model in which the trains can return to already visited machines, is to be called a *job-shop scheduling problem with recirculation* (Pinedo, 2005, p. 80).

The Job-shop Scheduling Problem (JSSP) is one of the classic problems in operations research, first described by Manne (1960). It can be formulated in terms of a disjunctive graph model as for example Brucker (2007) did. According to Strotmann (2007), a disjunctive graph G consists of a set V of nodes, a set C of directed arcs (conjunctions), and a set D of undirected arcs (disjunctions). Considering the JSSP, Strotmann (2007) defines the disjunctive graph $G = (V, C, D)$ as follows:

- The set V of nodes represents the set of all operations. In the case of the rail freight yard, V denotes the stations (e.g. the loading tracks).
- The set C of conjunctions represents the set of precedence constraints between consecutive operations of the same job. In the case of the rail freight yard, each job is a train and each train possesses an ordered set of operations (a sequence of processes in which it is involved).
- The set D of disjunctions represents the different orders in which jobs on the same machine may be scheduled. Applied to the rail freight yard it denotes the order of trains to be handled on the same track.

JSSPs get very complex with an increasing number of machines and jobs. Only a few special job shop problems can be solved polynomially. In particular, $J_i||C_{max}$ with $i \geq 3$, i.e. a JSSP with more than three machines, is NP-hard (Jaehn and Pesch (2014, p. 28 ff.) or Brucker (2007, p. 41 ff.)) which means that it is very unlikely to be solved in polynomial time. Larger optimisation problems are therefore generally solved with a branch-and-bound method or heuristics-based methods. Usually, optimisers search for ways to simplify an optimisation problem. It will be shown in the following section that the characteristics of the rail freight yard also allow the decomposition of the problem. Thus, the scheduling space can be reduced and the optimisation problem simplified.

5.2.3 Hierarchical planning and reduction of scheduling space

At larger rail freight nodes (but also in general), the generation of schedules follows priority rules. As initially stated in section 3.4.2, freight terminal operators assign fixed time slots for the loading and unloading processes of trains. From practice it is known that these slots can hardly be moved because they rely in turn on terminal-internal process constraints such as the preparation of the right loading units or the organisation of the storage areas. The railway undertakings can book those slots and have to make sure that their trains reach the booked slots on time. The assignment of the terminal handling slots is therefore regarded as a priority. This means that the other railway-related activities have to be planned secondarily.

Directly connected to the terminal's loading processes are the shunting operations. Consequently, they are very inflexible, too, since the delivery of trains must be carried out timely before and after the cargo transshipment operations. The scheduling of the shunting processes must be adapted to serve the terminal slots.

Further, connected to the shunting movements are the processes in the formation yard. They are bound on the other side to the train arrivals and departures and thus depend on both, the yard internal and the external influences/deviations. Consequently, the formation yard slots are the most crucial point in the overall slot plan scheduling to which particular attention must be devoted.

The fixation of the terminal transshipment slots is also adopted as a rule in the slot scheduling model. The scheduling space for the transshipment slots is optimised by the terminal manager and is thus outside of the model. Consequently, the rail yard manager does not have any possibility to modify schedules in these tracks. This fact reduces the scheduling space significantly. However, under certain conditions *slot swaps* will be introduced. They can be understood as the exchange of transshipment slots between two trains. The optimisation efforts can be put on the important optimal allocation of resources in the formation yard. The ultimately available scheduling space of one train is depicted in Figure 5.9. The formation tracks give the relevant

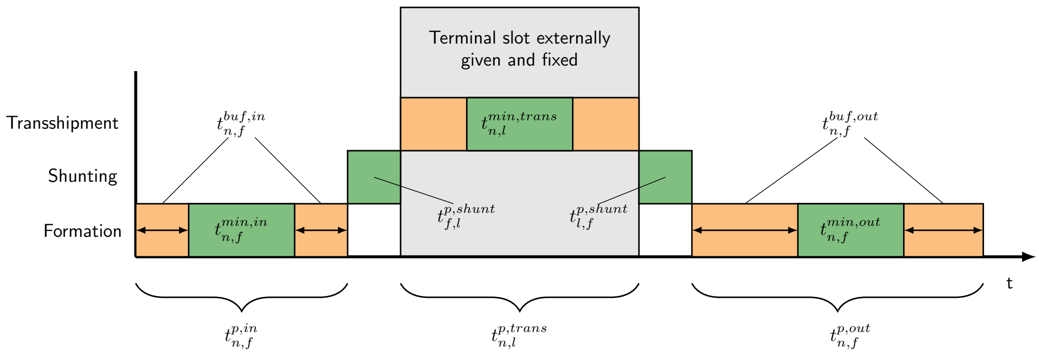


Figure 5.9: Scheduling Space under the assumption that the transshipment slots are externally provided by the terminal operator.

scheduling space for the rail yard manager. With the fixation of the terminal slots, the scheduling problem can now be reduced to a *parallel machine problem* which is faster to solve than a job-shop scheduling problem. When reducing the model to a parallel machine problem, it is important to consider that trains appear twice in the model – once as incoming and once as outgoing train. This interrelations between the two connected processes must be obeyed, especially when modelling timetable deviations as it will become important at a later stage of this thesis.

The shunting operations must be scheduled separately. This can easily be handled in a vehicle scheduling model that generates schedules for the shunting locomotives. Nonetheless of the simplifications, two major scheduling rules for the terminal slots will be introduced in order to determine the model behaviour in case of train delays:

1. Booked terminal slots can be cancelled or rescheduled at the latest two hours before the slot start time. Otherwise missed slots expire and a new one must be booked (and paid).
2. Trains that miss their assigned terminal slot due to arrival delays will be scheduled to another one.

The first scheduling rule allows the terminal to assign a slot that will be missed to another train. Two hours in advance are hereby required to prepare the correct cargo units (locate them in the storage area and transfer them to the intermediate storage) and to plan the processes inside the freight terminal accordingly. The second scheduling rule controls the slot allocation of delayed trains and determines the action strategies of the terminal. A first slot design with a set of trains can now be analysed in the next section.

5.3 Time slot scheduling in the non-delayed case

With the reduction of the optimisation space, the scheduling of slot plans concentrates on the optimal allocation of available resources, such as shunting crews or wagon technicians, to the freight trains. The object of the optimisation problem is now to organise the minimum process times $t_{n,i}^{min}$ of a series of considered trains in such a way that the available resources are optimally used but not over-used (capacity constraints must be respected).

The analysis is started without assuming arrival delays. In the non-delayed case it is assumed that all scheduled trains arrive on time and reach their slots accordingly. The behaviour with arrival deviations is evaluated in sections 5.4 and 5.5.

5.3.1 Prerequisites

Considered be a set of $i = 16$ parallel tracks in a formation yard as initially defined. Attached to it be an intermodal terminal with six parallel tracks. The terminal's transshipment slots are provided by the terminal manager as follows: The planned dwell time of a transshipment slot $t_{n,l}^{p,trans} = 300$ min. To avoid peaks and conflicts during shunting movements, the slots start in the parallel loading tracks with a time lag of 30 min. This results in a slot plan in the transshipment terminal with a recurring pattern, as shown in Figure 5.10. With the given slot plan for the loading tracks, the latest end time for the delivery and the earliest start time for the pick-up of the wagons by the shunting locomotives can be determined.

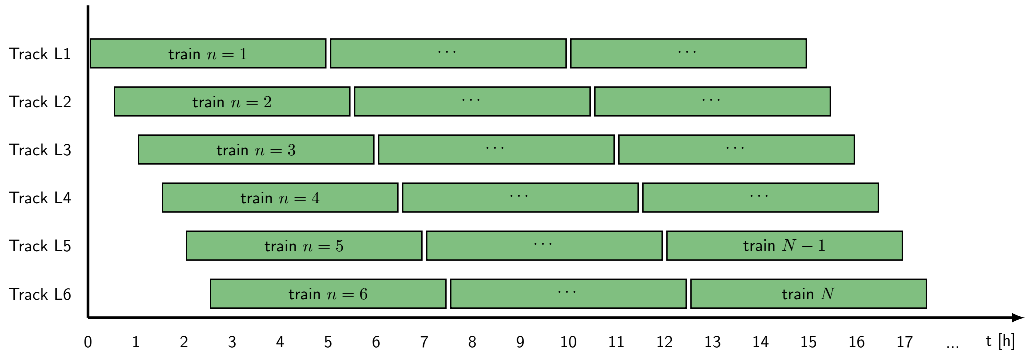


Figure 5.10: Slot plan pattern provided by the intermodal terminal as an input parameter for the scheduling approach.

The shunting movements determine in turn the latest end time of the train arrival process and the earliest start time of the train departure process. Both take place in the formation yard. With these information, all critical points are known to set up a first slot plan also for the formation yard. The well-known cycle through the node consisting of the five stages *Arrival - Shunting - Transshipment - Shunting - Departure* (cf. Figure 2.9) is accordingly mapped to an initial slot plan by using the previously determined planned process times $t^p_{n,i}$. This initial slot plan is listed as a table in Appendix A.1[20]. It contains 25 slots. Indicated are the starting times of the slots in the respective stations. So far, no trains have been assigned to the slots. How far this plan can be held, will now be determined by processing trains through the scheduling model.

5.3.2 The scheduling model

The scheduling model $Pi|p_n, r_n, d_n, prec|C_{max}$ used for the upcoming analyses is a parallel machine model minimising the maximum makespan, i.e. minimising the overall runtime for the model to execute all jobs. The optimisation criterion C_{max} is mostly suitable for optimising the schedules of a complex freight yard because it allows the processing of a given set of trains in the shortest possible time and simultaneously ensuring a good workload balancing on all machines.

The implemented optimisation model, differs from the queueing simulation model of chapter 4. In the queueing simulation, the rail tracks were considered as stations or machines. Now, the number of parallel machines I must indicate the number of

[20]The first slot does not start at $t = 0$, because the warm-up phase has already been taken into account and the analysis is carried out under full load.

resources that are available to handle the trains in the yard. Relevant resources are thus the teams of wagon technicians. This should not be confused with the number of available formation tracks. The number of formation tracks is a capacity constraint which determines how many train sets can be stored in the yard at the same time but not each track can be regarded as a working machine since the works can not be carried out on all tracks at the same time (unless the number of technician teams equals the number of formation tracks). The scheduling problem on parallel machines is considered with the following assumptions:

- A machine can process at most one job (train) at a time.
- A setup time is inserted between two succeeding trains on the same track to indicate eventual concerns, e.g. of the train control system.
- Preemption of the jobs is not allowed.
- No machine breakdowns occur.
- All machines can process both, incoming and outgoing trains.

The mathematical model is given as a linear programming formulation:

Parameters:

N	number of jobs (trains), $n = \{1, ..., N\}$
I	number of parallel machines, $i = \{1, ..., I\}$
D	dummy job index
M	a big number
p_n	processing time of train n
r_n	release time of job n
d_n	due time of job n
s_n	setup time of job n

Continuous variables:

ST_n	starting time of job n
C_{max}	maximum completion time (makespan)

Binary variables:

x_{ni}	is 1 if job n is assigned to machine i, otherwise 0
y_{noi}	is 1 if job o is processed after job n on the machine i, 0, otherwise
z_{dni}	is 1 if job n is processed as the first job on machine i, 0, otherwise

Equations:

$$min\ z = C_{max} \tag{5.11}$$

s.t.

$$ST_n + \sum_{i \in I} \Big(\sum_{\substack{m \in N \\ m \neq n}} s_n y_{mni} + p_n x_{ni} \Big) \leq C_{max} \qquad \forall\, n \in N \tag{5.12}$$

$$ST_n + \sum_{i \in I} \Big(\sum_{\substack{m \in N \\ m \neq n}} s_n y_{mni} + p_n x_{ni} \Big) \leq ST_o + M\Big(1 - \sum_{i \in I} y_{noi}\Big) \qquad \forall\, n, o \in N,\ o \neq n \tag{5.13}$$

$$\sum_{i \in I} x_{ni} = 1 \qquad \forall\, n \in N \tag{5.14}$$

$$\sum_{o \in N} z_{doi} = 1 \qquad \forall\, i \in I \tag{5.15}$$

$$\sum_{\substack{o \in N \\ o \neq n}} z_{oni} + z_{dni} = x_{ni} \qquad \forall\, n \in N; \forall\, i \in I \tag{5.16}$$

$$\sum_{\substack{o \in N \\ o \neq n}} z_{oni} = x_{ni} \qquad \forall\, n \in N; \forall\, i \in I \tag{5.17}$$

$$ST_n \geq r_n \qquad \forall\, n \in N \tag{5.18}$$

$$ST_n + p_n \leq d_n \qquad \forall\, n \in N \tag{5.19}$$

$$x_{ni},\ y_{noi},\ z_{dni} \in \{0, 1\} \qquad \forall\, n, o \in N; \forall\, i \in I \tag{5.20}$$

$$ST_n \geq 0 \qquad \forall\, n \in N \tag{5.21}$$

A dummy notation D is introduced to define a virtual first job and to define the sequence of jobs on each machine. Equation 5.11 is the objective function. Equation 5.12 calculates the makespan so that it is not smaller than the completion time of the last job. Equation 5.13 defines the precedence relations of jobs on each machine. Equation 5.14 ensures that each job has to be processed on exactly one machine. Equation 5.15 ensures that the dummy job O is placed as the first job of a machine sequence. Equation 5.16 defines the precedence relation of a job if it is assigned to a machine. Equation 5.17 describes that a job can be succeeded by at most one job. Equations 5.18 and 5.19 define that a job is executed between its release time and its due time. Equations 5.20 ensure the integrality of the variables and equation 5.21 satisfies the non-negativity condition.

5.3.3 Input data and expected results

A first timetable is initiated and applied to the model. It is a cyclic schedule with continuously arriving trains and is conform to the initial slot plan given in Appendix A.1. The depicted arrival times will be incorporated as the release times $r_n, \forall N$. In the same way, the departure times correspond to the due times $d_n, \forall N$. The cyclic structure has accordingly no peaks. Deviations and peaks will be considered at a later stage of the analysis.

The timetable is provided in Appendix A.2. Only the lower 25 slots will be used. In that way, the warm-up period of the model is already considered and the yard acts as already in full operation from the beginning of the optimisation run. The execution of the model provides important KPIs for measuring the performance of the produced slot plans. Such KPIs are:

- **Makespan** C_{max}: The overall process time of the whole schedule
- **Maximum tardiness** T_{max}: Delay of the train with the largest slot conflict (in minutes)
- **No. of late jobs** $\sum U_n$: Number of trains suffering a slot occupation conflict
- **Total tardiness** $\sum T_n$: Sum of all slot conflicts (in minutes)

The stability of the schedule is measured by the number of disrupted jobs ($\sum U_n$). A disrupted job denotes a train that is not processed within the provided planned time slot, i.e. outside of the interval $\{r_n, d_n\}$ and thus conflicts with other slots. A slot plan that contains disrupted jobs is not feasible and cannot be considered as a successful approach.

5.3.4 Computation results

The compliance of the planned slots depends on the number of resources available for handling trains. They determine, how many trains can be handled in parallel. The teams of employed wagon technicians are the important required resources here. A first optimisation run with one shunting team is strongly violating the slot plan by exceeding the planned dwell times. The result is illustrated in the graphical slot plan of Figure 5.11. The grey areas represent the planned process times $t^p_{n,i}$, i.e. the time slots

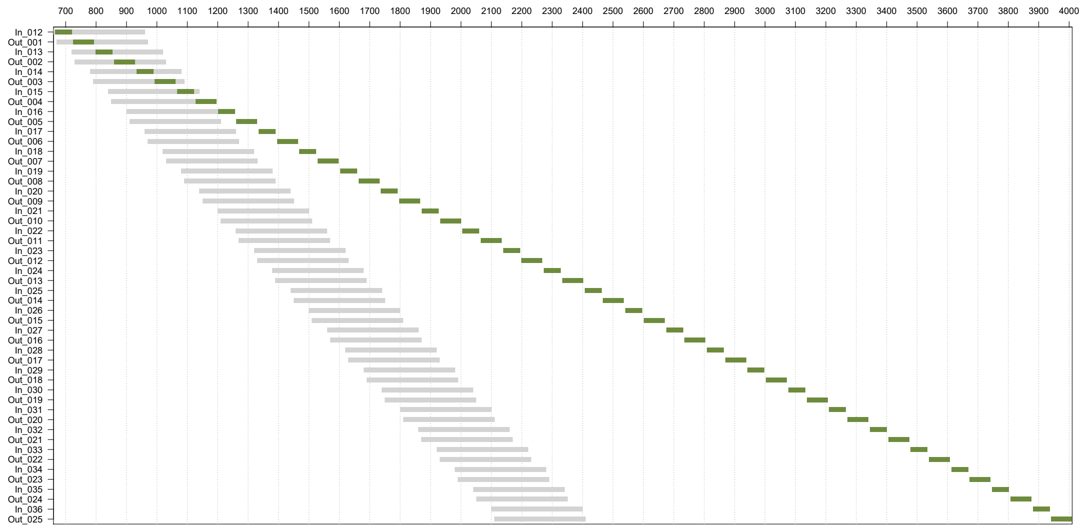


Figure 5.11: Slot plan for the formation yard having one technician crew handling the trains (abscissa: time in min, ordinate: train numbers, grey: assigned time slot, green: actual time of train handling). With only one technician crew, the working speed is insufficient to meet the slot plan. See Appendix A.3 for larger version.

in which the technician teams shall handle the trains. The green coloured areas indicate the period during which a technician team is actually performing the required train handling. It quickly becomes clear that one shunting team is not sufficient to handle the desired number of trains in the yard in the given time slots.

Employing a second team of technicians improved the situation considerably as illustrated in Figure 5.12. Although all the considered trains are being now handled within the planned processing time, it becomes evident upon closer inspection that two technician teams still do not lead to a steady state of the system. It can be recognised that the train handling takes place more and more at the end of the slot time with the later arriving trains. If the schedule would continue in the same way, it would come to delays at a later time. The solution with two technicians is thus not stable in the long term. The employment of a third technician team would stabilise

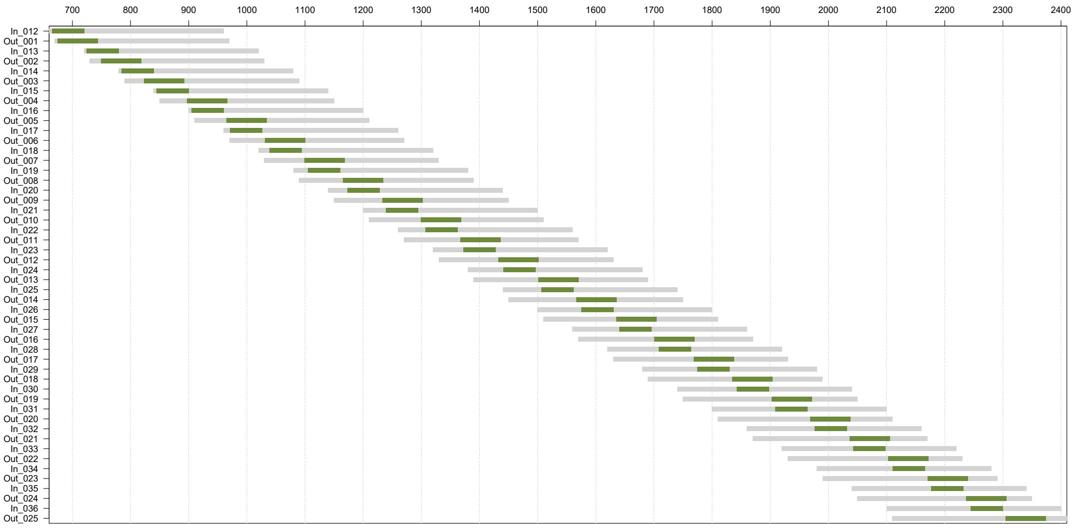


Figure 5.12: Slot plan for the formation yard having two technician crews handling the trains (abscissa: time in min, ordinate: train numbers, grey: assigned time slot, green: actual time of train handling). With two technician crews the working speed improved. See Appendix A.4 for larger version.

the system but it also leads to a lower productivity of the yard and thus leads to higher costs. All productivity figures of the comparison are summarised in table 5.1.

	1 tech team	**2 tech teams**	**3 tech teams**
max. required tracks:	29	10	10
Makespan C_{max}	3,350 min	1,714 min	1,524 min
Max. tardiness T_{max}	1,600 min	0 min	0 min
No. of late jobs $\sum U_n$	43	0	0
Total tardiness $\sum T_n$	34,822 min	0 min	0 min

Table 5.1: Comparison of the yard productivity in the different scheduling runs in the non-delayed case.

It is noticeable first, that the processing with one technician team is not even possible due to a violation of capacity constraints: It would require 29 tracks in the formation yard but only 16 are available. With the appointment of a second team, the number of the parallel occupied tracks can be reduced to 10. The remaining six tracks could therefore be available for special tasks or dispatching needs. The employment of the second team is also reducing the makespan by about 51% and it prevents the emergence of slot violations for the given schedule. The produced schedule is thus feasible. The employment of a third technician team obtains only a small reduction in the makespan but leads to a system which is stable in the long term.

All optimisation runs were carried out with the EDD priority rule (see section 5.2.1). The same results were obtained using the FCFS rule. This is because in the non-delayed case holds $p_n = d_n - r_n, \forall n$. Thus, the first arriving train has also the earliest due date. The consideration of priority rules is not significant yet but will become important when investigating the model behaviour on train arrival delays. For the sake of completeness it should be noted that the use of the SPT rule requires 11 parallel tracks for the same operations within a makespan of 2,382 min. The LPT and the WSPT rule usually did not lead to a useful result because the outgoing trains were always preferred. This led to bottlenecks in the handling of arriving trains.

