

Transport Infrastructure Slot Allocation

Kaspar Koolstra

Transport Infrastructure Slot Allocation

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. J.T. Fokkema
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 30 juni 2005 om 15:30 uur
door Kaspar KOOLSTRA
doctorandus in de Technische Planologie
geboren te Amsterdam

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. P.H.L. Bovy

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. P.H.L. Bovy	Technische Universiteit Delft, promotor
Prof. Dr.-Ing. I.A. Hansen	Technische Universiteit Delft
Prof. dr. ir. R.E.C.M van der Heijden	Radboud Universiteit Nijmegen
Prof. mr. dr. ir. S.C. Santema	Technische Universiteit Delft
Prof. dr. W.E. Walker	Technische Universiteit Delft
Prof. N.H.M. Wilson	Massachusetts Institute of Technology
Dr. ir. R. van Nes	Technische Universiteit Delft

TRAIL Thesis Series no. T2005/6, The Netherlands TRAIL Research School

TRAIL Research School

PO Box 5017

NL-2600 GA Delft

The Netherlands

Telephone: +31 15 27 86046

Telefax: +31 15 27 84333

E-mail: Info@rsTRAIL.nl

ISBN: 90-5584-066-1

Copyright © 2005 by Kaspar Koolstra.

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the author.

PREFACE

For the last 15 years, the design, analysis, and evaluation of dynamic traffic management have been major research topics in transportation science. Many traffic models have been developed, including dynamic traffic assignment models, queuing models, timetable optimization models, etc. New traffic management instruments have been introduced, mainly in the road sector, for example ramp metering and dynamic speed limits. The majority of research, however, is sector specific, and relates to operational traffic management only. Research on capacity management with respect to planned traffic is largely limited to timetable optimization.

This thesis focuses on transport infrastructure slot allocation, which is a specific type of infrastructure capacity management at the planning level. More specifically, it focuses on the highest level of slot allocation, i.e. selection slot allocation. The research project has been conducted at the Faculty of Civil Engineering and Geosciences, Transportation Planning and Traffic Engineering Section, of Delft University of Technology. The project was part of the interfaculty research center Design and Management of Infrastructures, and has been embedded in the dynamic traffic management research program of TRAIL Research School for Transport, Infrastructure and Logistics.

The empirical basis for this thesis has been provided by research on the current state-of-practice of slot allocation, combined with a number of interviews with a number of experts from the field. I would like to thank Prof. Ernst ten Heuvelhof and Helen Stout for our fruitful cooperation in our joint research project studying the current state-of-practice of infrastructure capacity management in the Netherlands. I would also like to express my gratitude to all experts who were willing to provide me with information about their current experience with slot allocation and the desired characteristics of slot allocation systems.

During my research on this subject in the past years I learned a lot of my colleagues at the Department of Transport and Planning and at the Design and Management of Infrastructures interfaculty research center. I would like to thank all these colleagues for helping me with my work by providing feedback, inspiring me with discussions, etc. Specifically, I would like to thank my thesis supervisor Piet Bovy for offering the opportunity to join the Department of Transport and Planning and for his encouragement and inspiration throughout these years. Furthermore, I would like to express my special gratitude to my daily supervisors Theo Schoemaker, who helped me in setting up my research project in the first years until his retirement, and Rob van Nes, who helped me with writing my Ph.D. thesis in the final stage. Writing this thesis was not easy, and it has taken a lot more time than I expected to write down my ideas and findings in a consistent way. Together with Piet, Rob has been a great help in shaping this thesis.

Finally, I would like to express my appreciation of the warm support of my relatives and friends, and especially of my dear wife Petra. And even little Jelinka has done her part in the last months by quietly playing when daddy was busy writing, and of course by distracting daddy when it was time for a break. Together you have helped me a lot with successfully finishing this thesis.

CONTENTS

1	Introduction	1
1.1	Background	1
1.1.1	Transport infrastructure capacity scarcity	2
1.1.2	Slot allocation	3
1.1.3	Scientific and societal relevance of slot allocation research	4
1.2	Focus and perspective	6
1.2.1	Focus on selection slot allocation at single bottlenecks	6
1.2.2	Normative, rational, and substantive perspective	7
1.3	Research setup and thesis contents	9
1.3.1	Research objective and research approach	9
1.3.2	Main contributions of this thesis	10
1.3.3	Structure of this thesis	11
2	Conceptual framework	13
2.1	The transportation system	14
2.1.1	Transportation system model	14
2.1.2	Layers and markets	15
2.1.3	Roles within the transport and traffic services layers	19
2.1.4	Conceptual model of slot allocation	22
2.2	Capacity	23
2.2.1	Definition of capacity	24
2.2.2	Determinants of capacity	28
2.2.3	Capacity and quality-of-service	29
2.3	Capacity management	33
2.4	Slot allocation	35
2.4.1	Objectives of slot allocation	35
2.4.2	Definition of slot allocation	36
2.4.3	Specification of slot borders	37
2.4.4	Levels of slot allocation	40
2.5	Transport service planning	43
2.5.1	Types of transport services	43
2.5.2	Specification of transport service planning levels	44
2.5.3	Transport service planning and slot allocation levels	48
2.6	Synopsis	49
3	Current practice with slot allocation	51
3.1	Current and potential applications of slot allocation	52
3.1.1	Rail traffic	52
3.1.2	Air traffic	54
3.1.3	Road traffic	57
3.1.4	Navigation	61

3.2	Current slot allocation regimes in Europe	63
3.2.1	Rail traffic slot allocation	63
3.2.2	Air traffic slot allocation regime	69
3.3	Main issues	72
3.3.1	Stability and flexibility	72
3.3.2	Competition and diversity	73
3.3.3	Heterogeneity	76
3.4	Conclusions	78
4	Selection slot allocation system design	83
4.1	Demand for selection slots	84
4.1.1	Desired validity of selection slots	84
4.1.2	Slot holdership	86
4.2	Slot allocation dynamics	88
4.2.1	Static and dynamic slot allocation	88
4.2.2	Proposal for selection slot allocation dynamics	90
4.3	Specification of selection slots	92
4.3.1	Specification of slot borders	92
4.3.2	Acceptability criteria	95
4.4	Conclusions	98
5	Specification of traffic markets	101
5.1	Congestion theory	102
5.1.1	Traffic supply and demand curves	102
5.1.2	System optimum and user equilibrium	104
5.2	Main characteristics of traffic markets at the selection level	106
5.2.1	Specification of traffic selection markets	106
5.2.2	Shapes of traffic supply and demand curves	107
5.2.3	Levels of infrastructure capacity scarcity	108
5.3	Infrastructure tolls	111
5.3.1	Congestion tolls	111
5.3.2	Scarcity tolls	112
5.4	Conclusions	115
6	Identification of traffic supply	117
6.1	Identification of bottlenecks	118
6.1.1	Types of traffic network elements	118
6.1.2	Capacity constraints	120
6.2	Capacity constraints from primary traffic processes	123
6.2.1	Following and queuing	123
6.2.2	Overtaking	127
6.2.3	Counter-flowing, crossing, and merging	130
6.2.4	Queuing, parking, and stopping at terminals	132
6.3	Capacity constraints due to secondary processes	134
6.3.1	Switching processes	134

6.3.2	Traffic control	136
6.3.3	Traffic externalities	137
6.4	Specification of capacity constraint parameters	138
6.4.1	Types of capacity constraints and corresponding parameters	139
6.4.2	Quality of service variables	140
6.4.3	Determination of optimal capacity constraint parameter values	144
6.5	Examples of bottlenecks in current practice	147
6.5.1	Rail traffic bottlenecks	148
6.5.2	Air traffic bottlenecks	153
6.5.3	Road traffic bottlenecks	157
6.5.4	Navigation bottlenecks	159
6.6	Conclusions	161
7	Selection slot allocation decision problems	165
7.1	Rational decision-making applied to selection slot allocation	166
7.1.1	Characteristics of rational decision-making	166
7.1.2	Example of rational slot allocation decision process	168
7.2	Specification of slot allocation objectives	170
7.2.1	Assumptions about objective formulation	170
7.2.2	Primary objectives of slot allocation	172
7.2.3	Secondary objectives of slot allocation	174
7.2.4	Specification of objective functions for the selection problem	176
7.3	Specification of optimum selection problems	179
7.3.1	Specification of the generic selection problem	179
7.3.2	Type 1: linear objective, single basic slot, single linear constraint	181
7.3.3	Type 2: linear objective, multiple basic slots, single linear constraint	182
7.3.4	Type 3: linear objective, multiple linear constraints	183
7.3.5	Type 4: non-linear objective or constraints	184
7.4	Solution procedures	184
7.4.1	Exact optimization algorithms	185
7.4.2	Approximation algorithm based on efficiency	186
7.5	Extensions to selection slot allocation decision approach	194
7.5.1	Basic slot allocation decision approach	194
7.5.2	Extension of solution approach to include satisficing constraints	196
7.5.3	Extension of solution approach to network problems	200
7.6	Conclusions	202
8	Practical implications of selection slot allocation	205
8.1	Example: Metropolitan Airport	206
8.1.1	Introduction	206
8.1.2	Specification of traffic supply and demand	208
8.1.3	Solving the selection slot allocation decision problem	214
8.2	Consequences of implementation of selection slot allocation	219
8.2.1	Main changes given implementation of selection slot allocation	219

8.2.2	Legal competence and slot allocation	221
8.2.3	Main consequences of implementation of selection slot allocation	222
8.3	Synopsis	223
9	Conclusions and recommendations	225
9.1	Short summary	225
9.2	Main conclusions of this thesis	227
9.3	Main contributions of this thesis	229
9.4	Recommendations	230
9.4.1	Recommendations about slot allocation in practice	231
9.4.2	Recommendations for further research	233
	References	235
	List of symbols	249
	Glossary	251
<i>Appendix A</i>	Main transport service types per sector	253
<i>Appendix B</i>	Identification of banks of Amsterdam Airport Schiphol	259
<i>Appendix C</i>	Slot availability data for London airports	261
<i>Appendix D</i>	Duplication of transport services	265
<i>Appendix E</i>	Overview of interviews	267
	Summary	269
	Samenvatting	275
	About the author	281

1

INTRODUCTION

This introductory chapter describes the background, objectives, and approach of this thesis. The first section provides the backgrounds of this thesis by discussing the structural nature of transport infrastructure capacity scarcity and the desirability of an objective and rational approach towards slot allocation. Slot allocation, in particular higher-level slot allocation, is the subject of this thesis. The first section also gives a brief overview of the state-of-the-art of slot allocation research. This state-of-the-art appears to be fragmented, which leads to the conclusion that this thesis may contribute by formulating a theoretical framework to analyze slot allocation problems. The second section reviews the focus and perspective of this thesis. Our focus will be on single bottleneck problems, rather than network problems, while our perspective may be characterized as normative, rational, and substantive. Furthermore, this thesis focuses on selection slot allocation, rather than timetable-related slot allocation, which implies that the central issue is which traffic can use scarce available capacity at infrastructure bottlenecks. The next step is the formulation of the objective and the research questions of this thesis in section 1.3. Based on these research questions, the research setup is formulated. Given the objective, i.e. to formulate a theoretical framework to analyze slot allocation problems, this thesis is largely devoted to the formalization of the selection slot allocation problem, based on information from the literature and a limited empirical survey. Next, the main contributions of this thesis are reviewed, and this chapter ends with an overview of the structure of this thesis.

1.1 Background

The allocation of infrastructure capacity is a major issue in various transport infrastructure sectors. More capacity for newcomers at congested airports, the introduction of toll

lanes and HOV-lanes (lanes reserved for high occupancy vehicles) on freeways, and giving priority to passenger trains over freight trains during peak hours are just three examples of plans and ideas that have been discussed in parliaments as well as the media in Europe. Infrastructure capacity allocation is the subject of this thesis. This section discusses why infrastructure capacity scarcity is a structural problem and how slot allocation may help to improve capacity utilization and to reduce the negative effects of capacity scarcity such as traffic congestion. Furthermore, it briefly reviews current issues with respect to existing slot allocation regimes, which appear to be related to the liberalization of transport markets within the European Union.

1.1.1 Transport infrastructure capacity scarcity

Different transport modes correspond with different types of infrastructure elements. For instance, air traffic uses airports, road traffic uses roads and freeways, rail traffic uses railways and railway stations, and water traffic (navigation) uses waterways and ports. Although mode-specific transport infrastructure networks are different in appearance, they share a number of common characteristics. By definition, infrastructure is the basic structure enabling systems to work (see Cambridge International Dictionary of English, 1995), and hence infrastructure, including non-transport infrastructure, is of key importance to our society. Transport infrastructure systems facilitate transport of persons and freight, which is necessary for our socio-economic system to function.

A characteristic of transport infrastructure elements is their limited capacity. Infrastructure capacity defines the maximum flow or load of traffic it can serve within a period, depending on prevailing circumstances. Given infrastructure capacity scarcity, traffic cannot flow freely, resulting in delays on the network (queuing) or hidden delays (different departure time, different route). In our daily lives, we may encounter these effects in the form of road congestion, delayed planes, and inconveniently scheduled rail services.

Persistence of infrastructure capacity scarcity

Infrastructure capacity scarcity is a structural problem. Although it is generally possible to solve traffic congestion problems by providing more infrastructure, there are many examples of structural traffic congestion with a history of years or even decades. Furthermore, we should realize that transport infrastructure is generally costly, occupies a lot of space and often has negative environmental impacts. Consequently, providing sufficient infrastructure capacity to meet the peak-hour demand level is in many cases not the socio-economic optimal solution. Furthermore, even if infrastructure construction is decided to relieve traffic congestion problems, many years may go by between the decision to investigate possible solutions to relieve congestion problems and the opening of a new road, railway, runway, etc.

In some cases it is possible to increase infrastructure capacity without increasing the 'bandwidth' of infrastructure, i.e. by providing more tracks or lanes. For instance, improvements of the railway traffic control system may allow shorter headways and consequently higher traffic flows. Another example is that if environmental capacity constraints are binding, infrastructure capacity may be increased by reducing vehicle emissions. However, the potential for higher capacities is not always used, for instance because this would result in a much lower quality-of-service. Furthermore, the increase in capacity is always limited, and often insufficient to eliminate capacity scarcity altogether. Consequently, the potential for higher utilization of infrastructure does not eliminate the structural character of infrastructure capacity scarcity.

Capacity allocation

Given the persistency of infrastructure capacity scarcity, it is important that the existing infrastructure capacity is used in an optimal way. The effect of capacity scarcity on traffic flows depends on the mechanism of capacity allocation. Capacity allocation is the allocation of infrastructure capacity to potential infrastructure users, in advance or in real-time, in such a way that capacity constraints are satisfied. An example of capacity allocation is the alternate allocation of the usage of a road intersection controlled with traffic lights to traffic from different directions. By optimizing the capacity allocation system, the negative effects of infrastructure capacity scarcity on traffic can often be reduced. The allocation of infrastructure capacity can be optimized if available capacity is allocated to individual infrastructure users and if capacity is reserved in advance. This approach to capacity allocation is called slot allocation. The next sub-section elaborates on a number of current issues with respect to slot allocation.

1.1.2 Slot allocation

In the absence of traffic control measures, capacity scarcity manifests itself as congestion. An effective way to prevent structural congestion on the network is slot allocation. Slot allocation is a type of traffic planning, of which the key characteristic is that infrastructure users have to reserve a 'slot' on the network before departure. The total number of users admitted to each bottleneck per period is limited, depending on its capacity. For instance, railway timetables currently are designed to be conflict-free. Therefore, if all trains run on time, no train will have to queue before any bottleneck.

The main objective of slot allocation is to solve capacity conflicts beforehand in planning, and not on the network where congestion would be the result (see section 2.4). However, slot allocation may also be used as an instrument to serve other objectives, for instance to stimulate competition in the transport market by giving priority to entrants on the traffic market or to reduce externalities of traffic by giving priority to environmentally friendly traffic. Such objectives are considered as secondary objectives in this thesis.

Slot allocation is currently applied in the railway and aviation sectors. Slot allocation is a public issue in these sectors, because many railways and airports are used by different infrastructure users. Shared usage of infrastructures is currently the dominant situation in the road, railway, aviation and navigation sectors. However, in situations where infrastructure is owned and exclusively used by a single user, slot allocation is just an internal problem of this transport company. For instance, this is the case for metropolitan railway networks, such as the London Underground and the Métro and RER in Paris. Consequently, slot allocation is no public issue for congested London Underground railway lines.

Given the EU objective of equal access to infrastructure networks and the increasing scarcity of infrastructure capacity, the propagation of fairness and efficiency of slot allocation in the railway and aviation sectors has become a major issue. For instance, the airport slot allocation regime is being criticized for discriminating against entrants and for not stimulating efficient usage of airport slots, because it gives priority to existing flights and hence hampers the introduction of new flights by new competitors (Nordic Task Force on Airline Competition, 2002). Furthermore, a separation of infrastructure access and train operation activities has been realized in Europe in recent years to enable equal access of competing railway companies on the rail freight transport market (EC, 2001c). Consequently, fair and efficient slot allocation is required to facilitate fair competition in the transport market.

1.1.3 Scientific and societal relevance of slot allocation research

Slot allocation is an instrument to solve conflicting traffic demands corresponding with different infrastructure users. This includes the allocation of infrastructure capacity to different carriers, which is an increasingly important policy issue in Europe, given the liberalization policies of the European Union. Liberalization of transport markets has resulted in an increasing competition between carriers for infrastructure capacity. These carriers may be direct competitors, which may be unequal in size and power. An objective and rational approach towards slot allocation is crucial in these situations.

A theoretical framework on slot allocation can support rational decision-making on slot allocation. Such a theoretical framework, as well as research on the functioning of current slot allocation regimes, can be used to support the design of slot allocation procedures in practice. A solid theoretical framework, with solid definitions, may help to structure discussions about slot allocation, and to generate ideas about alternative approaches towards slot allocation. This theoretical framework should, amongst others, indicate the essence of the slot allocation problem, and provide a methodology to solve slot allocation decision problems.

State-of-the-art of slot allocation research

The current state-of-the-art of slot allocation research does not provide such a theoretical framework with respect to slot allocation. In fact, there is no well-established state-of-the-art of slot allocation research, because it is mainly confined to a few specific topics.

The establishment of markets for infrastructure capacity, including markets for slots, is one of the topics studied by transportation economics. Various authors have proposed auctioning of airport slots, for instance Grether *et al.* in 1981, and more recently Starkie in 1998. Airport slot pricing is discussed by Hong & Harker (1992), while a similar method has been developed for railway slots by the same authors (Harker & Hong, 1994). The economically optimum allocation of railway slots has been studied by, amongst others, Brewer & Plott (1996), and Lerz (1996). The economic literature on slot allocation will be reviewed in more detail in chapter 5.

Slot allocation has also received some attention by operations researchers. The optimization of traffic patterns given infrastructure capacity scarcity has been studied at three different levels. At the lowest level, the problem is to reschedule traffic in a situation where the timetable has already been established. An example is the operational scheduling of arriving flights at an airport (see e.g. Beasley *et al.*, 2000). At the intermediate level, the problem is to establish a timetable given infrastructure capacity constraints. For instance, Szpigel (1973) formulated a railway timetable optimization model with capacity constraints for a single line. However, the purpose of this model is to assist the timetable design process of carriers, not to solve the slot allocation problem, assuming that a single carrier uses the infrastructure. A related problem is the optimal dynamic allocation of car traffic given infrastructure capacity constraints, which has been studied by various authors (e.g. Lafortune *et al.*, 1993). The genuine slot allocation problem at the timetable level, given desired timetables of different carriers, has been studied by Brännlund *et al.* (1998) for (small) railway networks. Finally, the problem at the highest level is to select which desired transport services will be accommodated in a situation of capacity scarcity, and which are rejected. No references to this problem have been found in the literature.

A third category of research on slot allocation is policy-oriented, focusing on the question of how current slot allocation regimes may be improved. Although evaluations of slot allocation regimes can often be found in government papers such as the British consultation document on air transport (Department for Transport, 2001), these can rarely be found in the international scientific literature. Furthermore, policy analyses on slot allocation are largely limited to pragmatic solutions to local sector-specific issues, without attaining a broader perspective. An exception to the latter is a generic analysis of capacity management in the Netherlands by Ten Heuvelhof *et al.* (2001), describing and analyzing the capacity management regimes of 10 infrastructure sectors.

Conclusions

A theoretical framework to analyze slot allocation problems is required to enable the design of rational slot allocation decision processes. However, the current state-of-the-art of slot allocation research does not provide this framework. Current contributions are restricted to either general economic analysis, or operations research tools for specific sub-problems, or applied policy analyses. Furthermore, most operation research contributions focus on low levels of slot allocation, in particular dynamic rescheduling, while no references have been found in the literature to the highest-level slot allocation problem, i.e. the problem to select which desired traffic should be accommodated by the bottleneck under consideration, given that traffic demand exceeds traffic supply.

1.2 Focus and perspective

In the previous section, the topic of this thesis was introduced, i.e. slot allocation in situations of infrastructure capacity scarcity. This section elaborates on the focus and perspective of this thesis. It first explains the focus on selection slot allocation, and the single-bottleneck approach of this thesis. Furthermore, the normative-rational perspective is explained. Finally, it focuses on the differences between substantive and procedural perspectives on slot allocation problems.

1.2.1 Focus on selection slot allocation at single bottlenecks

This thesis focuses on slot allocation, and more specifically it focuses on selection slot allocation. Selection slot allocation is the highest level of slot allocation, focusing on traffic patterns rather than timetables or daily traffic planning (see section 2.4). Lower levels of slot allocation such as timetable-related slot allocation and daily traffic planning have received much more attention in the scientific literature than selection slot allocation, while the latter slot allocation decisions generally have a larger impact. Selection slot allocation decisions are the more fundamental, not because they are valid for a longer period, but primarily because they deal with the question of who can use infrastructure bottlenecks, while later slot allocation stages mainly focus on traffic timing decisions within the framework outlined by selection slot allocation decisions.

The main assumption of this thesis is that infrastructure supply is given, and consequently capacity constraints are given. The scope of this thesis is not restricted to the sectors currently applying slot allocation, i.e. the railway and aviation sectors, but also includes potential areas of application, i.e. the road and navigation sectors. However, because slot allocation is currently applied in the first two sectors only, these sectors receive significantly more attention in this thesis.

This thesis focuses on single bottlenecks rather than network problems. The main reason is that restriction to single bottlenecks facilitates a thorough analysis of the essence

of slot allocation, without complicating this with problems specific for networks. However, we will also discuss the issue how the approach developed in this thesis can be extended to networks of bottlenecks.

Infrastructure bottlenecks may be found in different sectors, and may be different in nature. This thesis studies bottlenecks in four infrastructure sectors, i.e. air, rail, road, and navigation. Furthermore, capacity constraints may have different origins, including primary traffic processes (e.g. minimum headways), traffic service processes (e.g. traffic control), and traffic externalities (e.g. noise regulations). In air traffic, runway capacity may be a main limiting factor, but also the number of gates at terminals or the passenger handling capacity of the terminals may be binding. Furthermore, the handling capacity of air traffic control centers is often a bottleneck for air traffic. Examples of bottlenecks in road traffic are urban freeways and freeway bridges and tunnels, as well as urban parking facilities. In railway networks, many infrastructure elements may be bottlenecks, including railway links, railway nodes, and railway stations (number of platforms). In contrast, navigation infrastructure is often not capacity-constrained, but possible candidate bottleneck types are ports (sea terminals) and locks. The identification and classification of bottleneck types will be discussed in the sixth chapter of this thesis.

1.2.2 Normative, rational, and substantive perspective

The perspective of this thesis can be characterized as being normative, rational, and substantive. Normative means that we primarily focus on the question of how slot allocation decisions should be made, as opposed to the empirical question of how slot allocation decisions are made. Rational means that the preferences of the decision-maker are assumed to be consistent with the principles of logic, taking into account available information that is relevant to the decision problem. Finally, the decision problem itself is the subject of a substantive perspective, while the decision-making procedure is the subject of a procedural perspective.