5.4 Time slot scheduling with basic arrival time deviations

The model design in this section is unmodified to the previous one. It will further be simulated with two technician teams as limiting resources. The timetable is now being modified with arrival deviations. The study of the slot plan behaviour is again carried out carefully by changing only one variable at a time. The explanations will therefore focus directly on the results. In the following, different delay scenarios and their effects on the slot plan are discussed.

5.4.1 The arrival delay of a single train

First, train 15 (with the ID In_15) is progressively delayed in the range from 30 to 300 min. Its effect on the slot plan quality is examined. The disturbance of the ideal plan generates initially an increase in the makespan C_{max}. It rises from 1,714 min in the non-delayded case to 1,746 min if the train is for example 60 min late but the greater the delay, the closer C_{max} approximates to its minimum value. With 300 min delay, $C_{max} = 1,720$ min. The outcome of the entire simulation run is shown in Figure 5.13. This complex graph shows the relevant KPIs aggregated. It compares the delay of train 15 with the evolving change of the makespan. The size of the points indicates the total tardiness $\sum T_n$ and the colour indicates the number of late trains $\sum U_n$.

With only one delayed train, the consequences on subsequent handling processes can be absorbed. Only train 15 suffers a slot conflict induced by its own delay. This induces a higher infrastructure occupation. No other train suffers a processing delay that could cost him his terminal slot. The scheduling result depends of course also on the used priority rule. The shortest makespan of $C_{max} = 1,700$ min was achieved

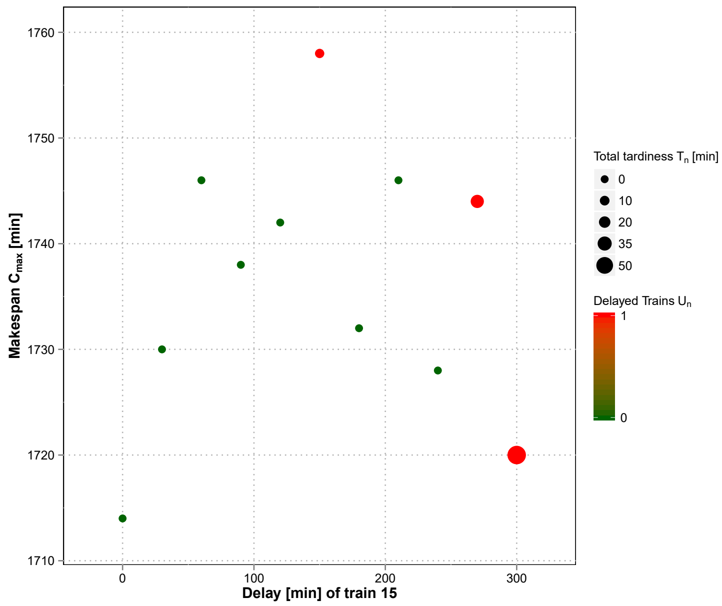


Figure 5.13: Makespan C_{max}, the number of delayed trains U_n and the total tardiness T_n with increasing delay of train 15 under the EDD priority rule: Slight arrival delays induce first an increase of the makespan but no processing delays. With an increase of delays, a swap of two slots can happen which reduces the makespan but induces processing delays.

using the MST rule. Both, SPT and FCFS achieve a nearly similar makespan. SPT but caused delays in some later appearing outgoing trains and consumes therefore two additional formation tracks during the peak time. The EDD rule takes the highest makespan. However, this is with 1,746 min vs. 1,700 min not a significant increase of 2.7%. For smaller delays (60 min) FCFS performs better. When it comes to larger delays (300 min) EDD is the better choice. The most important key figures of the priority rule comparison are summarised in the following table:

Priority rule	**MST**		**FCFS**		**EDD**		**SPT**	
Delay	60 min	300 min	60 min	300 min	60 min	300 min	60 min	300 min
No. of tracks	10	11	10	11	10	11	12	12
C_{max} [min]	1,700	1,718	1,724	1,724	1,746	1,720	1,722	1,722
T_{max} [min]	0	60	0	66	0	60	78	78
$\sum U_n$	0	1	0	1	0	1	10	10
$\sum T_n$ [min]	0	60	0	60	0	60	416	416

5.4.2 The arrival delay of several trains

In this situation, train 16 (In_16) is kept constantly delayed by 60 min. Additionally, train 15 is progressively delayed as before in a range from 30 to 300 min. The effects on the slot plan are as follows: Depending on the used priority rule, the slot plan performance differs in the moments when both delayed trains are being processed simultaneously. Let the situation be considered when train 15 is 150 minutes delayed: In this situation, the EDD rule prioritises the two delayed trains over others because their due dates are regarded as the critical criterion. This leads to the situation shown in Figure 5.14 where the technician teams are running idle for a moment while they are waiting for the two delayed trains to arrive. The result is a longer makespan as

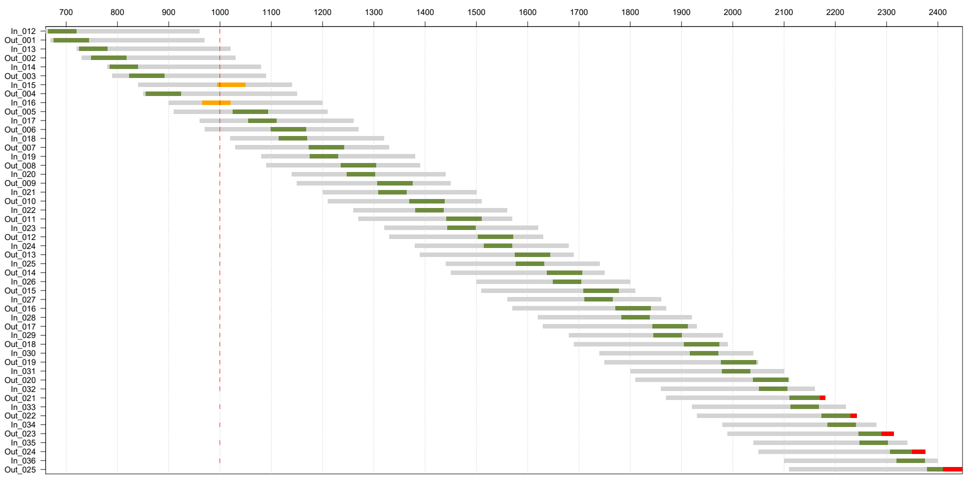


Figure 5.14: Scheduling with the EDD rule when train 15 is 150 min late. Waiting for the delayed trains caused idle time for the technicians and makes later arriving trains to be handled out of their slots (red markers). See Appendix A.5 for larger version.

well as a further delay and possible loss of the terminal slot for later arriving trains. The red markers in the Figure indicate the trains whose slot times were violated. Solving the model with the MST or the FCFS rule has avoided the problem. With a further increase of the delay of train 15, C_{max} and T_{max} increase under the EDD rule, see Figure 5.15. A bigger delay of train 15 induces impacts on more other trains and increases the total tardiness.

Summarising the analysis of scheduling delayed arriving trains, the EDD rule produced the longest makespan. Best results were achieved with the MST rule (lower makespan and lower tardiness). The FCFS rule is in this case also suitable for mi-

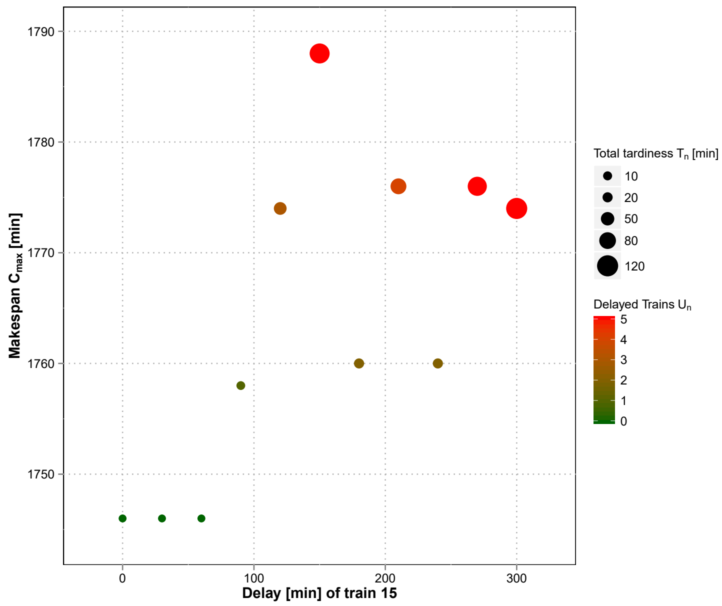


Figure 5.15: Makespan C_{max}, the number of delayed trains U_n and the total tardiness T_n with increasing delay of trains 15 and 16 under the EDD priority rule.

nor delays but performs worse with long ones. The results are summarised in the following table:

Priority rule	**MST**		**FCFS**		**EDD**	
Delay	**60 min**	**300 min**	**60 min**	**300 min**	**60 min**	**300 min**
Max. no. of tracks	10	11	10	11	10	11
C_{max}	1,700	1,718	1,724	1,746	1,746	1,774
T_{max}	0	60	0	88	0	60
No. of late jobs $\sum U_n$	0	1	0	1	0	5
Total tardiness $\sum T_n$	0	60	0	88	0	120

5.4.3 The early arrival of a single train

In this situation, the arrival time of train 21 (In_21) was progressively set to earlier states in a range from 30 to 300 min. The early arrival of trains does not lead to any positive effects in yard productivity but has some drawbacks. As it is shown in Figure 5.16, the early arrival has no positive influence on the job scheduling. The scheduling model makes no changes in the primary schedule. The technician team handles the train at a later suitable time. If a train arrives before its slot time starts, the slot gets

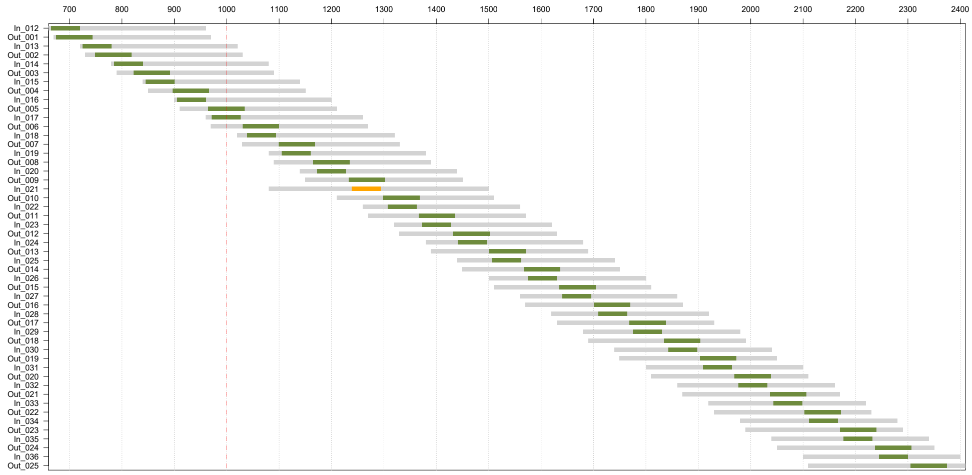


Figure 5.16: Train 21 arrives 120 min early. The designated formation track is therefore longer occupied than planned. Logistically, the earliness has no advantage. The technicians work according to the original schedule and the terminal slot of train 21 is not advanced, too.

prolonged which results in additional capacity occupation in the formation yard. The envisaged track could still be occupied and an alternative track must be found by the yard manager.

Also from a logistics point of view, the advantage of an early arrival is limited. The train's terminal transshipment slot will not start earlier because of its early arrival. Only in a very specific situation the early arrival could have an advantage: Terminal slots between an early and a delayed train could be swapped. Then, the exact arrival times of both trains must be known at least two hours in advance (see rules described in section 5.2.3) so that the terminal operator can react and swap the two slots accordingly.

The choice of the priority rules has also less important effects on early arrivals than it had on delayed trains. The shortest makespan was again achieved with the MST rule with $C_{max} = 1,692$ min. The FCFS and the EDD were equally good and close to MST with $C_{max} = 1,714$ min. Some improvement was only achieved by applying a shifting bottleneck heuristic minimising C_{max} to the scheduling problem but it led in return to the delay of four other trains. All values are illustrated in the following table:

Priority rule	MST	FCFS	SB-Heuristic min. C_{max}	EDD
Max. no. of tracks	11	11	12	11
C_{max}	1,692	1,714	1,686	1,714
T_{max}	0	0	606	0
No. of late jobs $\sum U_n$	0	0	4	0
Total tardiness $\sum T_n$	0	0	1,208	0

5.4.4 The early arrival of several trains

In this last analysed scheduling situation it is assumed that train 21 arrives 60 min earlier. At the same time, the arrival time of train 22 was progressively set to earlier states in the range from 30 to 300 min. Similarly to the observations made in section 5.4.3, also the early arrival did not lead to any positive effects.

The makespan and the other key figures did not change significantly when applying different priority rules. The shifting bottleneck heuristic could perform slightly better in this situation. It was confirmed that the number of additionally required formation tracks rises with every early train:

Priority rule	MST	FCFS	SB-Heuristic min. C_{max}	EDD
Max. no. of tracks	12	12	12	12
C_{max}	1,692	1,714	1,686	1,714
T_{max}	0	0	606	0
No. of late jobs $\sum U_n$	0	0	3	0
Total tardiness $\sum T_n$	0	0	1,148	0

Summarising, the earliness of arriving trains brings no benefits to the productivity of the freight hub and no acceleration of the transport and transshipment processes. Thus, no advantages derive from early arrivals for logistics.

5.5 Time slot scheduling with arrival uncertainty

While in the previous section systematic timetable deviations were examined, now random early and late arrivals of multiple trains will be applied to the original schedule. Random arrival times were generated by means of a Monte Carlo simulation[21].

[21] The Monte Carlo method can simulate problems with random behaviour and generate required data with certain distribution properties (cf. for example Carsey and Harden, 2013).

As an input for calculating randomness served the probability distribution function of the arrival times known from section 4.3.2. It was envisaged that the data reflect a realistic punctuality of today's rail freight transport. By applying the arrival time probabilities to the original timetable, the timetable deviations for the 25 incoming trains were generated so that the altered arrival times emerged as shown in table 5.2. Of the 25 incoming trains, 16 obtained a delay and 7 arrive earlier. Two trains con-

ID n	**Planned Arrival** $t_{n,f}^{p,arr}$	**Random delay** $t_{n,f}^{del,arr}$	**New arrival** $t_{n,f}^{arr}$
In_012	660	24	684
In_013	720	16	736
In_014	780	108	888
In_015	840	0	840
In_016	900	-10	890
In_017	960	-12	948
In_018	1,020	250	1,270
In_019	1,080	29	1,109
In_020	1,140	118	1,258
In_021	1,200	167	1,367
In_022	1,260	-28	1,232
In_023	1,320	-62	1,258
In_024	1,380	-128	1,252
In_025	1,440	153	1,593
In_026	1,500	302	1,802
In_027	1,560	61	1,621
In_028	1,620	213	1,833
In_029	1,680	233	1,913
In_030	1,740	0	1,740
In_031	1,800	-54	1,746
In_032	1,860	127	1,987
In_033	1,920	98	2,018
In_034	1,980	281	2,261
In_035	2,040	207	2,247
In_036	2,100	-49	2,051

Table 5.2: Random arrival times generated during a Monte Carlo simulation following the probability distribution function of arrivals from section 4.3.2.

tinue to arrive on time. The new target arrival time for each train n is defined as the sum of its planned arrivals and the random delay:

$$t_n^{arr} = t_n^{p,arr} + t_n^{delay} \tag{5.22}$$

The new timetable has been applied to the scheduling model and the existing slot structure. In this section it is assumed that the arrival deviations are not known in advance and for the moment no rescheduling of the booked time slots takes place. It will be examined, to what extent the scheduling model can absorb the timetable deviations with the available resources. Also here, different priority rules were taken into account. The main results are depicted for comparison in Figure 5.17. It was

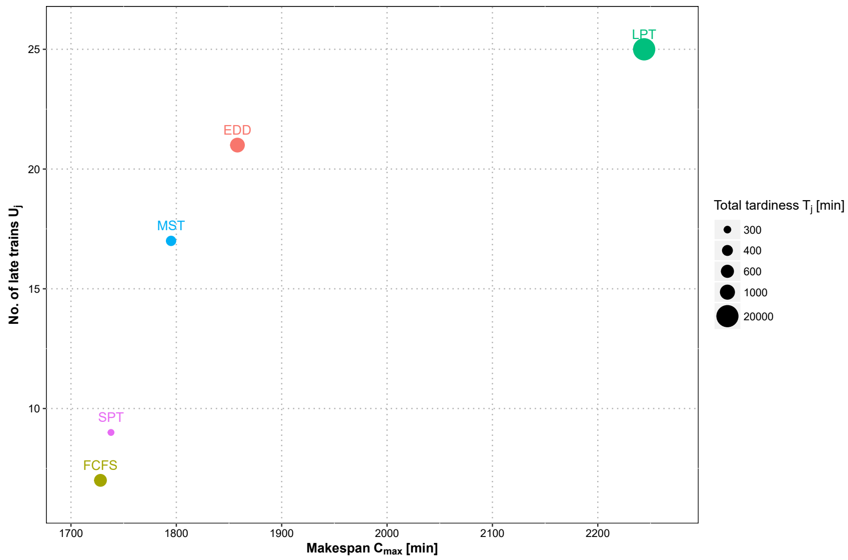


Figure 5.17: Performance of different priority rules with random train arrivals

found that the use of the FCFS rule delivers the best results, i.e., requiring the shortest makespan and producing the lowest number of delayed trains. The SPT rule performs similar since it privileges the incoming over the outgoing trains. It could thus perform better than in the non-delayed case. The MST and the EDD rule perform with a medium makespan but are delaying significantly more trains, although each train has only a small delay. Especially the MST rule performs worse than in the case of planned single deviation. The LPT rule was not able to be considered for further examination at all.

The experiments were repeated with a couple of different randomly generated schedules. Again, the different priority rules were compared. Although the EDD rule was expected to provide better performance, the FCFS rule was most suitable since it generated the lowest makespan. However, it delayed more trains in relative and absolute terms. The numerical results of a typical experiment are summarised in the following table:

Priority rule	EDD	FCFS	LPT	MST	SPT
Max. no. of tracks	13	13	30	13	13
Makespan C_{max}	1,858	1,728	2,244	1,795	1,738
Max. Tardiness T_{max}	108	116	1,224	62	65
No. of late jobs $\sum U_n$	21	7	25	17	9
Total tardiness $\sum T_n$	921	480	21,960	369	297

Particularly noticeable is that now 13 instead of 10 formation tracks are required at best. Also the number of late jobs and the total tardiness differs widely among the results. For further decision-making and assessment, a look at the graphical slot plan is helpful. Exemplarily, the slot plan of the FCFS rule is shown in Figure 5.18. Seven

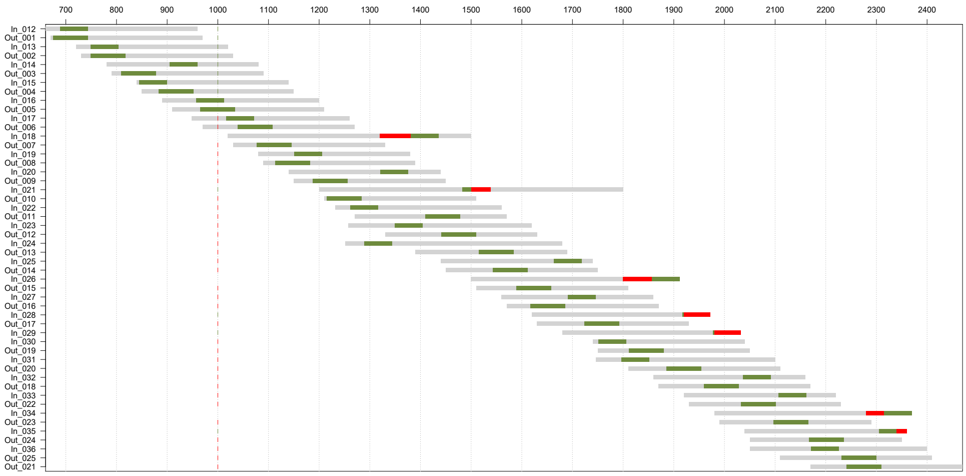


Figure 5.18: Slot Plan with randomly deviant arrivals: The trains In_18 and In_21 were assigned new terminal slots and thus new departure times. Thus, the dwell time in the formation yard is extended. See Appendix A.6 for larger version.

trains did not receive the incoming check within their slot time. They have to apply for a new terminal slot. For the formation yard this means that their dwell times will be prolonged further until new terminal slots have been assigned and that these trains will suffer an additional delay in the node as well as an eventual departure delay.

Exemplarily, the situation of train 18 shall be analysed: The incoming train *In_18* arrives with $t^{del,arr}_{In18} = 250$ min late. After being scheduled for the formation yard handling, the train violates his slot time by 116 min and therefore loses its terminal slot. In the best case, the terminal operator could now assign the slot of the also delayed train 21 to it. The result would be the following: Train 18 is at $t = 1,436$ min ready for entry into the terminal but since the new terminal slot starts at $t = 1,500$ min, the dwell time extends for another 64 min. On all the subsequent processes (shunting to the terminal, unloading, loading and shunting back to the formation yard), train 18 gets the slots of train 21 and thus a departure time $t_{Out18}^{dep} = 2,170$ min. Assuming that the train *Out_18* can exit the node at this time, it leaves with a delay of $t_{Out18}^{del,dep} = 180$ min. Its total turnover time was decreased from $\hat{T}_n = 970$ min to $\hat{T}_n = 900$ min. Hence, the initial delay could even be reduced. Whether delays can be reduced, depends on the freight terminal and its possibilities to provide new transshipment slots.

In the case of the next delayed train, *In_21*, the situation is slightly different. It could be allocated to the terminal slot of train *In_26* which is the following train that is going to miss his slot. Due to the long waiting time between trains 21 and 26, the total dwell time increased to $\hat{T}_n = 1,103$ min. Thus, the train arrived with a delay of $t_{In21}^{del,arr} = 167$ min but leaves with $t_{Out21}^{del,dep} = 300$ min.

From the illustrated example it becomes apparent, which coordination effort it entails if freight trains arrive unexpectedly at the node. Due to the delays, unknown in advance, the prearranged slot plans became infeasible. Major drawbacks of the situation are longer dwell times in the formation yard and thus a longer occupation of the infrastructure resources. This is reflected in a higher number of required formation tracks. Thus, the efficiency of the entire node is lowered.

5.6 Rescheduling time slots

The dynamic reaction on train arrival deviations can have a crucial impact on the efficient operations of freight nodes. However, the timely notification of train arrivals is an essential precondition for it. For the analysis in this section it is now assumed that the deviations in the arrival times are sufficiently known in advance and thus the rail freight node can dispatch its internal processing accordingly. With the help of appropriate rescheduling measures, the orders of trains can be rearranged so that the slot conflicts are minimised. In particular, the section addresses the questions:

- How does the yard control system react when disturbances arise?
- When and how is it more convenient to reschedule trains?
- Which algorithm performs best in terms of delay minimisation?

Several approaches to eliminate the disrupted jobs are possible subject to the constraint of minimising the total flow time. Depending on the rescheduling method, not only the conflicting but also the other trains can be affected by the rescheduling in varying degrees. Three rescheduling methods are considered in the following. Their objective is to integrate the conflicted trains back into a feasible slot plan:

FCFS rescheduling rule: [22] All incoming trains are sorted newly by their actual arrival time. The originally planned arrivals are rather ignored. The trains are then being assigned to the existent slot plan. The first arriving train is put into the first slot, the second train to the second slot and so on.

Minimisation of the maximum delays: All incoming trains were assigned to new slots that match their newly estimated arrival time best. The highest priority have trains with the biggest delay. They were assigned first. Then the early arriving trains were planned and after them all the rest.

Minimum slot swap: First, trains are getting scheduled according to the initial slot plan. Then, only the trains that provoked allocation conflicts in the original schedule, induced by their unpunctual arrival, are moved by one time slot (in the regarded case by $\pm$ 60 min). All other trains retain their original slots or are moved slightly in case of new emerged slot conflicts.

The performance of the three rescheduling methods can be compared by investigating relevant KPIs. The efficiency of a schedule is measured by the flow time and the makespan C_{max} as before. The flow time of a job corresponds to the time that a train spends in the system. A number of cost criteria are closely related to the total flow time of a schedule, e.g. the total waiting time, the total lateness and the number of late jobs. Of the variety of possible KPIs, the following key figures were considered as relevant for the analysis of the rescheduling methods:

- Number of late outgoing trains compared to the original schedule
- Number of early outgoing trains compared to the original schedule
- Number of trains that suffered an additional delay
- Average dwell time of the trains in the system
- Average reduction of delays

[22] not to be confused with the FCFS priority rule of the optimisation model

The three rescheduling rules have been applied to the delay scenario from table 5.2. It should be borne in mind again that at best 7 of the 25 trains caused a slot conflict. The analytical results of the rescheduling action with the three presented methods on the given delay scenario are illustrated in the following subsections.

5.6.1 First come first served (FCFS rescheduling rule)

Regardless of the planned arrival time and the initially scheduled slots, incoming trains are sorted according to their actual arrival time. Afterwards they are assigned to slots by following the existing slot structure one-to-one in this order. The simple rule is illustrated by the pseudo code of Algorithm 1.

Algorithm 1 FCFS Rescheduling Rule

```
Read actually determined arrival times
Sort arriving trains by t^{arr}_{n,f} |∀f, n downwards in a list i
for i = 1 to N do                              ▷ Assign trains to slot numbers
    Assign train_i → slot_i
end for
```

In contrast to the original schedule (Appendix A.2), the slot plan is now as shown in this table[23]:

No.	Train ID	New arrival time	Slot start time	Departure time
1	In_012	684	660	1,630
2	In_013	736	720	1,690
3	In_015	840	780	1,750
4	In_014	888	840	1,810
5	In_016	890	900	1,870
...				

The changed train order is directly apparent. Train 14, for example, is now scheduled after train 15 because of his arrival delay of 108 min. The newly created slot plan must be checked again for feasibility. This is carried out by applying it to the slot scheduling model. Again, several priority rules were tested. The FCFS and the MST rule produced also here the best results by scheduling all trains without inducing new delays and with a very low makespan C_{max}. Besides, the MST rule required the usage of only maximally 11 parallel tracks of the formation yard:

[23] Shortened slot plan. The full slot plan is found in Appendix A.7.

Priority rule	EDD	FCFS	LPT	MST	SPT
Max. no. of tracks	11	12	30	11	12
Makespan C_{max}	1,810	1,728	2,244	1,750	1,738
Max. Tardiness T_{max}	60	0	1,224	0	273
No. of late jobs $\sum U_n$	9	0	25	0	4
Total tardiness $\sum T_n$	264	0	21,960	0	512

Only the MST and the FCFS priority rule generated feasible schedules with zero slot violations. Much more interesting than the comparison of the priority rules at this stage is the performance of the rescheduling method. Of the 25 incoming trains, 16 had an initial arrival delay and 7 arrived earlier. After rescheduling, only 8 trains leave the node with a delay. The others could reduce their delay during the stay. 9 trains are ready for departure earlier than needed. The early readiness is not optimal since these trains occupy infrastructures in a non-productive time until they can leave the freight node. However, the application of the FCFS rescheduling rule has thus contributed to a better efficiency of the freight node and also improved the stability of rail freight traffic in the overall network.

5.6.2 Minimisation of the maximum delays (MMD rescheduling rule)

The second scheduling method was designed for reducing the maximum, hence the extreme, delays. Accordingly, the trains with the biggest arrival delays were assigned to new slots first. Regardless of the original schedule, they get assigned to the time slots that are closest to their estimated arrival time. Hence, multiple slots can be skipped. Second, the early arriving trains and then the other trains were assigned to slots. The method is outlined as pseudo code in Algorithm 2.

The method caused some difficulties since the delayed trains get usually assigned to slots that were already occupied by other trains which then have to get rescheduled, too. This may lead to new slot conflicts. The rescheduling efforts are significantly high for this method.

The final slot plan of the simulation run is given in Appendix A.8. It can be seen that two original slots could not be filled. The corresponding trains have been assigned to new slots at a later time at the bottom of the list instead. It is important to note that creating such gaps in the slot plan lowers the terminal performance and it raises the idle time for the transshipment resources.

Algorithm 2 MMD Rescheduling Rule

Read actually determined arrival times
Sort arriving trains by $t_{n,f}^{del,arr} | \forall f, n$ downwards in list i
for $i = 1$ to $N|_{t_{n,f}^{del,arr} > 0}$ **do** ▷ Loop over the the late arriving trains
    $slot_i = |t_{slot_i}^{start} - t_{i,f}^{arr}| \rightarrow min$ ▷ Assign slot with start time $t_{slot_i}^{start}$ closest to train arrival time $t_{i,f}^{arr}$
    Remove train n from list
end for
Sort arriving trains by $t_{n,f}^{del,arr} | \forall f, n$ upwards in list i
for $i = 1$ to $N|_{t_{n,f}^{del,arr} < 0}$ **do** ▷ Loop over the the early arriving trains
    $slot_i = |t_{slot_i}^{start} - t_{i,f}^{arr}| \rightarrow min$ ▷ Assign slot with start time $t_{slot_i}^{start}$ closest to train arrival time $t_{i,f}^{arr}$
    Remove train n from list
end for
for $i = 1$ to N **do** ▷ Loop over the the remaining trains
    $slot_n = |t_{slot_n}^{start} - t_{n,f}^{arr}| \rightarrow min$ ▷ Assign slot with start time $t_{slot_n}^{start}$ closest to train arrival time $t_{n,f}^{arr}$
    Remove train n from list
end for

The feasibility check showed that the new schedule again performs best using either the FCFS or the MST priority rule. Compared to the first rescheduling method, the results are not optimal since minor delays occur:

Priority rule	**EDD**	**FCFS**	**LPT**	**MST**	**SPT**
Max. no. of tracks	12	12	29	12	12
Makespan C_{max}	1,784	1,728	2,244	1,744	1,738
Max. Tardiness T_{max}	34	80	1,164	44	79
No. of late jobs $\sum U_n$	6	5	25	2	7
Total tardiness $\sum T_n$	92	202	20,714	46	270

In the best case the MST rule generated still two conflicting trains. They need to be rescheduled a second time in order to make the slot plan feasible. For comparison, the FCFS rules produced five conflicting trains as illustrated in the table above.

The second rescheduling method performs at the first glance not as good as the FCFS rescheduling rule: After rescheduling, still 10 trains leave the node delayed and one train is too early ready for departure. However, the method reduces the extreme delays of 120 minutes or more significantly. The critical disadvantage of the method is that some trains (six in this example) received an additional lateness in the node through no fault of their own.

5.6.3 Minimum slot swap (MSS rescheduling rule)

The third rescheduling method, the minimum slot swap rule received its name from its aim to change as few of the originally planned slots as possible. It is based on an iterative procedure in which only the trains than do not reach their envisaged time slot will be rescheduled. Here, the conflicting slot is shifted by exactly one slot unit (= 60 min). The procedure is described in Algorithm 3.

Algorithm 3 MSS Rescheduling Rule

Generate initial slot plan
while $\sum U_j > 0$ **do** ▷ Check for any disrupted slots
    for $\forall\, n|_{T_n>0}$ **do** ▷ For all disrupted slots
        $slot_n = slot_{n+1}$ ▷ Assign train to next slot
    end for
    for $\forall\, n|_{T_n=0}$ **do** ▷ For all non-disrupted slots
        $slot_n = slot_n$ ▷ Keep current slot
        if $slot_n : occupied$ **then** ▷ If current slot is already occupied
            if $t_n^{del,arr} > 0$ **then** ▷ If train arrived late
                while $slot_n : occupied$ **do** ▷ Move to next free slot
                    $slot_n = slot_{n+1}$
                end while
            end if
            if $t_n^{del,arr} \leq 0$ **then** ▷ If train arrived early or on time
                while $slot_n : occupied$ **do** ▷ Move to previous free slot
                    $slot_n = slot_{n-1}$
                end while
            end if
        end if
    end for
end while

Compared to the MMD rescheduling rule, the need for rescheduling non-conflicting slots is much lower with this method because from the beginning on, fewer slots must be relocated. Thus, the rescheduling process is much simpler and faster. The final slot plan for this method is given in Appendix A.9. The feasibility evaluation produced the following results after the first iteration:

Priority rule	EDD	FCFS	LPT	MST	SPT
Max. no. of tracks	12	12	30	12	12
Makespan C_{max}	1,812	1,728	2,244	1,742	1,738
Max. Tardiness T_{max}	62	80	1,224	2	79
No. of late jobs $\sum U_n$	10	5	25	1	6
Total tardiness $\sum T_n$	296	224	21,960	2	236

The MST priority rule was again best performing. It produced a schedule containing only one conflicting train with a tardiness $T_{In26} = 2$ min. The result is not optimal since the schedule is methodologically not feasible yet. Taking into account the prevailing emergences of timetables deviations in rail freight transport, these two minutes of lateness may, however, be regarded negligible. After performing a second iteration, a feasible and conflict-free schedule was generated.

The MSS rescheduling method provided very good results regarding the delay recovering. Of the 25 trains, only seven leave the yard late and five are ready before their envisaged departure. Particularly, the extreme arrival delays could be compensated since no train leaves the node with a delay of more than 60 min.

5.6.4 Comparison and evaluation of rescheduling measures

Three rescheduling methods have been applied to the given set of delayed arriving trains. All three methods could generate feasible slot plans for the regarded set. Through the rescheduling, the late trains could also recover from delays. It means that they left the freight node with a smaller delay than they had at their arrival. This is not only relevant for the efficiency of the freight node but especially important for the stable vehicle scheduling of the railway undertakings and thus the stability of rail freight traffic in the overall network.

The rescheduling rules have worked differently well. For comparison, their important KPIs are summarised in the following table:

Rescheduling Method	FCFS Rule	MMD Rule	MSS Rule
Number of late outgoing trains:	8	10	7
Number of early outgoing trains:	9	1	5
Number of trains that suffered an additional delay:	2	6	0
Average Dwell time [min]:	888.24	967.44	888.24
Average reduction of delay:	54%	23%	69%

The first method (FCFS rescheduling rule) is an easy to implement method that produced already good and feasible rescheduled slot plans. It could reduce the delays quite well in relative and absolute numbers. The average dwell time could be reduced from 970 min to 888.24 min. This shows that delays were reduced by the rescheduling action.

The second rescheduling rule (MMD rescheduling rule) allowed a good reduction of extreme arrival delays. This, however, had notable negative impacts on the scheduling of the other trains: Six trains suffered an increase of their delays in the analysis. The low amount of delay reductions is also expressed in the low reduction of the average dwell time.

The third investigated rescheduling rule (MSS rescheduling rule) delivered the best results for the given timetable. It achieved the highest reduction of delays in both, relative and absolute numbers. The MSS rules is also in favour since in generated the least rescheduling efforts and therefore the least changes from the original schedule[24]. Interestingly, also the extremely late arriving trains could achieve a significant reduction of their delays. Two iterations were necessary. After the first iteration, no train had to leave the yard with a delay of more than 60 min. After the second iteration, one train had a maximum outgoing delay of 120 min. The impact on the non-disrupted slots was also very low, meaning that no train suffered an additional delay through no fault of its own. Hence, there was no train or RU discriminated by rescheduling. As the only one, the MSS rescheduling rule considers an existing slot plan and detects conflicts therein. The other two rescheduling algorithms try to prevent slot deviations in the first place without knowing where they may arise precisely. The MSS rule is accordingly based on a broader data base which allows it to make more target-oriented decisions than the other two rules.

The slot plans produced by the three rescheduling methods have been tested for stability using the scheduling model which was introduced in section 5.2. Each schedule has been calculated using several priority rules. The FCFS priority rule delivered in all three cases the minimum makespan, meaning that it enabled to process all trains in the shortest time period. The MST priority rule delivered also a comparably very short makespan but additionally could achieve a lower number of simultaneously occupied tracks and a lower number of late jobs. Thus, using the MST rule,

[24]Generating new schedule sequences usually will result in some nervousness which should be avoided (cf. Stadtler and Kilger, 2005, p. 197 ff.). Therefore, schedules with few changes are preferred over others.

the need for a second rescheduling cycle is less likely. Due to the experiment design, the LPT rule usually produced the least desirable results.

Rescheduling trains in a freight node has not only an influence on the production system of the yard itself, it has also an impact on the rail network and on other freight nodes. Especially, the shifting of slots has an impact on outgoing trains and their booked track slot to the rail network. Early leaves could become a problem since they will have to wait for their scheduled departure time in the formation yard and thus occupy scarce infrastructure resources. Late trains might have lost their departure slot and will have to wait for another one. Generally, the analysed rescheduling methods help to counterbalance early and late arrivals and thus bring back trains to their planned timetables.

5.7 Chapter summary

The chapter dealt with the development, design and optimisation of slot plans in formation yards. First, the minimum required process times as well as the planned process times at different stations of the yard model were determined. Thereupon, it has been defined, how a time slot is internally structured. The single time slots have been tested for robustness against observation data.

Building on these definitions, time slot sequences for the processing of trains through the freight node were defined. The slot plans of a couple of trains were created using a linear scheduling model which minimises the total makespan of all train dwells. It was found that due to the constraints in the yard production system, the optimisation problem could be reduced to a parallel machine problem of the formation yard and thus it could be significantly simplified. Several scenarios with different arrival deviation patterns were defined. For the undisturbed case (without train arrival delays), the optimisation model created stable schedules with sufficiently calculated planned slot times and the envisaged buffer times. The scheduling model was further tested for robustness by adding train arrival deviations. The impacts were analysed by, first, deviating one train from its schedule and, later, deviating multiple train arrivals.

In all analyses, the applicability of different priority rules (as illustrated in section 5.2) was tested. The MST rule performed the best results when no arrival delays occurred. Contrarily, the SPT rule produced the best results when delays occurred. The EDD rule provided the best results when only a few but long delays emerged. The FCFS rule generated in the most cases good results which was supported by the fact that the release dates r_n and the due dates r_n of the slots are relevant constraints in the optimisation model. No rule could prevail as optimal. Hence, it is advisable to always test several priority rules in the slot plan development.

An analysis of the yard performance after the introduction of slot management strategies shows a significant performance increase compared to the conventional production method. With the given slot plan, the maximum time in the system is 950 minutes per train. Without slot management, it was in the most realistic scenario 3b (cf. page 89 f.) at 1,595.70 min. This equals to a reduction of ca. 40.5%.

The utilisation of the loading tracks could also be increased with slot management compared with the conventional production method. In section 4.4.5, a productivity of 81.76% per loading track was calculated. Under the same assumptions of a 24/7 operation, it is with slot management:

$$48 \text{ LU per train} \cdot \frac{24 \text{ h per day}}{5 \text{ h per Slot}} \cdot 365 \text{ days} = 84,096 \text{ LU / year} \tag{5.23}$$

The realistic transshipment capacity of a loading track amounts accordingly:

$$\frac{84,096 \text{ LU}}{100,498 \text{ LU}} = 83.68\% \tag{5.24}$$

Further, the utilisation of the formation tracks can now be determined more accurately, as it is known, at what time shares the trains are actually involved in active processes. In the previous simulative analysis (chapter 4) this was not possible due to the lack of information in the input data. It was only possible to detect an overload based on blockages and congestion phenomena. Such phenomena do no longer occur after the creation of an optimal slot plan.

The last section of this chapter was concerned with the rescheduling of slot plans in unsteady environments. Train arrival deviations make the initial slot plans infeasible. During an analysis with randomly deviated arrival times, rescheduling mechanisms have been introduced to examine the slot stability under real-life situations. Three different rescheduling algorithms were studied. It was found that the time slot management approach is a powerful tool that keeps the yard production system in a steady state. It even allowed reducing the dwell times of trains with arrival delays and thus reduced their overall delays. The slot management approach does not only optimise the processes of the freight hub but also those in favour of other stakeholders in the transport chain, such as freight forwarders, shippers, or railway undertakings. Although they performed successfully, the three presented rescheduling algorithms may not be all-embracing. When applying them to other environments, they must be examined on an individual basis. The algorithms provide a solid foundation for continuing research in the field freight hub scheduling.

6 Requirements and incentives to implement slot management

This chapter focusses on the question of how the developed slot management methods can become accepted by the node users (i.e. the railway undertaking and freight forwarders). It covers two main areas which are important for a successful implementation of slot management strategies in rail freight yards: The focus in section 6.1 is placed on the provision of train arrival information as input parameters for slot scheduling in order to plan the freight node's internal processes. Section 6.2 deals with issues of economics and regulations, in order to provide incentives for an efficient use of the created slot management structures.

6.1 Prediction of train arrival times

The successful analysis on rescheduling measures in section 5.6 showed how important it is to provide sufficient information about train arrival times in order to establish an efficient slot management approach. The capacity utilisation of the modelled freight node was improved significantly when train arrival information were available and the rescheduling rules could be applied. Already Lüthi (2008, p. 351) stated that "One of the most distinctive features of an integrated real-time rescheduling system is that it immediately communicates a dynamically changeable schedule to all actors...". This has not yet been achieved throughout in practice.

The prediction of the estimated time of arrival (ETA) has been researched already under various aspects. New approaches become possible with the use of GPS data and new information technology systems. In today's rail freight transport, delay information are often passed by phone or fax (Schönemann, 2014) which is too slow for efficient dispatching measures and a common source of errors. The conventional methods are insufficient as input data for the rescheduling mechanisms. Rather, automated, continuous information provision is required. Continuous information about the location of trains on the network can usually already be obtained from the train control systems of the national infrastructure manager. Data can be bought by stakeholders and accessed via proprietary interfaces. However, the location data do not provide predicted arrival times directly. Further, their use is often not desired, mistrusted or simply not affordable for some stakeholders so that independent solutions are often preferred. This chapter focuses on methods for the determination of the ETA that are both, preferably applicable for the optimised production planning of rail freight hubs and independent from the information available from other infrastructure managers.

A short section on the recent advances on train delay estimation gives a first insight into the research work on the topic. Thereupon, appropriate independent methods for the arrival time estimation at rail freight hubs are presented. The pursued methods rely on public available data and those providable by railway undertakings or freight forwarders. Thus, they offer an alternative way of ETA calculation compared to those based on train control system data.