Normative rational perspective

Although the (empirical) question of how slot allocation is performed in practice is relevant, it is not the main issue analyzed in this thesis. We will primarily focus on the *normative* question of how slot allocation decisions ideally should be made. How slot allocation decisions are currently made is also relevant (see chapter 3), but is only of secondary importance. A normative perspective is required to advise slot allocation decision-makers. Given this normative perspective, it is reasonable to assume *rational* decision-making. Although rational decision-making often considered as an unattainable ideal, it is clear that prescribing irrational decision-making, for instance decisions that are inconsistent with the objectives of the decision-maker and the available information, is highly undesirable (see French, 1989). Empirical evidence suggesting that

decision-making is often not rational is not relevant here, given our normative perspective.

An exact definition of rational decision-making is not needed here; a short description of the concept will suffice. For a brief discussion of the concept of rationality, the interested reader may refer to Simon (1955), while French (1986) elaborates on the key relationship between uncertainty, information, and rational choice. The main characteristic of rational decisions is that the best solution is chosen given the available information. The deduction of the best alternative is made following the principles of logic. Furthermore, rational decisions are based on objective characteristics of the alternatives, while irrelevant information is not taken into account. However, information is generally not complete, for instance probabilities are estimated and not known exactly. Consequently, rational decisions may still partly be based on *a priori* assumptions (see French, 1986). Nonetheless, empirical information should always dominate *a priori* views.

Substantive versus procedural perspective

Slot allocation involves different decision types, including meta-decisions on the sequence and timing of decisions. In general, two perspectives on decision processes can be distinguished (Faludi, 1973). The object of a *substantive* perspective is the future situation, while a *procedural* perspective deals with the decision procedures that are or should be followed. Consequently, a substantive perspective on slot allocation focuses on the questions of what kind of slots should be specified and which slot should be allocated to which user, while the specification of the decision procedure, including the definition of decision timings, decision criteria, etc., is the main procedural problem. For example, a consultant who has been contracted by the Ministry of Transport to advise on the opening times of a railway bridge from a substantive perspective may propose a schedule of bridge opening times given an analysis of current traffic demand of barges and trains. From a procedural perspective, the consultant may propose that the Ministry of Transport first determines the minimum bridge opening frequency, while the railway companies decide on the exact timings of bridge openings by consensus.

Analyzing the slot allocation procedure from a substantive perspective is required as a background to decide rationally on slot allocation procedures, and therefore the dominant perspective in this thesis is substantive. However, the sequence of slot allocation decisions has to be specified first, and this requires a procedural perspective. For instance, if in the last example the suggestion of the second consultant is followed, the Department of Transport will first be faced with the substantive problem of determining the frequency of bridge openings, while the scheduling of bridge openings will be decided on in a later stage by the joint railway companies. Therefore, the structure of the slot allocation decision process will be determined first in this thesis, which implies a

procedural perspective, while the main part of this thesis will focus on the selection slot allocation problem from a substantive perspective.

1.3 Research setup and thesis contents

This thesis adopts mainly a normative, rational, and substantive perspective, as was argued in the previous section. However, empirical and procedural issues will also be discussed in this thesis. This section elaborates on the research setup, including the balance between empirical and normative research questions, and between procedural and substantive issues. Sub-section 1.3.1 first defines the general objective of this thesis, and then this objective is translated into specific research questions. Finally, these research questions are used to specify the research setup. The second sub-section summarizes the main contributions of this thesis, including scientific and societal contributions. Finally, the structure of this thesis is reviewed in sub-section 1.3.3.

1.3.1 Research objective and research approach

This thesis studies transport infrastructure slot allocation, focusing on selection slot allocation. The main objective of this thesis is to formulate a theoretical framework to analyze slot allocation problems. This theoretical framework includes the specification of slot allocation levels, of which selection is the highest level. It also includes the specification of desired characteristics of selection slot allocation. Furthermore, this theoretical framework includes a method to formalize selection slot allocation decision problems as optimization problems. This requires a proper definition of traffic supply, i.e. slots and capacity constraints, and an analysis of slot demand, i.e. slot allocation objectives and slot values. Finally, an objective of this thesis is to provide a solution method for the selection slot allocation decision problem.

Research questions

To structure the establishment of this theoretical framework regarding slot allocation, a number of research questions have been formulated:

- Which slot allocation levels may be distinguished, and how are these related to transport service planning? (chapter 2)
- What are the current and potential applications of slot allocation, what is the structure of current slot allocation regimes, and what are the main issues with respect to these slot allocation regimes? (chapter 3)
- What are the desired characteristics of selection slot allocation with respect to slot validity, slot holdership, slot margins, and the slot allocation procedure? (chapter 4)
- What types of capacity scarcity can be distinguished and for which type of capacity scarcity is selection slot allocation desirable? (chapter 5)

- How can traffic supply and demand be modeled, and what is the resulting selection slot allocation decision problem? (chapters 6 and 7)
- What is the best way to solve the selection slot allocation decision problem? (chapter 7)

Research approach

These research questions will be dealt with sequentially in this thesis, as is explained in more detail in sub-section 1.3.3. In order to answer these questions, the following steps have been taken. In the first place, a conceptual framework has been developed to provide definitions of slot allocation related concepts, key actors, etc. Furthermore, the current state-of-practice of slot allocation has been determined in an empirical study, which largely relied on official documents (laws, decrees) regarding slot allocation, but also on information from the literature and information from interviews. These interviews were part of a second empirical study, which aimed to retrieve information about the desired characteristics of slot allocation systems. To this end, interviews were held with experts working at a number of carriers, i.e. of Transavia Airlines, Tulip Air, Railion, and NS Reizigers. Additionally, experts from ProRail/Railned (the body responsible for railway capacity allocation in the Netherlands) have been interviewed (see Appendix E for an overview of all interviews). Given this information, an inductive approach has been followed to formulate the main desired characteristics of selection slot allocation. Induction means that desired characteristics are hypothesized, which are consistent with this information. Next, an approach has been developed to model traffic supply and demand, based on the current scientific literature. Analytical models of the selection slot allocation problem have been developed, and finally a method has been chosen to solve the resulting selection slot allocation problem. Given our objective of formulating a theoretical framework to analyze slot allocation problems, the formalization of the selection slot allocation problem is the key contribution of this thesis.

1.3.2 Main contributions of this thesis

In accordance with the main objective of this thesis, it contributes to the establishment of a theoretical framework to analyze slot allocation problems. In the first place, it provides a conceptual framework with concepts and their definitions, which can be used to describe and analyze slot allocation systems in a systematic way. Furthermore, this conceptual framework has been applied to analyze the current (European) practice with slot allocation in the air and rail sectors. Finally, main issues with respect to the design of slot allocation systems have been formulated, which appear to cover the dominant current discussions about desired improvements of slot allocation systems. The main issue appears to be the optimal balance between stability and flexibility. The theoretical framework established in this thesis may be used to find a balanced solution to these issues. For instance, separating selection and scheduling slot allocation, as is proposed in this thesis, enables both a high level of stability with respect to long-term slot alloca-

tion decisions, and a high level of flexibility with respect to short-term slot allocation decisions.

This thesis contributes to the state of the art of (analytical) transportation research by introducing a unified approach to formulate capacity constraints of different types of bottlenecks that applies all kinds of transportation sectors. This multiple user-class approach can deal with heterogeneous traffic, and it can be applied to formalize transportation problems such as the optimal selection slot allocation decision problem, which is another main contribution of this thesis. The most likely instances of the selection problem have been specified as 'standard' optimization problems, which implies that any exact (or approximate) solution algorithm described in the literature may be applied to these problems. However, it is argued in this thesis that a greedy approximation algorithm is most suitable for this type of problem. In situations with only a single capacity constraint, a basic greedy algorithm described in the literature can be used. Additionally, an extended greedy algorithm is developed in this thesis that can be applied in situations with two or more capacity constraints.

Finally, this thesis may contribute to discussions about slot allocation systems in practice. In the first place, the framework developed in this thesis may contribute to a transparent discussion about objectives, alternatives, complications, etc. In current discussions, the objectives are not always clear, and the arguments in favor of the chosen approach are often fragmented. A first contribution of this thesis is distinguishing different slot allocation problems at different levels. Explicitly distinguishing selection from scheduling slot allocation, as proposed in this thesis, enables decision-makers to make different choices with respect to slot validity, priorities, etc., at different levels of slot allocation. Furthermore, this thesis introduces a systematic approach to evaluate alternative slot allocation decisions, which may be used as a guideline to redesign slot allocation systems. The main strength of this approach is that it is relatively simple and easy to interpret, while it is also suitable for bottlenecks that are often considered as relatively complex. Contrary to the simple priority rules applied in current practice, it offers a systematic approach to deal with heterogeneity of traffic and with different types of capacity constraints applying to the same bottleneck.

1.3.3 Structure of this thesis

The structure of this thesis is as follows. After this introductory chapter, our focus is on the establishment of a conceptual framework. This conceptual framework specifies the position of slot allocation in the transportation system, provides definitions of key roles and key concepts such as capacity and slot allocation, and specifies types and levels of slot allocation. The third chapter focuses on the current practice with slot allocation. It first reviews the current and potential applications of slot allocation, and then analyzes the structure of current slot allocation regimes. Finally, it discusses a number of issues with respect to these slot allocation regimes. Drawing on the conclusions from current

practice, chapter 4 reviews the desired characteristics of selection slot allocation systems. The main topics are slot validity, slot holdership, slot margins, and the slot allocation procedure. Chapter 5 discusses the main characteristics of traffic markets requiring selection slot allocation. It describes the essence of capacity scarcity situations requiring slot allocation, and discusses how traffic supply and demand should be properly defined in the context of this thesis. Based on this traffic market specification, levels of capacity scarcity are defined, and economic approaches to deal with capacity scarcity are reviewed. While a fixed capacity value, which does not take into account heterogeneity of traffic, is assumed in chapter 5, a more diverse approach to traffic supply is introduced in chapter 6. It reviews various types of bottlenecks, and discusses how different types of capacity constraints can be formulated as a result of different types of traffic processes and traffic externalities corresponding with these bottlenecks. This is an important step for the formalization of selection slot allocation decision problems, which is the main topic of chapter 7. Chapter 7 first discusses the structure of rational slot allocation decision processes. Next, it discusses the specification of objectives, which is followed by the formal specification of four different instances of the selection slot allocation decision problem. Although in most cases exact optimization methods may be applied to solve these problems, this thesis suggests the application of approximation algorithms. In the final sections of chapter 7, these solution algorithms are reviewed, and additional issues such as the extension of this approach to networks with several bottlenecks are discussed. The implementation of the approach towards slot allocation proposed in this thesis is discussed in chapter 8. First, this approach is applied to a hypothetical metropolitan airport, including the application of the solution procedure proposed in chapter 7. Furthermore, the main consequences of application of this approach in practice are discussed. Finally, chapter 9 summarizes the main conclusions and recommendations of this thesis, and provides some suggestions for further research.

2

CONCEPTUAL FRAMEWORK

According to the specification of the thesis objectives in the previous chapter, this thesis aims to develop a theoretical framework with respect to slot allocation. The first step in the development of this theoretical framework is the specification of a conceptual framework with a number of key notions. As a starting point, a three-layer transportation system model is specified in section 2.1. This model is used to position the slot allocation problem, i.e. at the traffic market. Furthermore, the main roles that are relevant to the slot allocation problem are specified in this section and a conceptual model of slot allocation is presented. Section 2.2 focuses on the definition of capacity, a key concept in this thesis. Furthermore, determinants of capacity are briefly reviewed, and the relationship between capacity and quality-of-service is discussed. Section 2.3 briefly discusses capacity management, because this thesis focuses on a specific type of capacity management. Three levels of capacity management are specified, i.e. strategic, tactical and operational. Slot allocation, which is the subject of section 2.4, is a type of tactical capacity management. This section provides definitions of slot and slot allocation. Furthermore, three levels of slot allocation are introduced, corresponding with different types of traffic planning decisions. This thesis focuses on the highest level of slot allocation, i.e. selection slot allocation. In section 2.5, six types of transport services and three levels of transport service planning are identified which appear to be related with the three levels of slot allocation specified in the previous section. It is concluded that the desirability of each level of slot allocation depends on the types of transport services using the infrastructure. The final section overviews the concepts introduced in this chapter. These concepts will be used extensively in the next chapters. For instance, the levels of slot allocation introduced in section 2.4 will be used for the description of current practice with slot allocation in chapter 3, and the relationship between capacity and quality-of-service will be elaborated on in chapter 6.

2.1 The transportation system

This section first introduces a three-layer model of the transportation system. A multi-layer transportation system model is useful to understand the position of slot allocation within the transportation system. Furthermore, this type of model is used to identify key roles and actors within the transportation system. Finally, the transportation system model is translated into a conceptual model of slot allocation. The transportation system model introduced in this section is generic, i.e. it is independent of transport sector or geographical area.

2.1.1 Transportation system model

Various alternative layered transportation system models have been proposed in the literature. Schoemaker *et al.* (1998) proposed a three-layer model that has been used at Delft University of Technology for a number of years. A similar model was adopted by Van Nes (2002), distinguishing activities, transport services, and traffic services. Also similar, but based on different definitions of transport and traffic services, is the layer model introduced by Schaafsma (1997, 2001). Finally, Van Binsbergen & Visser (2001) proposed a transportation system model that is specifically suited for freight transportation, consisting of 7 layers.

In this thesis, a three-layer model is used (figure 2.1). This three-layer model is almost identical to the models proposed by Schoemaker *et al.* and Van Nes. These models are confined to three essential layers, while the other models introduce extra layers that are not relevant to the analysis of slot allocation.

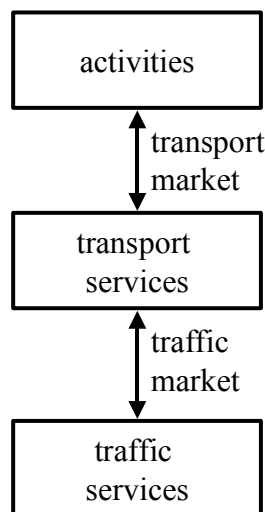


Figure 2.1: Multi-layer transportation system model

The transportation system model proposed in this thesis consists of the following layers (figure 2.1):

- activities;

- transport services;
- traffic services.

Between these layers, this model distinguishes markets where the supply of services by the lower layer and the demand for services by the upper layer are balanced (which is indicated by the arrows in figure 2.1):

- transport market;
- traffic market.

The next sub-section provides a characterization of these layers and markets.

2.1.2 Layers and markets

Layers are characterized by the services they provide to higher layers and/or require from lower layers in the transportation system. Table 2.1 summarizes the main characteristics of the layers of the transportation model, i.e. objects, functions, and actors. The objects are the entities required to deliver the functions, which again are required to produce the activities or services that characterize the layer. For instance, without persons, there is no travel, and travel is required to be able to perform activities at different locations. Finally, table 2.1 lists the main actor types associated with each layer.

Table 2.1: Objects and functions associated with layers

layer	objects	functions	main actor types
activities	<ul style="list-style-type: none"> • persons • freight 	<ul style="list-style-type: none"> • production • consumption • travel 	<ul style="list-style-type: none"> • end users: • traveler • shipper • receiver
transport services	<ul style="list-style-type: none"> • vehicles 	<ul style="list-style-type: none"> • carriage of passengers and cargo • propulsion • steering and routing • travel information 	<ul style="list-style-type: none"> • carrier
traffic services	<ul style="list-style-type: none"> • infrastructure elements 	<ul style="list-style-type: none"> • support of vehicle • guidance of vehicle • separation of traffic • energy supply • traffic information 	<ul style="list-style-type: none"> • allocation body • traffic controller • infrastructure provider

Activities

Activities are essential components of the socio-economic system. Because activities are performed at different locations, transport of goods and persons is required to link these activities. Goods may be the input for or output of production and consumption

processes, and production and consumption related activities are performed by persons. Consequently, the objects that are moved are either persons or freight.

The people or organizations whose activity patterns result in transport demands are referred to as *end users*. Three types of *end users* are distinguished here:

- travelers;
- shippers;
- receivers.

Persons using transport to combine activities or as an activity in itself are called *travelers*. *Shippers* are individuals or companies sending freight, while *receivers* are individuals or companies receiving freight.

Transport services

A transport service is the provision of functions required for transportation of persons and freight. Drawing on Van Binsbergen *et al.* (1995), four main transport functions can be identified:

- carriage of passengers and cargo;
- propulsion of passengers and cargo;
- routing of passengers and cargo;
- information to travelers and shippers.

The carriage function is to provide a vehicle or carriage conveying passengers or freight. Vehicles are the means used to produce these transport functions, together with the staff required to operate these vehicles. In this thesis, all kinds of transport means, including trains, planes, ships, etc. that are used to transport passengers and goods, are referred to as vehicles. Given that transport implies movement over a network, transport services include propulsion and routing of passengers and cargo, which is performed by steering and propulsion of vehicles. Finally, travelers and shippers generally need information about the transport service, including advance information (e.g. transport service timetables) and real-time information (e.g. on delays). Carriers use transport service timetables to announce the time and space characteristics of scheduled transport services, including both passenger and freight transport services.

A *transport network* is a network of transport services. Traffic is required to produce transport networks, implying that traffic patterns and transport networks are similar. However, traffic may also serve logistic purposes, for instance ferrying of empty vehicles, which means that some traffic does not have a directly associated transport service. In this thesis, actors responsible for providing transport services are referred to as *carriers*.

Traffic services

A traffic service is the provision of functions required for transportation and parking of vehicles. The following functions are distinguished in this thesis:

- support of vehicles;
- guidance of vehicles;
- energy supply to or traction of vehicles;
- separation of traffic;
- traffic information.

While vehicles are needed to provide transport services, infrastructure is needed to provide traffic services. *Infrastructure* facilitates traffic by producing traffic services. For instance, road infrastructure provides physical support to cars. Moreover, infrastructure may guide vehicles physically, e.g. in the case of rail traffic. Another type of traffic function is to ensure safe separation of traffic (traffic control). Full separation of traffic may be provided by a block system (rail traffic), while partial separation of traffic is provided by traffic lights (road traffic). Furthermore, the infrastructure may facilitate energy supply (e.g. contact wire) or even directly provide traction to vehicles (e.g. traction cable). Lastly, providing traffic information is also a traffic service.

Infrastructure networks are shaped by physical infrastructure elements such as roads, railways, stations, etc. However, the supply of traffic services may be spatially and temporally different from the corresponding infrastructure network. On one hand, the availability of infrastructure elements can be restricted in time. For instance, a landing strip can be closed at night, and hence it is not part of the traffic network for that particular period of time. On the other hand, air corridors constitute an infrastructure network that does not directly correspond to physical infrastructure elements. In this thesis, *traffic network* denotes a network of traffic services, which is largely shaped by the infrastructure network. The building blocks of traffic networks are *traffic network elements*. Finally, we define *bottleneck* as a traffic network element experiencing capacity scarcity, and an infrastructure element experiences capacity scarcity in a certain period if traffic demand exceeds traffic supply for that period.

Different types of actors are responsible for providing traffic services. An *allocation body* is responsible for slot allocation, i.e. it allocates infrastructure capacity among carriers. *Traffic controllers* are responsible for all traffic control tasks, including traffic information and the operation of infrastructure elements such as signals and switches. Finally, *infrastructure providers* are responsible for ensuring the availability of the infrastructure by coordinating infrastructure maintenance and construction activities.

Transport and traffic markets

In the transport system model, a market concept is used to describe the interactions between the three layers. The interactions between layers are denoted in figure 2.1 by

bi-directional arrows. Please note, however, that adaptations of supply and demand will generally not take place instantaneously, and may have different dynamics. Two markets are distinguished:

- transport market;
- traffic market.

In the transportation system, supply of and demand for transport and traffic services are balanced in the transport and traffic markets respectively. These markets may operate as 'free' markets in an economic sense, but it is also possible that these markets are, for instance, governed by administrative rationing systems. Generally, these markets do not function as 'normal' markets as described in economic theory. Supply and demand can be balanced in many ways other than using a market price, for instance by using a priority list (which user groups have priority in receiving a service) or by letting chance decide who will be supplied and who not. Of course, combinations are also feasible, for instance a reservation fee combined with a priority list.

Transport and traffic markets act as regulators of supply of and demand for transport and traffic. Requirements posed by the upper system levels determine the demand for transport and traffic services, and possibilities offered by the lower system levels determine the supply of transport and traffic services. The transport market balances transport supply and demand, while the traffic market balances traffic supply with traffic demand. Given that the production of transport services requires traffic services, transport supply implies traffic demand.

This thesis focuses on slot allocation, which is an approach to balance traffic supply and demand. The traffic services layer determines the supply of traffic services, including the allocation of infrastructure capacity to transport services. The demand for traffic services stems from the transport services layer. Therefore, the next sections focus on the interaction between the transport services and traffic services layers in general and the supply of traffic services in particular.

Regulation

Transportation systems are subject to transportation policy and regulation. Authorities at various geographical scales may interfere with transportation for various reasons, for instance to propagate safety, to reduce negative environmental effects, etc. Transportation policy and regulation may interfere with the transportation system at different levels, i.e. at each layer and market. For instance, the allocation of capacity by allocation bodies is often governed to a large extent by regulations imposed by the authorities. Sub-section 2.1.4 elaborates on the interaction between allocation bodies and authorities.

2.1.3 Roles within the transport and traffic services layers

In the previous section, a number of actor types have been introduced, including carriers, allocation bodies, and traffic controllers, which are associated with the transport and traffic services layers. These actor types are responsible for performing one or more *roles*. Roles are specific functions within the transportation system that can be performed by different actors (e.g. different organizations). Furthermore, the same role may be performed by different actors for different areas or sub-networks, and the size of these sub-networks may be different for each role. Consequently, a transport service provider may cooperate with different transport service producers, while a transport service producer may cooperate with different traffic service providers, while a transport service producer may cooperate with different traffic service providers.