6.1.1 Recent advances in delay estimation

Following Bonsra and Harbolovic (2012, p. 16), the potentially utilised methodologies for the estimation of train arrivals and delays can be summarised as:

1. Simulation Models
2. Queuing Models
3. Stochastic Models
4. Regression Models

A significant fundamental research on train delay estimation was accomplished by Chen and Harker (1990); Hallowell and Harker (1996, 1998). Törnquist and Davidsson (2002); Törnquist and Persson (2005) give an approach on the train delay handling on the Swedish rail network. The calculation of ETAs is based on the observed traffic on a network branch and the application of few simple dispatching rules thereon. Results were produced in a simulation environment and mainly meaningful for the infrastructure manager.

Yuan (2006) developed a stochastic model for the propagation of delays in a Dutch passenger railway station. Based on historic train running data, a delay function was determined which estimates the knock-on delays of trains caused by route conflicts and late transfer connections.

Wendler (2007) described an approach to predict the scheduled waiting times using a semi-markovian queuing model. He continued thus the works initiated by Schwanhäußer (1974). Flier et al. (2009) analysed delay data of the Swiss rail network in order to find dependencies between trains and track occupations which have a significant impact on daily operations. They present regression models which allow detecting two of the most important types of dependencies, those due to resource conflicts and those due to maintained connections. The focus was also here on passenger transport.

Berger et al. (2011) propose a stochastic graph model for delay propagation and the calculation of arrival and departure time distributions applicable to public transport. The initial experiments were carried out with simple artificial distributions for the travel time deviations. An update on the approach is given by Lemnian et al. (2014).

All the publications illustrated here strive for the prediction of train arrival times in order to manage the trains on the network in such a way that an optimal network utilisation is established. For the coordination of a freight hub, this is of course only marginally relevant. Only location information for trains coming towards the hub are necessary. The subsequently developed approaches have primarily the objective of serving the hub manager for the optimisation of the hub-internal production planning.

Another drawback of the existing approaches is that they all largely depend on external data, e.g. those of the infrastructure managers. Undoubtedly, these data are mostly comprehensive and easy to analyse but, as stated above, not simply available for everybody. For this reason, rail freight nodes have an interest in independent methods for determining train arrival times in daily business which facilitate direct communication with customers and railway undertakings.

6.1.2 Independent methods for the estimation of arrival times

In the following subsections, alternative methods for estimating the arrival times of freight trains in freight terminals will be introduced. The methods rely on input data independent from the infrastructure manager. The input data must thus be available from public sources or owned by the railway undertakings or terminal operators. Potential for the reasonable independent ETA have the following methodologies:

(a) Estimating arrival delays from the distance travelled

(b) Estimation of delays from public train delay data

(c) Heuristics for estimating delays from positioning data

(a) Estimating arrival delays from the distance travelled

A simple but basic way of estimating arrival times can be achieved by analysing the relation between the travelled distance and the delay of incoming trains. More specifically, the travel distance to the destination freight node is to be compared with the deviation of the arrival time from the timetable. In the present study, a correlation analysis and a regression analysis of arrival time record samples lead to the following conclusions: The likelihood of an arrival delay increases with the train's travel distance but only very slightly as Figure 6.1 shows. The left graph represents the relation between the average arrival delay and the travel distance. A linear regression indicates that delays are likely to increase with the travel distance from the origin. However, the significance of the regression model is low. The R-squared, the statistical measure that describes how well the data fit the regression line, is $R^2 = 0.06125$, (adjusted R-squared: $R^2_{adj} = 0.0237$). On the right side of the picture can be seen that the standard deviation of the average arrival delay rises with the distance, too. Thus, the greater the distance between origin and destination, the more difficult it is

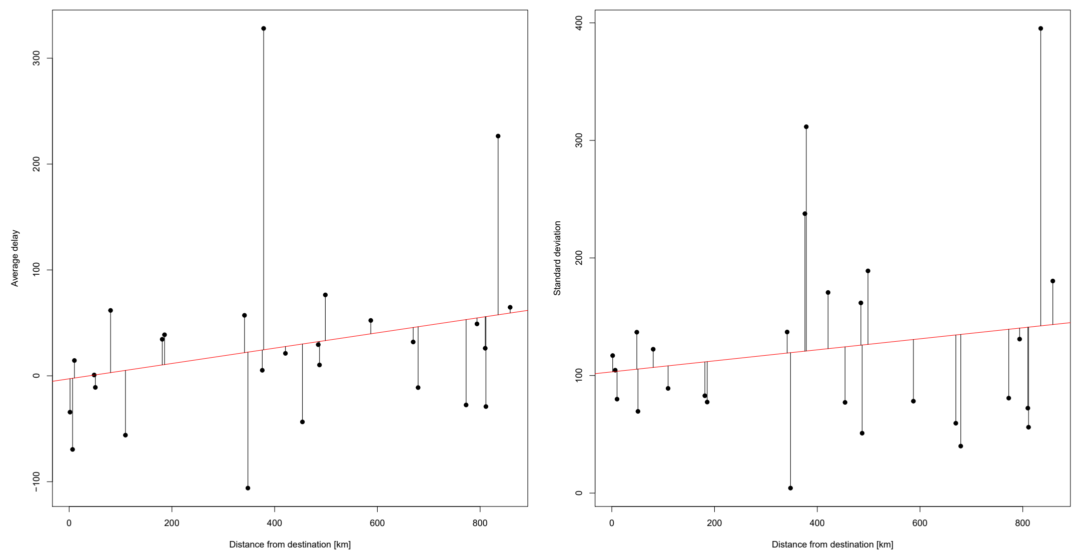


Figure 6.1: Estimated arrival time depending on the train origin. Left: Average arrival delay depending on the distance from the departure station. Right: Standard deviation of the arrival delay values.

to predict the arrival time deviation. Generally, the standard deviations differ widely among the nodes in the sample data. The nodes having the largest average delays have consequently also the largest variances in the data. It is thus possible to identify nodes, from which regularly delayed trains emerge, whereas those coming from other origins are more often punctual. One might conclude that trains coming from certain nodes, are subject to a certain higher risk of delay.

With an increasing distance between source and destination station, the predictions regarding the delay probability are therefore to be judged more and more critical. Accordingly, the regression model can give a rough orientation on the arrival time deviations of the trains but cannot give general evidence for all arrival deviations. However, specific origin nodes producing notorious arrival time deviations can be identified and trains coming from those nodes can be considered in the slot plans accordingly.

(b) Estimation of delays from public train delay data

A more complex method for the estimation of arrival times is possible by reproducing and simulating relevant train traffic on the feeder routes of a freight node in order to deduce general conclusions on the likelihood of delays thereof. Benning and Zimmermann (2014) developed a combined simulation and stochastic method based on an initial work of Emde (2013). In this work, the train traffic on the example relation between Berlin and Hamburg has been mapped in a simulation environment. Following the hierarchical planning approach, the passenger trains of a certain period were planned first. Freight train paths were added in the free spaces thereupon. Delays of passenger trains were incorporated with the help of historical delay data available from the Zugfinder website (Schubert, 2015)[25]. An automatic dispatching decision support system considers the delay probabilities of the passenger trains and assigns new train paths to the freight trains. Accordingly, the delays of the trains and their estimated arrival times could be calculated.

The work followed initially a simulative approach, although, it does not rely on undisclosed data (as for example in Törnquist and Davidsson, 2002). Key result is the determination of delay probabilities for the nodes along the rail line and the automated rearrangement of train paths based on commonly used dispatching rules. By comparing delays of passenger and freight trains, conclusions on the interaction of the two train types were drawn. These in turn allow the forecast of how freight trains will behave in similar operational situations and what delays they will probably have at their destination. The results can be depicted as probability distribution functions. Figure 6.2 shows an exemplary accumulated probability distribution that a freight train suffers a delay on the regarded railway line.

[25] Zugfinder.de is a website that provides statistics about train punctualities in Germany.

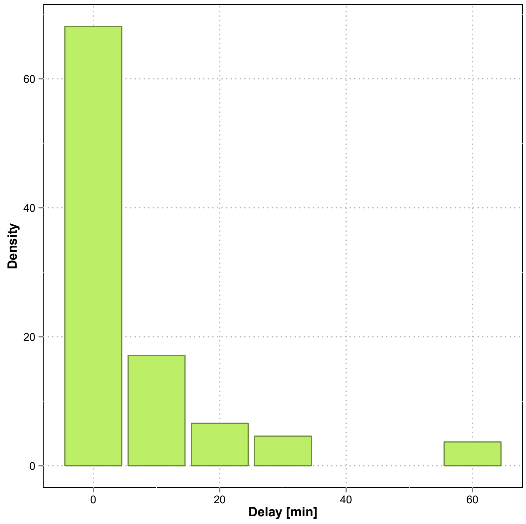


Figure 6.2: Histogram of the probability distribution that a freight train suffered a delay on the observed railway line (based on Benning and Zimmermann, 2014).

The method allows the identification of recurring delay patterns in the traffic data and the derivation of delay forecasts from the succession of trains. By comparing the current traffic with similar events in the historic data, predictions on the course of the actual operational situation can be made. Benning and Zimmermann (2014) show an example of how the probability of a delay of a freight train can be calculated based on the current delay situation of other trains travelling ahead on the same track. The forecast results of one train can be stated as summarised in the following table:

Delay of the preceding train	**Delay of the regarded freight train**	**Probability of occurrence**
< 141 s	no influence	
141 - 626 s	Growing with passenger train delay $0 < t_n^{del} < 1,476\ s$	90%
633 - 2,198 s	Delay of 1,476 s	92%
> 2,198 s	no forecast possible	

In this example it was planned according to the timetable that a passenger train leaves a station followed shortly by a freight train. With a small delay of the passenger train, this results in no knock-on delay. However, if the passenger suffers a greater delay, the freight train is first allowed to pull out. In the further course, the passenger train will, due to its higher speed, catch up with the freight train. This in turn is ordered to wait in a passing loop to allow the passenger train to pass. From a sufficient amount of historical data with the same scheduling patterns it is thus possible th calculate the probabilities for delays.

(c) Estimation of delays from positioning data

The most advanced method for estimating the train arrival times in the context of this thesis is based on GPS data. In the research project *ViWaS* (Schönemann et al., 2015), co-funded by the European commission, a set of GPS data records was analysed in terms of arrival forecasting. The data were available in a comma-separated values (CSV) file which contained location and time information collected by a GPS module mounted on a rail freight wagon. It recorded the wagon movements within a period of about two months. At the time of testing, the algorithms of the GPS module were still under development, so that it came partially to incomplete data records. In particular, it is difficult to determine the exact start of a journey from the given data since it was the recognition of movements which was initially inadequately configured. For the purpose of arrival time estimation this is a minor drawback when the trains journey is long enough. Fortunately, the freight wagon was in service on an almost 700 km long relation. Thus, a set of long-running journeys was recorded. The details of the data analysis are as follows:

With the detection methods currently available, the GPS records of a single train trip enable no sufficiently accurate arrival forecasts, since the detection of position data occurs too roughly (e.g. once every hour). This interval is too infrequent, in order to generate reliable traffic flow information as they are for example known from road traffic analysis. However, multiple repetitive recordings of the same journey delivered already some valuable results for a traffic flow analysis. Figure 6.3 shows several courses of the train on the defined route. The black dots define all places wherever a position has been recorded over the set of all records. The course of the railway line can thus be identified quite well. In contrast, the three coloured lines represent the records of three trips. They indicate clearly that the position of the recording intervals are quite coarse and rather randomly. Hence, from an individual train journey neither the exact route nor a possible arrival time can be determined because the time and location of a GPS measuring event is unpredictable in the first place. Thus it can be treated as a stochastic process.

More information about the course of the trains can be obtained by considering multiple recorded journeys in a time-distance diagram. It produces the image shown in Figure 6.4. Obviously, many journeys take place without major delays. They accu-

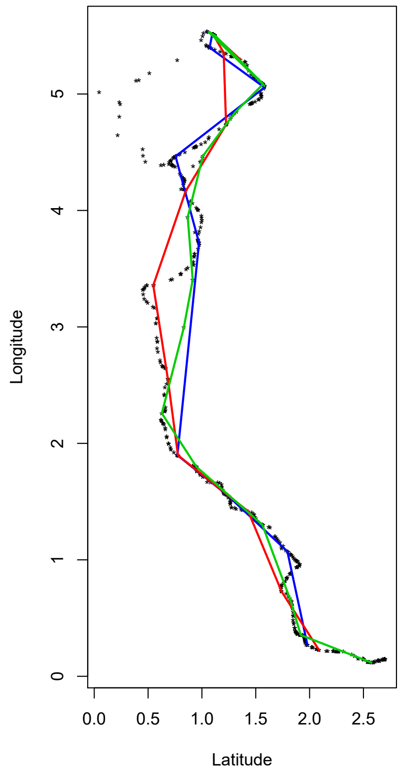


Figure 6.3: GPS records of three trains running on the same route. With current methods, the measurement points are very rough and rather random.

mulate at the bottom of the graph. However, there seems to be a large scatter. Some extreme delays can be noticed in various locations. Particularly striking is that a number of trains interrupt the journey shortly before reaching the destination node. The delay phenomena were examined more detailed within a correlation and regression analysis.

The regression analysis examines the relationship between the distance and remaining driving time to the final destination. The regression is with $R^2 = 0.844$ quite well explained. Also a correlation analysis explained a strong positive coherence between the arrival time and the distance to the destination node. The test data delivered a correlation coefficient of $r = 0.919$ with a 95% confidence interval $0.893 < r_{0.95} < 0.939$. Figure 6.5 shows the 95% confidence band of the analysed values. The confidence band indicates that, with a probability of 95%, the train arrival at the correspondent cross-section lies within the time space specified by the band. The

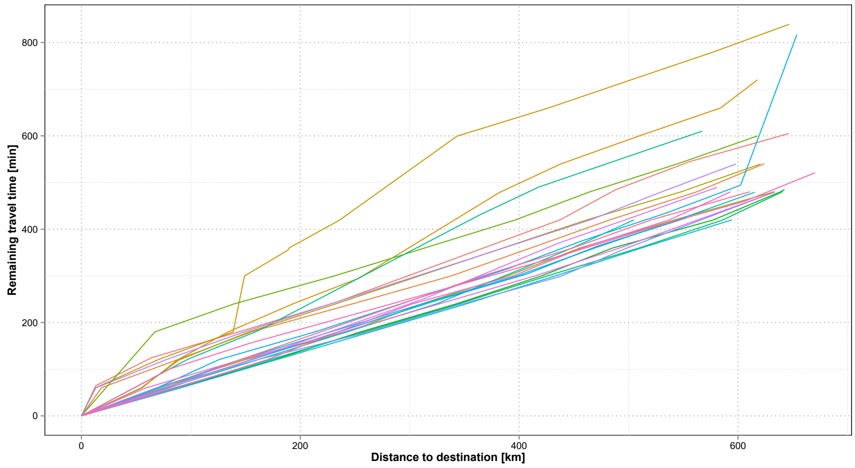


Figure 6.4: Path-time diagram of the trains travelling the considered route.

95% confidence band is quite narrow, but not constant. It increases at the rear (greater distance to destination) but also slightly in the area near the destination. It follows logically that with larger distances to the destination the arrival time is subject to greater variations. But, interestingly also with short distances (last mile) the arrival uncertainty rises again. This phenomenon, however, may derive from a systematic error as a result from the hub production system: Incoming trains might be held in a preliminary queue, i.e. they wait in a freight yard prior to the final destination node before they can enter it.

Delays induced by the network nodes on the train journeys could be investigated in an Analysis of Variance (ANOVA). The ANOVA is used to test the differences between the averages of three or more samples for significance. In the present case study, different subsets of the GPS data records are to be understood as samples. The present data were divided into distance classes with a width of 10 km. This has consequently resulted in the definition of 68 clusters with corresponding members of measurement points (location x time). The null hypothesis of the ANOVA is

$$H_0 = \mu_1 + \mu_2 + \cdots + \mu_{68} \tag{6.1}$$

that the mean values of the groups are identical and that no differences in the arrival probability exist depending on the distance. To illustrate the results, the box plots in Figure 6.6 were prepared.

It is apparent that the variance of the remaining travel time tends to decrease with the distance to the destination. Therefrom one can conclude that the time of arrival

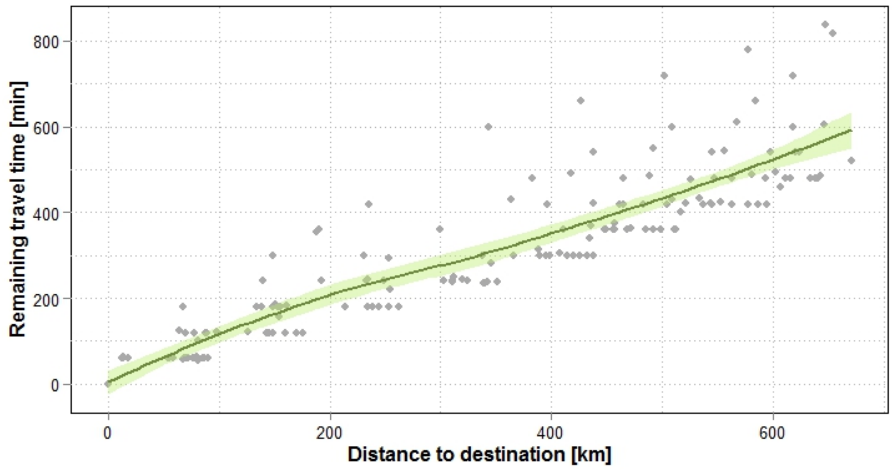


Figure 6.5: Regression and 95% confidence intervals of the GPS data points

can be predicted more accurately, the closer the train comes to his destination yard. Notably, the measurement points have very different statistical spreads. The box plots with large variances indicate position measuring events which were recorded in railway nodes and thus denote possible siding or waiting processes. The box plots with smaller variances indicate in contrast position measuring events on the open track (moving train). From the position records, which were taken while the wagon dwellt in a node, consequently a greater likelihood of arrival time deviation can be assumed.

Unfortunately, it appears that the amount of available data for the ANOVA is still insufficient. With a larger sample of data records, more accurate statements on the delay probability per route and node could be created. However, the here presented first approach is a promising start. To improve the forecasts, it is advisable for future research to work with a self-learning knowledge-based system. Thus, the arrival time predictions can become more accurate with an increasing data basis.

Another contribution to improve the data base is to reduce the randomness of the data points. In the used GPS module, the measuring events are time-related, i.e. one measuring event is recorded every time unit (e.g. one hour). Better results would be expected if the measurements would be location-dependent. This is, however, hardly possible with today's battery-powered GPS modules, as it requires a permanent position monitoring which would reduce the battery life dramatically. A permanent position monitoring is today only imaginable with a permanent power supply which again requires the mounting of GPS modules on locomotives or on electrified freight

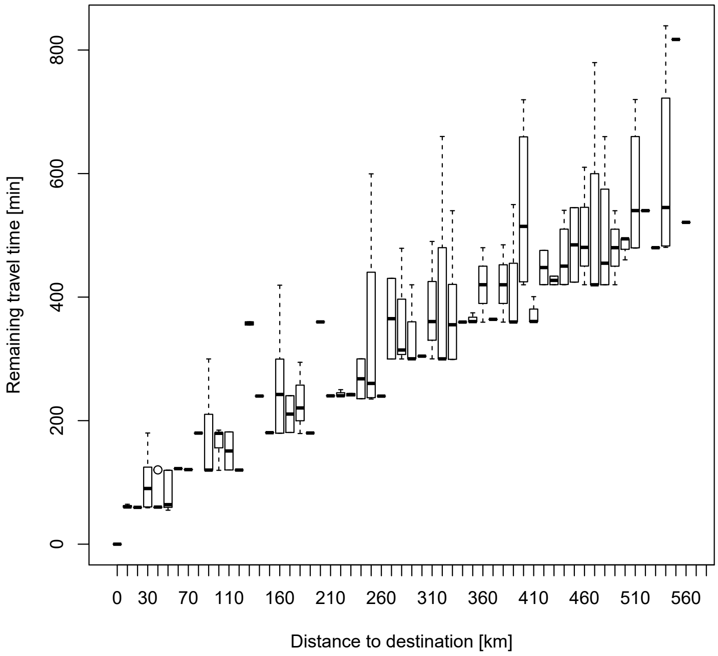


Figure 6.6: Boxplots of the ANOVA

wagons. Since it is not expected that freight wagons receive a pan-European compatible electrical connection in the near future, the presented approximate solution for ETA calculation from GPS data remains as the best independent solution.

6.1.3 Summary on the prediction of train arrival times

It was identified that the ETA is one of the most important input parameters for a flexible slot management system and the efficient rescheduling of production processes in rail freight nodes. In daily practice, however, the information and dissemination of the real arrival times is not always given, so that freight terminals cannot plan their internal processes efficiently. This section proposed appropriate methods for estimating the arrival times of trains approaching a freight node. In particular, emphasis was placed on forecasting methods that make use of data from independent sources.

The first, very basic, approach aimed at estimating the arrival delays from a train's origin. Using a regression model, a relation between the travel distance and the arrival delay could be verified, although it was quiet low. It could, however, be noted that certain source nodes can be identified as notorious delay originators. This can therefore be taken into account as important information in the slot planning.

The second approach, using a combined simulation-stochastic model, estimated freight train delays from public available train delay data. From historical train traffic data and the identification of typical delay patterns, conclusions about the likelihood of freight train delays could be drawn. Advantage of this method is the active consideration of the current traffic situation and the possibility to derive appropriate dispatching measures thereof.