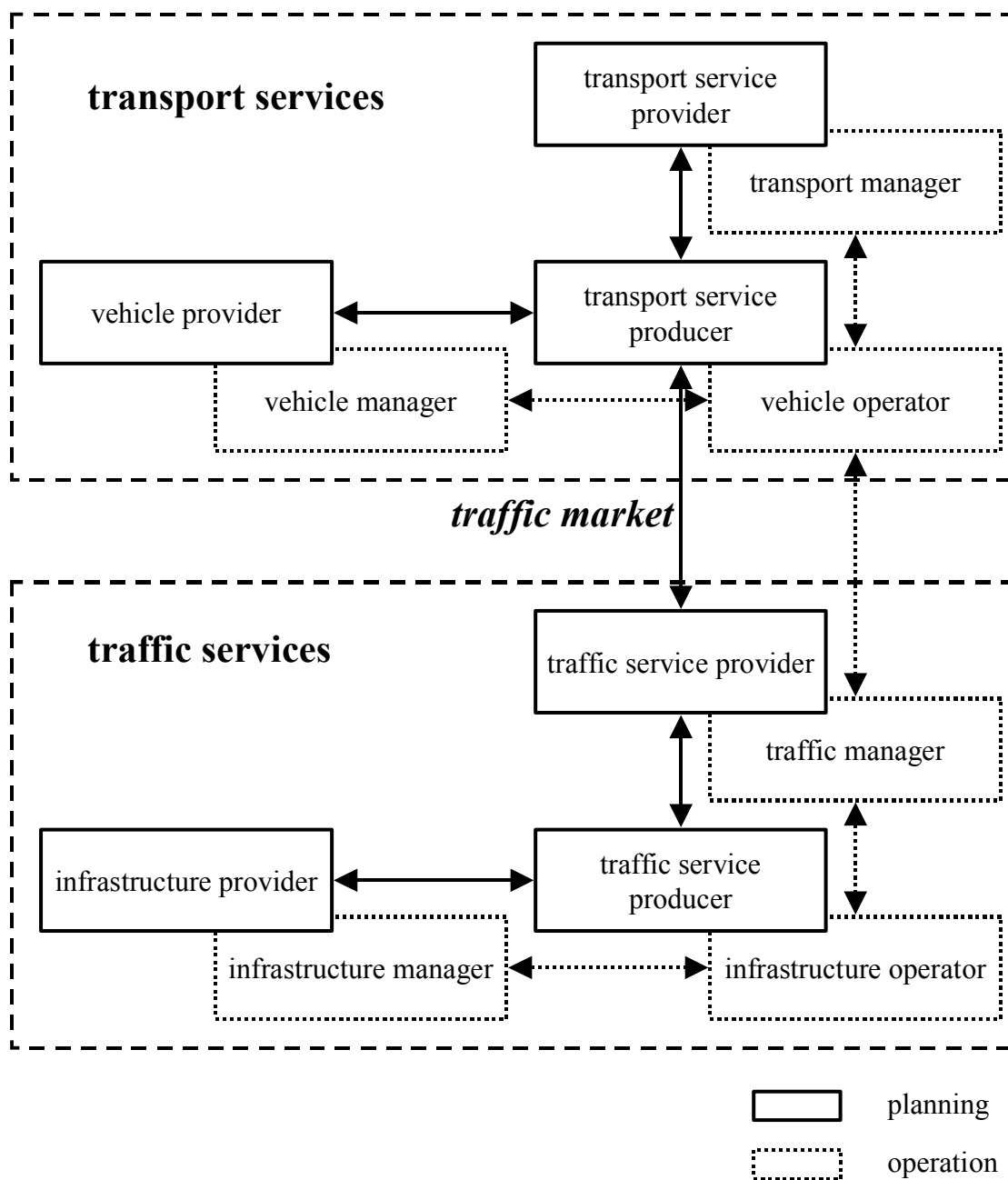


Figure 2.2: Roles within the transport and traffic planning layers

Figure 2.2 visualizes the roles that are distinguished in the traffic and transport layers, including the relationships between these roles. Planning roles are visualized as solid boxes, while operational roles are visualized as dotted boxes. For instance, transport service producers are concerned with planning the production of transport services, while vehicle operators are concerned with the production of transport services in daily practice. The 'upper' roles of each layer are responsible for providing services, while the 'lower' roles are responsible for production activities. Direct interactions between roles are indicated by arrows. For instance, slot allocation is an interaction between transport service producers requesting slots and traffic service providers allocating slots. However, as we will see in section 4.1, also 'higher' roles may request and hold slots, e.g. transport service providers or shippers. Not explicitly included in the figure are the interactions between planning and operations, e.g. the interaction between traffic service producers and infrastructure operators.

Some roles interact so strongly that separation of roles is hardly feasible in practice. In particular, production and service provision at the operational level are hard to separate because changes in production have direct consequences for the service level. For instance, train conductors and bus drivers are responsible for both passenger hosting (transport management) and operation of doors (vehicle operation). Similarly, railway traffic control centers are responsible for traffic management as well as infrastructure operation.

Roles within the transport services layer

Within the transport services layer, figure 2.2 distinguishes the following roles:

- transport service provider;
- transport manager;
- transport service producer;
- vehicle operator;
- vehicle provider;
- vehicle manager.

Transport service providers supply transport services to end users; their main responsibility is to sell transport services to end users and to determine the demand of transport services. *Transport managers* are responsible for the daily management of transport services, including informing travelers and shippers. The production planning of transport services is the responsibility of *transport service producers*. Transport service producers determine the production needs, for instance forecasting the future vehicle requirements. Furthermore, transport service producers determine the demand for traffic services, which means that they are responsible for reservation of slots if this is applicable. *Vehicle operators* are responsible for the operational production of transport services, i.e. operating vehicles. *Vehicle providers* are responsible for the availability of vehicles, which includes vehicle maintenance. In the British rail transport sector, for

instance, rolling stock companies (ROSCOs) lease trains to train operating companies (TOCs). Finally, operational maintenance of vehicles is the responsibility of *vehicle managers*.

Roles within the traffic services layer

Within the traffic services layer, figure 2.2 distinguishes the following roles:

- traffic service provider;
- traffic manager;
- traffic service producer;
- infrastructure operator;
- infrastructure provider;
- infrastructure manager.

Traffic service providers are responsible for the supply of traffic services to transport service producers. Their main task is to determine the available capacity and to allocate slots in the case of a slot allocation regime. The *traffic manager's* task is to provide real-time traffic information. Furthermore, traffic managers are responsible for real-time traffic management decisions. *Traffic service producers* are responsible for coordinating the availability of infrastructure, including the determination of the available capacity given scheduled maintenance activities. Furthermore, they are responsible for the planning of infrastructure construction. The real-time operation of infrastructure elements is the responsibility of *infrastructure operators*. Infrastructure operators implement the traffic management decisions of traffic service providers (planning) and traffic controllers (operations). Finally, *infrastructure providers* are responsible for the availability of infrastructure, which includes the organization of infrastructure maintenance and construction. Infrastructure maintenance and construction activities are carried out by *infrastructure managers*.

Actors

The division of roles among actors is situation specific. Every role may be performed by a different actor, but it is also feasible that a single actor is responsible for multiple or even all transport service as well as traffic service roles. A single actor may even be responsible for roles in different layers. In the previous sub-section, a couple of main actor types have been introduced. These actor types will frequently be referred to in the remainder of this thesis. Before reviewing the actors currently providing traffic services in the Dutch rail and air sectors, we briefly review which actor types are responsible for which roles. A *carrier* has both the roles of transport service provider and transport service producer, which means that carriers are also responsible for balancing transport supply with traffic demand. An *allocation body* is responsible for slot allocation, i.e. it has the role of traffic service provider in situations that slot allocation is applied. Finally, a *traffic controller* has both the roles of traffic manager and infrastructure operator, implying that traffic controllers can directly implement the consequences of traffic

management measures. To simplify the generic description of actor types, we assume in the remainder of this thesis that indeed transport service provision and production are integrated and performed by carriers, and that traffic management and traffic operation are also integrated and performed by traffic controllers.

To illustrate the current division of responsibilities in the traffic services layer, we review the current situation in the Dutch railway and aviation sectors. In the Dutch railway and aviation sectors, different actors are responsible for traffic services (see table 2.2). Until recently, most traffic service roles in the Dutch railway sector were performed by subsidiaries of Dutch Railways. As of January 1st 2003, these organizations (Railned, Verkeersleiding and Railinfrabeheer) are departments of an independent organization called ProRail. *ProRail/Railned* is responsible for slot allocation, *ProRail/Verkeersleiding* performs traffic management and operates the infrastructure, and *ProRail/Railinfrabeheer* is responsible for railway construction and maintenance. Traffic service provision is the joint responsibility of the ProRail departments *Railned* and *Railinfrabeheer*. Similarly, slot allocation at Dutch airports is the responsibility of *Airport Coordination Netherlands* (ACN), air traffic management is the responsibility of *Air Traffic Control Netherlands* (ATCN), and planning and management are largely the responsibility of the airport management companies, e.g. *Amsterdam Airport Schiphol Ltd.* In both sectors, operational maintenance activities are generally performed by private contractors, for instance *BAM NBM Infra*.

Table 2.2: Examples of current actors in the Dutch railway and aviation sectors providing traffic services

role	rail actor	air actor
traffic service provider	ProRail/Railned	ACN
traffic manager	ProRail/Verkeersleiding	ATCN
traffic service producer	ProRail	Schiphol Ltd.
infrastructure operator	ProRail/Verkeersleiding	ATCN
infrastructure provider	ProRail/Railinfrabeheer	Schiphol Ltd.
infrastructure manager	BAM NBM Infra	BAM NBM Infra

2.1.4 Conceptual model of slot allocation

In this section, a transportation system model has been specified, and roles within the transport and traffic layers have been defined. Based on the transportation system model, a conceptual model of slot allocation decisions is formulated here. The conceptual model presented by figure 2.3 gives an overview of the main actor types involved with slot allocation. It can be used as a basis to formalize the slot allocation decision problem, which is the subject of chapter 7. The downward arrows indicate the determination of the framework of choice options for the lower levels, while the upward arrows indicate the demand for decisions by upper levels.

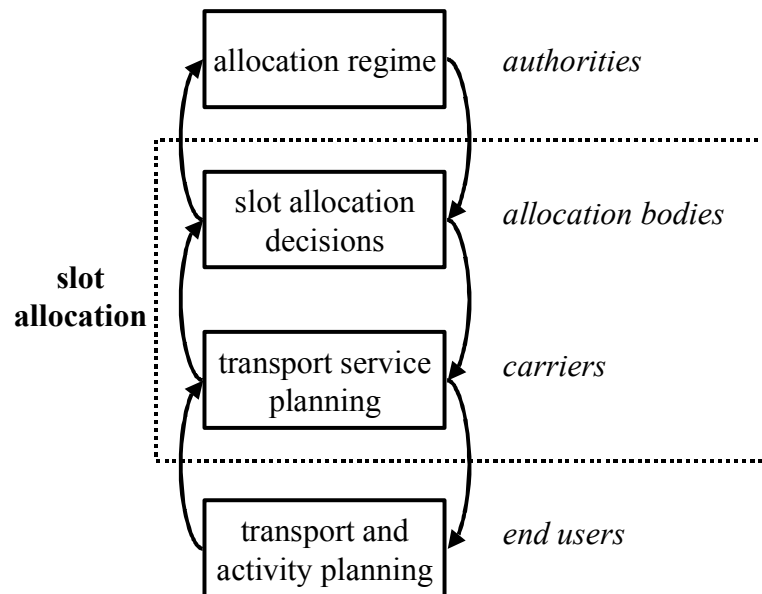


Figure 2.3: Conceptual model of slot allocation and related actors

The conceptual model of slot allocation (figure 2.3) focuses on the interaction between slot allocation decisions and transport service planning. It includes four decision levels, i.e. allocation regime, slot allocation decisions, transport service planning, and transport and activity planning. The corresponding actor types are authorities, allocation bodies, carriers, and end users. This conceptual model is similar to the transportation system model introduced in the first sub-sections of this chapter, i.e. the lowest three decision levels of the conceptual model correspond with the three layers of the transportation system model, however in inverted sequence.

Allocation bodies are responsible for slot allocation decisions, while carriers are responsible for planning the production of transport services. Carriers adapt the supply of transport services given slot allocation decisions, and allocation bodies allocate slots given the traffic demand expressed by carriers. Two secondary interactions are included in figure 2.3. The slot allocation regime, i.e. the regulations determining how slots are allocated, which carriers are entitled to request and hold slots, etc., is determined by authorities. Slot allocation decisions are adapted to the slot allocation regime, while authorities may adapt the slot allocation regime if slot allocation decisions are structurally different than desired. Another secondary interaction is the adaptation of transport patterns and activity patterns given the supply of transport services. Again, carriers may adapt transport service patterns given changes in transport demand.

2.2 Capacity

This section focuses on the definition of capacity, which is a key concept in this thesis. We first discuss which aspects are included in our definition of capacity, and review which type of capacity definition is appropriate for this thesis, given the stochastic nature of traffic and traffic service processes. In the second sub-section, the dependency

of capacity on traffic conditions and external factors is discussed. Finally, the relationship between capacity and quality-of-service is reviewed in the third sub-section.

2.2.1 Definition of capacity

In this thesis, the capacity of traffic network elements is simply referred to as *capacity*. Capacity is our key measure of traffic supply, and therefore the definition of capacity is the main subject of this section. The first step of this section is to investigate the essence of the capacity concept. To this end, we first review a generic and a sector-specific definition of capacity. In general, *capacity* can be defined as 'the total amount that can be contained or produced' (Cambridge International Dictionary of English, 1995, p. 191). A specific definition of road infrastructure capacity is provided by the Highway Capacity Manual. According to the HCM, *capacity* is 'the maximum hourly rate at which vehicles reasonably can be expected to traverse a point or a uniform section of a lane or roadway during a given time period under prevailing roadway, traffic, and control conditions' (TRB, 2000, p. 2-2). Because the first definition is too generic and the second is too specific, a new definition will be derived in this section.

Maximum flow or maximum load

In the HCM definition, capacity is defined as a maximum flow. *Maximum flow* is the maximum number of vehicles that enter or exit an infrastructure element during a certain period. According to the Cambridge definition, capacity can also be a maximum amount that can be contained, i.e. a maximum load. *Maximum load* is the maximum number of vehicles that infrastructure elements may contain. As will be argued in chapter 6, whether capacity can best be defined as a maximum flow or a maximum load depends on both the type of traffic process under consideration and the specification of periods. Therefore, capacity may be either a maximum load or a maximum flow in this thesis.

While in the HCM definition maximum flow is defined as a rate, it is defined as a number (of vehicles) in this thesis. The main reason is that this results in the same measure for maximum flow and maximum load. Hence, capacity κ restricts the maximum traffic volume on a certain section of time-space. Flow capacity $\kappa_{s,p}$ restricts the traffic volume q which can pass location s within period p :

$$q_{s,p} \leq \kappa_{s,p} \quad (2.1)$$

Similarly, load capacity $\kappa_{a,t}$ restricts the maximum traffic volume q that can be contained simultaneously in area a at time t :

$$q_{a,t} \leq \kappa_{a,t} \quad (2.2)$$

Figure 2.4 illustrates the difference between maximum flow and maximum load, using vehicle trajectories in a space-time diagram. Maximum flow capacity constraints relate

to the maximum number of vehicles passing location s in period p , while maximum load capacity constraints relate to the maximum number of vehicles present in area a on time t .

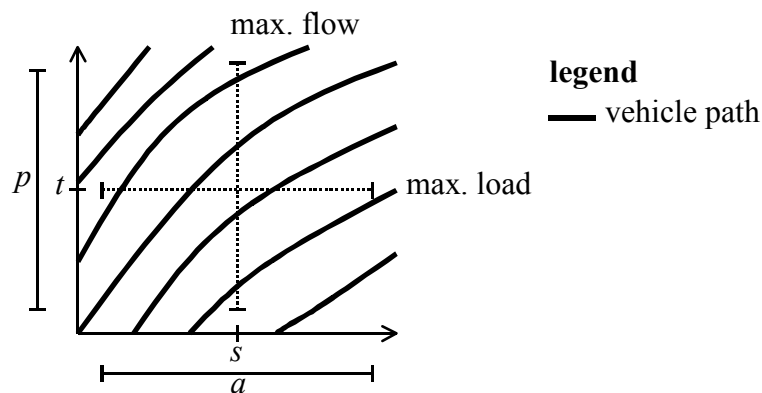


Figure 2.4: Maximum flow and maximum load

Heterogeneity of traffic and condition-dependency of capacity

Based on an analysis of infrastructure capacity notions in four transport infrastructure sectors, Kreutzberger & Vleugel (1992) conclude that, as a rule, capacity is condition-dependent. For instance, freeway capacity depends on the percentage of trucks and buses, because these types of vehicles are larger and have a lower maximum speed (TRB, 2000). Similarly, lock capacity depends on the types of ships that are to be locked through. However, according to the HCM definition of capacity, capacity is not only dependent on traffic conditions, but also roadway and control conditions. For instance, capacity decreases given adverse weather conditions such as rain and fog. Furthermore, the capacity of controlled road intersections depends on the traffic control system that is applied.

A related aspect is the possible dependency of capacity on the level of heterogeneity of traffic. For instance, runway capacity depends on the percentage of instances that small planes follow large planes, due to the wake vortex turbulence caused by large planes and the sensitivity to these wake vortices of small planes (Nolan, 1994). Another example is the dependency of capacity on heterogeneity with respect to speed, in situations with severe overtaking limitations. For instance, De Ruiter *et al.* (1990) illustrate the dependency of capacity on speed differences for double-track railway links. In figure 2.5, space-time trajectories are shown in situations with capacitated traffic. It appears that the maximum flow given only express trains or only local trains may be several times higher than the maximum flow given an alternating mix of express and local trains. Finally, the effect on capacity of a small traffic network expansion in the form of an intermediate pass-track is illustrated in right-hand diagram of figure 2.5.

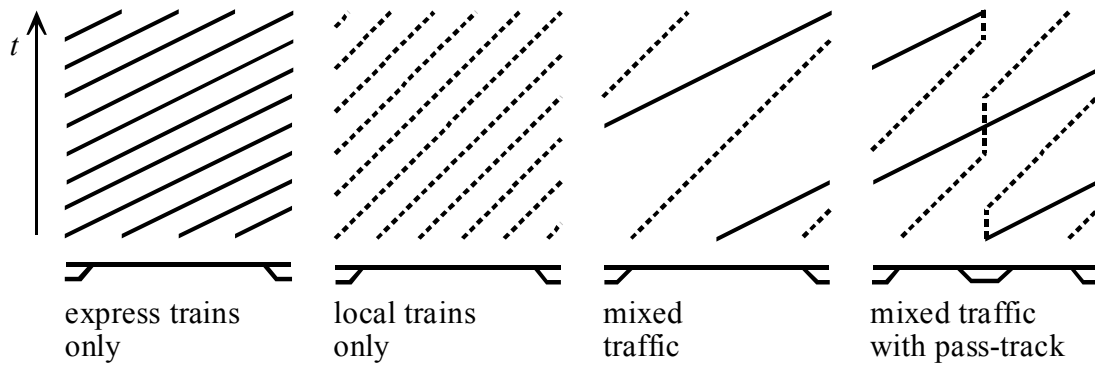


Figure 2.5: Capacitated rail traffic given different transport service concepts (De Ruijter et al., 1990).

In this thesis, the condition-dependency of capacity is incorporated by assuming standard composition-of-traffic and standard traffic conditions, which enables the specification of a single standardized capacity value (see chapter 6). This implies that in situations where traffic conditions or composition-of-traffic are significantly different than assumed for the specification of capacity, maximum traffic volumes may be higher or lower than the standardized capacity.

Capacity and quality-of-service

In the example of figure 2.5, capacity is not only dependent on composition-of-traffic, but also on minimum quality-of-service that is considered acceptable. *Quality-of-service* is a generic notion describing the level of correspondence between desired and realized traffic characteristics. For instance, the highway capacity manual uses density as main characteristic to specify the level-of-service of car traffic flows (TRB, 2000). Another frequently applied quality-of-service variable is expected delays, e.g. by Horonjeff (1994) to specify airport capacity. Finally, heterogeneity is also a quality-of-service variable as far as the correspondence between desired and realized traffic characteristics is concerned. In each of these examples, capacity depends on minimum quality-of-service. A higher maximum traffic volume can only be attained given a higher traffic density, a higher delay probability, or by homogenizing traffic characteristics. The relationship between capacity and quality-of-service will be discussed in more detail in sub-section 2.2.3.

Maximum, actual, and available capacity

Various factors determining capacity have a random component. For instance, the composition-of-traffic is usually not known exactly in advance, and the traffic context includes stochastic variables such as the actual weather conditions. Furthermore, other factors determining capacity, such as minimum headways, have a random component, for instance due to differences in driving styles (Minderhoud et al., 1997). Finally, breakdown and repair of infrastructure elements, and of vehicles causing infrastructure elements to be blocked, are random processes resulting in stochastic variations of capacity (see e.g. Bär et al., 1988).

In the context of the stochastic nature of capacity, we may distinguish different types of capacity. If an infrastructure element is functioning with the maximum effort that is theoretically feasible and if no service time is lost between the arrival of new vehicles, its maximum flow or maximum load corresponds with the *maximum capacity* (maximale Leistungsfähigkeit according to Bär *et al.*). Bär *et al.* (1988) also introduced the *actual capacity* of infrastructure elements (*momentane Leistungsfähigkeit*), which is a stochastic capacity variable with a value range between zero and the maximum capacity. A zero actual capacity corresponds with the complete failure of the infrastructure element, for instance the failure of a railway switch, or the blocking of a runway due to severe weather conditions. However, maximum capacity and actual capacity are not suitable as capacity notions for traffic planning purposes. The maximum capacity is a theoretical maximum, which is generally higher than the actual capacity, and the actual capacity is the realized level of capacity, which is unknown at the planning level. Therefore, a third capacity definition is introduced here, i.e. available capacity. The *available capacity* of infrastructure elements is the capacity that is assumed to be available for planning purposes, i.e. it is the number of vehicle paths that can be assigned a slot. Available capacity is the type of capacity considered in this thesis, and whenever 'capacity' is used without further explanation, this refers to the available capacity.

Our definition of capacity

Given the considerations outlined in this section, (available) *capacity* is defined here as the maximum planned traffic volume on a traffic network element, either simultaneously or per unit of time, given the acceptable minimum quality-of-service, assuming standard composition-of-traffic and traffic context. The symbol used to denote capacity is κ . In this section, indices a , p , s , and t have been introduced to indicate which area, period, location, or moment is considered. In the remainder of this thesis, however, such indices will not be used, unless several areas, periods, etc. are considered simultaneously.

In this thesis, allocation bodies (having the role of traffic service providers) are assumed to determine the desired minimum quality-of-service and the corresponding capacity values. As will be illustrated in the next sub-section, the capacity specification may include both physical limitations and externally imposed limitations on traffic volumes. Consequently, traffic service providers have a relatively large freedom to determine these values.

A related concept is the transport capacity of traffic network elements. Transport capacity is the maximum transport volume (i.e. persons, goods), rather than vehicles, that may be conveyed on traffic network elements. The transport capacity of a bottleneck depends on the (traffic) capacity of this bottleneck, but also on the types of vehicles being used. Different types of vehicles may have a different maximum load, and hence com-

position of traffic with respect to maximum vehicle load is one of the determinants of transport capacity.

2.2.2 Determinants of capacity

The capacity of traffic network elements may be restricted by various factors. Primary restrictive factors include the bandwidth of traffic network elements, and the required minimum safe headways. The *bandwidth* of a traffic network element is the number of available parallel *traffic servers*, which is the smallest possible unit of a traffic network element. Examples of traffic servers are platforms (railway stations), tracks (railways), runways (airports), berths (ports), and lanes (freeways). Secondary restrictive factors include the limited handling capacity of traffic control centers, and environmental capacity.

Primary factors

The maximum load of traffic elements determines the maximum number of vehicles that may use it simultaneously, which largely depends on its bandwidth. For instance, a one-lane road of 100 meters long may contain 20 queuing cars if the minimum headway distance, given queue speed, is 5 meters. Furthermore, a large platform in a railway station can often accommodate either one long train or two small trains simultaneously. Finally, the number of ships that can be locked through simultaneously depends on both the length and width of ships.

Traffic flows are restricted by physical limits due to the fact that safe and stable traffic flows require vehicles to be separated with sufficient headways (see also section 6.2). Minimum headways not only depend on the type of traffic processes under consideration, but also on the applied traffic control systems, safety norms, maximum acceptable delays, etc. Furthermore, the maximum number of vehicles served per unit of time can also be restricted by other service processes. For instance, the flow capacity of a terminal platform or gate depends on the time required for boarding and alighting, for opening and closing of doors, and the time needed to check whether the doors have closed properly. Finally, the service time of parking accommodations and buffers is determined by the period that a vehicle needs to be buffered or parked.

Secondary factors

Capacity can also be limited by secondary factors. Three types of secondary factors are distinguished in this thesis:

- traffic service processes;
- traffic control processes;
- environmental constraints.

In the first place, traffic service processes may limit capacity in a number of different ways. For instance, the headway between vehicles may be reduced because of time required to switch an infrastructure element between two states. Examples of different states of switches are straight and branching (railway switch), high and low level (lock), or open and closed (bridge).

In the second place, traffic control processes may be binding, because traffic control systems may have a limited handling capacity. For instance, the maximum workload of traffic controllers is limited and traffic control centers often have limited staff, implying that possibly some peaks in traffic demand cannot be accommodated. Increasing the number of controllers or control centers is not always a solution, because this increases communications between traffic controllers. In the aviation sector, for instance, every traffic controller is responsible for his own sector of airspace, and therefore every airplane in a traffic control sector may be seen as occupying a place in a virtual infrastructure server (traffic control server). Indeed, European air traffic is currently limited by the handling capacity of air traffic control centers (Majumdar & Polak, 2001).