The third method showed how tracking data from GPS modules mounted on freight wagons can be used for the estimation of arrival times. Data of a recurring train journey were evaluated in an analysis of variance (ANOVA). A combination of a regression and a stochastic model was chosen in order to obtain the desired conclusions. In the results, the estimated arrival times could be illustrated in a 95% confidence band.

All three approaches use input data that are independent from those available from the infrastructure manager. This causes, of course, a degree of uncertainty in arrival prediction quality which cannot be avoided due to the fact that the available data describe the traffic flows only incomplete. The best results for arrival time estimation can surely be obtained by combining the approaches (b) and (c). The second approach provided the best data base and its biggest benefit is that it considers interactions with other moving (passenger) trains on the network. The third approach contributes the most accurate information on freight train positions. Both systems together may picture the situation of rail freight traffic the most accurate and predict suitable dispatching measures as well as arrival forecasts of freight trains. The affected node can respond with appropriate rescheduling policies and adjust its internal processes to the situation, so that its performance is maintained on a high level.

6.2 Economic issues and pricing schemes

In addition to the technical framework of a successful slot management system, incentives and financial regulations are important control mechanisms to ensure that the railway undertakings respect the specified slot strategy and organise the arrivals and departures of their trains accordingly. Various elements of pricing schemes can affect the occupation time of the infrastructures. Accordingly, compliance with the slot rules can be stimulated through the price system that the freight node is using.

The issue is examined in this section from two angles – the yard manager's and the railway undertaking's perspective.

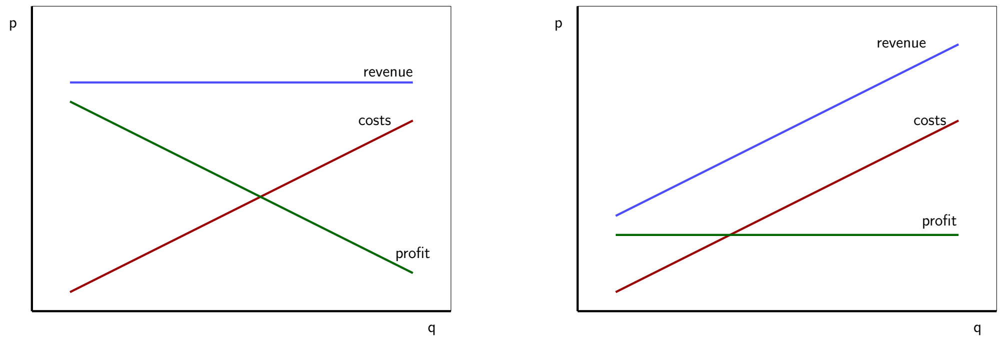
In addition to the design of the pricing schemes, the problem of planning uncertainty, which the railway undertakings are facing, will be considered from an economic perspective. It will be discussed, how the pricing schemes and further regulations can support the reduction of dwell time uncertainties.

6.2.1 The yard manager's perspective

In marshalling yards, the fees for the storage of wagons and trains are usually time-based. A fixed price per time unit (unit price) is to be paid by the railway undertaking. It depends on only a few criteria, such as the track category. It is possible that the pricing system contains an incentive-based component with monotonously decreasing time-dependent costs. This allows a kind of discount on longer sojourn periods (for example in Wustermark, 2015).

The maximum revenue of such time-based pricing schemes with unit prices is limited. Considering the situation when the yard is maximally occupied, i.e. all available rail tracks are booked all the time, the earnings for the yard manager are maximal (ignoring for simplicity a possible quantity discount). In this case it is irrelevant how many trains per time unit (e.g. per day) actually pass through the yard. If all available tracks are maximally occupied, the revenue will always remain on the constant maximum level. Accordingly, it is appealing to a yard manager to attract clients which stay longer. Fewer trains reduce the coordination efforts and increase its profit. This argument is illustrated in Figure 6.7a. If the number of trains handled per time unit increases (q is rising), the expenses increase (e.g. due to increased dispatching efforts) and therefore the costs for the yard manager. From the perspective of the yard manager, an increase of train handlings per time unit leads to no advantage because with uniform prices he does not receive a financial benefit for the increasing efforts.

This behaviour, however, does not meet the desired goal of increasing the efficiency of complex freight nodes and is hardly compatible with the interests of the adjacent cargo terminals and the operators of sidings. An incentive for the acceptance of more traffic per time unit would be achievable for example through price discrimination approaches which would gain higher profits for the yard manager for each additional train. In the long term, an effective pricing system must produce at least a profit function as illustrated in Figure 6.7b. With an increasing number of trains handled per time unit (q increases), the revenue increases, too. In the given example, the



(a) Constant unit-based pricing scheme **(b)** Quantity-dependent pricing scheme

Figure 6.7: Left: Constant unit-based pricing schemes gain limited revenues. Costs increase with the number of handled trains and therefore it is attractive for the yard manager to reduce the number of handled trains q per time unit. Right: A pricing scheme with price discrimination. The rail yard manager's profit does not decrease when the number of handled trains increases.

manager's profit is constant which means that he is at least indifferent to the number of handled trains.

More constructive for increasing the efficiency and the throughput rate are pricing schemes that generate increased profits for the yard manager with the increasing number of trains. Such pricing systems are usually composed of two stages: (a) a basic price covering the yard manager's unit costs for a train which allows the sojourn of the wagons for a defined time and (b) time-based penalties for exceeding the intended maximum length of stay. With regard to the slot management approach, the allowed sojourn time would correspond to the duration of one slot. Such or similar pricing systems have emerged in highly utilised rail freight nodes, such as sea ports (see for example Freie Hansestadt Bremen, 2015; Hamburg Port Authority, 2015).

The application of a new pricing system does not immediately incur that railway undertakings always obey the slot management specifications. Of course, they are aiming to avoid extra costs for the slot exceedance. However, one has to keep in mind that the railway undertakings are still depended on the decision making of other stakeholders in the entire logistics chain and the resulting process uncertainty. One will quickly discover that the railway yard and the cargo terminals should be regarded as complementary goods and therefore subject to special market mechanisms. From the perspective of the railway undertakings, this results in some peculiarities.

6.2.2 The railway undertaking's perspective

From the perspective of the railway undertakings, the cargo terminal and the rail yard are complementary elements, which are to be regarded as a common functional unit. The terminal will not work without the functions of the marshalling yard and vice versa. For a short analysis it may be considered a cost function for the use of a slot in the rail formation yard of the type

$$c(t) = \begin{cases} c_{slot_n} & 0 \leq t_{dwell} \leq t_{slot_n}^{end} \\ c_{slot_n} + yt & t_{dwell} > t_{slot_n}^{end} \end{cases} \tag{6.2}$$

This cost function represents a typical two-part slot tariff as introduced above. It is illustrated in Figure 6.8. For a booked slot, a lump sum c_{slot_n} will be due. When

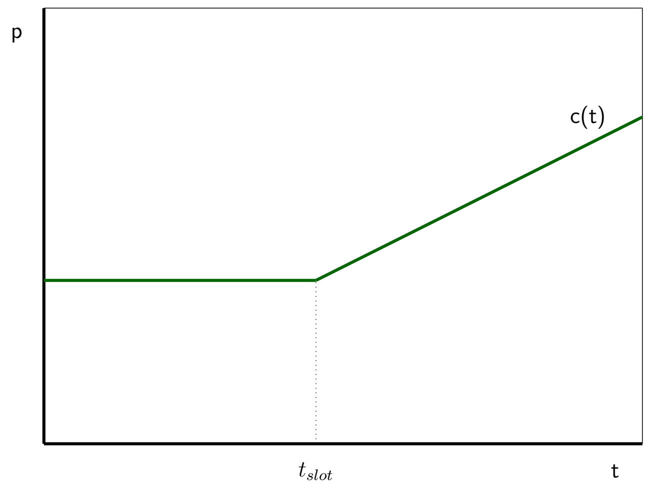


Figure 6.8: Two-stage linear cost function $c(t)$.

exceeding the time slot, a time-dependent extra charge of yt is to be paid. For the use of the freight terminal infrastructures, generally no time-depended charges incur (see e.g. Hamburger Hafen und Logistik Aktiengesellschaft, 2014). The terminal fees comprise the costs of the transshipment of freight units. They do not change with the length of stay and are therefore not crucial. Nevertheless, missing a terminal slot can lead to increased expenses by re-registering for new slots. Thus, administration costs for the railway undertaking incur and because of delays, penalties for non-performance might to be paid by the railway undertaking to its customers. The overall costs for the use of the railway yard and the freight terminal thus vary depending on the punctuality of the incoming train. In the case of a punctual train arrival, the following overall costs C_p incur to the railway company:

$$C_p = c_{slot}^{arr} + c_{terminal} + c_{slot}^{dep} \tag{6.3}$$

The railway undertaking has to pay the costs of the marshalling yard for the processes of the incoming c_{slot}^{arr} and the outgoing train handling c_{slot}^{dep} as well as the administrative costs of the terminal slot $c_{terminal}$. With a punctual arrival, no penalties incur. Of course, further costs, e.g. for the shunting service arise. However, these can be considered as constant and are therefore not relevant for the calculation.

In the case of an early train arrival, the following overall costs C_e arise to the railway company:

$$C_e = (c_{slot}^{arr} + yt) + c_{terminal} + c_{slot}^{dep} \tag{6.4}$$

Due to the longer dwell time of the arriving train, additional time-based costs yt in the rail yard arise. The other costs will remain the same.

In the case of a late train arrival, the following overall costs C_l incur to the railway company:

$$C_l = c_{slot}^{arr}[+yt] + 2 \cdot c_{terminal} + c_{slot}^{dep}[+yt][+ \text{ additional track access fees}] \tag{6.5}$$

If the terminal slot was missed, a second one must be applied for. When the new terminal slot can not be assigned immediately (due to a scheduling conflict), this may again lead to longer waiting periods for the arriving and the departing train. If the departing train can not reach his exit route into the public rail network on time, this can possibly result in the payment of additional track access charges. A late arriving train would consequently cause the highest, whereas a punctually arriving train causes the lowest overall handling costs. The costs depending on the arrival punctuality can thus be ordinally ranked:

$$C_p < C_e < C_l \tag{6.6}$$

The decision of the railway undertaking, which cost strategy should be chosen, depends now not only on the handling costs itself. The compliance of the railway undertaking with the promised service quality to its customers must be secured, too. From Chapter 4 it is known that in badly coordinated freight nodes it can regularly come to process time overruns. The railway companies are in those cases exposed to external risks which they cannot influence.

Undoubtedly, the risk of a schedule disruption r of the operations at the node is the highest for trains which arrive late (r_l). If a freight train reaches its destination punctually, the risk of a disruption r_p during the yard processing is to be expected intuitively as lower. The minimum possible risk of disruptions is achieved by implementing additional (unnecessary) buffer times which is equivalent to risk of early arrivals r_e. The risks depending on the train arrival can therefore be ordinally ordered as:

$$r_e < r_p < r_l \tag{6.7}$$

An early arrival causes the least risk of a disruption. It incurs higher but predictable slot costs. The decision, at what costs a freight node is processed, thus depends on the overall risk situation. The sole introduction of penalties for extended stays will not necessarily make the railway undertaking to shorten the train dwell times. With the calculable risk, penalty payments to customers, annoyed customers, etc., which might result in higher costs than the slot penalties yt, are prevented. This risk-minimising behaviour, however, results in a less efficient transshipment system.

6.2.3 Slot management as an enabler to reduce uncertainties

One aim of the slot management approach is that railway undertakings shall follow the defined slot rules, i.e., by planning shorter turnaround times for their processes in the freight node. To make this possible, it must be ensured that the operational risks, which the railway undertakings are exposed to, will be abolished. This applies particularly to

(i) the uncertainty of a loss of terminal slots and

(ii) the uncertainty to miss the track access slot when leaving to the rail network.

The first uncertainty (i) has to be covered by the consortium of railway yard and cargo terminal by not charging any penalties or slot losses, if the railway undertaking transmits the estimated time of arrival duly. An incentive to cooperate in the ETA information transmission can thus be accomplished with the guaranteed preservation of the scheduled time slot. As seen in chapter 5, the yard has manager has the ability to reorganise the slot plans of delayed trains quickly without prolonging their dwell time by using efficient slot management and rescheduling strategies.

The scheduling of the exit routes of leaving trains to the national rail network (ii) is basically a bilateral matter between the rail network manager and the railway undertakings which is particularly volatile. If shorter turnaround times within the freight node shall be achieved, it must be ensured that the outgoing trains can meet their already booked exit slots, also and especially when trains arrived late. With the continuous slot management in the freight node, it becomes possible to determine a train's expected departure time already with his arrival and an arrival delay can accordingly be considered. Any new calculated departure time can be reported to the rail infrastructure manager in good time in order to negotiate an appropriate exit route. The continuous compliance with the slots provides space for the reduction of unnecessary buffer times.

In the long term, the better linking of slots in the freight node with exit routes of the outgoing trains is a task to be worked on in co-operation with the rail infrastructure managers. Particularly, a stronger process-related linkage between the incoming and the outgoing trains has a potential to reduce buffer times. The present thesis goes already towards this ambitious objective by having established the slot management approach. Further process integration requires, however, further research which is beyond the scope of this thesis.

6.2.4 Further pricing elements influencing the hub efficiency

Besides the development of a general slot pricing scheme, there are further mechanisms which have an influence on regulating the slot demand from an economic perspective. For the sake of completeness, some important pricing elements are reviewed in this section. Reference is made to the relevant existing literature. There are two major topics addressed, peak pricing and complementary prices:

Peak pricing: In the analytical part of this thesis it has been demonstrated that the best utilisation rate for railway nodes and especially for transshipment facilities can be achieved if trains arrive and leave continuously distributed over time. The same is also true for other modes of transport. For road, for example, recent scheduling software is emerging that controls the flow of trucks into a freight node (e.g. Kouloumbis, 2015) and thus schedules the transshipment slot allocation. To achieve the same for European rail freight transport, is, however, often difficult. Many freight trains travel overnight, leaving the originating station in the evening and arriving at the destination station in the morning. Many manufacturing companies arrange their production processes accordingly. The daily production will be sent in the evening by train to customers. Likewise, goods and raw materials are available in the morning with the start of production. For the railways, this mode of production has also a distinct advantage – the route conflicts with the passenger trains travelling during daytime can be largely reduced.

For the processes in logistics nodes, this production system has significant drawbacks. In the morning and the evening it leads to an overload of the available infrastructure resources (peaks). It might even cause a node to be expanded in terms of infrastructure for the maximum peak, which is required then only a few hours per day.

By assigning time slots to trains, the freight node can control the customer's demand according to the capacity in the node and thus control the inflow properly. In addition to the slot management, there is also the possibility of increasing access charges during the high demand hours until the optimal level of congestion is reached (peak-pricing, congestion-pricing). An alternative is the already mentioned price discrimination – in slot shortage the access costs are rising. In the field of airport management, the economic impact of such regulatory instruments on hubs has already been discussed. They shall be examined here for their transferability to rail freight nodes.

Following Czerny (2010), in theory, slot trading and congestion pricing as regulation instruments can both lead to an optimal welfare result. Applying his findings to rail freight transport requires, however, that hub authorities possess perfect information about the benefits and costs of the railway undertakings, which is not true in reality. Another study that deals with the comparison of regulatory mechanisms for infrastructure capacities comes from Basso and Zhang (2010). The authors compared the welfare of slot-allocation mechanisms (slot trading, slot auctioning) and congestion pricing. In both publications the scientists pursue the question what is the right choice of regulation instruments under the current specifications which includes uncertainty about customer benefits? Czerny (2010) provided evidence that under certain circumstances both instruments can be beneficial at congested airports. It is, however, not clear yet, how the proposition can be applied to rail freight transport because rail freight demand is not comparable in flexibility to the aviation industry. Particularly Basso and Zhang (2010), for example, assumed in their study that demand is perfectly elastic. Rail freight transport demand, however, is known as being rather inelastic (cf. Andersson and Elger, 2012). In this respect, the transfer of the results to rail freight transport is an interesting question for future economic research.

Complementary prices: In section 6.2.2 it has already been mentioned that a logistics hub is to be understood as a structure of complementary elements from an economic point of view. Unlike in aviation, where only one infrastructure manager, the airport, is present, in a land transportation hub often more than one infrastructure enterprise is involved (e.g. the rail yard and one or several freight terminals). Railway undertakings consider the costs of calling at a hub as a whole and thus the costs that arise for using the facilities of both enterprises depend on each other.

If the rail yard and the freight terminal are complements, the critical feature is that the demand for each of the two units depends on the prices of both. When forming its slot charges, the railway yard must therefore notice the price formation of the freight terminals, too. As it is very well described e.g. by Varian (2010), an increase of the prices of one of the complementing enterprises would reduce the demand on both enterprises.

Moreover, it appears generally that a merger of two companies offering complementary products, results in lower prices, compared with an independent determination of prices by both companies. Varian (2010) refers this statement on two complementary monopolies. It is not wholly applicable to the area of rail freight because also freight hubs are competing with each other. However, significant hubs, e.g. major sea ports, can often be considered as regional monopolies. In the monopoly case, an integrated company, which demands a total price for both, the rail services and the cargo handling, would appear more reasonable from an economic perspective.

6.2.5 Summary on economic matters and pricing schemes

The design of the pricing strategies in rail freight yards has a certain impact on the feasibility of slot management strategies. Conventional time-based systems that possibly grant quantity discounts, are not effective for the enforcement of time slots. It was analysed how slot pricing systems can be designed in order to foster shorter train dwell times. It has been stressed that the pure introduction of monetary penalties for the exceedance of slot times will not necessarily make the railway undertakings to obey the slot rules. Rather, the risks that arise from disruptions of the process chain must be reduced. Such disruptions, which can incur further delays, are for example the loss of the terminal slot or the loss of the track access of the departing train. Today, railway undertakings avoid risks in freight nodes by planning rather an early than a punctual arrival. They accept higher costs for extended dwell times than risking a process disruption.

The issue is reminiscent of the chicken or the egg dilemma: Trains have a long dwell time with large buffers, because the handling in the freight node is too opaque and often takes longer than necessary. Due to the long dwell times, trains occupy infrastructure unnecessarily long, which results in occupancy conflicts and increased coordination efforts for the yard manager. Due to the higher conflict potential, it then comes to longer processing times. The long unnecessary dwells can thus be reduced only if the risk of disruptions for the railway undertakings are reduced, too. To solve the problem, an appropriate hub-wide processing time must be promised as well as a guarantee towards the railway undertakings that the trains are ready for departure at the scheduled time. The slot management and an effective rescheduling mechanisms developed in chapter 5 foster this approach and should thus be supported by appropriate pricing schemes.

In addition to the general pricing schemes, other pricing elements, which have an impact on the efficiency of railway systems in complex freight nodes, were examined. It has been questioned, how fluctuations in demand and peaks in traffic volumes can be controlled over prices. In other transport sectors (aviation and road freight), related approaches exist. Whether and how they can be applied to rail freight transport, is not fully clear, yet. Since a further investigation would go beyond the scope and context of this thesis, this is an interesting area for further research.

Finally, it was recognised that railway yard and cargo terminals should be regarded as economic complements whose efficiency is influenced mutually by the respective pricing strategies of both entities. Moreover, the section raised questions for future research on how to regulate congested rail freight nodes.

7 Conclusions and closing remarks

The present thesis examined the experimental implementation of a slot management scheduling method as a measure for the process flow optimisation in complex rail freight hubs. The approach seeks to unite the interests of rail operations and logistics in order to achieve a better node efficiency and to accelerate train and cargo throughput. This chapter summarises the key findings of the study and compares them with the research questions defined in chapter 1. Besides, the used methodology is critically examined. Outstanding issues, which offer opportunities for further research, are being discussed at the end.

7.1 Resume of the main achievements

The thesis was motivated by the prevailing but unsatisfying production system of rail freight yards which is to a large extent based on spontaneous action and improvisation. As a result, "the rail yards are the major sources of delay in the rail freight networks" (Marinov and Viegas, 2011b). By studying the processes in the railway premises of several European freight hubs, current problems were identified which led to the definition of the main research objective and three associated research questions. The main research objective was formulated as: "Develop a combined simulation and optimisation model that identifies sophisticated production methods for rail freight yards which lead to a better operational performance at complex logistics hubs and which have a positive effect on the reliability of the rail freight operations in such hubs." To achieve this research objective, the three research questions are answered; primarily in the chapters 4 and 5. Before they could be addressed, the essential characteristics of rail freight operations in Europe have been discussed in chapter 2. The infrastructural design and the exchange operations in complex freight nodes were illustrated, based on the results of a detailed process analysis. The main finding of chapter 2 is that the tactical infrastructure utilisation planning in such nodes is currently poorly coordinated. The quality of process coordination depends on the optimisation strategies of the individual involved actors. Due to the absence of a superior coordinator it is expected that there is a significant potential in the short-term planning and in the dispatching strategies for increasing the efficiency of such nodes. The process analysis thus emphasises the relevance of the research questions.

Chapter 3 introduced slot management as a planning mechanism from a scientific perspective and gave examples of successful implementations. It was identified that the slot management processes of airport ground operations can be compared as the closest with those in rail freight yards. The chapter provides the foundation for the development of an appropriate slot management approach. After discussing the fundamentals, the next chapters are concerned with answering the three research questions:

Q1: How can an improved process synchronisation increase the yard performance and reduce unproductive times?