Finally, traffic externalities may also result in the imposition of capacity constraints. Environmental constraints are administrative traffic restrictions that are imposed to restrict the effects of traffic on the environment. Environmental constraints are related to either noise, pollution, or external hazards. Authorities may limit noise emissions to protect residents living near airports, railways, etc. For example, noise regulations restrict the maximum yearly number of flights that is allowed at the Dutch national airport Schiphol and the frequencies of heavy trains at some Dutch railway links (Stout, 2001; Railed, 2000a). Furthermore, authorities may impose limits on hazards to the public due to traffic accidents such as collisions, crashes, and derailments.

In this sub-section, we have briefly reviewed a number of different factors that may restrict capacity. The identification of types of bottlenecks and associated capacity constraints will be discussed at length in chapter 6. This section continues with a discussion of the relationship between capacity and quality-of-service.

2.2.3 Capacity and quality-of-service

In sub-section 2.2.1 it has been noted that capacity values are determined by traffic service providers, depending on the minimum quality-of-service that is considered acceptable. In general, capacity and quality-of-service are negatively related, i.e. higher traffic volumes can only be attained at the cost of a lower quality-of-service. Which minimum quality-of-service is required, or which balance between capacity and quality-of-service is desired, is a normative issue. Given a properly specified objective function, it is possible to determine the optimal balance between capacity and a quality-of-service variable. An example is the balance between capacity and expected delays of railway traffic analyzed by Hertel (1994).

According to Morlok, (1978), quality-of-service may include various aspects, including speed, reliability, safety, and comfort. This sub-section reviews a few examples of the negative relationship between capacity and quality-of-service. This thesis focuses on the following quality-of-service variables:

- reliability;
- travel time;
- scheduling freedom.

Reliability

According to the argumentation of sub-section 2.2.1, actual capacity is a stochastic variable. Given the stochastic nature of traffic processes, capacity is related to reliability. A possible quality-of-service dependent definition of capacity κ is that it is the maximum traffic volume that may be realized with probability r , given stochastic capacity K (Persaud and Hurdle, 1991):

$$r = P(K \geq \kappa) \quad (2.3)$$

We refer to this probability as the reliability level. Figure 2.6 illustrates the relationship between the specification of the available capacity κ , given stochastic capacity K , and reliability. The curve exemplifies the probability density p as a function of capacity. In figure 2.6, the reliability level given the choice of available capacity κ can be derived by estimating which share of the area under the probability density curve is at the right hand side of the available capacity (the shaded area in figure 2.6). Clearly, increasing κ corresponds with a reduction of the shaded area, and hence in a reduction of reliability.

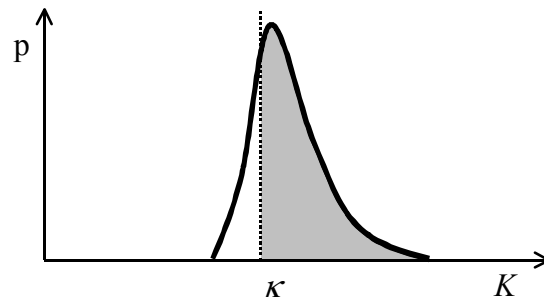


Figure 2.6: Specification of available capacity given stochastic capacity

Travel time

Given its prominent place in the literature, travel time may be regarded the most important quality-of-service variable. The relationship between traffic volume and maximum delay has been studied extensively in various situations. Both traffic arrival processes and the service processes of infrastructure elements may be modeled as random processes. For instance, Wakob (1985) has developed a generic approach to analyze random queuing processes in railway bottlenecks, and Jain & Smith (1997) have proposed a method to model road traffic processes as random queuing processes. Furthermore,

Horonjeff & McKelvey (1994) review a number of simulation studies regarding queuing processes on airports.

Given the random nature of queuing processes, Horonjeff & McKelvey (1994) propose a delay-related capacity definition, given the (normative) choice of the acceptable average aircraft delay level (see figure 2.7). In this figure, $E(d)$ denotes expected delays, and q denotes the volume of traffic arriving at the bottleneck in the period under consideration. Figure 2.7 illustrates the relationship between the flow rate and the expected steady state queuing delay. The expected waiting time in the static state is a monotonically increasing function of the arrival rate to capacity ratio, which approaches infinity when the arrival rate approaches the service rate. Please note that given our definition of capacity, the available capacity κ will be lower than the service rate of the bottleneck, which corresponds with the maximum capacity. The available capacity is defined given the condition of an acceptable quality-of-service, which implies that expected delays should be restricted to an acceptable level.

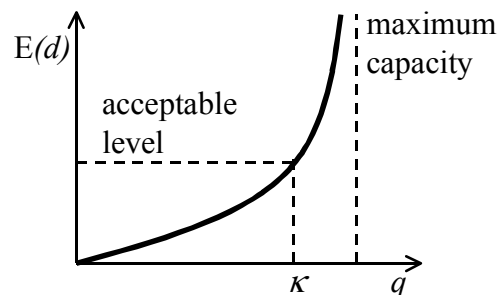


Figure 2.7: Delay-related definition of capacity (Horonjeff & McKelvey, 1994)

Travel time also depends on the maximum speed that can be realized. Regarding speed as an independent variable, maximum flow is generally speed-dependent. Figure 2.8 exemplifies how maximum flow and speed are commonly related. Maximum flow q^{\max} , as a function of speed v , attains its maximum at the optimum speed v^{opt} . The maximum speed, i.e. the speed that is attained if traffic can flow freely, is denoted by v^{\max} . Similar relationships have been specified for road traffic (e.g. TRB, 2000), rail traffic (e.g. Schwanhäusser, 1994), and inland navigation (Groenveld, 1999). The spatial separation of air traffic is not speed dependent, and therefore the speed flow relationship of air traffic has a different shape (see Nolan, 1994).

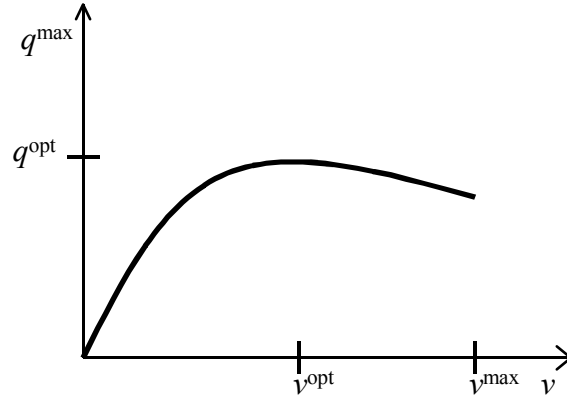


Figure 2.8: Typical form of speed-flow relationship

A consequence of the speed-flow relationship shown in figure 2.8 is that, given that maximum speed v^{\max} is higher than v^{opt} , a balance has to be found between accepting a lower capacity than q^{opt} and accepting a lower speed than v^{\max} . This is again an example of a negative relationship between capacity and a quality-of-service variable.

Another example is the relationship between noise emissions and speed. Noise emissions are positively correlated with speed. Consequently, given that noise and toxic emissions are bounded by an environmental capacity constraint, capacity may be increased by reducing maximum speed. For instance, speed limits have been imposed to a freeway running through a quarter of the Dutch city of Rotterdam with the objective of reducing noise and toxic emissions to acceptable levels.

Scheduling freedom

Finally, we review the relationship between planned traffic volume and scheduling freedom. This relationship is relevant in situations where slot allocation is applied, and traffic is scheduled in advance, for instance in a timetable. The available capacity determines which traffic volume has to be scheduled on a certain traffic network element in a certain period. This scheduling step implies that slots are allocated with a higher level of precision, i.e. with smaller temporal margins. When traffic volume q approaches the number of available scheduling slots κ^s , the probability increases that timetables of different carriers are in conflict with each other.

A possible definition of scheduling freedom is the probability P^{free} that any scheduling period p still has one or more free scheduling slots (see sub-section 2.4.4). We assume that each scheduling slot has equal probability of being free, and that traffic is homogeneous. The selection period under consideration consists of j scheduling periods of equal size. The total number of free scheduling slots n is given by:

$$n = \kappa^s - q \quad (2.4)$$

We can model the number of free scheduling slots within a scheduling period p as the number of successes in a random sample of size n drawn from a population of size κ^s .

Hence, the number of free scheduling slots in a scheduling period follows a hyper-geometric distribution. However, if n is relatively small compared with the population size κ^s , we may approximate this hyper-geometric distribution with a binomial distribution. As a result, the probability of free scheduling slots P^{free} can be approximated as:

$$P^{\text{free}} \approx 1 - \left(\frac{j-1}{j} \right)^n \quad (2.5)$$

Given equation 2.5, scheduling freedom may be increased by increasing n , i.e. increasing the difference between κ^s and q , which may be achieved by decreasing the maximum traffic volume q , but also by increasing the maximum number of scheduling slots κ^s , or decreasing the number j of scheduling periods, i.e. increasing the margins of scheduling slots. The relationship between traffic volume q and scheduling freedom P^{free} , assuming given values of κ^s and j , is illustrated by figure 2.9.

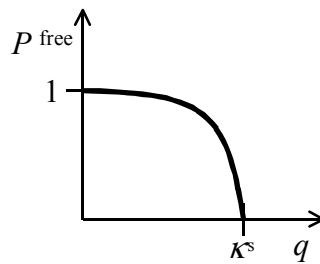


Figure 2.9: Relationship between traffic volume and scheduling freedom

Capacity specification

In this sub-section, the relationship between capacity and quality-of-service has been briefly reviewed for a few variables. Each possible capacity value is associated with a certain minimum quality-of-service. Hence, given quality-of-service criteria, and given that the relationship between capacity and quality-of-service is known, the resulting capacity may be derived. Section 4.3 elaborates on the question of how the minimum acceptable quality-of-service may be derived. The optimization of the balance between traffic volume and quality-of-service is elaborated on in chapter 5. Finally, section 6.4 elaborates on the specification of capacity constraint parameter values given quality-of-service criteria. This chapter continues with a brief overview of capacity management levels.

2.3 Capacity management

In section 2.2, capacity has been defined and the issue of infrastructure capacity scarcity has been introduced. Given infrastructure capacity scarcity, there are different ways to mitigate the effects. *Capacity management* is the application of measures to control the allocation of infrastructure capacity to alternative users and uses. Ten Heuvelhof & Kuit

(2001) note that this may also include non-decisions and decisions that do not explicitly aim at managing infrastructure capacity.

As in many infrastructure sectors, three different levels of transport infrastructure capacity management are distinguished, i.e. strategic, tactical and operational (see e.g. Watson, 2001; Ten Heuvelhof & Kuit, 2001; Schaafsma, 2001). The main differences between these three capacity management levels are summarized by table 2.3.

Table 2.3: Capacity management levels

	spatial scope	temporal scope	main actor
strategic	generic	long-term plan	authorities
tactical	specific	specific plan	traffic service providers
operational	specific	real-time	traffic controllers

Strategic capacity management

The strategic level of capacity management deals with long-term questions of capacity supply. Strategic capacity management is defined as capacity management through generic long-term decisions on traffic market access. Furthermore, tactical and operational capacity management objectives and strategies may be determined at the strategic capacity management level. In general, national governments are responsible for strategic capacity management decisions. An example of strategic capacity management is that in Europe only carriers holding specific permits have access to the railway network (Koolstra, 2001a). Furthermore, bilateral agreements between countries determine which carriers have access to the mutual international airports.

Tactical capacity management

Tactical capacity management is capacity management through time and space specific decisions on traffic network access that are made in advance. Traffic service providers are responsible for tactical capacity management. An example of tactical capacity management is the restriction of railway bridge operation times to specific time intervals. Given these restricted intervals, both land and water traffic may adapt their traffic patterns. Another example is the introduction of time and location specific infrastructure pricing, which acts as a financial incentive to avoid rush hours on congested highways, railway lines or airports. Finally, an example of a more rigid kind of tactical capacity management is slot allocation, which is the subject of section 2.4. For instance, airlines need to have a slot to operate an air service to fully coordinated airports such as Heathrow airport in London and La Guardia airport in New York.

Operational capacity management

Operational capacity management is capacity management in real-time, reacting on actual traffic situations. While strategic and tactical capacity management is proactive

(i.e. anticipating traffic), operational capacity management is reactive, i.e. managing traffic as it takes place. Traffic controllers are responsible for operational capacity management. Please note that operational capacity management systems may operate automatically as well as manually. For instance, traffic-responsive traffic lights can be used to restrict infrastructure access in real-time. Another example is operational railway capacity management. In case of delays or other deviations from the traffic plan, traffic controllers may decide to let a delayed train proceed and delay another train, or do the opposite.

2.4 Slot allocation

Slot allocation is the focus of this thesis. The main characteristic of slot allocation is that traffic access to the infrastructure network is restricted. This restriction is made in advance and is time and space specific, which implies that slot allocation is a type of tactical capacity management. This section first briefly reviews the objectives of slot allocation. Furthermore, this section provides definitions of key concepts such as slot and slot allocation, as well as definitions of concepts such as vehicle path, slot period, and bandwidth. Finally, it introduces two different types of slot allocation (exclusive and shared) and three different levels of slot allocation (selection, scheduling, and *ad hoc* slot allocation).

2.4.1 Objectives of slot allocation

Slot allocation may serve various purposes. For instance, slot allocation may help to avoid traffic congestion. In general, overall transportation costs may be reduced by adapting routes and departure times as an alternative to congestion on the network (see e.g. Arnott *et al.*, 1990). Furthermore, queues may block upstream links, which may result in a lower capacity being available for traffic that does not have to pass the original bottleneck. Other types of policy objectives may also be implemented using slot allocation, for instance maximization of competition between carriers, or minimization of noise hindrance. Maximization of competition means that capacity is allocated in such a way that competition between transport services is stimulated. Minimization of noise hindrance may be stimulated by, for instance, giving relatively 'quiet' traffic a higher priority or value in the slot allocation process.

Section 7.2 elaborates on the specification of objectives for the slot allocation problem, distinguishing between primary and secondary objectives. Objectives that are only indirectly related to the traffic market, such as maximization of competition or minimization of noise, are considered secondary objectives in this thesis. The most likely primary objective of slot allocation is to maximize infrastructure utilization. This objective will be elaborated on in section 7.2. Maximizing infrastructure utilization means maximizing the benefits - as perceived by the allocation body - associated with a slot allocation, given limited available capacity.

2.4.2 Definition of slot allocation

The term *slot allocation* has two constituent parts, i.e. slot and allocation. We first discuss the definition of slot. In general use, a *slot* is 'an amount of time which is officially allowed for a single event in a planned order of activities or events' (Cambridge International Dictionary of English, 1995, p. 1355). Applying this to traffic, a traffic event is the usage of an infrastructure element by a vehicle, or more formally an intersection in time-space of a vehicle path and an infrastructure element. For instance, 'slot' is currently used in the air traffic sector to denote the departure or arrival time-window of an airplane. Slots may overlap, but every individual vehicle path corresponds with an individual slot. In this thesis, *slot* is defined as a time-space domain on an infrastructure element designed to facilitate a specific vehicle path with a certain minimum quality-of-service. This definition contains two notions that will be explained later in this section, i.e. time-space domain and vehicle path. The concept of quality-of-service has already been introduced in section 2.2, and, for instance, may include maximum travel time, reliability, and minimum and maximum speed.

In the definition of this thesis, slots are elementary slots, i.e. specified on a single traffic network element, which is in accordance with the single-bottleneck approach of this thesis. However, a sequence of elementary slots (or *network slot*) is needed to facilitate an entire vehicle path. Figure 2.10 provides an example of a network slot with the corresponding vehicle path, with s denoting space and t denoting time.

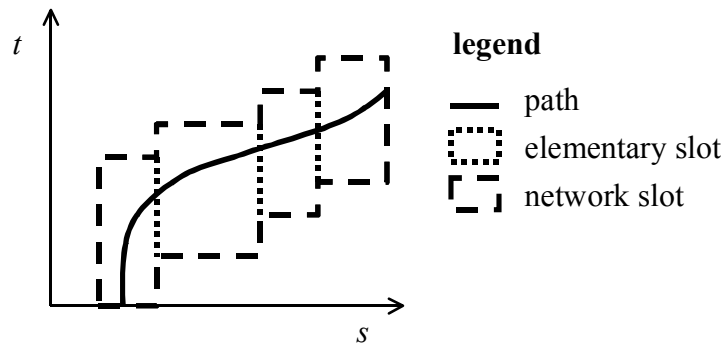


Figure 2.10: Vehicle path with corresponding slots

In this section, slot has been defined as a specific type of time-space domain. A time-space domain is 'a closed subset of time-space, which is characterized by a common administrator, a common user, etc.' (Hägerstrand, 1970). With respect to slot allocation, the spatial dimensions are restricted by the infrastructure network. A slot is but one example of a time-space domain; access restrictions also define time-space domains. For instance, the access of lorries to a city center may be restricted to specific periods. However, the latter type of allocation implies the allocation of time-space domains to a class of vehicles, while slot allocation implies allocation to individual vehicle paths.

Vehicle path is another term that has been used in the definition of slot. A *vehicle path* is a trajectory in time-space that is followed by a vehicle. Taking into account vehicle

length, this is by nature a two-dimensional band in a two-dimensional representation of time-space. For the sake of simplicity, however, vehicle paths are usually represented by a one-dimensional line.

Finally, *slot allocation* is defined as the allocation of slots to slot requests in such a way that capacity constraints are satisfied. The requirement that capacity constraints are satisfied implies that at least one traffic pattern can be specified that respects the slot margins and satisfies capacity constraints. If the vehicle path is realized according to plan, the part of the vehicle path on the infrastructure element under consideration is a subset of its slot.

Slot holders and slot requests

According to the definition provided in this section, slots are allocated to slot requests. A slot request refers to a single slot and includes information about the identity of the slot holder and the intended usage of the slot, including the desired characteristics of the vehicle path. Slot requests are submitted by (intended) slot holders. According to the specification of roles in section 2.1, slot requests are submitted by transport service producers. Hence, we assume that slots will usually be requested and held by carriers. However, it is also possible to allow other interested parties to submit slot requests, for instance authorities issuing concessions, transport service providers, or even shippers. In the latter case, the actor holding the slot is entitled to choose the carrier that may use this slot. Section 4.1 elaborates on the question of which actors may reserve slots and to which extent slots may be exchanged between carriers or between uses.

Besides that slots are reserved to facilitate traffic, it is also possible to allow infrastructure providers to reserve slots for maintenance activities. According to the model of roles in the traffic layer introduced in section 2.1, traffic service providers are responsible for decisions about availability of infrastructure. However, the current situation in the Dutch railway sector is that the responsibility to decide on the availability of infrastructure is integrated with the allocation body. Hence, maintenance activities require slots in the railway sector. However, this thesis does not elaborate on the effects of maintenance on the available capacity and on the optimum allocation of infrastructure capacity to maintenance activities.

2.4.3 Specification of slot borders

A slot is a type of time-space domain, which means that spatial size and temporal size are specified. An exact specification of slot borders is required to define the difference between 'on time' and 'late' and between 'correct location' and 'other location'. Vehicle paths are expected to stay within the spatial and temporal margins specified by their slots. For instance, the arrival time of a train at a station can only be planned with finite precision, and the same holds for the arrival location of the same train. Furthermore,

specifying larger slots leaves more room for optimization in later planning stages. For instance, it is not necessary to specify which train will use which platform in the allocation timetable of a railway station; platform allocation may be left to operational traffic control. Alternatively, it is possible to allocate long-term slots specifying station access and short-term slots specifying the allocation of platforms to trains.

In the first place, slot dimensions are limited by the size of the bottleneck under consideration. We assume that a bottleneck is a linear element. The main bottleneck size variables are bottleneck length l^0 , which is determined by the location parameters s_1 and s_2 , and bottleneck bandwidth w^0 , which is determined by the number of parallel traffic servers that are included in the bottleneck (see figure 2.11). Parallel traffic servers are numbered r_1 to r_w (with $w = 3$ in figure 2.11).

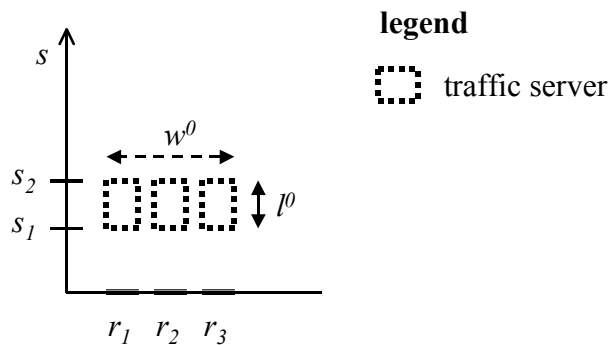


Figure 2.11: Specification of bottleneck

Having specified the bottleneck under consideration, slots may be specified. The specification of slot size is illustrated by figure 2.12. Slots have a temporal dimension, a longitudinal dimension, and a transversal dimension. In this thesis, the temporal size of slots is referred to as the slot duration m , which is determined by the earliest entry time t_1 , and the latest exit time t_2 . It is assumed that no latest entry time or earliest exit time is specified. The longitudinal size is the slot length l , which is assumed to be equal to the bottleneck length l^0 . Finally, the transversal size is referred to as the *slot bandwidth*, specifying the number of parallel infrastructure servers (lanes, platforms, etc.) included in a slot. Slot bandwidth w may be smaller than bottleneck bandwidth w^0 , i.e. not all traffic servers need to be included in each slot. In this thesis, slots have rectangular shapes, i.e. the slot period does not differ between different locations in the bottleneck.

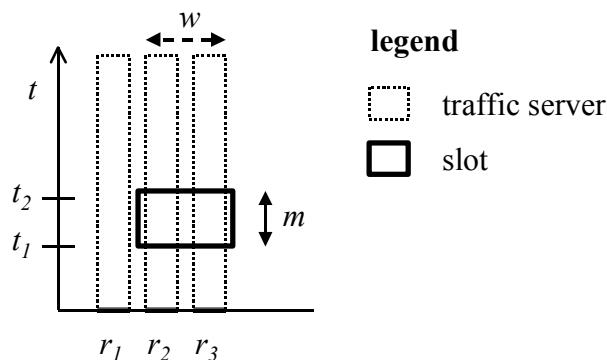


Figure 2.12: Specification of slot size

Exclusive and non-exclusive slots

A slot is a time-space domain that is allocated to a single vehicle path. However, this does not necessarily imply that slots of different vehicle paths may not overlap; they may even be identical. The specification of exclusive slots implies that slots are separated by a positive time buffer b (see figure 2.13). A time buffer is a buffer between the end of the previous slot and the beginning of the next slot. If all vehicle paths are confined within their slots and if the buffers are larger than the minimum headways, there are no conflicts between vehicle paths. In this case, capacity constraints are satisfied because each vehicle path is allocated sufficient service time on the bottleneck. Furthermore, this buffer reduces the probability that the vehicle path that has been assigned this slot is delayed by previous traffic.

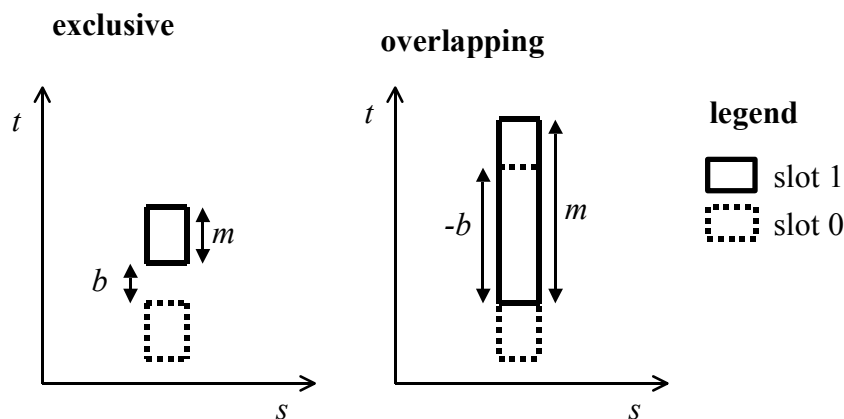


Figure 2.13: Specification of exclusive and non-exclusive slots

The main disadvantage of exclusive slot allocation is that the total number of slots that can be allocated may be significantly lower than the available capacity, especially when slot periods have a duration m that is much larger than required minimum headways. Relatively large slots may be required when the timing of vehicle paths is uncertain. Therefore, the alternative of non-exclusive slot specification is preferred when relatively large slots are required. Non-exclusive slots have a negative time buffer, which means that large slots are possible without negative effects on the total number of slots that can be allocated. The only restriction is that capacity constraints are satisfied for each slot

period. The sequence of vehicle paths is not planned in advance, which may be a disadvantage when capacity depends on the sequence of vehicle paths.