The answer to the first research question is initiated in chapter 4 in which the processes in rail freight yards were examined by means of a queueing simulation. The required input parameters have been obtained from observations in rail freight nodes and were implemented as generalised probability distributions to the simulation. In a number of simulation scenarios it became clear what influences on the vehicle dwell time derive from various factors. The analysis of the rail freight terminal productivity has shown that the utilisation of available infrastructures under realistic conditions is very low due to low synchronisation between several processes. The simulation model made the relationships between the individual processes apparent so that the requirements on a successful process synchronisation could be defined.

Chapter 5 formed the core of the slot management approach and delivers the concluding answer to question Q1. It was shown how slot management strategies can be applied to rail freight terminals. Further, it could be pointed out that they contribute to the acceleration of terminal processes and thus increase the productivity of hubs. The establishment of a slot management approach contributes accordingly to the process synchronisation, improves the performance and reduces unproductive dwell times.

Q2: How can freight nodes flexibly respond to stochastic (random) external conditions, e.g. arrival time deviations? What is the effect of deviations on the yard performance?

With the introduction of the slot rescheduling approach it has been shown how a freight node can respond to external arrival deviations which have been identified as the most relevant source of disturbance and delay. The developed slot rescheduling increases the degree of flexibility on the yard operations remarkably without lowering their reliability. With the analysis of three different rescheduling strategies, their influence on the yard performance was studied. It was identified that the rescheduling of non-punctually arriving trains can reduce the overall delays and thus can support the stabilisation of vehicle schedules. Additionally, it was possible to reschedule the slots of unpunctual trains in such a way that the impact on other trains was kept minimal. The slot rescheduling approach answers extensively research question Q2.

Q3: What effects does a higher degree of organisation in freight hubs have on the external railway network? How can it contribute to a more reliable work flow?

Chapter 5 answers also in depth research question Q3. A higher degree of organisation is achieved with slot planning and the slot rescheduling. This in return enables more stable operations in the freight node so that the train movements are better predictable. By assigning the trains to slots throughout all process steps in a freight node, the temporal course of the train dispatch can be determined more accurately. Thus, the time at which a train will be ready for departure is more predictable. This has a positive effect on the external rail network and supports the production system transparency for the involved stakeholders (railway undertakings, infrastructure managers, etc.).

Chapter 6 emphasised technical and economic recommendations in order to lead slot management in complex freight node to a success. It was clearly pointed out that the production quality of freight nodes will not simply increase with the introduction of new production methods such as slot management. Rather, the environment, i.e. the procedures of the other stakeholders must be adapted. On the technical side, especially the most accurate prediction of the arrival time will lead to successful slot management. Only then, slot management is able to contribute to the increase of hub efficiency. Several independent solutions were presented to calculate the ETA. From an economic perspective, some notions for the design of the price system, which fosters slot management, were discussed. It was emphasised how prices and regulations influence the operational stability of rail freight yards and how they contribute to an increased planning reliability for the railway undertakings so that these have an incentive to reduce the buffer times in their process organisation.

By successfully answering the research questions also the main research objective was fulfilled. In the combined simulation and optimisation model, the processes of freight handling in complex nodes were analysed, structured, and improved. The slot management approach was identified as a sophisticated production method which increased the performance of the freight hub and additionally increased its reliability and transparency towards other stakeholders in the transshipment process. With the selected slot regime, the average train dwell times could be reduced compared to the observed dwell times. The utilisation of the formation and the loading tracks was thereby also positively influenced. The slot system approach holds still room for further improvement, since it is flexibly changeable by modifying several degrees of freedom. It is, for example, possible to achieve shorter slot times in the formation tracks by employing more wagon technicians. However, when changing such parameters it must be assured that no new bottlenecks arise elsewhere in the system.

Finally, it is yet essential to integrate the slot management system into the general rail freight production process which was illustrated in section 3.4.1. The annual national timetable should serve as a basis for the development of an initial long-term slot plan. Transport announcements on short notice are to be integrated during the season. The final fixed allocation of a train's loading slot takes place when the train leaves the departing station. Simultaneously, it will be taken in an iterative control process, as depicted in Figure 7.1. In this process, the estimated time of arrival ETA is monitored continuously. Within a rolling horizon approach the next hours of yard

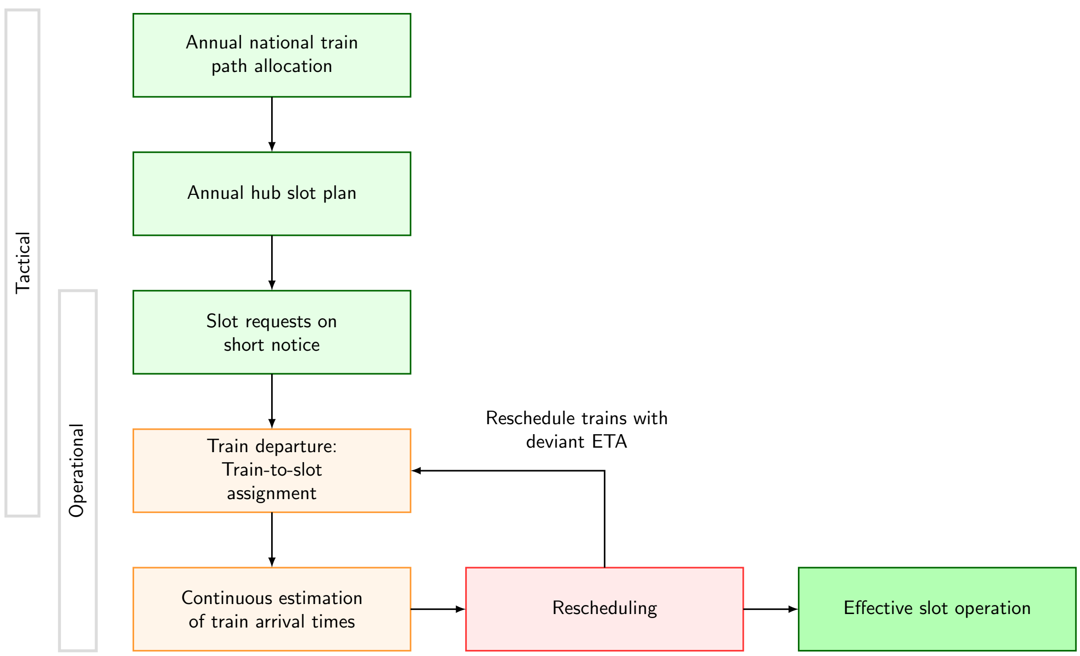


Figure 7.1: The integration of the slot management in the operational planning process

operations can be scheduled in advance. Periodically, for example every 2-3 hours, a rescheduling process is triggered which compares the current slot plan with the ETA of incoming trains and, if necessary, takes corrective action.

7.2 Concluding critics to the used model and methodologies

The combined simulation-optimisation model of a railway-based transshipment node is a simplified reproduction of the reality and therefore can not reflect all the specifics of freight hubs. It aims to present a general overview on the topic of slot management in complex freight hubs. When applied to a certain node, site-specific adjustments are necessary. In the model considered so far, its capacity utilisation was optimised. Some operational and regulatory issues were however abstracted or simplified. They should now be addressed for completeness.

A first issue concerns the engagement of staff in the model. In the optimisation it was assumed that the wagon technicians are always available. Working time arrangements, breaks, etc. were ignored at the moment. Much more important for the smooth workflow in practice, however, is the fact that wagon technicians usually are not provided by the yard manager. Rather, each railway undertaking has its own staff or appointed subcontractors. Herein lies another factor of uncertainty. Since the yard manager has very little influence on the staff scheduling, the compliance with the contracted slot times can not be fully ensured.

Another issue concerns the unified model train that was defined for this thesis. In the simulation, as well as in the optimisation phase, no distinction was made between the requirements of different railway undertakings. All trains and cargo units have been considered as equal. The unified model train is indeed representative for a large majority of transports. In practice, however, typically the few deviations from the regularity or special service requirements lead to disturbances in the process flow. As for this thesis existed no specific use case, a consideration of such interferences was not necessary, but should be regarded when in comes to implementation.

The arrival process undergoes a critical contemplation as well. It was assumed as uniform, so that there was a steady inflow of trains over time. For any volume fluctuations in the course of time, additional capacities would be necessary. Since peaks can be reduced only slowly, they cause a loss of performance for the hub which is to be taken into account in the slot plans. The assumption of uniformly distributed arrivals is, however, not unrealistic for large and complex nodes. If it comes to capacity shortages in a node, trains wait in practice in marshalling yards ahead of the complex freight node and will be delivered at the right time.

Equally distributed train arrivals can, however, not always be achieved. Due to the increase of regional passenger railway traffic (suburban railways) in some densely populated European urban areas (e.g. Ruhr area in North Rhine-Westphalia in Germany) in the recent years, rail freight transport is more and more pushed into the night hours. On some relations there are almost no routes for freight trains available during daytime. This applies not only to transports to and from the affected region but also to transit traffic. The shift to early morning or night slots causes pressure on the freight terminals since it induces peak hours which can lead to congestion and violation of slot plans.

The interaction between neighbouring rail network operators was considered only marginally. Public European railway infrastructure service operators are obliged to grant access to their network to all authorised railway undertakings. Due to the prohibition of discrimination, an unpunctual arriving train must not be refused to enter a facility. For marshalling yards and similar nodes it means, that early arriving trains extended the dwell times and may violate assigned time slots. This issue was always considered in the studies of the previous chapters. The coordination of handing over trains from one infrastructure service provider to the next was always assumed to be ideal. This may not always be the case in practice. The buffering function of a railway node is therefore still irreplaceable, although it has a negative effect on its productivity.

For the successful slot management it can, especially when rescheduling, be helpful if completed trains could leave the node prior to their scheduled departure time in order to free the formation yard for incoming traffic. Such a procedure would require an intensive agreement with the neighbouring infrastructure network operators. Through the continuous improvement of exchange processes, many complex nodes have already achieved a relaxation of the situation with neighbouring infrastructure managers, which often leads to more flexibility for both, the freight node and the neighbouring network operators.

7.3 Future research potentials

The establishment of slot management strategies in complex freight hubs enables room for further improvements in rail freight services. Future research potentials were identified in both the technological and the organisational fields.

7.3.1 New role of the hub manager in the transport chain

It became clear at several points that the freight hub manager should play a more significant role in the transport chain if slot management strategies shall be implemented successfully. His initially addressed passive role (section 3.4.2) is to be dissolved. A hub production system using slot management induces that the hub operator takes up a more leading role in the transport chain where he not only reacts on other chain member's decisions but aims to enforce the optimisation of his own process. This requires an intensive communication between the hub operator, the neighbouring railway infrastructure operators and the railway undertakings in order to negotiate required process and information flows. The new role of the hub manager can be found in his position as the forerunner of a collaborative yard planning approach. The necessity for collaborative planning was also addressed in very recent research projects, e.g. EcoHubs (cf. Nestler et al., 2014). The port of Bremen shall be mentioned in this case again as an example from the observations in practice. They made already much efforts in the role as an overall node coordinator by initiating an overall yard optimisation with relevant stakeholders (cf. page 57). Nonetheless, to bring slot management to a full success, more efforts are required in future economic research of collaborative planning.

7.3.2 Simplification of processes

Slot management contributes already to a certain degree of process flow standardisation in rail freight yards. During the analysis of the process flows it has been noted that there is still a considerable potential for further simplification and reduction of redundant processes, especially when comparing the railway-related processes with those of other transport modes. For the further standardisation of the transshipment operations from and to rail it can therefore be advisable to simplify certain related processes.

The analysis of current processes revealed for example, that some steps (e.g. the check of the wagon order) are undertaken several times by different stakeholders, although the information were or could be transmitted electronically with the freight documents. An extended error probability analysis of realistic transshipment yard processes should identify such redundant measures and determine thus their necessity.

7.3.3 Effects of energy-efficient train driving on yard performance

The precise indication of free capacities and thus the better reliability and transparency in freight nodes, allows for the first time to define an optimal freight train arrival time. The knowledge of the optimal train arrival time, allows modifying the driving strategy of the trains. It is thus possible to implement better methods of energy-efficient train control strategies on freight trains. The main advantage of this strategy would be to avoid early arrivals. This would have a positive impact on both the hub processes and the energy consumption of the trains. In hubs, the untimely occupation of entrance track is avoided. Through energy-efficient or slower driving, the railway undertaking would reduce the traction power consumption.

For the implementation of the approach, further studies are necessary because a number of specific questions need to be clarified. First, it must be examined to what extent the target arrival time can be determined more accurately. Second, the energy-efficient driving regimes have an impact on the line and terminal capacity which must be determined and negotiated with the infrastructure managers. Undoubtedly for this approach, an intensive coordination between the freight hub manager, the national infrastructure manager and the railway undertakings is required.

A Appendix

The appendices on the following pages contain large tables and figures which was referred to in the text.

A.1 Initial slot plan
A.2 Planned time table
A.3 Slot plan with one technician team
A.4 Slot plan with two technician teams
A.5 Slot plan with delay of two trains
A.6 Slot plan with random delays
A.7 Rescheduled slot plan using the FCFS rule
A.8 Rescheduled slot plan Minimising the Maximum Delays
A.9 Rescheduled slot plan using the Minimum Slot Swap Rule

A.1 Initial slot plan

Slot no.	Train arrival in formation yard $t_{n,f}^{p,in}$	Shunting to terminal $t_{f,l}^{p,shunt}$	Terminal handling starts $t_{n,l}^{p,load}$	Shunting to formation yard $t_{l,f}^{p,shunt}$	Departure from formation yard $t_{n,f}^{p,dep}$
1	660	960	980	1,280,	1,630
2	720	1,020	1,040	1,340	1,690
3	780	1,080	1,100	1,400	1,750
4	840	1,140	1,160	1,460	1,810
5	900	1,200	1,220	1,520	1,870
6	960	1,260	1,280	1,580	1,930
7	1,020	1,320	1,340	1,640	1,990
8	1,080	1,380	1,400	1,700	2,050
9	1,140	1,440	1,460	1,760	2,110
10	1,200	1,500	1,520	1,820	2,170
11	1,260	1,560	1,580	1,880	2,230
12	1,320	1,620	1,640	1,940	2,290
13	1,380	1,680	1,700	2,000	2,350
14	1,440	1,740	1,760	2,060	2,410
15	1,500	1,800	1,820	2,120	2,470
16	1,560	1,860	1,880	2,180	2,530
17	1,620	1,920	1,940	2,240	2,590
18	1,680	1,980	2,000	2,300	2,650
19	1,740	2,040	2,060	2,360	2,710
20	1,800	2,100	2,120	2,420	2,770
21	1,860	2,160	2,180	2,480	2,830
22	1,920	2,220	2,240	2,540	2,890
23	1,980	2,280	2,300	2,600	2,950
24	2,040	2,340	2,360	2,660	3,010
25	2,100	2,400	2,420	2,720	3,070

Table A.1: Initial slot plan used in the scheduling model of chapter 5.3. All times in minutes.

A.2 Planned time table

Slot No. $slot_n$	Train-ID n	Planned Arrival $t^{p,arr}_{n,f}$ [min]	Train-ID n	Planned Departure $t^{p,dep}_{n,f}$ [min]
			Out_001	970
			Out_002	1,030
			Out_003	1,090
			Out_004	1,150
			Out_005	1,210
			Out_006	1,270
			Out_007	1,330
			Out_008	1,390
			Out_009	1,450
			Out_010	1,510
			Out_011	1,570
1	In_012	660	Out_012	1,630
2	In_013	720	Out_013	1,690
3	In_014	780	Out_014	1,750
4	In_015	840	Out_015	1,810
5	In_016	900	Out_016	1,870
6	In_017	960	Out_017	1,930
7	In_018	1,020	Out_018	1,990
8	In_019	1,080	Out_019	2,050
9	In_020	1,140	Out_020	2,110
10	In_021	1,200	Out_021	2,170
11	In_022	1,260	Out_022	2,230
12	In_023	1,320	Out_023	2,290
13	In_024	1,380	Out_024	2,350
14	In_025	1,440	Out_025	2,410
15	In_026	1,500		
16	In_027	1,560		
17	In_028	1,620		
18	In_029	1,680		
19	In_030	1,740		
20	In_031	1,800		
21	In_032	1,860		
22	In_033	1,920		
23	In_034	1,980		
24	In_035	2,040		
25	In_036	2,100		

Table A.2: Initial timetable used in the scheduling model of chapter 5.3.

A.3 Slot plan with one technician team

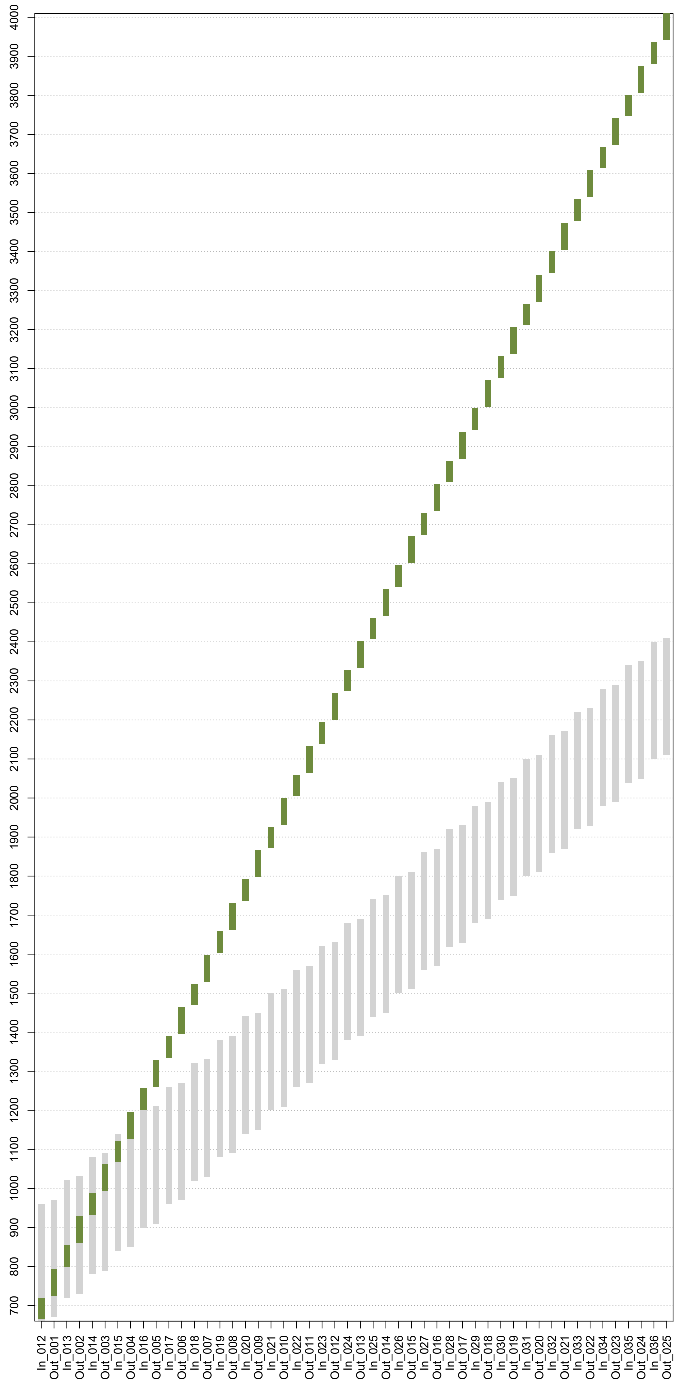


Figure A.1: Slot plan created with the scheduling model in chapter 5.3 when one technician team is employed (abscissa: time in min, ordinate: train numbers).

A.4 Slot plan with two technician teams

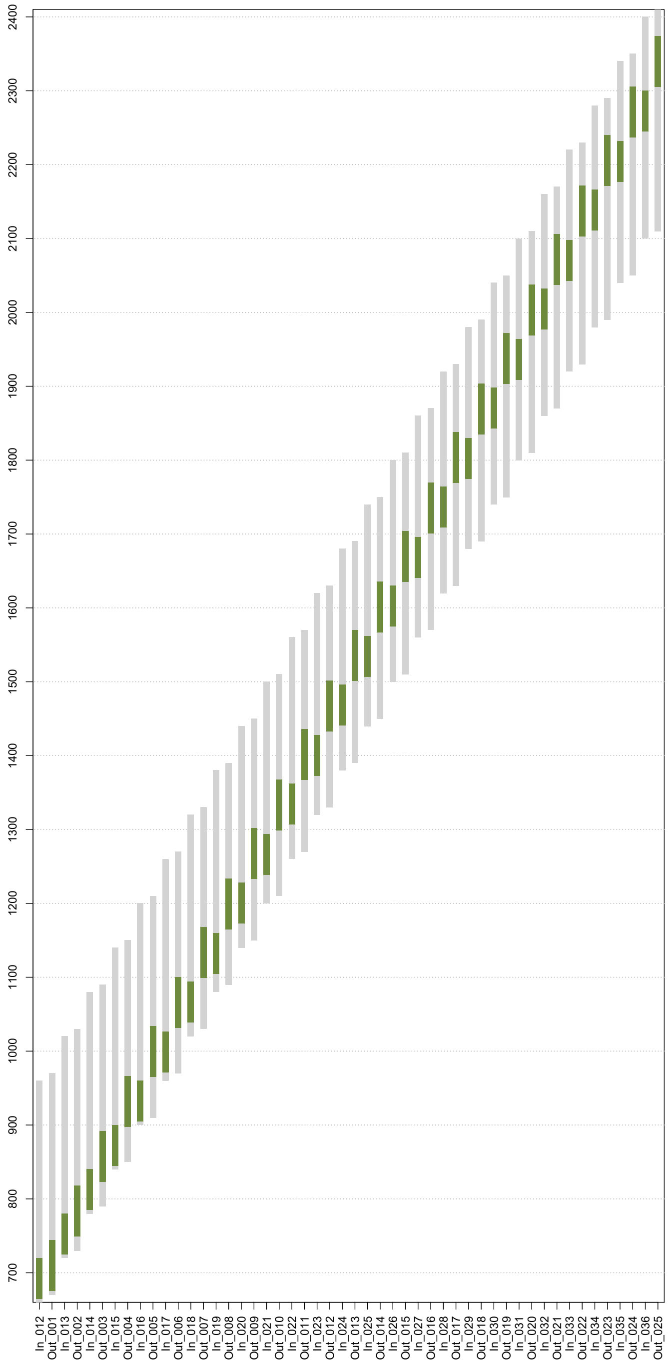


Figure A.2: Slot plan created with the scheduling model in chapter 5.3 when two technician teams are employed (abscissa: time in min, ordinate: train numbers).