Non-exclusive slot specification will be the best solution in most cases. Exclusive slots are only used when the reliability of arrival and departure times is high relative to the minimum headways, and when there are particular reasons to plan the sequence of vehicle paths in advance. This is the case with scheduling and *ad hoc* slot allocation in the railway sector, where the sequence of trains for each infrastructure server is indeed included in the allocation timetables because trains can only overtake each other at fixed locations. In the absence of the particular conditions mentioned above, non-exclusive slot allocation is applied. For instance, airport landing and take-off slots at Schiphol airport uses slot periods of 20 minutes (Vos, 2002), and ATC slots have a slot period of 10 minutes (Van Hoorn, 2002), while the minimum headway between airplanes is only a few minutes, depending on aircraft types and local circumstances (Nolan, 1994). Furthermore, Schaafsma (2001) favors the introduction of non-exclusive slot allocation for scheduling slot allocation in the railway sector to enable the optimization of bottleneck usage at the operational level. His views are supported by De Jong (2002), who notices that slot periods may be significantly larger for cargo trains than currently applied.

2.4.4 Levels of slot allocation

Slot allocation includes a range of possibilities from long-term decisions regarding concessions valid for several years to short-term decisions that are made just before departure. Here, three levels of slot allocation are proposed:

- selection slot allocation
- scheduling slot allocation
- *ad hoc* slot allocation

The motivation for distinguishing these levels will be discussed in the next section. The main argument is that these levels of slot allocation correspond with different levels of transport service planning. Which levels of slot allocation are demanded depends on the type of transport service. Some transport services may result in slot requests from all three levels, while others may only require one level of slot allocation. The main characteristics of these three levels of slot allocation and the relationships with the demand for slot allocation of carriers are discussed below.

Selection slot allocation

Selection slot allocation is the selection of slot requests with the purpose of determining which slot requests can be accommodated and which have to be rejected. Additionally, although this is strictly speaking not slot allocation according to our definition, selection slots may be reserved for specific purposes such as freight transport, without already

specifying the associated vehicle path or carrier. The specification of slot periods is based on acceptability criteria, i.e. only non-acceptable periods are excluded (see section 4.3), resulting in relatively large slot periods. Selection slot allocation is related to the design of transport service networks by carriers. Selection slot allocation decisions are preferably valid for relatively long periods (several years), and selection slots are reserved relatively long in advance (see section 4.1).

Scheduling slot allocation

Scheduling slot allocation is the allocation of time and space specific slots, given slot requests, with the purpose of determining which transport service timetables are feasible given infrastructure capacity constraints. The slot size is restricted by quality-of-service criteria stating the temporal precision of the allocation timetable, and scheduling slots are usually much smaller than selection slots. Scheduling slot allocation is related to the design of transport service timetables by carriers. Scheduling slot allocation decisions are preferably valid for the same period as the associated transport service timetable, for instance a year.

Ad hoc slot allocation

Ad hoc slot allocation is the allocation of time and space specific slots, given slot requests, with the purpose of determining the feasibility of vehicle paths in short notice. The slots are of comparable size as scheduling slots. The main difference is that scheduling slot allocation takes place well in advance, while *ad hoc* slot allocation takes place at short notice. Short-notice changes in both traffic supply and demand can be incorporated in *ad hoc* slot allocation decisions. Furthermore, selection and scheduling slot allocation decisions are valid for long periods and this enables an elaborate procedure to optimize the allocation decisions. In contrast, *ad hoc* slot allocation decisions usually evaluate only one slot request each time, and comprehensive evaluation of alternative allocations is neither feasible nor practical at this short timescale. *Ad hoc* slot allocation is associated with the planning of demand-driven and *ad hoc* transport services and the planning of logistic operations such as ferrying of empty vehicles (see next section).

The following example illustrates the difference between these levels of slot allocation. A railway company is considering to introduce a container shuttle service between Rotterdam and Duisburg. Before it makes financial commitments with respect to this service, for instance by ordering new locomotives, it may want to reserve selection slots. The railway company needs to be sure it can operate this transport service for a number of years in order to recover its investments. Selection slots ensure the availability of sufficient capacity of sufficient quality to be able to run this service. Furthermore, the carrier may want to reserve scheduling slots before it publishes the transport service timetable. Finally, it may want to reserve *ad hoc* slots to be able to run additional trains at short notice or when it needs additional slots for logistic reasons, for instance to move locomotives.

Hierarchical relationship of slot allocation levels

In two different ways, the three levels of slot allocation distinguished in this thesis have a hierarchical relationship. In the first place, selection slot allocation decisions are used as a basis for more detailed selection slot allocation decisions or *ad hoc* slot allocation decisions. Selection slots are relatively large, which primarily means that these have a large period. Scheduling slots are smaller, and are a subset of the corresponding selection slots. The allocation of more specific slots is sometimes left to the *ad hoc* level (see next section) and in this case selection slot allocation and *ad hoc* slot allocation have this type of hierarchical relationship. In the second place, lower levels of slot allocation always have to respect the decisions made at higher levels. For instance, it is possible to accept scheduling slot requests that are not associated with an accepted selection slot, if only the previous selection slot allocation decisions are not violated.

The hierarchical relationship between the three levels of slot allocation is illustrated by figure 2.14. Furthermore, the related decision types at the strategic and operational level are added to provide a complete overview. Selection slot allocation decisions should comply with strategic decisions on network access, while operational traffic control decisions should respect slot allocation decisions as much as possible.

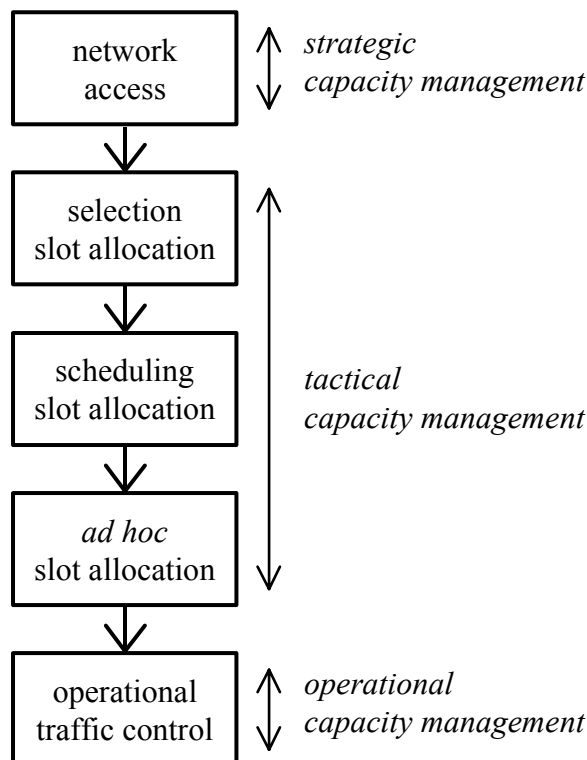


Figure 2.14: *Generic slot allocation decision levels*

Selection slot allocation is the highest level of slot allocation. Selection slot allocation decisions are crucial because they determine which transport services are feasible. Given that a slot request has been accepted, the availability of scheduling or *ad hoc* slots of sufficient quality is guaranteed. The influence of scheduling and *ad hoc* slot alloca-

tion on the quality of traffic supply is relatively small, given that selection decisions have been made. Therefore, this thesis focuses on selection slot allocation. Nevertheless, the other slot allocation levels will still be included in the next chapter discussing current practice with slot allocation.

2.5 Transport service planning

In the previous section, three levels of slot allocation have been introduced. These levels of slot allocation are closely related to transport service planning levels. Planning transport services implies planning traffic, which means that slots have to be reserved if slot allocation is applied. Because the transport service planning process may differ considerably between types of transport services, the first step of this section is to classify transport services. Next, four levels of transport service planning are specified, based on the different transport service planning needs of different types of transport services. Finally, the relationship is established between these four levels of transport service planning and the three levels of slot allocation specified in the previous section. Information from general handbooks (i.e. Ceder, 2001) and interviews with carriers (see Appendix E for an overview) have been used to specify these levels of transport service planning.

2.5.1 Types of transport services

To explain differences in traffic demand dynamics, this thesis distinguishes between different types of transport services. Six classes of transport services are distinguished based on the following criteria:

- public or private;
- structural or *ad hoc*;
- supply-driven or demand-driven.

In the first place, we distinguish between private and public transport services. *Public* transport services are offered by carriers and are available to the general public. In contrast, *private* transport services are organized and produced by the traveler or shipper, or, for instance, by a relative or subsidiary company of the traveler or shipper, and are not available to the general public. For instance, driving a car to collect a friend from the railway station is a private transport service, while taxis offer public transport services to the general public.

Both public and private transport services may be either structural or *ad hoc*. *Structural* transport services are provided with regular intervals. In contrast, *ad hoc* transport services are not offered with regular intervals, but offered at an individual basis. An example of a structural public transport service is a line service running according to a fixed timetable. Private transport services can also be structural, for instance commuting

travel or regular dedicated shipments, i.e. of car parts, chemicals, etc., for large manufacturers.

With respect to public transport services, we may again distinguish between different types of services. For instance, taxis and line busses both offer public transport services, but have very different dynamics of transport and traffic planning. An important difference between taxi and line services is that taxi services are demand-driven and line services are supply-driven. *Supply-driven* transport services are designed by carriers and are collective, i.e. they can be used simultaneously by several travelers or shippers. Supply-driven transport services are often scheduled, which means that the supply of transport services can be made known to the public in advance in the form of a transport service timetable. In contrast, *demand-driven* transport services are tailored to serve specific transport demands, because these services are supplied based on agreements between carriers and end-users. On-demand services can still be collective, e.g. shared taxi services. Private transport services are demand-driven by definition.

Examples

Table 2.4 summarizes a few examples of different types of public transport services. An indication of occurrence rates of these types of transport services in different sectors is provided in Appendix A. Scheduled transport services are generally periodic, e.g. common rail passenger transport services, airline services, etc. However, scheduled transport can also be *ad hoc*, for instance flights that are offered by carriers or tour operators to transport supporters to major sports events. On-demand transport services are often provided on an *ad hoc* basis, for instance road and air taxi services. However, on-demand transport can also be periodic, which is often the case with supplies to manufacturing companies.

Table 2.4: Types of public transport services (examples)

	supply-driven	demand-driven
structural	scheduled public transport services	periodic transport services (manufacturing supplies)
ad hoc	extra transport services for special events	taxi services

2.5.2 Specification of transport service planning levels

In the previous sub-section, six types of transport services have been defined. The next step is to identify which transport service planning levels can be distinguished for each type of transport service. Because demand-driven public transport and private transport are probably planned in a similar way, only four types of transport services are considered here:

- structural supply-driven;

- structural demand-driven;
- *ad hoc* supply-driven;
- *ad hoc* demand-driven.

Four different levels of transport service planning are introduced in this section:

- transport service patterns;
- seasonal timetables;
- *ad hoc* timetables;
- daily traffic plans.

An overview of these transport service planning levels and the relationship with transport service types is provided in the remainder of this section. This relationship is summarized by table 2.5.

Table 2.5: Relationship between transport service types and types of selection and scheduling decisions

transport service type		levels of transport planning
structural	supply-driven	<ul style="list-style-type: none"> • transport service patterns • seasonal timetables • (daily traffic plans)
	demand-driven	<ul style="list-style-type: none"> • transport service patterns • daily traffic plans
<i>ad hoc</i>	supply-driven	<ul style="list-style-type: none"> • <i>ad hoc</i> timetables • (daily traffic plans)
	demand-driven	<ul style="list-style-type: none"> • daily traffic plans

Transport service patterns

Transport service patterns describe which kind of transport services (e.g. line services, taxi services) is offered in which area, and which locations are connected with which frequency. The transport service pattern of a supply-driven transport service can be described as a line network. Main characteristics of transport service patterns are the direct services offered (lines), stops connected per line, connections between lines, and frequency per line.

Structural transport services have a stable transport service pattern, while *ad hoc* transport services are planned on a one-by-one basis. Structural transport services are offered with more or less regular intervals, while *ad hoc* transport services are generally not repeated hourly, daily or weekly, at least not outside the limited period of a specific event. The transport service patterns of structural transport services are usually more persistent than their timetables, and therefore the establishment of transport service patterns is specified here as a separate transport service planning level for structural transport services. The reasons for this stability will be discussed in section 4.1.

Both supply-driven and demand-driven structural transport services have fixed transport service patterns. A difference, however, is who initiates the planning process of transport services. The transport service network of supply-driven services is determined by the carrier, while the main characteristics of demand-driven services, e.g. which connections are offered when, are determined by the end users. Consequently, if selection slots have to be requested to enable these services, it is reasonable to expect that slots for supply-driven services will be requested by carriers, while slots for demand-driven services will be requested by or on behalf of end users, for instance shippers.

Seasonal and ad hoc timetables

Although we may assume that expected departure and arrival times of all types of transport services are known some time before departure, only some types of transport services work with an explicit timetable. Publishing a timetable is only important for supply-driven transport services, not for demand-driven transport services. Timetables inform potential travelers and shippers about the supply of transport services. Structural supply-driven transport services generally have a fixed timetable, while *ad hoc* transport services are scheduled on a one-by-one basis. Based on these differences, we distinguish between seasonal (transport service) timetables and *ad hoc* (transport service) timetables.

Structural supply-driven services are usually scheduled in seasonal timetables. The timetable season is generally a full year for trains and buses, while airline timetables are commonly valid for a half-year. Seasonal timetables are preferred to making *ad hoc* timetables for each day or even each run to reduce design and communication costs. However, there are a few exceptions to this rule. For instance, flights offered to leisure destinations by tour operators are often not scheduled in the seasonal timetable, despite their structural nature. Instead, these flights are scheduled in an *ad hoc* timetable that is communicated to only the travelers that already have booked their trips. Furthermore, publishing a seasonal transport service timetable is unimportant for transport services with high frequencies. Given maximum intervals between service runs of less than ten minutes, few passengers check timetables in advance, and therefore many urban transit systems such as subways do not publish transport service timetables. In this case, only the service frequency for each time period should be determined in advance, scheduling decisions may be left to operational traffic control.

Ad hoc supply-driven services require a timetable for the same reason as structural services, i.e. to attract demand. The deadline for the planning of *ad hoc* supply-driven transport services is determined by the time needed to advertise this service to the public. *Ad hoc* services are usually scheduled when the seasonal timetable has already been established. For instance, the participants of major sports events such as the Champions League finals are only known a number of weeks in advance, and therefore extra flights

to convey soccer supporters from their home region to the city where the finals will take place are not included in the seasonal timetables.

Daily traffic plans

Because demand-driven services are planned in cooperation with the end user, time-plans do not have to be published in advance. Since flexibility is the main advantage of demand-driven transport, these services may be planned or altered at short notice. However, users of demand-driven services also require information with respect to the availability of transport services and the expected departure and arrival times, just as supply-driven transport services. Furthermore, vehicle and crew scheduling decisions need to be made, including availability of vehicles and staff. Therefore, we assume that demand-driven services are planned at short notice before departure and this type of short-term plans are referred to here as *daily traffic plans*.

Daily traffic planning accommodates the need for short-term planning of demand-driven transport services. Furthermore, logistic planning decisions of supply-driven transport services may be left to this level of transport service planning. Logistic plans describe the usage of vehicles, including ferrying of empty vehicles, and use of staff that is required to produce transport services. According to Ceder (2001), vehicle and crew scheduling are the final steps of transport service planning. Furthermore, decisions about the choice of alternative routes may be left to this planning level when both routes are feasible alternatives to deliver the same transport services. However, sometimes logistic planning is integrated with timetable design, and sometimes it is left to the daily traffic planning level.

An example of integrated timetable design and logistic planning is the planning process of NS Reizigers, currently the principal Dutch railway passenger carrier. According to Stellingwerff (2002), the annual logistic plan of NS Reizigers is made directly after the draft timetable has been established. This enables adaptations to the timetable that are required to design a feasible and efficient logistic plan. On the other hand, the flexibility of logistic planning of supply-driven services may be enhanced by making logistic planning decisions at short notice instead of integrating logistic planning with the design of seasonal timetables. For instance, most freight transport services offered by Railion are scheduled in an annual timetable, but decisions about which locomotive is used for traction of which train are left to the daily traffic planning level, because that leaves more possibilities to optimize the utilization of these relatively expensive transport means (De Jong, 2002). Because freight transport service timetables usually have much less precision than passenger transport timetables, there is more room for adaptations at short notice. Table 2.5 lists this level of transport service planning between brackets because only some carriers leave logistic decisions to daily traffic planning.

2.5.3 Transport service planning and slot allocation levels

In the previous sub-section, four levels of transport service planning have been introduced. In this sub-section, the correspondence between levels of transport service planning and levels of slot allocation is analyzed. The first step is to review the differences in validity and scope of levels of transport service planning, with the aim to assess whether this results in different levels of demand for slot allocation.

Table 2.6 summarizes the characteristics of the four levels of transport service planning with respect to validity and scope. *Validity* is the period that transport services are operated based on the same planning decision. Transport service pattern decisions are generally valid for several seasons, seasonal timetables are valid for one season, and *ad hoc* timetables are valid for a single occasion. Daily traffic plans can remain valid for a single occasion to one day, depending on the type of transport service under consideration. *Scope* indicates whether decisions are made simultaneously for the entire network or individually for each transport service. Transport service patterns and seasonal timetables typically have a network scope. The long validity of these decision types justifies that allocation decisions are based on an evaluation of alternatives. In contrast, *ad hoc* timetable decisions are generally made for specific transport services. Daily traffic plans will usually be made individually for demand-driven services, because comprehensive planning for the entire network is usually not possible given the short time-scale. However, daily traffic plans of supply-driven transport services will probably be made for the entire network simultaneously, because the same means and staff are used for different transport services.

Table 2.6: Transport service and slot allocation planning levels

planning levels	validity	scope
transport service pattern planning	several seasons	network
seasonal timetable planning	one season	network
<i>ad hoc</i> timetable planning	single occasion	individual
daily traffic planning	single occasion, one day	individual, network

Given that transport service patterns and seasonal timetables have a significantly different validity than *ad hoc* timetables and daily traffic plans, and given the differences in scope, it is reasonable to assume that different levels of slot allocation are involved. However, *ad hoc* timetable planning and daily traffic planning both require a flexible slot allocation system that allows slot allocation on a short timescale without much time for comprehensive optimization on a network level, and therefore these levels of transport service planning are proposed to correspond with a single level of slot allocation (see table 2.7). The correspondence between transport service pattern planning and selection slot allocation is clear, because both focus on the question of which transport services and which corresponding traffic can be facilitated given the capacity scarcity of

bottlenecks. Furthermore, both seasonal timetable planning and scheduling slot allocation focus on the establishment of seasonal timetables, i.e. transport service timetables and allocation timetables. Finally, *ad hoc* timetable planning and daily traffic planning both require traffic planning at short notice, without time (or need) for comprehensive evaluation of alternative allocation decisions.

Table 2.7: Transport service and slot allocation planning levels

transport service planning levels	slot allocation levels
transport service pattern planning	selection slot allocation
seasonal timetable planning	scheduling slot allocation
<i>ad hoc</i> timetable planning	<i>ad hoc</i> slot allocation
daily traffic planning	

2.6 Synopsis

This chapter has defined slot allocation and specified its position in the transportation system. A consistent conceptual framework has been defined, based on a three-layer model of the transportation system. The focus of this thesis is on the traffic market, which is the interaction between the transport services and traffic services layers of the transportation system. The traffic market balances the supply of and demand for traffic services. Infrastructure networks, including control systems, are required to produce traffic services. The supply of traffic services is limited by capacity constraints, defining a maximum traffic load or traffic flow. The capacity of traffic network elements may depend on several factors, including composition-of-traffic and traffic context. Furthermore, capacity depends on the desired minimum quality-of-service. Usually, a higher capacity can be attained given a lower quality-of-service and *vice versa*. Hence, a key task of traffic service providers is to determine the optimum balance between capacity and quality-of-service.

Capacity management is the application of measures to control the allocation of infrastructure capacity to alternative users and uses. Three planning levels of capacity management have been defined, i.e. strategic, tactical and operational. The strategic planning level deals with network access questions, while planning of traffic services takes place at the tactical planning level. The operational planning level deals with real-time traffic control.

Slot allocation is a type of tactical capacity management, i.e. it aims at a time and place specific balance between traffic supply and demand before traffic actually flows. A specific characteristic of slot allocation is that a guaranteed quality-of-service may be specified by requiring users to plan their traffic. Slot allocation can be used to optimize the utilization of infrastructure capacity, for instance by encouraging traffic with a high occupancy level. Consequently, slot allocation may also influence the transport market.

Different types of slot allocation decisions may be distinguished, corresponding with different types of transport service planning decisions. The structure of the transport service planning process depends on the type of transport service under consideration. Three levels of slot allocation have been identified:

- selection slot allocation;
- scheduling slot allocation;
- *ad hoc* slot allocation.

At each level, a different kind of slots is allocated. These slot allocations have a hierarchical relationship. These levels of slot allocation are related to four levels of transport service planning:

- transport service pattern planning;
- seasonal timetable planning;
- *ad hoc* timetable planning;
- daily traffic planning.

The first two levels of transport service planning correspond with the first two slot allocation levels. Furthermore, both *ad hoc* timetable planning and daily traffic planning also facilitated by *ad hoc* slot allocation.

Next chapters

Because the quality of traffic supply is largely determined by selection slot allocation decisions, this thesis focuses on selection slot allocation. However, the overview of current practice with slot allocation in the next section considers all levels of slot allocation in order to get a complete overview. The next chapter not only provides an overview of current applications of slot allocation in the aviation and railway sectors, but also discusses the potential application of slot allocation in other sectors. The structure of the selection slot allocation problem will be analyzed in chapters 4, 5, and 6. Chapter 4 focuses on the demand for selection slot allocation and the specification of slot sizes, chapter 5 focuses on the specification of selection slot markets, and chapter 6 focuses on the specification of traffic supply.

3

CURRENT PRACTICE WITH SLOT ALLOCATION

In the previous chapter, a conceptual framework has been provided which can be used to analyze current slot allocation systems. Our next step is to review the current applications of slot allocation in the railway and aviation sectors and the potential applications of slot allocation in the road and navigation sectors. Although the next chapters focus on selection slot allocation, this chapter provides a broad overview of slot allocation systems, including all planning levels. After a brief description of current and potential applications of slot allocation in section 3.1, this chapter focuses on the current slot allocation regimes in Europe (situation of 2002, updated until 2004), including the institutional settings, procedures and priorities (section 3.2). It appears that the current levels of slot allocation are similar to or combinations of the levels proposed in the previous chapter, with the notable exception that a separate level of selection slot allocation does not exist in the aviation sector and has only been introduced recently in the railway sector. In section 3.3, three main issues are introduced which are of key importance in designing slot allocation regimes. The first issue is the balance between stability and flexibility of slot allocation, and the other issues are the balance between competition between and diversity of transport services, and the issue how to deal with heterogeneity of traffic. One of the main conclusions of this chapter is that distinguishing a separate selection slot allocation level is desirable for a number of reasons, including that it enables an optimal balance between stability and flexibility at each level. The results of this chapter will be used in the next chapter to elaborate on the desired characteristics of selection slot allocation procedures.

3.1 Current and potential applications of slot allocation

This section focuses on the question of in which transport infrastructure sectors slot allocation is currently applied and which other applications are possible in the future. This section subsequently discusses the railway, aviation, road, and navigation sectors with respect to the application of slot allocation. Slot allocation is currently applied for railways and for some airports. Furthermore, slot allocation at a short-term planning level is applied to *en route* air traffic and to a number of ports (navigation). This section includes a brief description of the history and institutional setting of slot allocation in these sectors. This section also reviews the spatial and temporal scope of the current slot allocation regimes. Given this, potential extensions of these slot allocation systems are discussed. Finally, this section discusses new application areas of slot allocation that have been proposed in the literature, for instance to avoid congestion on freeways. A more detailed description of current slot allocation regimes, focusing on slot allocation procedures from an institutional perspective, is left to section 3.2.

3.1.1 Rail traffic

Probably the main reason to apply slot allocation in the railway sector is to avoid congestion on the rail tracks in situations where railway networks are used by more than one carrier. When a railway company exclusively uses its part of the railway network, capacity scarcity problems can be solved internally. However, when several railway companies share the same infrastructure, independent slot allocation is required to solve capacity conflicts. In the Netherlands, the number of regular railway companies using the national railway network has been between 5 and 10 in the last years, while it was just 1 ten years ago.

We may distinguish several types of bottlenecks in the railway infrastructure. For instance, railway links are often used by different types of trains with different riding characteristics, and uncoordinated usage would easily result in congestion on the rail tracks. Furthermore, the number of available platforms of main railway stations may be lower than desired, which means that the arrival and departure times of trains should be distributed more evenly over time. Finally, many railway stations lack separate reversing facilities, which means that reversing trains have to use platform tracks, which are often also the tracks used by through trains. As a consequence, the possible choice of terminal locations for railway lines is limited by the supply of platform tracks and reversing tracks. A more detailed overview of types of railway bottlenecks and the current level of railway capacity scarcity will be provided in section 6.5.

Demand for slot allocation in the rail traffic sector

From the introductory remarks of this sub-section we may conclude that there is a strong case for slot allocation in the railway sector. According to the overview in Appendix A, railways are predominantly used by structural, supply-driven transport ser-

vices that usually run according to a timetable, and hence there is demand for slot allocation, at least at the scheduling level (see sub-section 2.4.4 for definitions of slot allocation levels). Furthermore, because traffic running according to a timetable is already planned in advance, the application of slot allocation does not require major adaptations of the transport service planning process. Finally, the allocation regime can easily be enforced given the rigid traffic control system and the limited number of carriers using the system.

The demand for selection slot allocation is a different issue. Assuming that homogenization of railway traffic is considered acceptable by the decision-makers, few examples can be found of traffic demand exceeding traffic supply. However, it can be expected that an optimal balance between traffic quantity and traffic quality requires traffic volume restrictions. In section 6.5, a few examples will be provided of situations where this is probably the case. Consequently, there is demand for selection slot allocation, at least in regions with a relatively high railway traffic demand, or a relatively low traffic supply, for instance due to single-track railway links. This situation applies at least to significant parts of the Dutch and German railway networks.

History of railway slot allocation

Until the late 1990's, European countries commonly had no explicit railway slot allocation regime. Because railways were commonly operated, managed and exclusively used by a single national railway company, traffic planning and transport service timetable design were integrated problems that were solved by the planning departments of national railways. In the Netherlands, for instance, a situation with competing railway companies existed until the 1920's. In principle, railway companies used (and maintained) their own railway networks, but carriers had the right to connect their network with other railway networks and to use the railway infrastructure of other companies within the framework of the old Railway Act (*Spoorwegwet*, 1875, arts. 4/5). In 1937, the two remaining competitors (HIJSM and SS) officially merged to create the national railway company NS after years of *de facto* integration of both companies (Veenendaal, 1998). Between 1937 and 1996, NS was responsible for all traffic and transport roles in the railway system, including the role of traffic service provider.

In the early 1990's, a profound restructuring operation of the European railway sector was launched. 1991 data on the subsidy and debt rates of European railway companies revealed that except British Railways, all EC national railway companies were either subsidized for more than 25% or had deficits of more than 25% (UIC, 1991). One of the measures to improve this situation was to gradually reduce the monopoly position of national railway companies. In 1991, the European Commission passed Directive 91/440 ordering the financial separation from the national railway companies of infrastructure construction and management (EC, 1991). While some European countries, for instance France, Belgium, and Austria, have only implemented a financial separation

within the national railway company, other countries have gone beyond these minimum obligations. In some countries, infrastructure management and slot allocation is the responsibility of an independent organization, for instance Banverket in Sweden, Pro-Rail in the Netherlands, and Railtrack in Great Britain. In 2001, stricter provisions have been introduced, requiring that an independent regulator should be created if not an independent organization is created for infrastructure management and slot allocation (EC, 2001a). The latter option has been chosen in most European countries, including Germany and France.

The separation of traffic and transport services was needed to facilitate a more open traffic market. Currently, the rail freight transport market has been liberalized in a number of European countries, and in a number of countries the national railway companies compete with other carriers for concessions to operate regional passenger trains. Independent slot allocation is necessary to ensure a fair distribution of rail capacity between passenger and freight transport and between different passenger and freight transport companies. However, external slot allocation is not required for railway networks that are exclusively used for suburban and metropolitan transport, such as the RER and Métro networks of Paris and the S-Bahn and U-Bahn networks of Berlin. Because these are exclusively used by a single carrier, capacity allocation is an internal problem of the carriers that may be integrated in the transport service timetable design process. To conclude, the railway sector has a long tradition of internal slot allocation, but only a short history of external slot allocation.

Completeness of railway slot allocation

To conclude this sub-section, we focus on the question of whether the current railway slot allocation system is complete. The first aspect reviewed here is spatial completeness. According to EU regulations, independent slot allocation should be applied to most railway lines, at least those that are used by more than one carrier (EC, 2001c). Another aspect is whether slots can be reserved at all planning levels. Slots can be reserved about a year in advance in the annual allocation timetable, but also up to one or two days in advance. After this point, traffic control takes over the responsibility of allocating slots. Furthermore, framework agreements enable the reservation of long-term selection slots, i.e. slots that do not specify an exact timetable but specify the minimum quality-of-service guaranteed for a period of up to seven years. More information about levels of railway slot allocation is provided in section 3.2. To conclude, the railway slot allocation has a comprehensive spatial and temporal scope.

3.1.2 Air traffic

Slot allocation is not only currently applied in the railway sector, but also in the aviation sector. Actually, two different slot allocation systems are employed in this sector at different timescales. Airport slots apply to airports only, and airport slots are allocated

between a day and a year in advance. Air traffic control (ATC) slots apply to the entire flight from the airport of origin to the destination airport, and ATC slots are generally allocated a few hours before departure. Both slot allocation systems are discussed in this sub-section.

Probably, the main reason to apply slot allocation in the aviation sector is to avoid (airborne) delays. In the first place, airborne delays are more costly than ground holding, for instance due to higher fuel costs. Furthermore, large airborne delays are not accepted for safety reasons. Consequently, there is demand for slot allocation in the aviation sector. Finally, slot allocation is applied to avoid overload of air traffic controllers. Enforcement of slot allocation is not a problem given the limited number of carriers and the existing air traffic control system.

Demand for slot allocation in the air traffic sector

Slot allocation is frequently applied to airports because of their limited physical capacity (see section 6.5). Runway capacity is a frequent bottleneck, but also the available number of gates. Finally, environmental restrictions can be a major bottleneck, which is the case for Amsterdam Airport Schiphol. Given that capacity scarcity is peak-period related, as is often the case, this results in a demand for scheduling slot allocation. According to Appendix A, most air transport services are scheduled services, and hence it is reasonable to expect that air carriers desire to check the feasibility of their timetables. However, airport capacity scarcity may attain such levels that not all air traffic demand can be accommodated with a reasonable quality-of-service, which implies traffic volume restrictions and hence demand for selection slot allocation. Finally, to avoid overload of air traffic controllers, it is important to optimize the timing and routing of traffic. Because en route congestion seems not very well predictable in advance, *ad hoc* slot allocation is required in this case.

Airport slot allocation

As will be discussed in more detail in section 6.5, a number of airports, including large hubs and small regional airports, face structural congestion problems. Furthermore, airports are usually used by several airline companies. Even in countries with only one national carrier, international services may usually be offered by carriers from both countries. Consequently, competition between carriers for infrastructure capacity is not relatively new in the aviation sector, contrary to the railway sector. For instance, more than 100 different air carriers are currently holding slots at Amsterdam Schiphol Airport, which is the Dutch national airport.

Airlines coordinate their schedules in order to avoid congestion problems and to optimize connections. The international association of air carriers IATA has organized biannual Schedules Conferences since 1947. The second objective (to optimize connections) was dominant in the first years, but since the 1960 the objective of avoiding con-

gestion has grown in importance. The IATA slot allocation system is based on voluntary cooperation of air carriers and consequently compliance to this system cannot be enforced by law (IATA, 2002). According to Stout (2001), the largest carrier of an airport usually has the responsibility of allocation body, at least before the introduction of independent slot allocation bodies in Europe and the USA.

In 1993, the European Union introduced its own scheduling guidelines for slot allocation at community airports in Regulation 93/95¹ (EC, 1993). The provisions of this European Regulation are based on the IATA Scheduling Procedures with respect to fully coordinated airports. The main difference with the IATA system is that EU regulations can be enforced by (inter)national law (Stout, 2001). The main objective of EEC Regulation 93/95 is to avoid distortion of competition between European carriers. According to art. 4, allocation bodies should be able to make allocation decisions independently. To ensure this, independent slot allocation bodies have been established in European countries, such as ACL in Great Britain and SACN in the Netherlands.

Slot allocation is only applied where necessary. These so-called coordinated airports are assigned by national governments according to art. 3 of Regulation 95/93. Airports may be assigned as allocated by the initiative of the national government or if this is requested by air carriers or airport managers.

ATC slot allocation

In the aviation sector, the most important reason to apply slot allocation is safety. ATC slot allocation is applied to avoid congestion in the air, and to ensure a reasonable workload for ATC controllers. Aircraft operators submit flight plans, describing the planned route and schedule of the flight, which also function as ATC slot requests. To avoid excess traffic in ATC sectors, flight plans can be changed by air traffic control, for instance by restricting flight speeds, or by delaying aircraft at departure airports (ground holding). Ground holding is also used in situations of capacity scarcity at destination airports due to weather conditions, which is a major source of air traffic delays in the USA (Milner, 1995).

The obligation to submit flight plans to the air traffic control system prior to departure has been introduced in the 1930's to ensure safe separation of IFR traffic. IFR (Instrumental Flying Rules) traffic is air traffic relying on on-board instruments and traffic control information to ensure a safe flight, contrary to VFR (Visual Flying Rules) traffic relying on visual information (Nolan, 1994). According to Dutch regulations, flight plans are currently required for any controlled IFR or VFR flight (Wet luchtvaart, 1999, art. 5.9). Consequently, all flights using ATC services are not allowed to depart before their flight plans have been approved.

¹ This regulation has been revised recently by Regulation 793/2004 (EC, 2004).

Completeness of airport and airway slot allocation

We conclude this overview of slot allocation in the aviation sector with a brief review of the spatial and temporal scope of the slot allocation system. Airport slot allocation is not applied to all airports, but just to congested airports that have been assigned the status of coordinated airport. Furthermore, the airport slot allocation regime may only be applied to those periods that congestion may be expected (EC, 1993, art. 3). On the other hand, slots may be reserved at different timescales, between a day in advance up to a year in advance. In the first place, slots are allocated biannually corresponding with the establishment of summer and winter transport service timetables. Selection and scheduling slot allocation are integrated, and the long-term stability is ensured by the historicity principle (see section 3.3). Finally, slots may be reserved *ad hoc* up to a day in advance. After this point, air traffic control takes over the responsibility of allocating airport slots (Van Hoorn, 2002). To conclude, airport slot allocation is only applied where necessary and the airport slot allocation regime encompasses all timescales.

The ATC slot allocation regime, based on the obligation to submit flight plans, is generally applicable to all air traffic using air traffic control services. Similar to airport slots, regular scheduled services may submit a repetitive flight plan to avoid having to submit almost identical flight plans every day for the same flight (Regeling vliegplannen, 1998). However, although flight plans may be submitted long in advance, ATC slots are allocated only a few hours before departure (Van Hoorn, 2002). The question emerges whether there is demand for airborne slot allocation at the same timescale as airport slot allocation. For instance, air taxi services are frequently delayed with 20 to 30 minutes because no earlier ATC slot is available. Sometimes, rescheduling taxi services to an earlier departure time is possible, but this currently requires the submission for approval of a new flight plan, which takes time that may not be available (Van Hoorn, 2002). However, ground delays due to limited ATC handling capacity are probably not structural enough to be significantly reduced by adaptations of the seasonal timetables, which would eliminate the main support for a long-term airborne slot allocation system. At least, no indications have been found in the literature and in the interviews with Transavia and Tulip Air (Vos, 2002; Van Hoorn, 2002) that airlines would like to be able to reserve ATC slots longer in advance. Furthermore, flight routes and speeds may be adapted subject to weather conditions and these are not known in advance. However, if ATC handling capacity increasingly becomes a structural problem in the future, then the introduction of a long-term airborne slot allocation system would be an interesting solution, because this enables carriers to minimize delays and to optimize transport service timetables given known capacity constraints.

3.1.3 Road traffic

Bottlenecks in the road traffic networks can easily be recognized by observing structural congestion on freeways and urban roads. Frequent bottleneck types are freeway merges, interchanges, and crossings of urban roads. Although road congestion is a common

phenomenon in urbanized regions, slot allocation is currently not applied in this sector. Part of the explanation may be that contrary to the air and road transport sectors, only a small percentage of traffic corresponds with transport services running according to a fixed timetable, which means that there is little demand for traffic planning at this time level. Furthermore, road traffic consists of many private carriers, instead of a few public transport companies, complicating the interaction between carriers and allocation bodies (see Appendix A). Finally, the road traffic control system is largely limited to automatically operating signals, while rail and air traffic controllers communicate with individual vehicle operators.

Demand for slot allocation in the road traffic sector

Although there are a number of arguments against the application of slot allocation in the road traffic sector, it is still an interesting option to consider, because slot allocation might be helpful to reduce the negative effects of capacity scarcity. In the first place, congestion costs will likely be higher than the departure and arrival time adaptation costs required to eliminate congestion. Congestion on the road increases driver and fuel costs, while slot allocation enables the optimization of activity patterns given the necessary departure and arrival time adaptations. Secondly, slot allocation enables giving priority to certain types of traffic, without the need to create separate lanes for high occupancy vehicles, trucks, etc. The problem with these special purpose lanes is that the target group should be small enough to avoid congestion, but large enough to avoid significant underutilization of these lanes (see e.g. Dahlgren, 1998). Finally, slot allocation avoids structural blocking back problems. Blocking back is caused by queues blocking vehicle paths that do not pass the original bottleneck, for instance because they leave the freeway before the bottleneck (see figure 3.1). As a result of blocking back, the capacity of upstream elements of road networks decreases. Blocking back is the result of a lack of buffer capacity upstream of the primary bottleneck, and therefore blocking back can be avoided by increasing buffer capacity, for instance by increasing the number of lanes upstream of the bottleneck (Broeren & Westland, 1998).

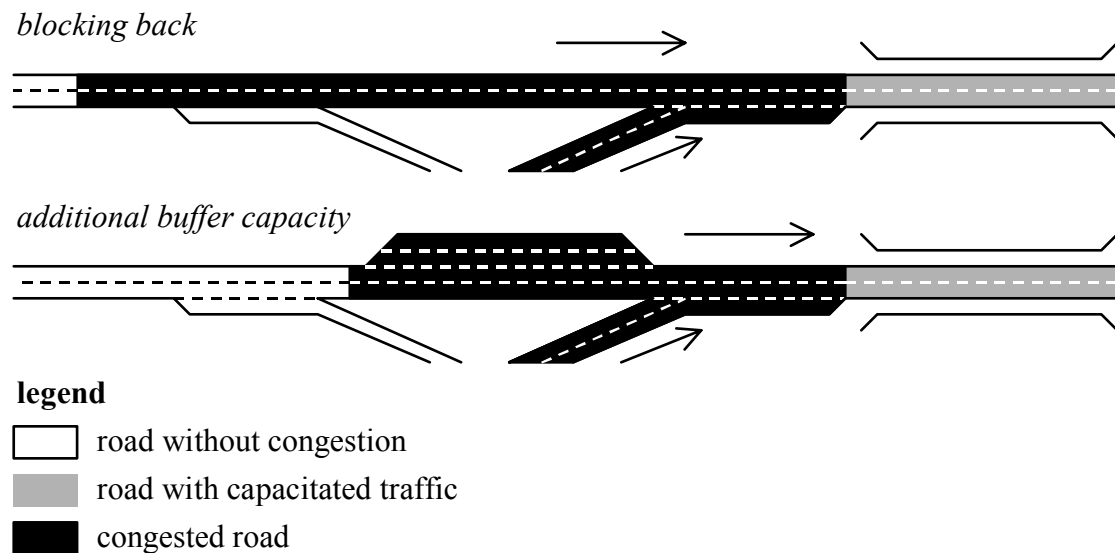


Figure 3.1: Reduction of blocking back problem by increasing buffer capacity

Applying slot allocation to the entire road network would be infeasible and impractical. The required investments in traffic control infrastructure would be too high, while the benefits of the application of slot allocation outside main bottlenecks would be virtually absent. Consequently, it should be limited to parts of the road network experiencing structural congestion. Furthermore, it is reasonable to limit the temporal scope to peak periods only, since having to reserve a slot in the absence of capacity scarcity is an unnecessary burden for road users.

Given the dominance of private transport services on the road, the main slot allocation stage will be *ad hoc* slot allocation. Additionally, selection slot allocation may be desired. Commuter travel, which is usually a major factor in peak-hour road traffic, has often a structural nature. Finally, scheduling slot allocation may be desired by public transport services such as regular bus services. However, it should be noted that the latter type of transport services is usually only a minor factor in road traffic, and an alternative would be to exclude them from the obligation to reserve *ad hoc* slots.

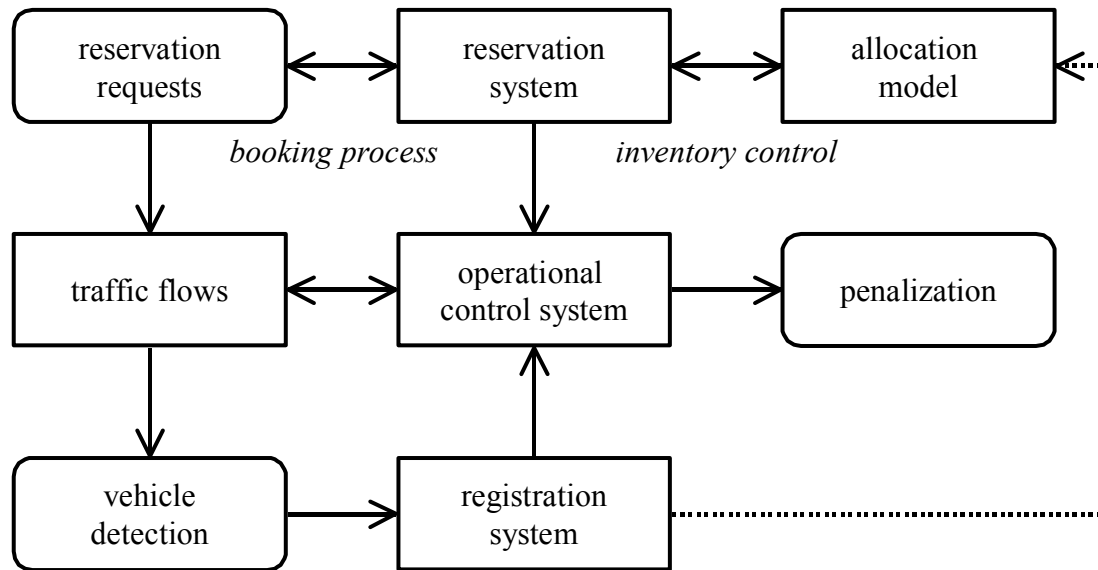
In the literature, the application of slot allocation has been proposed for freeways and urban parking places. Furthermore, application to dedicated bus lanes and busways is also feasible and easier to implement because the users are more similar to the users of rail and air traffic networks. However, given that bus lanes usually have sufficient capacity to accommodate all traffic demand, there is probably no demand for busway slot allocation.

Freeway slot allocation

Probably the first to publish a paper about the possible application of slot allocation to freeway traffic was Wong (1997). According to Wong, four system components are required to enable slot allocation on roads:

- booking process;
- inventory control;
- operation control;
- vehicle detection.

These system components are not specific for road traffic, but are required for any slot allocation system serving large numbers of users. Figure 3.2 illustrates how these components interact.



Legend

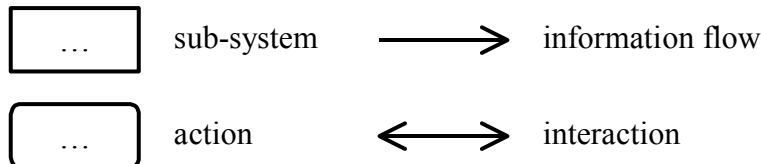


Figure 3.2: Control structure of freeway slot allocation system (see Wong, 1997)

The booking process is the communication between freeway users having reservation requests and the *reservation system* resulting in slot allocation decisions. Given that the rate of slot reservation requests is much higher for freeways than for railways or airports, it is important that this process is automated. Different ways of communication can be used, including the Internet, telephone, and special purpose reservation points near freeway on-ramps (Koolstra, 1999). The *reservation system* checks whether the requested booking can be accepted or not using an *allocation model*. This interaction is also referred to as inventory control. Because travel times cannot be known exactly in advance, the propagation of traffic through the network has to be estimated. Please notice that road traffic lacks individual traffic control at the operational level, contrary to rail traffic and air traffic. The reservation system may use information from the operation control system to adjust the assumptions about available capacity in real time.

Traffic flows are not only influenced by the (acceptance of) reservation requests, but also by actions of the operational control system. The *operational control system* adjusts the actual capacity in real time and registers violators of the reservation regime. These data may be used for penalization of violators. The operational control system includes registration and penalization of violators of the allocation system. The operation control uses the data from the *registration system*, which is based on vehicle detection data. Individual vehicle identification and registration is required near on-ramps and off-ramps and - in case of route choice obligations, at some freeway sections.

To apply freeway slot allocation in an acceptable way, a number of practical issues have to be solved. For instance, a question is to which extent vehicles are supposed to confine with the margins of their slots. Given the absence of individual traffic control and given that *en route* delays will be largely outside the responsibility of the driver, it is only reasonable to expect car traffic to be on time when entering the freeway network. A consequence is that, to enable cars arriving too early to wait for their slots, buffers would be required near the on-ramps. Another issue is how vehicles are to be identified (Koolstra, 1999).

Parking reservation

A slot allocation regime for parking places has been proposed by Minderhoud *et al.* (1995). In urban areas, the demand for parking accommodations is often much larger than the supply, which means that there may be demand for slot allocation in this sector. According to Minderhoud *et al.*, enabling car drivers to reserve a parking place in advance reduces the time needed to find a parking place. As a consequence, inner-city streets will be less congested with cars trying to find a parking place. Implementation of a parking reservation system also enables more differentiation of parking tariffs, for instance to avoid that potential visitors are scared away by high parking fees, while a large percentage of parking places is still available. The set up of the parking reservation system proposed by Minderhoud *et al.* is similar to the system proposed by Wong (1997).

3.1.4 Navigation

As will be discussed in section 6.2, the current level of capacity scarcity in the navigation sector is relatively low. Main bottlenecks are terminals, locks and bridges; outside these nodes vessel traffic usually does not encounter any congestion. Given that terminals, locks and bridges may be bottlenecks, these are potential subjects of slot allocation. However, slot allocation is currently only applied to main terminals. Furthermore, the opening times of some bridges are scheduled, and in principle this could be extended to a slot allocation system.