A.5 Slot plan with delay of two trains

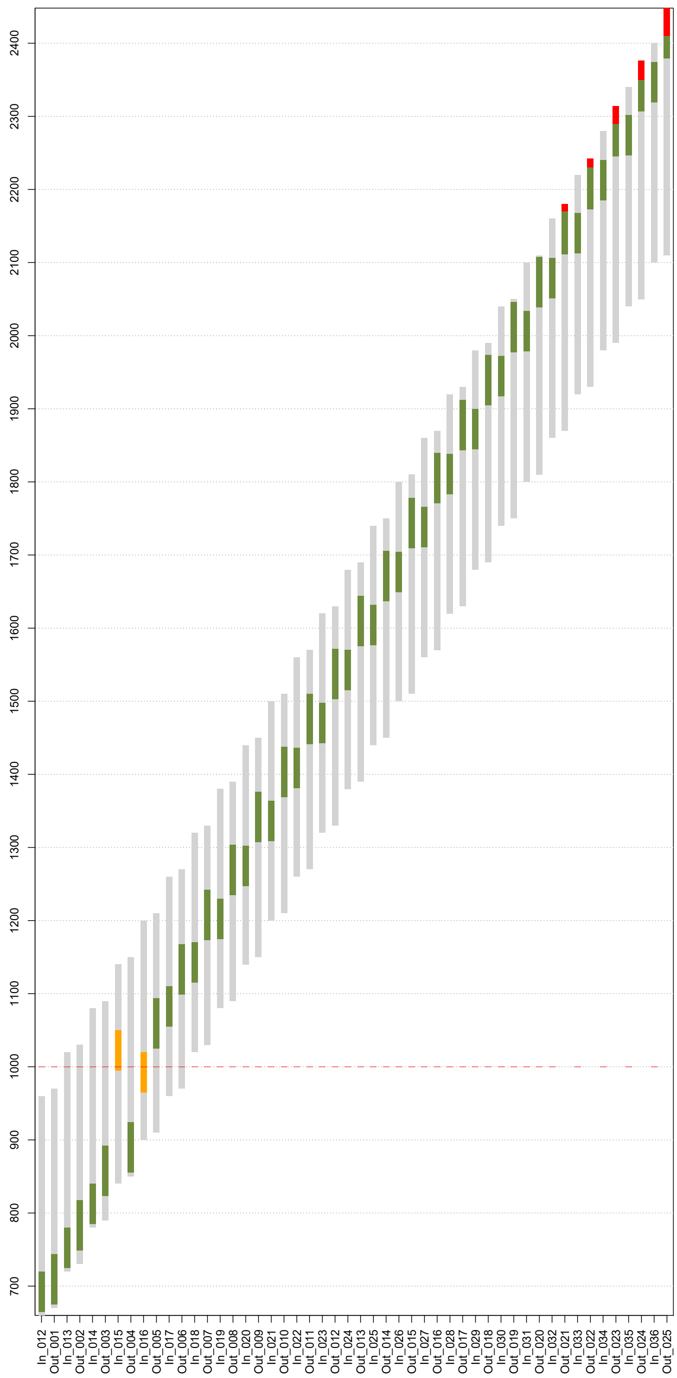


Figure A.3: Slot plan created with the scheduling model in chapter 5.4 when two trains arrive with a delay. (abscissa: time in min, ordinate: train numbers)

A.6 Slot plan with random delays

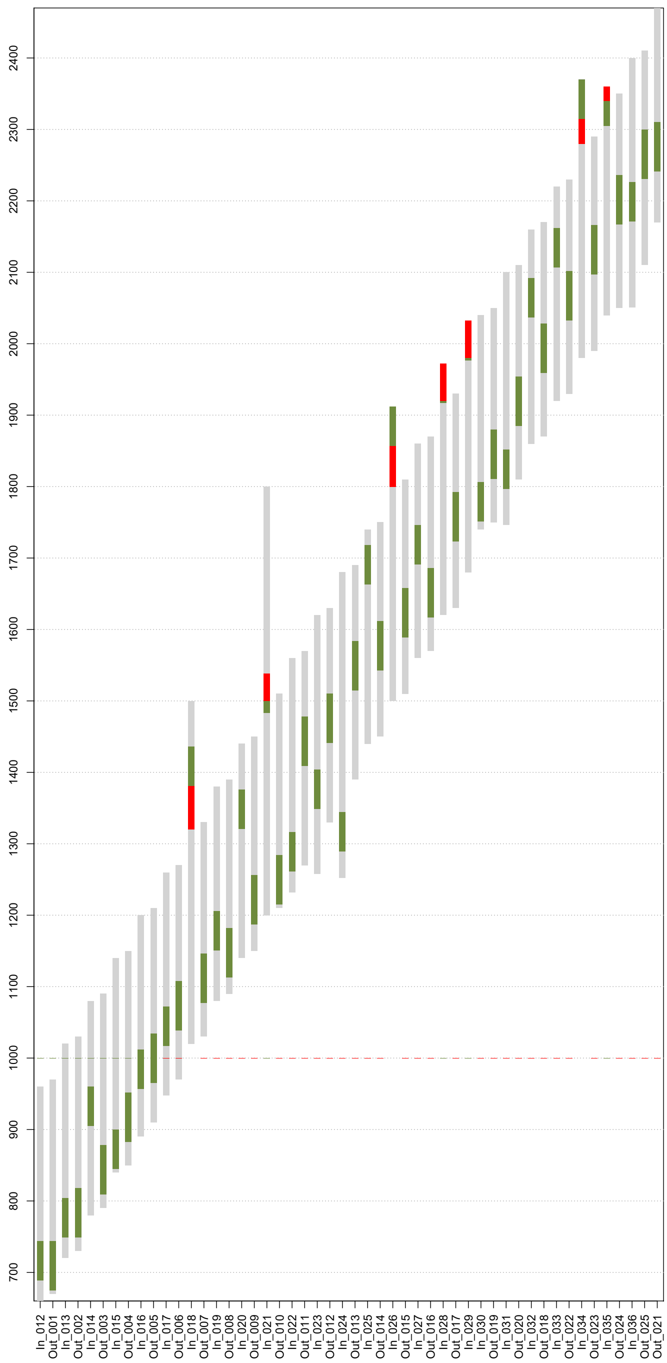


Figure A.4: Slot plan created with the scheduling model in chapter 5.5 when all trains arrive with a random delay (abscissa: time in min, ordinate: train numbers).

A.7 Rescheduled slot plan using the FCFS rule

No.	Train ID	New arrival time	Slot start time	Departure time
1	In_012	684	660	1,630
2	In_013	736	720	1,690
3	In_015	840	780	1,750
4	In_014	888	840	1,810
5	In_016	890	900	1,870
6	In_017	948	960	1,930
7	In_019	1,109	1,020	1,990
8	In_022	1,232	1,080	2,050
9	In_024	1,252	1,140	2,110
10	In_020	1,258	1,200	2,170
11	In_023	1,258	1,260	2,230
12	In_018	1,270	1,320	2,290
13	In_021	1,367	1,380	2,350
14	In_025	1,593	1,440	2,410
15	In_027	1,621	1,500	2,470
16	In_030	1,740	1,560	2,530
17	In_031	1,746	1,620	2,590
18	In_026	1,802	1,680	2,650
19	In_028	1,833	1,740	2,710
20	In_029	1,913	1,800	2,770
21	In_032	1,987	1,860	2,830
22	In_033	2,018	1,920	2,890
23	In_036	2,051	1,980	2,950
24	In_035	2,247	2,040	3,010
25	In_034	2,261	2,100	3,070

Table A.3: Rescheduled slot plan using the FCFS Rule in chapter 5.6. All times in minutes.

A.8 Rescheduled slot plan Minimising the Maximum Delays

No. $slot_n$	Train ID n	New arrival time $t^{arr}_{n,f}$	Slot start time $t^{start}_{slot_n}$	Departure time $t^{dep}_{n,f}$
1	In_012	684	660	1,630
2	In_013	736	720	1,690
3	In_014	888	780	1,750
4	In_015	840	840	1,810
5	In_016	890	900	1,870
6	In_017	948	960	1,930
7			1,020	
8	In_019	1,109	1,080	2,050
9	In_020	1,258	1,140	2,110
10	In_022	1,232	1,200	2,170
11	In_018	1,270	1,260	2,230
12	In_023	1,258	1,320	2,290
13	In_024	1,252	1,380	2,350
14	In_021	1,367	1,440	2,410
15	In_025	1,593	1,500	2,470
16	In_027	1,621	1,560	2,530
17			1,620	
18	In_026	1,802	1,680	2,650
19	In_030	1,740	1,740	2,710
20	In_031	1,746	1,800	2,770
21	In_028	1,833	1,860	2,830
22	In_029	1,913	1,920	2,890
23	In_032	1,987	1,980	2,950
24	In_033	2,018	2,040	3,010
25	In_036	2,051	2,100	3,070
26	In_035	2,247	2,280	3,250
27	In_034	2,261	2,340	3,310

Table A.4: Rescheduled slot plan using the MMD Rule in chapter 5.6. All times in minutes.

A.9 Rescheduled slot plan using the Minimum Slot Swap Rule

No.	Train ID	New arrival time	Slot start time	Departure time
1	In_012	684	660	1,630
2	In_013	736	720	1,690
3	In_014	888	780	1,750
4	In_015	840	840	1,810
5	In_016	890	900	1,870
6	In_017	948	960	1,930
7	In_019	1,109	1,020	1,990
8	In_018	1,270	1,080	2,050
9	In_020	1,258	1,140	2,110
10	In_022	1,232	1,200	2,170
11	In_021	1,367	1,260	2,230
12	In_023	1,258	1,320	2,290
13	In_024	1,252	1,380	2,350
14	In_025	1,593	1,440	2,410
15	In_027	1,621	1,500	2,470
16	In_026	1,802	1,560	2,530
17	In_030	1,740	1,620	2,590
18	In_028	1,833	1,680	2,650
19	In_029	1,913	1,740	2,710
20	In_031	1,746	1,800	2,770
21	In_032	1,987	1,860	2,830
22	In_033	2,018	1,920	2,890
23	In_036	2,051	1,980	2,950
24	In_034	2,261	2,040	3,010
25	In_035	2,247	2,100	3,070

Table A.5: Rescheduled slot plan using the MSS Rule in chapter 5.6. All times in minutes.

Bibliography

Adamko, N. and Klima, V. (2008). Optimisation of Railway Terminal Design and Operations Using Villon Generic Simulation Model. *Transport*, 23(4):335–340.

Albrecht, M. (2009). *Supply Chain Coordination Mechanisms: New Approaches for Collaborative Planning*. Springer, New York.

Andersson, F. N. G. and Elger, T. (2012). Swedish Freight Demand: Short, Medium, and Long-term Elasticities. *Journal of Transport Economics and Policy*, 46(1):79–97.

Arcot, V. C. (2007). *Modeling Uncertainty in Rail Freight Operations: Implications for Service Reliability*. Master's thesis, Academy of Maryland.

Arnold, P., Peeters, D., and Thomas, I. (2004). Modelling a rail/road intermodal transportation system. *Transportation Research Part E: Logistics and Transportation Review*, 40(3):255–270.

Baldassarra, A., Impastato, S., and Ricci, S. (2010). Intermodal terminal simulation for operations management. *European Transport \ Trasporti Europei*, (46):86–99.

Baron, B. (2011). *Analyse der Umschlagvorgänge Seeschiff-Schiene in modernen Containerterminals*. Seminar paper, Technische Universität Berlin, Fachgebiet Schienenfahrwege und Bahnbetrieb, Berlin.

Basso, L. J. and Zhang, A. (2010). Pricing vs. slot policies when airport profits matter. *Transportation Research Part B: Methodological*, 44(3):381–391.

Benning, R. and Zimmermann, U. (2014). *Ausbreitung und Prognose von Verspätungen im Schienengüterverkehr*. Bachelor Thesis, Technische Universität Berlin, Berlin.

Berger, A., Gebhardt, A., Müller-Hannemann, M., and Ostrowski, M. (2011). Stochastic Delay Prediction in Large Train Networks. Technical Report 1, Martin-Luther-Universität, Halle (Saale).

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

Bonsra, K. and Harbolovic, J. (2012). *Estimation of run times in a freight rail transportation network*. Thesis (M. Eng. in Logistics), Massachusetts Institute of Technology.

Bontekoning, Y. M. (2006). *Hub exchange operations in intermodal hub-and-spoke operations: Comparison of the performances of four types of rail-rail exchange facilities*. Netherlands TRAIL Research School, Delft, The Netherlands.

Boysen, N., Fliedner, M., and Kellner, M. (2010). Determining fixed crane areas in rail–rail transshipment yards. *Transportation Research Part E: Logistics and Transportation Review*, 46(6):1005–1016.

Boysen, N., Jaehn, F., and Pesch, E. (2011). Scheduling Freight Trains in Rail-Rail Transshipment Yards. *Transportation Science*, 45(2):199–211.

Braun, I., König, S., and Schnieder, E. (2005). Multi Agent Systems for Rail Transport. Technical report, Technische Universität Braunschweig. URL: https://www.researchgate.net/publication/228644400_Multi_agent_systems_for_rail_transport [cited 24.06.2015].

Brucker, P. (2007). *Scheduling Algorithms*. Springer, Berlin Heidelberg, 5. ed edition.

Büker, T. and Seybold, B. (2012). Stochastic modelling of delay propagation in large networks. *Journal of Rail Transport Planning & Management*, 2(1-2):34–50.

Cacchiani, V., Huisman, D., Kidd, M. P., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, and Joris (2013). An Overview of Recovery Models and Algorithms for Real-time Railway Rescheduling. Technical Report EI2013-29, Erasmus Universiteit Rotterdam. URL: http://repub.eur.nl/pub/50112/EI-Report-2013-29-1-.pdf [cited 20.02.2015].

Cambridge University Press (2015). Cambridge Dictionary of British English. URL: http://dictionary.cambridge.org/dictionary/british/ [cited 16.06.2015].

Carey, M. (1999). Ex ante heuristic measures of schedule reliability. *Transportation Research Part B: Methodological*, 33(7):473–494.

Carey, M. and Carville, S. (2000). Testing schedule performance and reliability for train stations. *Journal of the Operational Research Society*, (51):666–682.

Carey, M. and Crawford, I. (2007). Scheduling trains on a network of busy complex stations. *Transportation Research Part B: Methodological*, 41(2):159–178.

Cargotec Finland Oy (2014). Kalmar Gloria DRG420-450 reachstackers: Gloria Options brochure. URL: http://www.kalmarglobal.com/equipment/reachstackers/container-handling/ [cited 06.07.2014].

Carsey, T. M. and Harden, J. J. (2013). *Monte Carlo simulation and resampling*.

Carteni, A. and De Luca, S. (2009). Simulation of a container terminal through a discrete event approach: literature review and guidelines for application. In Association for European Transport, editor, *European Transport Conference*.

Carteni, A. and De Luca, S. (2010). Analysis and modeling of container handling equipment activities. *European Transport \ Trasporti Europei*, (46):52–71.

Chen, B. and Harker, P. T. (1990). Two Moments Estimation of the Delay on Single-Track Rail Lines with Scheduled Traffic. *Transportation Science*, 24(4):261–275.

Chen, L., Schmid, F., Dasigi, M., Ning, B., Roberts, C., and Tang, T. (2010). Real-time train rescheduling in junction areas. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 224(6):547–557.

Chmielewski, A., Naujoks, B., Janas, M., and Clausen, U. (2009). Optimizing the Door Assignment in LTL-Terminals. *Transportation Science*, 43(2):198–210.

Chu, F. and Oetting, A. (2013). Modeling capacity consumption considering disruption program characteristics and the transition phase to steady operations during disruptions. *Journal of Rail Transport Planning & Management*, 3(3):54–67.

Civil Aviation Authority (2009). *CAP 694: The UK Flight Planning Guide*. The Stationery Office, Norwich.

Concannon, K. (2003). *Simulation modeling with SIMUL8*. Visual8 Corp., Mississauga, Ontario.

Conte, C. (2007). *Identifying Dependencies Among Delays*. PhD Thesis, Georg-August-Universität Göttingen, Göttingen.

Crainic, T. G. (2000). Service network design in freight transportation. *European Journal of Operational Research*, 122(2):272–288.

Crainic, T. G. and Kim, K. H. (2005). Intermodal Transportation. Technical report, Centre de recherche sur les transports (C.R.T.), Montreal, Canada.

Crainic, T. G. and Laporte, G. (1997). Planning models for freight transportation. *European Journal of Operational Research*, 97(3):409–438.

Czerny, A. I. (2010). Airport congestion management under uncertainty. *Transportation Research Part B: Methodological*, 44(3):371–380.

D'Ariano, A., Pacciarelli, D., and Pranzo, M. (2007). A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research*, 183(2):643–657.

DB Netze AG (2014). Richtlinie 402.0202. Trassenmanagement. Planungsprocedere; Trassenanmeldungen.

DB Netze AG (2015). Richtlinie 402.0203. Trassenmanagement. Planungsprocedere; Aufgaben und Abläufe im Planungsprocedere für den Netzfahrplan.

de Neufville, R. and Odoni, A. R. (2003). *Airport systems: Planning design, and management*. McGraw-Hill, New York.

Deflorio, F. P., Gonzalez-Feliu, J., Perboli, G., and Tadei, R. (2012). The Influence of Time Windows on the Costs of Urban Freight Distribution Services in City Logistics Applications. *European Journal of Transport and Infrastructure Research*, 12(3).

Deutsches Institut für Normung e.V. (2001). System of road vehicles - Vocabulary of power-driven vehicles, combinations of vehicles and towed vehicles.

Deutsches Institut für Normung e.V. (2007). Swap bodies - Non-stackable swap bodies of class C - Dimensions and general requirements.

Emde, A. (2013). *Auswirkungen Differenzierter Halteregime auf die Streckenkapazität*. Master's Thesis, Technische Universität Berlin, Berlin.

Ethernet Working Group (05.06.2015). IEEE 802.3. URL: http://www.ieee802.org/3/ [cited 05.06.2015].

European Parliament (2001). Directive 2001/14/EC of the European Parliament and of the Council of 26 February 2001 on the allocation of railway infrastructure capacity and the levying of charges for the use of railway infrastructure and safety certification. Technical report, European Parliament.

European Union (1993). Council Regulation (EEC) no. 95/93 on common rules for the allocation of slots at Community airports.

European Union (2010). Allocation of time slots. URL: http://europa.eu/legislation_summaries/transport/air_transport/l24085_en.htm [cited 02.10.2012].

Ferschl, F. (1964). *Zufallsabhängige Wirtschaftsprozesse*. Physica-Verlag, Wien, Würzburg.

Flier, H., Gelashvili, R., Graffagnino, T., and Nunkesser, M. (2009). Mining Railway Delay Dependencies in Large-Scale Real-World Delay Data. In Ahuja, R. K., Möhring, R. H., and Zaroliagis, C. D., editors, *Robust and online large-scale optimization*, pages 354–368. Springer, Berlin and , New York.

Freie Hansestadt Bremen (2015). Entgelte für die Nutzung der Bremischen Hafeneisenbahn. Entgeltgrundsätze und Liste der Entgelte. URL: http://www.bremenports.de/standort/bremische-hafeneisenbahn/nutzung-und-entgelte [cited 02.06.2015].

Gasser, H. (2010). Innovative Umschlagtechnik kann neue Marktsegmente erschließen. *Rail Business Spezial - Kombinierter Verkehr in Europa*, (1):12–13.

Gille, A. and Schönemann, R. (2009). Kapazitätsreserven und Betriebsabwicklung bei der Hafenbahn Hamburg nach Umstellung auf die Fahrplanfenster-Strategie Slotmanagement (unpublished). Technical report, unpublished.

Graham, R. L., Lawler, E. L., Lenstra, J. K., and Kan, A. (1979). Optimization and Approximation in Deterministic Sequencing and Scheduling: a Survey. In Korte, B. H. and Johnson, E. L., editors, *Discrete Optimization I*, volume 5 of *Annals of Discrete Mathematics*, pages 287–326. Elsevier, Burlington.

Gudehus, T. (1976). Staueffekte vor Transportknoten. *Zeitschrift für Operations Research*, 20(6):B207–B252.

Gudehus, T. (2012). *Logistik 1: Grundlagen, Verfahren und Strategien*. Springer Berlin Heidelberg, Berlin and Heidelberg, 4 edition.