Container terminals

Container terminals are usually operated and managed by a single company, for instance ETC. According to Thomas (2001), terminal operations are usually planned in advance, although the actual situation may be different than scheduled due to various reasons. Especially container terminals are able to schedule their operations in advance, because container ships usually operate according to schedule. For instance, the schedule of Maersk/Sealand container shipment services of up to a year in advance may be obtained from their website, and similarly the timetables of arrivals and departures of container ships at the ECT terminals in Rotterdam is published on the Internet. This system may be regarded a type of scheduling slot allocation. Additionally, terminal managers may apply *ad hoc* slot allocation for short-term planning of terminal usage.

Slot allocation at container terminals is not subject to specific European or national legislation. Terminal managers conclude contracts with carriers that wish to use the terminal. If necessary, these contracts may function as a kind of selection slot allocation. In principle, they are free to determine the conditions of terminal access and to determine how they want to allocate terminal capacity. Consequently, slot allocation on barge and ship terminals is a private affair of the terminal operators.

Bridges and locks

Slot allocation is usually not applied to bridges and locks. Although the opening times of many railway bridges are scheduled in advance, this is not slot allocation according to our definition, because these 'slots' are not allocated to individual vessels. In a true bridge opening slot allocation system, inland navigation carriers should be able to reserve slots themselves, but in current practice they cannot influence the scheduling of bridge opening times. Because scheduled bridge openings are in principle always sufficient to accommodate all ships and only few ships run according to a fixed timetable, there is no real need to enable the bridge slot reservation by vessels.

Most bridges and locks are operated on demand. An information system is being developed to enable vessel operators and lock operators to communicate earlier on to optimize lock operation decisions, but this is still part of the operational traffic control system (AVV & TNO Inro, 2001). Given that vessels have priority at most bridges, these are no bottlenecks for inland navigation. Lock congestion may give rise to some congestion, and consequently it would be interesting to investigate the pros and cons of lock slot allocation. Lock slot allocation means that locking operations are to be scheduled in advance, and that any user can reserve a slot in a locking cycle. Given the application of lock slot allocation, vessel operators may reduce the negative effects of lock capacity scarcity by:

- traveling faster to catch an earlier locking cycle;
- traveling slower to save fuel;
- choosing an alternative route.

Given the nature of these adaptations, lock slot allocation can best be implemented at the *ad hoc* level. In this case, vessel operators may reserve locking slots just before departure or even *en route*. Demand for scheduling slot allocation is not very probable, except maybe in the case that the lock under consideration is used for scheduled passenger transport services. Finally, given the current low level of capacity scarcity of locks, there will be no demand for selection slot allocation.

3.2 Current slot allocation regimes in Europe

In the previous section, the prevailing rail and air traffic slot allocation regimes have been briefly described. A more detailed description of the current European slot allocation regimes in the railway and aviation sectors is provided in this section. The description of these slot allocation regimes applies to the entire European Union, as far as these are determined by European legislation. For the aspects determined by national legislation, the Dutch situation is described. This section elaborates on all decision stages of these slot allocation regimes, including relevant capacity management decisions at strategic and operational planning levels. This overview includes a description of the decision criteria that are currently applied.

The description of slot allocation regimes in this section is based on European and Dutch legislation with respect to slot allocation. The basis of European legislation is formed by the treaties establishing the European Communities (e.g. the Treaty of Rome, 1967). Based on these treaties, the European Council of Ministers can issue directives and regulations. Since a few years, this legislative power is shared with the European Parliament. Regulations are directly applicable to every member state of the European Union, while Directives have first to be implemented in the national legislation of the member states before coming into power. Laws are the basic form of national legislation, requiring approval of the national legislative powers, i.e. the parliament. Laws may delegate decision power to the ministers or even to other public bodies. Based on this decision power, ministers may issue decrees. Generally, laws provide general prescriptions, while decrees provide more detailed prescriptions.

3.2.1 Rail traffic slot allocation

The setup of the rail traffic slot allocation regime in the EU is largely determined by Directives 91/440 (EC, 1991), 95/19 (EC, 1995), and directives 2001/12 thru 2001/14 (EC, 2001a-c). In the first place, strategic decisions about network access determine which carriers are entitled to request slots. Directive 91/440 states that traffic network access rights should be given to a few categories of carriers providing international transport services, but national legislation and regulation determine the traffic network access rights for other types of transport services. An outline of the slot allocation decision procedure is provided by Directive 2001/14. Carriers may reserve selection slots by concluding a framework agreement with the allocation body. Based on this framework

agreement, carriers are entitled to sufficient slots of sufficient quality in the annual allocation timetables, at least for the duration of the agreement. The annual establishment of allocation timetables is a design process that may take some time to allow optimization of the transport service timetables of carriers. Given that in the railway sector most transport services are regular scheduled services (see Appendix A), most transport services will be scheduled in the annual timetables. However, given that a few percent of traffic corresponds with *ad hoc* transport services, there is also demand for *ad hoc* slot allocation. Finally, operational slot allocation is the responsibility of rail traffic control centers. Operational traffic control has the important function in the slot allocation regime to minimize the deviations between the actual traffic pattern and tactical allocation decisions. Figure 3.3 summarizes the main decision levels of rail traffic slot allocation. These decision levels will be discussed in more detail in the remainder of this section.

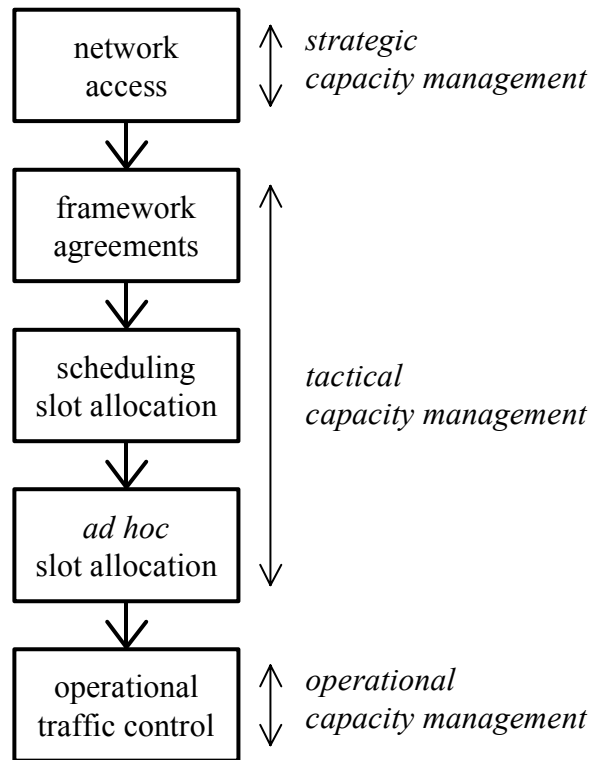


Figure 3.3: Main decision levels of rail traffic slot allocation

Traffic network access

The principal European legislation with respect to network access is Directive 1991/440 (EC, 1991). Article 10 of this directive assures traffic network access rights to two categories of international rail transport:

- international combined transport of goods;
- international transport performed by international groupings of railway companies.

Carriers planning to provide one of these types of transport services should obtain a license first. If a carrier complies with the criteria of Directive 1995/18 (EC, 1995;

modified by EC, 2001b), it is entitled to an operation license issued by the authorities in its home country. In the Netherlands, the Ministry of Transport is responsible for issuing operation licenses. Operation licenses are valid in the entire European Union.

Access criteria for other types of transport services are determined by national legislation and regulation. An overview of railway network access criteria in the Netherlands is provided by Koolstra (2001a). In a number of European countries, the rail freight transport market has been liberalized. In the Netherlands, for instance, any railway company is allowed to operate freight trains on the Dutch railway network if it satisfies a number of basic criteria that are mainly related to safety (V&W, 1999a). To be able to operate through freight trains in Europe without the need to handle trains over to other carriers, the cargo departments of Dutch, German and Danish national railways have merged into a single company called Railion. In its home market, Railion has many small competitors such as Rail4Chem, ERS, ACTS, and Connex subsidiaries such as CCL and NordWestCargo (Latten, 2002).

In Europe, the passenger rail transport market is usually not liberalized to the extent that license holders may operate, in principle, any service they want. However, in a number of countries, concessions to operate regional services are tendered, resulting in an increasing number of small regional carriers in countries such as Sweden and Germany. Moreover, Great Britain has no national carrier anymore, but only a large number of carriers operating regional or intercity services. As a consequence of the limited level of liberalization, there is generally no competition between passenger transport companies, apart from specific situations, for instance when two lines from different carriers coincide, or when alternative routes may be chosen. A notable exception is the Gera - Berlin - Rostock service introduced by Connex Germany in 2002, which freely competes with services of the national railway company DB (Latten, 2002).

Framework agreements

Framework agreements have been introduced by Directive 2001/14/EC (EC, 2001c) to provide for the desire to ensure the continuity of desired transport services within a period of several years of carriers and authorities tendering concessions. In this way, carriers and authorities may ensure that the conditions of concessions may be met, or carriers may ensure that their investments in new services can be paid back by a sufficient period of operation. Framework agreements may have be valid up to 10 years, unless exceptional circumstances justify longer commitment, for instance large-scale, long-term investments covered by contractual commitments (EC, 2001c, art. 17).

Not only carriers may be allowed to apply for a framework agreement, but national legislation may also allow other parties, such as public authorities, shippers, etc., to apply for a framework agreement. The allocation body decides about the conclusion of framework agreements. The decision process to conclude a framework agreement is

neither coupled with decisions about other framework agreements nor with scheduling slot allocation decisions.

As of 2003, the possibility to conclude framework agreements with the allocation body has been implemented in the Netherlands in the new Railway Act (Spoorwegwet, 2003, art. 57 & 60). The allocation body (ProRail) has been granted the responsibility to conclude framework agreements with either carriers or authorities issuing concessions. In the Dutch system, framework agreements are directly related to concessions to operate transport services (see e.g. Concessiewet personenvervoer per trein, 2000). However, as of May 2004, these provisions have not yet come into force. Currently, the minimum frequencies specified by decree in 2000 are still applied. These minimum frequencies have been specified for five types of transport services, i.e. regional/suburban passenger transport, national passenger transport, international passenger transport, conventional freight transport, and heavy freight transport (Interimbesluit capaciteitstoewijzing spoorwegen, 2000). These minimum frequencies largely correspond with the train frequencies realized in the 2000/2001 timetable and hence may help to ensure the continuity of existing transport services. However, these minimum frequencies are not carrier specific, which may be an issue given the situation of free competition between freight transport companies, because the question of which carrier is assigned which number of slots is left to the scheduling slot allocation decision level.

Scheduling slot allocation

In the railway sector, scheduling slot allocation decisions are often referred to as allocation timetables. The procedure to establish allocation timetables is broadly outlined by Directive 2001/14(EC, 2001c). According to this directive (Annex 3), the allocation timetable shall be established once per calendar year. This allocation timetable is based on the desired timetables of railway companies. The allocation timetable period of 1 year corresponds with the validity of transport service timetables.

Directive 2001/14 provides an outline of the allocation timetable design procedure. The submission deadline of requests to reserve capacity in the allocation timetable may be up to one year before the start of the timetable period. No later than eleven months before the start of the timetable period, a provisional design of international train paths (vehicle paths in time-space) is jointly made by the cooperating allocation bodies. Generally, these paths are not to be altered in the following steps of the establishment of the allocation timetables, unless there are compelling reasons, and consequently international trains have a higher priority than national trains. Next, carriers submit their scheduling slot requests. If there are conflicts between the requested scheduling slots, the allocation body will try to find a solution that is agreed with both parties. If, however, it is not possible to reach a compromise, a congestion fee may be applied, which may solve the conflict if one of the carriers decides to change the slot request. The final

step to resolve capacity conflicts is to apply priority rules. Finally, capacity for maintenance and for later *ad hoc* requests will be reserved in the annual allocation timetables.

These priority rules applied in case of capacity conflicts that cannot be resolved by negotiation should be included in a Network Declaration. A Network Declaration provides various information about railway infrastructure access, including information about which railway links are capacity-constrained (see e.g. Railned, 2001b). It is jointly published by the infrastructure provider and the allocation body (if these are separate bodies at all). In the Dutch situation, however, the allocation body is not free to determine the priority rules. These priority rules are provided by the aforementioned Interim Decree on Railway Capacity Allocation (*Interimbesluit capaciteitstoewijzing spoorwegen*, 2000). These priorities are not absolute, but are guidelines how the interests of different carriers should be weighed in the allocation decisions of the allocation body (see e.g. figure 3.4).

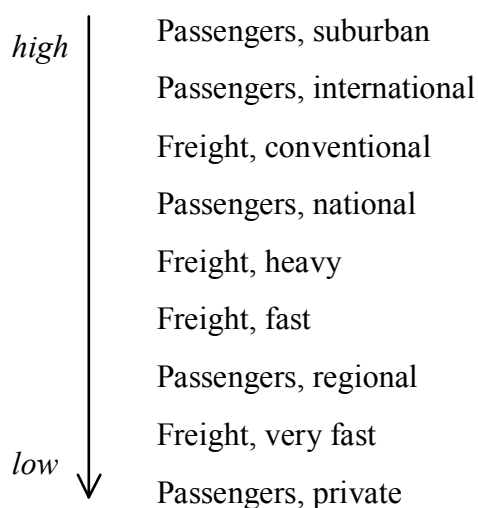


Figure 3.4: Priority list of types of transport services in the Netherlands (source: *Interimbesluit capaciteitstoewijzing spoorwegen*, 2000)

The Interim Decree on Railway Capacity Allocation is not clear on whether these priority criteria should be applied to scheduling decisions, selection decisions, or both (see section 2.4 for a definition of selection and scheduling slot allocation). Since most capacity conflicts at the selection level can already be resolved by applying the minimum frequencies, it is reasonable to assume that, in practice, these criteria will mainly be applied to capacity conflicts at the scheduling level. However, the arguments on which the chosen priority list is based seem to assume capacity conflicts at the selection level. For instance, the Dutch Minister of Transport argues that time-critical (fast and very fast) freight transport should have a lower priority than conventional and heavy freight transport, because only the latter are proven markets (see figure 3.4). This is a reasonable argument for selection slot allocation decisions, but for scheduling slot allocation decisions the most straightforward argument would be to give priority to time-critical freight transport, because hidden and on network delays are less acceptable for

this market segment by definition! It seems that selection and scheduling slot allocation are currently confused in the Dutch railway sector (Koolstra, 2001b). This confusion could be avoided in the future by explicitly distinguishing between selection and scheduling slot allocation.

Ad hoc slot allocation

The spare capacity in the annual allocation timetables is available for short-term, non-periodical requests. Furthermore, in the allocation timetable international freight paths are reserved that are not yet allocated to specific carriers. Freight carriers may request to use these paths on a first come, first served basis (Railned, 2001b). *Ad hoc* slot allocation is required for freight trains that cannot be scheduled in the annual allocation timetable, for instance demand-driven services that are chartered by large companies, for instance steel manufacturing industries, car manufacturing industries, chemical industries, etc. Furthermore, new services may be introduced during the season, different paths may be required to optimize the usage of locomotives, regular trains may be split up if demand is higher or combined if demand is lower, etc. (De Jong, 2002).

The allocation body should respond within 5 working days to any *ad hoc* request (EC, 2001c, art. 23). *Ad hoc* slots may be requested any moment within the timetable season. In the Dutch situation, slot requests with respect to international freight paths may be submitted to the allocation body up to 5 days in advance. Furthermore, last-minute slot requests or changes of traffic plans can be submitted to rail traffic control. Last-minute changes are also handled on a first-come, first-served basis (Railned, 2001b).

In the Dutch situation, *ad hoc* slot requests have a relatively low status, which may be explained by the dominance of passenger transport in the railway sector. Because planned trains that have not yet been assigned a locomotive are treated as 'optional' instead of 'regular' trains, Railion is more or less forced to make a logistic plan for the entire season (De Jong, 2002). However, as was noted in section 2.5, Railion would prefer to leave logistic decisions to the daily traffic planning level, and hence reserve the corresponding slots at an *ad hoc* basis.

Rail traffic control

Rail traffic controllers are responsible for operational slot allocation decisions and for the operation of the traffic control system. Given that usually exclusive slot allocation is applied, rail traffic control only has to make allocation decisions in case of delays or failure of infrastructure. According to Dutch regulation, traffic control decisions should be made in consultation with the carrier if only one carrier is involved. If train delays are in excess of a certain margin, which may differ between railway nodes, rail traffic controllers may have to plan a new slot in cooperation with the carrier. Finally, in case of unforeseen capacity restrictions, for instance due to failure of infrastructure elements,

rail traffic control will aim to optimize the utilization of capacity, while respecting the original allocation decisions as much as possible (Railned, 2001b).

3.2.2 Air traffic slot allocation regime

Given the national sovereignty of states with respect to air traffic, including *en route* air traffic flying over their territories, network access is largely determined by bilateral agreements between states. The tactical planning level of the air traffic slot allocation regime in Europe is largely determined by the provisions of Regulation 95/93 (EC, 1993). Additionally, the scheduling guidelines of IATA are used, but only as far as these are not in conflict with European and national legislation. The continuity of selection of transport services is ensured by historic rights, i.e. carriers have first choice to reuse the slots they used in the previous allocation timetable. Since most air transport services are scheduled (see Appendix A), a large percentage of traffic is included in the allocation timetable. Similar to the railway sector, there it is also possible to obtain *ad hoc* slots, which is important for *ad hoc* services. An important difference with the railway sector is that allocation timetable and *ad hoc* allocation only applies to airports. However, the next step, the allocation of ATC slots shortly before departure, applies also to *en route* traffic. Operational capacity management is handled by operational traffic control. Figure 3.5 summarizes the main decision levels of air traffic slot allocation.

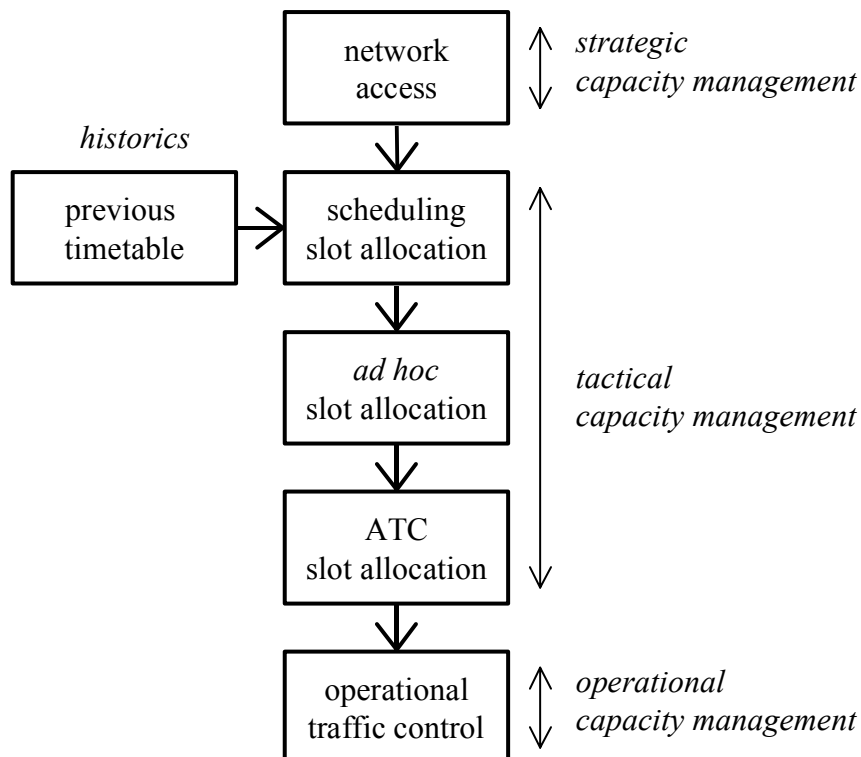


Figure 3.5: Main decision levels of air traffic slot allocation

Traffic network access

Access of international air traffic to the airports and airspace of states is governed by the Chicago Agreement and by bilateral treaties. In classic bilateral agreements, access is restricted to the national airline companies (*flag carriers*) for transport services with their home country as origin or destination. However, within the European Union and between countries with *open skies treaties* fewer restrictions are applied to foreign carriers (Schipper, 2001).

Two generic freedoms have been agreed in the Chicago Agreement, i.e. the right to pass over the territory of the signatory states without landing (first freedom) and the right to land in the territory of the signatory states for non-commercial reasons (second freedom). Other freedoms have to be agreed upon by bilateral treaties between states, for instance the right use an airport for commercial purposes for transport to and from the home country of the carrier (third and fourth freedoms), the right to perform transport services between two other countries (fifth to seventh freedoms), and the right to perform national transport services within the territory of another country (eighth freedom). All eight freedoms have currently been realized with respect to air traffic between member states of the European Union (EP, 2001). Free market access within the European Union has been achieved from 1 April 1997, based on the provisions of Regulation 2408/92 (EC, 1992b). There are, however, a few exceptions, for instance to allow states to protect the interest of newly introduced domestic services between regional airports (art. 6). Furthermore, states are allowed to make strategic decisions on the distribution of types of traffic among airports within a metropolitan airport system, for instance between the Paris airports Charles de Gaulle, Orly, and Le Bourget or between the Berlin airports Tegel, Tempelhof and Schönefeld (art. 7). Any carrier complying with the criteria of Regulation 2407/92 (EC, 1992a) can obtain a European operating license.

Seasonal airport slot allocation and historic rights

The procedure to allocate seasonal airport slots is broadly outlined by Regulation 95/93 (EC, 1993). Additionally, the provisions of the IATA Scheduling Guidelines (IATA, 2002) are used. Regulation 95/93 applies to community airports that have been granted the status of 'fully coordinated airport' in accordance with the provisions of this regulation. Allocation bodies, called coordinators in the Regulation, are appointed by the member states, usually by the Minister of Transport, and can be either legal or natural persons. In 1998, the Dutch national airport Schiphol was assigned as a fully coordinated airport (V&W, 1998a). The Dutch Minister of Transport has appointed Airport Coordination Netherlands as coordinator (V&W, 1998b).

The airline industry works with two timetable seasons per year, a summer season and a winter season. The summer season starts the last Sunday of March, while the winter season starts the last Sunday of October. To facilitate negotiations, Schedules Conferences are organized by the coordinators before the carriers establish their transport

service timetables. These Schedules Conferences are an initiative of IATA and do not only apply to fully coordinated airports, but also to airports where airlines coordinate their schedules voluntarily.

The first step of the slot allocation procedure is to determine the historic slot claims. A slot that has been operated by an air carrier in the previous year for the same timetable season entitles this air carrier to claim the same slot in the next season. More specifically, an air carrier can claim historic rights if it has operated a regular service on the same time of day and the same day of the week for more than 80% of the period within the last season that this slot has been allocated to this air carrier (EEC Regulation 95/93, 1993, art. 10). According to the IATA guidelines, carriers should jointly negotiate a solution if capacity reductions at an airport would result in a situation where not all historic rights may be respected. Another aspect is that carriers with historic rights requesting a different slot (i.e. earlier or later), have priority over new requests according to the IATA guidelines. Any slot that has not been used sufficiently, without a justified cause is collected in a slot 'pool'. Also slots given up by previous users are collected in this pool. New entrants have the first choice to request up to 50% of these slots. The remaining slots are allocated among the other air carriers (see figure 3.6).

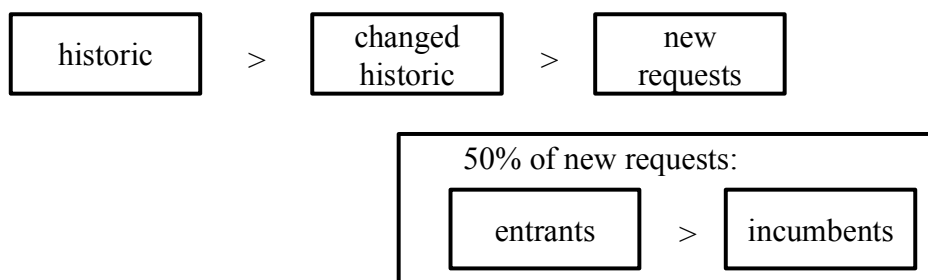


Figure 3.6: Priorities applied to establish allocation timetables

In principle, scheduling slots are allocated to a combination of carrier and route. However, after two years of operation, carriers have the right to use slots for different services or exchange slots with other carriers (EC, 1993, art. 8). Consequently, carriers are in practice free to optimize the usage of 'their' slots. This freedom is possible because slot only apply to nodes in the airline networks, enabling changes in traffic patterns without the need to involve the allocation body.