Günther, H.-O. (2005). Supply Chain Management and Advanced Planning Systems: A Tutorial. In Günther, H.-O., Mattfeld, D. C., and Suhl, L., editors, *Supply Chain Management und Logistik*, pages 3–40. Physica-Verlag, Heidelberg.

Günther, H.-O. and Kim, K. H. (2006). Container terminals and terminal operations. *OR Spectrum*, 28(4):437–445.

Hall, R. W. (1991). *Queueing methods: For services and manufacturing*. Prentice Hall International series in industrial and systems engineering. Prentice-Hall, Englewood Cliffs, NJ.

Hallowell, S. F. and Harker, P. T. (1996). Predicting On-Time Line-Haul Performance in Scheduled Railroad Operations. *Transportation Science*, 30(4):364–378.

Hallowell, S. F. and Harker, P. T. (1998). Predicting On-Time Performance in Scheduled Railroad Operations: Methodology and Application to Train Scheduling. *Transportation Research Part A: Policy and Practice*, 32(4):279–295.

Hamburg Port Authority (2015). Das Entgelt-System der Hafenbahn. Aufbau, Struktur und Nutzungsbeispiele. URL: http://www.hamburg-port-authority.de/de/hafenkunden/hafenbahn/ [cited 01.06.2015].

Hamburger Hafen und Logistik Aktiengesellschaft (2014). Nutzungsbedingungen für Serviceeinrichtungen. URL: http://hhla.de/fileadmin/download/HHLA_Nutzungsbedingungen_Serviceeinrichtungen_Bahn_Logistiker.pdf [cited 03.06.2015].

Hansen, I. A. (15.05.2013). Capacity Estimation Methods: 5th International Seminar on Railway Operations Modelling and Analysis - RailCopenhagen.

Hansen, I. A. (2004). Automated shunting of rail container wagons in ports and terminal areas. *Transportation Planning and Technology*, 27(5):385–401.

Hansen, I. A., editor (2010). *Timetable planning and information quality*. WIT Press, Southampton.

Hansen, I. A. and Pachl, J., editors (2014). *Railway Timetabling & Operations*. Eurail-Press.

Hartmann, S. (2004). Generating scenarios for simulation and optimization of container terminal logistics. *OR Spectrum*, 26(2):171–192.

Heister, G. (1978). Das Verspätungsniveau zweigleisiger Eisenbahnstrecken - eine statistische Untersuchung. Technical Report 11, Technische Universität Hannover, Hannover.

Henesey, L. E., editor (2006). *Multi-agent systems for container terminal management*. Blekinge Institute of Technology, Karlskrona.

Higgins, A., Kozan, E., and Ferreira, L. (1997). Heuristic Techniques for Single Line Train Scheduling. *Journal of Heuristics*, 3:43–62.

Homeyer, D. (2000). *Rangieren im Bahnbetrieb*. DB-Fachbuch. Eisenbahn-Fachverlag, Heidelberg.

Huynh, N. N. (2008). Analysis of Container Dwell Time on Marine Terminal Throughput and Rehandling Productivity. In *Proceedings of 87th Transportation Research Board Meeting*.

IATA (2012). Worldwide Slot Guidelines. 3rd edition. URL: www.iata.org/wsg [cited 01.10.2012].

IATA (2014). Worldwide Slot Guidelines. 6th edition. URL: http://www.iata.org/wsg [cited 19.06.2015].

Jaehn, F. and Pesch, E. (2014). *Ablaufplanung: Einführung in Scheduling*. Springer Gabler, Berlin, Heidelberg.

Jovanović, D. and Harker, P. T. (1990). A decision support system for train dispatching: An optimization-based methodology. *Journal of the Transportation Research Forum*, 31(1).

Kavička, A. and Klima, V. (1998). Simulation Model Of Marshalling Yard Linz VBF. Technical report, University of Žilina, Faculty of Management Science and Informatics. URL: http://www.simcon.sk/articles/genova1999.pdf [cited 10.09.2013].

Kendall, D. G. (1953). Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. *The Annals of Mathematical Statistics*, 24(3):338–354.

Kim, K. H. and Günther, H.-O., editors (2007). *Container Terminals and Cargo Systems: Design, Operations Management, and Logistics Control Issues*. Springer.

Klabes, S. G. (2010). *Algorithmic Railway Capacity Allocation in a Competitive European Railway Market*. PhD Thesis, Rheinisch-Westfälische Technische Hochschule, Aachen.

Koolstra, K. (2005). *Transport Infrastructure Slot Allocation*. PhD Thesis, Technische Universiteit Delft.

Kösters, D. (2010). *Planmäßige Verspätungen im Luftverkehr: Definition und Nutzung eines Qualitätsmaßes für die Flughafenkoordination*. Dissertation, Rheinisch-Westfälische Technische Hochschule, Aachen.

Kouloumbis, M. (2015). Ecohubs ECOTAURUS. Joint Dissemination Event of the European Projects EcoHubs, Swiftly Green, and GreCor, 28/04/2015, Brussels.

Kozan, E. and Preston, P. (2006). Mathematical modelling of container transfers and storage locations at seaport terminals. *OR Spectrum*, 28(4):519–537.

Kraay, D. R. and Harker, P. T. (1995). Real-time scheduling of freight railroads. *Transportation Research Part B: Methodological*, 29(3):213–229.

Kreft, H. (2015). Die Hamburger Hafenbahn: Herausforderungen für Europas größten Eisenbahnhafen: Seminar lecture. Technische Universität Berlin, Fachgebiet Schienenfahrwege und Bahnbetrieb, 02.02.2015.

Kreutzberger, E. (1999). Innovative Networks and New-generation Terminals for Intermodal Transport: Improving the Cost-quality Ratio by Bundling of Flows. In Bovy, P. H. L., editor, *Five years crossroads of theory and practice*, volume 99,6 of *TRAIL conference proceedings series*. Univ. Press, Delft.

Kroon, L. and Huisman, D. (2011). Algorithmic Support for Disruption Management at Netherlands Railways. Technical report, Erasmus University Rotterdam, Econometric Institute.

Landex, A. (2012). Reliability of Railway Operation. *Proceedings from the Annual Transport Conference at Aalborg University*.

Lee, B. K., Park, S. O., and Seo, J. H. (2006). A Simulation Study for Designing a Rail Terminal in a Container Port. In Perrone, L. F., Lawson, B., Liu, J., and Wieland, F. P., editors, *Proceedings of the 38th conference on Winter simulation*, pages 1388–1397. Winter Simulation Conference.

Lemnian, M., Rückert, R., Rechner, S., Blendinger, C., and Müller-Hannemann, M. (2014). Timing of Train Disposition: Towards Early Passenger Rerouting in Case of Delays. In *ATMOS 2014*, Wroclaw.

Limbourg, S. and Jourquin, B. (2009). Optimal rail-road container terminal locations on the European network. *Transportation Research Part E: Logistics and Transportation Review*, 45(4):551–563.

Lind, D., Hsieh, J. K., and Jordan, M. A. (2007). Tandem 40 Dockside Container Cranes and Their Impact on Terminals. In Watson, W., editor, *Ports 2007: 30 Years of Sharing Ideas: 1977-2007*.

Lusby, R. M., Larsen, J., Ehrgott, M., and Ryan, D. (2011). Railway track allocation: models and methods. *OR Spectrum*, 33(4):843–883.

Lüthi, M. (2008). Evaluation of energy saving strategies in heavily used rail networks by implementing an integrated real-time rescheduling system. In Allan, J. J., editor, *Computers in railways XI*, pages 349–358. WIT Press, Southampton.

Maersk Group (2013). Worlds largest ship. URL: http://www.worldslargestship.com/the-ship/#page/efficiency [cited 22.05.2014].

Malavasi, G. and Ricci, S. (2003). Performance of railway freight terminals: a generalised model. In *Proceedings of the European Transport Conference*, volume 2003 of *ETC Proceedings*.

Manne, A. S. (1960). On the Job-Shop Scheduling Problem. *Operations Research*, 8(2):219–223.

Marinov, M. V. and Viegas, J. M. (2009). Analysis And Evaluation Of Double Ended Flat-shunted Yard Performances Employing Two Yard Crews. *Mechanics, Transport, Communications*, (2).

Marinov, M. V. and Viegas, J. M. (2011a). A mesoscopic simulation modelling methodology for analyzing and evaluating freight train operations in a rail network. *Simulation Modelling Practice and Theory*, 19(1):516–539.

Marinov, M. V. and Viegas, J. M. (2011b). Tactical management of rail freight transportation services: evaluation of yard performance. *Transportation Planning and Technology*, 34(4):363–387.

Martin, U., Cui, Y., Hantsch, F., Chu, Z., and Li, X. (2015). *Knotenkapazität - Bewertungsverfahren für das mikroskopische Leistungsverhalten und die Engpasserkennung im spurgeführten Verkehr (RePlan)*, volume 8 of *Neues verkehrswissenschaftliches Journal NVJ*. VWI Verkehrswissenschaftliches Institut Stuttgart GmbH.

Nash, A. and Hürlimann, D. (2004). Railroad simulation using OpenTrack. In Allan, J. J., editor, *Computers in railways IX*. WIT Press, Southampton and England and Boston.

Nestler, S., Nobel, T., and Mackenthun, F. (2014). Best Practice TriCon intermodal terminal in the FV Nürnberg (Germany): EU Project EcoHubs (Environmental COherent measures and environmental interventions to debottleneck HUBS of the multimodal network favoured by seamless flow of goods). Technical report. URL: http://www.ecohubs.eu/default.aspx?articleID=5352 [cited 20.05.2015].

Neuman, U. M. and Atkin, J. A. D. (2013). Airport Gate Assignment Considering Ground Movement. In Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Pacino, D., Voß, S., and Jensen, R. M., editors, *Computational Logistics*, volume 8197 of *Lecture Notes in Computer Science*, pages 184–198. Springer Berlin Heidelberg, Berlin, Heidelberg.

Oberländer, K. (1959). Die betrieblichen Vorbereitungs- und Abschlußzeiten im Lokomotivdienst. *Der operative Dienst*, (8):175–177.

Oliver, K. R. and Webber, M. D. (1992). Logistics Catches up with Strategy. In Christopher, M., editor, *Logistics. The Strategic Issues*, pages 63–75. Chapman & Hall, London.

Pachl, J. (2008). Timetable design principles. In Hansen, I. A. and Pachl, J., editors, *Railway timetable & traffic*. EurailPress, Hamburg.

Pachl, J. (2013). *Systemtechnik des Schienenverkehrs: Bahnbetrieb planen, steuern und sichern*. Imprint: Springer Vieweg, Wiesbaden, 7., überarb. u. erw. aufl. 2014 edition.

Persson, J. (2005). *Improvement of Estimated Time of Arrival for Railroad Traffic*. Master's Thesis, Blekinge Institute of Technology, Blekinge.

Petersen, E. R. (1977a). Railyard Modeling: Part I. Prediction of Put-Through Time. *Transportation Science*, 11(1):37–49.

Petersen, E. R. (1977b). Railyard Modeling: Part II. The Effect of Yard Facilities on Congestion. *Transportation Science*, 11(1):50–59.

Pinedo, M. (2005). *Planning and scheduling in manufacturing and services*. Springer series in operations research. Springer, New York.

Pinedo, M. (2008). *Scheduling: Theory, algorithms, and systems*. Springer, New York, 3rd ed edition.

porttechnology.org (2012). Latest generation, hybrid automated guided vehicles in operation at Rotterdam. URL: http://www.porttechnology.org/news/latest_generation_hybrid_automated_guided_vehicles_in_operation_at_rotterda/ [cited 17.06.2015].

Potthoff, G. (1975). *Verkehrsströmungslehre: Band 5: Bedienungstheorie*. Transpress, Verlag für Verkehrswesen, Berlin.

Potthoff, G. (1977). *Verkehrsströmungslehre. Bd. 2. Betriebstechnik des Rangierens*. Transpress, Verlag für Verkehrswesen, Berlin, 3 edition.

Radtke, A. and Bendfeld, J.-P. (2001). Handling of railway operation problems with RailSys. In UIC, editor, *Word Conference on Railway Research*. UIC.

Rizzoli, A. E., Fornara, N., and Gambardella, L. M. (2002). A simulation tool for combined rail/road transport in intermodal terminals. *Mathematics and Computers in Simulation*, 59(1-3):57–71.

Rodrigue, J.-P., Comtois, C., and Slack, B. (2006). *The geography of transport systems*. Routledge, London and New York.

Rodrigue, J.-P., Comtois, C., and Slack, B. (2013). *The geography of transport systems*. Routledge, London and New York, 3 edition.

Rodriguez, J. (2007). A constraint programming model for real-time train scheduling at junctions. *Transportation Research Part B: Methodological*, 41(2):231–245.

Ropke, S. and Cordeau, J.-F. (2009). Branch and Cut and Price for the Pickup and Delivery Problem with Time Windows. *Transportation Science*, 43(3):267–286.

Rouse, M. (2007). Networking and Communications Glossary. URL: http://searchnetworking.techtarget.com/definition/slot-time [cited 05.06.2015].

Ryerson, M. S., Hansen, M., and Bonn, J. (2014). Time to burn: Flight delay, terminal efficiency, and fuel consumption in the National Airspace System. *Transportation Research Part A: Policy and Practice*, 69:286–298.

SBB Cargo AG (2015). Wagentypen-Suche: Sgns, Sgns(s). URL: http://www.sbbcargo.com/de/angebot/wagen-waren/wagentypen-suche/sgns.html [cited 11.06.2015].

Schönemann, R. (08.04.2014). Railway Operations at the Freight Village of Interporto Bologna: Interview with Giuseppe Dall'Asta, Project Manager at Interporto Bologna.

Schönemann, R. (09.06.2015). Prozessabläufe im Produktionsystem der Bremischen Hafeneisenbahn: Interview with Birgit Bierwirt, Jens-Uwe Krage and Iven Krämer, Freie Hansestadt Bremen, Bremische Hafeneisenbahn, Referat 31 - Hafenwirtschaft und -infrastruktur, Schifffahrt.

Schönemann, R. (19.01.2011). Freight Handling at the Intermodal Terminal Frankfurt (Oder): Interview with Sybille Rehse, Project Manager at the Investor Center Ostbrandenburg GmbH.

Schönemann, R., Riebe, E., Galonske, N., Fuchs, T., and Behrends, V. (2015). Report on advanced management procedures and ICT: Deliverable 8.1 of the research project Viable Wagonload Production Schemes (ViWaS). Technical report.

Schönknecht, A. (2009). *Maritime Containerlogistik: Leistungsvergleich von Containerschiffen in intermodalen Transportketten*. Springer, Berlin and Heidelberg.

Schubert, J. (2015). Zugfinder. URL: http://www.zugfinder.de/ [cited 12.05.2015].

Schwanhäußer, W. (1974). *Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn*. PhD Thesis, Rheinisch-Westfälische Technische Hochschule, Aachen.

Seidelmann, C. (2010). 40 Years Combined Transport Road-Rail in Europe. Technical report, UIRR, Brussels.

Siefer, T. and Radtke, A. (2006). Evaluation of Delay Propagation. In UIC, editor, *World Conference on Railway Research*. UIC.

Smith, L. D., Ehmke, J. F., Mattfeld, D. C., Waning, R., and Hellmann, L. (2014). Strategic Decision Support for Airside Operations at Commercial Airports. In González-Ramírez, R. G., Schulte, F., Voß, S., and Ceroni Díaz, J. A., editors, *Computational Logistics*, volume 8760 of *Lecture Notes in Computer Science*, pages 132–150. Springer International Publishing, Cham.

Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Springer, Wien, New York.

Stadtler, H. and Kilger, C., editors (2005). *Supply chain management and advanced planning: Concepts, models, software and case studies*. Springer, Berlin and , New York, 3 edition.

Stahlbock, R. and Voß, S. (2007). Operations research at container terminals: a literature update. *OR Spectrum*, 30(1):1–52.

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

Strotmann, C. (2007). *Railway scheduling problems and their decomposition*. Dissertation, Universität Osnabrück, Osnabrück.

The World Bank (2014). Container port traffic. URL: http://data.worldbank.org/indicator/IS.SHP.GOOD.TU?cid=DEC_SS_WBGDataEmail_EXT [cited 14.06.2015].

Tomii, N., Fukumura, N., and Zhou, L. J. (2001). Development of Station Shunting Scheduling Algorithm. *Quarterly Report of RTRI*, 42(4).

Törnquist, J. (2004). *Computer-based decision support for handling uncertainty in railway traffic and transportation*. School of Engineering, Blekinge Institute of Technology, Ronneby.

Törnquist, J. (2006). *Railway traffic disturbance management*. Dissertation, Blekinge Institute of Technology, Blekinge.

Törnquist, J. and Davidsson, P. (2002). A Multi-Agent System Approach to Train Delay Handling. In Timm, I., Schleiffer, R., Davidsson, P., and Kirn, S., editors, *Agent Technologies in Logistics: 15th European Conference On Artificial Intelligence (ECAI-2002)*. IOS Press, Lyon.

Törnquist, J. and Persson, J. A. (2005). Train Traffic Deviation Handling Using Tabu Search and Simulated Annealing. In Sprague, R. H., editor, *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. IEEE Computer Society Press, Los Alamitos, California.

United Nations Economic Commission for Europe (2001). Terminology on Combined Transport. Technical report, United Nations Economic Commission for Europe, New York and Geneva. URL: http://www.unece.org/trans/wp24/documents/term.pdf [cited 07.05.2014].

United States Maritime Administration and National Association of Stevedores (1993). *The U.S. stevedoring and marine terminal industry*. U. S. Maritime Administration.

Vacca, I. (2011). *Container Terminal Management: Integrated Models and Large-Scale Optimization Algorithms*. Ph.D. thesis, EPFL Lausanne, Lausanne.

Vacca, I., Salani, M., and Bierlaire, M. (2008). Two-stage column generation and applications in container terminal management. In *Swiss Transport Research Conference*.

van Dijk, N. M. (2000). On hybrid combination of queueing and simulation. In Joines, J. A., Barton, R. R., Kang, K., and Fishwick, P. A., editors, *Proceedings of the 2000 Winter Simulation Conference*.

Varian, H. R. (2010). *Intermediate Microeconomics*. Gardners Books, 8 edition.

Verein Deutscher Ingenieure (2013). VDI-Richtlinie 3633: Simulation von Logistik-, Materialfluss- und Produktionssystemen. Begriffe.

Verein Deutscher Ingenieure (2014). VDI-Richtlinie 3633: Simulation von Logistik-, Materialfluss- und Produktionssystemen. Blatt 1.

Wendler, E. (2007). The scheduled waiting time on railway lines. *Transportation Research Part B: Methodological*, 41(2):148–158.

Wendler, E. and Naehrig, M. (2004). Statistische Auswertung von Verspätungsdaten. In Verband Deutscher Eisenbahn-Ingenieure e.V., editor, *Eisenbahningenieurkalender*, pages 321–331. Tetzlaff - Verlag, Hamburg.

Wieberneit, N. (2008). Service network design for freight transportation: a review. *OR Spectrum*, 30(1):77–112.

Windt, K. (2001). *Engpaßorientierte Fremdvergabe in Produktionsnetzen*, volume 579 of *Fortschritt-Berichte VDI Reihe 2, Fertigungstechnik*. VDI-Verl., Düsseldorf.

Woxenius, J. (1998). *Development of Small-scale Intermodal Freight Transportation in a Systems Context*. PhD Thesis, Chalmers University of Technology, Göteborg.

Wu, Y., Luo, J., Zhang, D., and Dong, M. (2013). An integrated programming model for storage management and vehicle scheduling at container terminals. *Research in Transportation Economics*, 42(1):13–27.

Wustermark, R. (2015). Preisliste für die Nutzung von Serviceeinrichtungen. URL: http://www.rlcw.de/images/pdf/Zugangsbedingungen/150101_Preisliste_f%C3%BCr_die_Nutzung_von_Serviceeinrichtungen-_ZNBI_RLCW.pdf [cited 01.06.2015].

Wye River, A. and de Wit, J. (2007). Airport Slot Management in Europe: NEXTOR Workshop 2007. URL: http://www.isr.umd.edu/NEXTOR/Conferences/200706_Airport_Access_Rights/deWit.pdf [cited 13.08.2012].

Xin, J., Negenborn, R. R., and Lodewijks, G. (2013). Hierarchical Control of Equipment in Automated Container Terminals. In Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Pacino, D., Voß, S., and Jensen, R. M., editors, *Computational Logistics*, volume 8197 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin Heidelberg, Berlin, Heidelberg.

Yuan, J. (2006). *Stochastic Modelling of Train Delays and Delay Propagation in Stations*. PhD thesis, Technische Universiteit Delft, Delft.

Zampelli, S., Vergados, Y., van Schaeren, R., Dullaert, W., Raa, B., and Zampelli, S. (2013). The berth allocation and quay crane assignment problem using a CP approach. In Schulte, C., editor, *Principles and Practice of Constraint Programming*, volume 8124, pages 880–896. Springer, Berlin and Heidelberg.

Zeng, Q. and Yang, Z. (2009). Integrating simulation and optimization to schedule loading operations in container terminals. *Computers & Operations Research*, 36(6):1935–1944.

Zwaneveld, P. J., Kroon, L., Romejn, H. E., Salomon, M., Dauzère-Pérès, S., van Hoesel, S. P., and Ambergen, H. W. (1996). Routing Trains Through Railway Stations: Model Formulation and Algorithms. *Transportation Science*, 30(3):181–194.