Ad hoc airport slots

Unwanted allocated slots should be returned at least two months before the start of the season. Moreover, if a transport service ceases to operate during the season, the airline company should immediately return the allocated slots for the rest of the season, since these slots can be reallocated for *ad hoc* use. According to the guidelines with respect to ad hoc operations at Heathrow airport, the allocation of ad hoc slots starts after the slot return deadline. Heathrow airport uses a priority list in case of conflicting ad hoc slot requests. Moreover, a waiting list of outstanding slot requests is maintained and if two

conflicting requests have the same priority, the date of request determines the final decision. The ad hoc slot allocation process continues throughout the season (ACL, 2000).

ATC slots

Most types of air traffic are required to reserve ATC slots. ATC slots apply to the entire vehicle path from airport to airport. ATC slots are requested by submitting flight at least 1 hour in advance (Donohue, 2001; Regeling Vliegplannen, 1998) in advance. ATC slots are allocated on a first-come, first-served basis. In some countries, airport traffic control centers of coordinated airports deny permission to flights without an airport slot, but there are also countries where air traffic control centers do not perform such a check (Vos, 2002). In the latter situation, there is no hierarchical relationship between airport slot allocation and ATC slot allocation, contrary to the theory outlined in section 2.5.

Air traffic control

Both airport slots and ATC slots are shared and consequently traffic conflicts remain to be solved by air traffic controllers. Air traffic controllers do not operate according to priority criteria. The main criterion for their decisions is safety. Only in the second place, additional criteria may play a role, such as the question of whether flights are on schedule and minimization of environmental burden, which is especially relevant near airports.

3.3 Main issues

In designing slot allocation regimes, a balance has to be found between various aspects, often corresponding with different policy objectives. Proposals for revisions of legislation regarding slot allocation often reflect these issues. In this section, three key issues are discussed with respect to finding the optimum balance between desirable properties of slot allocation regimes. The first issue is the balance between stability and flexibility of slot allocation, the second issue concerns the balance between competition between and diversity of transport services, and the final issue is heterogeneity of traffic.

3.3.1 Stability and flexibility

The introduction of framework agreements in the railway sector and the current application of historic rights in the aviation sector are solutions to facilitate the demand for selection slot allocation. The main purpose of selection slot allocation is to enable stability of transport service patterns. Stability is required by (existing) carriers for various reasons and this demand has been explicitly indicated in interviews with air carrier Transavia and rail carrier NS (Stellingwerff, 2002; Vos, 2002). Section 4.2 elaborates on the reasons why stability is required. The highest level of stability is achieved by the

application of historic rights, as is currently the case in the aviation sector. However, this system is criticized for being too inflexible and obstructing fair competition between incumbents and newcomers and among incumbents (see e.g. Starkie, 1998; European Commission, 2001).

A disadvantage of the application of historic rights, or other measures promoting stability, is that it results in less flexibility and less chances for newcomers. Flexibility is achieved when slot allocation decisions are frequently reconsidered and when historic considerations are not taken into account. However, not only flexibility is required to promote the introduction of new carriers and new services, but also stability. Carriers invest in new services and uncertainty about available infrastructure capacity in the future may impair the introduction of new services. Consequently, lack of stability may also have negative impacts on the level-of-competition between transport services (see also next sub-section).

To attain a better balance between stability and flexibility in the aviation sector, the European Commission has proposed to change Regulation 95/93 in such a way that historic rights are limited to a maximum of 7 years (European Commission, 2001). After this period, slots would be added to the slot pool. In the railway sector, a different solution has been chosen. No historic rights are given to scheduling slots, only framework agreements can have validity of 7 years. The advantage of framework agreements is that it leaves more room for annual timetable optimization, because the framework agreements are much broader than scheduling slots.

Every solution of the problem to find an optimal balance between stability and flexibility will be some sort of compromise between these objectives. However, if selection and scheduling slot allocation are treated separately, a different balance may be chosen for each level of slot allocation. As will be illustrated in section 4.1, the demand for stability is usually higher at the selection level than at the scheduling level, while the demand for flexibility is usually higher at the scheduling level. Therefore, this thesis recommends to separate selection and scheduling slot allocation as proposed in section 2.4, enabling sufficient stability at the selection slot allocation level and sufficient flexibility at the scheduling slot allocation level.

3.3.2 Competition and diversity

While the previous issue was related to the choice of slot allocation levels and the application of historic rights, the current issue is related to the priorities to be applied. It appears that an optimum balance has to be found between promoting competition and promoting diversity. This is especially an issue in open transport markets, such as the European international passenger transport market. The promotion of competition in European transport markets is one of the transportation policy objectives of the European Union (see section 7.2). However, more competition in the transport market im-

plies more parallel services, which means that more traffic is needed to serve a similar transport demand with similar availability in time compared with a situation without competition. In contrast, a high level of diversity is attained when the distribution of transport services in time and space is optimized to serve many different end users. This requires transport services to be spread in time and space, avoiding duplication, but also resulting in a low level-of-competition between transport services of different carriers.

A high level-of-competition is attained when two or more carriers offer parallel transport services that are regarded by end users as equivalent substitutes. For instance, morning travelers between London Victoria station and Gatwick airport may choose between the 9:30 Gatwick Express train service and the 9:32 South Central train service (2003 weekdays timetable). Similarly, morning travelers between Amsterdam Airport Schiphol and Lyon Satolas Airport may choose between the 9:10 KLM airline service and the 9:20 Air France airline service (2003 summer and winter timetables). Appendix D indicates that a substantial percentage of internal EU flights departing from Amsterdam Airport Schiphol are duplicate flights of competing carriers with similar departure times. Slots used for these duplicate flights cannot be used to offer higher frequencies on other lines or to increase the number of destinations served.

Table 3.1 exemplifies how the supply of airline services between Amsterdam and Spain might look like given alternative timetables with a high level-of-competition and with many direct connections. The distribution of flights over time is visualized by figure 3.7. In the first example, it happens that two flights of different carriers depart with the same destination within less than an hour. In the second example, duplicate flights are eliminated to facilitate the addition of 5 other Spanish destinations. Furthermore, only one carrier operates direct flights to any destination. It appears that the current timetable is more similar to the competition alternative than the connections alternative.

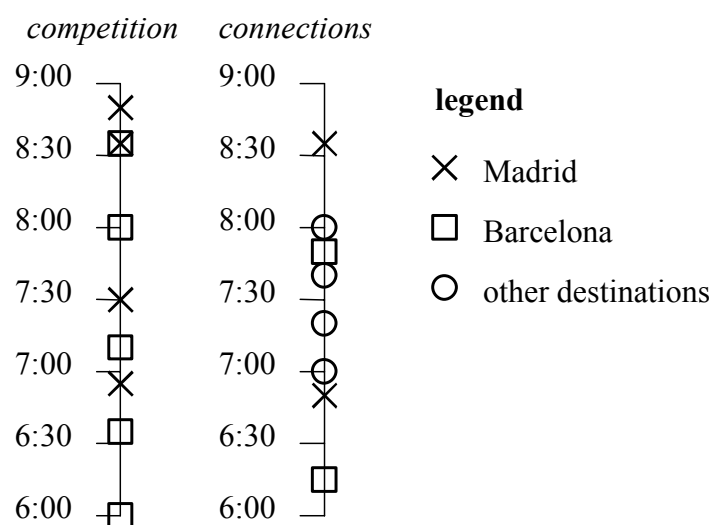


Figure 3.7: Distribution of departures over time given high level-of-competition and with many direct connections

Table 3.1: Example of alternative 6-9 a.m. departures of scheduled airline passenger services between Amsterdam and Spain (based on 2003 summer timetable, Fridays)

<i>high level-of-competition</i>			<i>many direct connections</i>			<i>current timetable</i>		
dept. time	destination	airline	dept. time	destination	airline	dept. time	destination	airline
6:00	Barcelona	TA *	6:20	Barcelona	KLM	6:00	Barcelona	TA
6:35	Barcelona	Iberia	6:55	Madrid	Iberia	6:35	Barcelona	Iberia
6:55	Madrid	KLM	7:00	Sevilla	TA	6:55	Madrid	KLM
7:10	Barcelona	KLM	7:00	Las Palmas	TA	7:10	Barcelona	KLM
7:30	Madrid	TA	7:20	Alicante	TA	7:30	Madrid	TA
8:00	Barcelona	TA	7:40	Malaga	TA	7:40	Malaga	TA
8:35	Barcelona	Iberia	7:50	Barcelona	KLM	8:00	Tenerife	TA
8:35	Madrid	Iberia	8:00	Tenerife	TA	8:35	Madrid	Iberia
8:50	Madrid	KLM	8:35	Madrid	Iberia	8:50	Madrid	KLM

* TA: Transavia Airlines

Currently, only the airport slot allocation system has an explicit rule promoting competition, i.e. the 50% rule regarding slot pools (see sub-section 3.2.2). This rule is intended to be compensation for the incumbent-friendly principle of historic rights. Furthermore, it is hard to assess how allocation bodies balance between competition and diversity in the allocation of pool slots. According to the website of Airport Coordination Netherlands (ACN), the allocation body of Amsterdam Airport Schiphol, frequency and competition are indicators used to decide on the allocation of pool slots. However, ACN has not been found willing to provide more details on the criteria it uses. In the railway sector, no explicit references have been found in EU legislation regarding slot allocation policy with respect to choices between more competition and more diversity. Dutch legislation is also relatively vague on this issue, giving only broad indications of the aspects that have to be taken into account, including the aim to obtain a fair distribution of adaptation costs between carriers (Interimbesluit capaciteitstoewijzing spoorwegen, 2000).

The problem to find the optimum balance between competition and diversity may be circumvented by allowing only competition for concessions to operate transport services, without free competition in the transport market. It is also possible to leave the establishment of this balance to the market, for instance by introducing slot auctions or free trade of slots. The later solution is favored for airports by Starkie (1998). However, some authors fear that large carriers might buy up all free slots to avoid competition (e.g. McGowan & Seabright, 1989). Section 5.3 discusses the pros and cons of slot auctions in more detail.

In this thesis, a generic selection slot allocation decision procedure is proposed which allows the inclusion of both competition characteristics and characteristics of slot re-

quests related to the added value of proposed transport services. However, it is up to the decision-makers to recognize the dilemma between promoting competition and promoting diversity and to decide on the preferred balance between these two objectives.

3.3.3 Heterogeneity

In the context of this thesis, heterogeneity of traffic is especially relevant as far as a different composition-of-traffic will result in a different capacity. Heterogeneity of traffic may be due to either transport service characteristics or vehicle characteristics. An example of heterogeneity largely due to service characteristics is mixed traffic of local and intercity trains. These services may be operated by the same train types, but local trains have a lower average speed due to the larger number of stops. Mixed operation of slow heavy freight trains and fast passenger trains and mixed traffic with small and large planes are examples of heterogeneity that is largely due to vehicle characteristics (see section 6.2).

Heterogeneity of traffic due to vehicle characteristics can only be reduced by changing the composition-of-traffic per infrastructure element per period. In a number of situations, heterogeneous traffic results in larger minimum headways between vehicles. For instance, a phenomenon called 'wake turbulence' arises in air traffic. A relatively large separation is needed for small aircraft following large aircraft to avoid that small aircraft become uncontrollable due to the turbulence caused by large aircraft (Nolan, 1994). Chapter 6 elaborates on the relationship between heterogeneity of traffic and minimum headways between vehicles. Capacity loss due to inhomogeneous traffic composition may be avoided by separating vehicle classes in time and space. For instance, operations of small aircraft may be concentrated on a separate runway or in specific periods.

A slightly different issue is heterogeneity of traffic due to heterogeneity of transport service characteristics, which is especially an issue in the railway sector. At the selection slot allocation level, the primary problem is that allocation bodies have to choose between facilitating heterogeneity with a lower capacity as the result, and stimulating the homogenization of traffic. Homogenization of traffic may not only be attained by giving priority to slot requests with similar vehicle and transport service characteristics, but also by stimulating carriers to adopt a transport service model resulting in less capacity loss due to heterogeneity. For instance, Zigterman *et al.* (1999) propose to facilitate more traffic on capacity-constrained railways in the Netherlands by replacing the traditional local and intercity trains by only intercity trains having one or two extra stops each to serve smaller stations previously served by local trains ('Inter-stop-city', see figure 3.8). However, it is not very likely that Dutch Railways will accept these proposals, given that studies from their product management department indicate that the percentage of customers experiencing negative effects is large compared with more conventional alternatives (Bouman *et al.*, 2001).

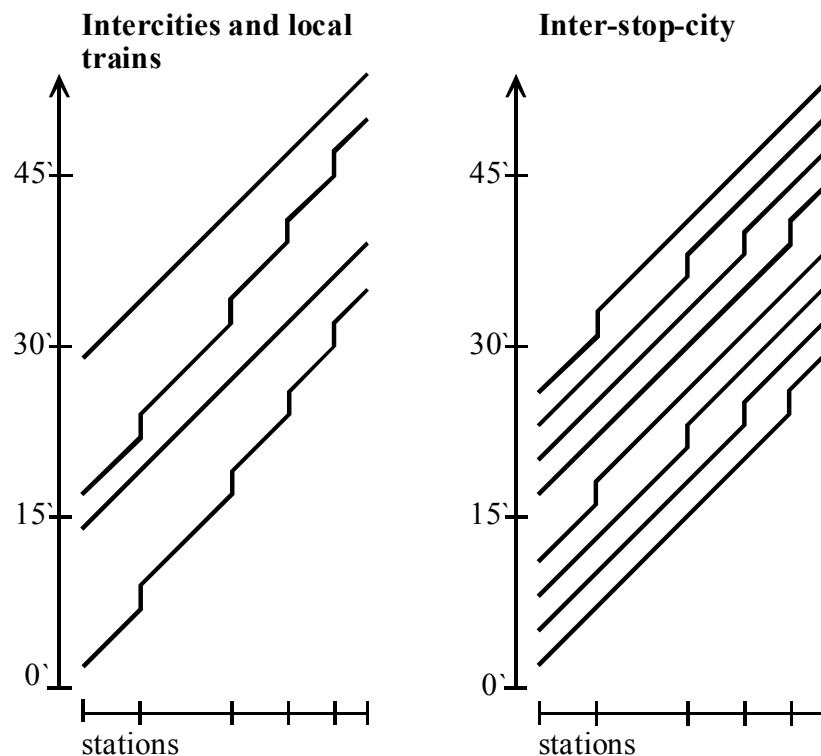


Figure 3.8: Illustration of capacity gain of homogenization of train traffic

An issue with respect to transport service homogenization is to which extend allocation bodies have the authority to promote the adoption of a different transport service model. In a reaction on proposals on the future distribution of railway capacity by the Dutch allocation body (Railned, 2000b), Mr. Van Dijk, representative of passenger carrier NS Reizigers, stated that the proposals in this report to homogenize railway traffic interfered with their entrepreneurial responsibilities. According to Mr. Van Dijk, determining the supply of transport services is solely the responsibility of carriers (Railned, 2000c). However, these arguments of NS Reizigers do not take into account that optimizing the utilization of infrastructure is the core responsibility of the allocation body, which implies that it is reasonable that allocation bodies actively investigate alternative infrastructure utilization models.

The current slot allocation framework in both the aviation and the railway sectors, as defined by European legislation, gives allocation bodies the opportunity to give higher priority to slot requests corresponding with an efficient usage of the infrastructure. For instance, a proposed new airline serving about 250 passengers each flight may be given priority over a proposed new airline serving about 100 passengers each flight. Furthermore, homogenization of transport service characteristics is possible in railway sector at the scheduling level. For instance, railway slot allocation bodies may give fast trains slower slots than requested in order to facilitate more railway traffic. Another possible alternative at the scheduling slot allocation level is to specify 'average' slots and to sell these for market price. Carriers will have to buy several of these standard slots to facilitate vehicle paths that are slower or faster than average. This principle is already applied

to allocate slots for the Channel Tunnel between France and England (Imhof, 2001). Finally, specialization of infrastructure is possible, for instance the designation of high-speed or freight railway lines, international or domestic airports, etc. It seems that, to a large extent, the choice of strategy to be followed with respect to homogenization is usually left to the allocation bodies.

The generic selection slot allocation decision procedure proposed in this thesis tackles the heterogeneity problem by calculating the efficiency of alternative slot requests given the expected average usage of the bottleneck (see chapter 7). However, given the focus of this thesis on selection slot allocation, the approach proposed in this thesis does not solve the homogenization problem at the scheduling slot allocation level. According to the approach followed in this thesis, the capacity at the selection level, and as a result the maximum degree of homogenization at the scheduling level, is determined by the chosen balance between capacity and quality-of-service. Optimizing this balance is the responsibility of allocation bodies.

3.4 Conclusions

In this chapter, we have explored the current practice with slot allocation in different infrastructure sectors. Furthermore, we have reviewed the desirability of slot allocation in various sectors. Finally, we have formulated a number of key issues with respect to slot allocation. From this overview of current practice and main issues, we may draw a number of general conclusions.

Application areas of slot allocation

Slot allocation is currently applied in the railway and aviation sectors, but not yet in the road sector and only scarcely in the navigation sector. Applications in road and navigation sectors have been proposed in the literature. However, the road sector currently lacks an appropriate control system, and the acceptance of traffic planning requirements may be less in a sector with dominantly private transport. The restriction of slot allocation in the navigation sector to container terminals is probably due the low level of capacity scarcity in this sector, at least outside the terminals. Introducing slot allocation to congested freeways, locks, etc., is possible, but a detailed evaluation of the pros and cons of slot allocation is required to assess the desirability of the introduction of slot allocation in these sectors.

Which slot allocation levels are needed in which sector depends on the types of transport services to be facilitated. For instance, if a bottleneck is dominantly used by *ad hoc* services, there is no demand for slot allocation at the selection and scheduling slot allocation levels. Given that the air and rail transport sectors have a large percentage of structural supply-driven services and a substantive minority of *ad hoc* services, all three slot allocation levels are required in these sectors.

In the railway sector, the slot allocation regime applies to virtually the entire infrastructure network. Excluded, however, are dedicated networks that are exclusively used by one carrier, for instance suburban and metropolitan railways, because these carriers may solve capacity conflicts internally in their own timetables. Slot allocation is possible at all planning levels, from multi-year framework agreements to short-term *ad hoc* slots. With respect to air traffic, only the short-term ATC slot allocation system applies to the entire network. Slot allocation at higher planning levels is only applied to congested airports. Although no suggestions have been found in the literature to extend the ATC slot allocation system to allow reservation of slots several days in advance or even at the timetable planning level, it would be advisable to consider these extensions if ATC handling capacity increasingly becomes a bottleneck in the future.

Current slot allocation regimes

The current European slot allocation regimes have been reviewed in this chapter. These appear to consist of several decision levels, largely similar to the decision levels proposed in sub-section 2.4.4. Apart from the slot allocation decision levels, we also focused on traffic network access (strategic capacity management) and traffic control (operational capacity management). At the tactical capacity management level, up to four different levels of slot allocation have been identified: framework agreements, scheduling slot allocation, *ad hoc* slot allocation, and (air) traffic control slot allocation. In the railway sector, the current levels of slot allocation largely correspond with the levels distinguished in theory, but the aviation sector lacks a separate selection level, and features an additional level in the form of ATC slot allocation.

In both the railway and aviation sectors, scheduling slot allocation is the dominant level of slot allocation, while selection slot allocation is relatively underdeveloped in these sectors. In the first place, structural supply-driven transport services are dominant in both sectors, which implies that most transport services run according to a timetable. Furthermore, the aviation sector lacks a separate selection slot allocation decision level and the system of framework agreements has not yet come into operation in the railway sector. In the aviation sector, selection slot allocation is completely integrated with scheduling slot allocation. The principle of historic rights ensures the stability of transport service patterns, but this principle is criticized for discouraging the entry of newcomers on the air transport and traffic markets. In the railway sector, the possibility to conclude framework agreements has been introduced, which, depending on national legislation, may become the basis of a selection slot allocation system. However, no explicit evaluation is envisaged with regard to these framework agreements, and the European legislation is vague about the issue to which extent framework agreements may be used to solve capacity conflicts at the selection slot allocation level.

Slot allocation at the *ad hoc* level is possible in both the aviation and the railway sector. The aviation sector even distinguishes between *ad hoc* airport slots and ATC slots,

which can be explained by the separation of airport slot allocation and ATC slot allocation and the weather-dependency of air traffic, resulting in short-notice planning. However, due to the lack of a separate, fully developed selection slot allocation system, stability of transport service patterns can only be ensured by reserving scheduling slots. However, this requires a level of precision in space and time that is usually not desired by carriers offering demand-driven transport services. An alternative would be to allow these carriers to reserve selection slots corresponding with their transport service patterns, and reserve *ad hoc* slots when these transport services are scheduled, while the scheduling slot allocation level is not used.

Main issues

In this chapter, three key issues with respect to the design of selection slot allocation regimes have been identified. Most discussions in policy documents and scientific papers about the desirability of adaptations in current slot allocation regimes are related to these issues. The first issue is stability *versus* flexibility. An important difference between the railway and aviation sectors is that the stability of transport service patterns is ensured in different ways. In the airport slot allocation regime, historic rights are applied giving carriers the right to claim the exact scheduling slots they used in the previous timetable season. In the railway sector, carriers may conclude framework agreements assuring the feasibility of the desired transport service pattern for a number of years. The application of historic rights is criticized for minimizing flexibility and reducing the opportunities for newcomers to obtain slots. A less rigid alternative is to ensure long-term stability at the selection level only, but not at the scheduling level. This approach has been chosen in the railway sector, where long-term stability can be ensured by concluding framework agreements.

Another issue is that decision-makers with respect to the design of slot allocation regimes should be aware that choices have to be made between competition and diversity, because stimulating competition between transport service providers means less room for diversity in situations of capacity scarcity. This is a conflict between different objectives of the decision-makers. It is the responsibility of these decision-makers to recognize this dilemma and to decide on the desired balance between diversity and competition.

The final issue discussed in this chapter is heterogeneity. The key problem is to find a balance between heterogeneous traffic demand and the fact that more demand can be facilitated if traffic is homogenized. This is a specific example of the balance between capacity and quality-of-service. Again, it is the responsibility of the decision-makers to decide on the desired balance between capacity and quality-of-service, given heterogeneous traffic demand.

Recommendations

Discussions about the design of slot allocation regimes, including discussions about the application of historic rights and about the priorities to be applied, would be much clearer if the difference between selection and scheduling decisions would be recognized explicitly. For instance, in the current discussion to restrict historic rights in the aviation sector, clarity is desired about whether historic rights are desired at the scheduling level or only at the selection level. Another example is that the motivation of priorities applied in the Dutch railway slot allocation system is implicitly based on selection decisions, while these priorities will mainly be used for scheduling decisions. Decisions on slot allocation procedures can only be made in a rational way if the alternatives are clear, which requires explicitly distinguishing between selection and scheduling.

Selection and scheduling slot allocation should not only be separated in discussions, but may best be treated as truly separate levels of slot allocation in practice. Separating selection and scheduling slot allocation would enable the application of different criteria to these levels of slot allocation, resulting in a better match between allocation decisions and objectives of slot allocation. Furthermore, it has been argued in this chapter that selection and scheduling slot allocation require a different balance between stability and flexibility, which is another reason to separate these levels of slot allocation. Finally, the flexibility of slot allocation could be enhanced by allowing carriers to reserve slots only at the selection and *ad hoc* levels, without having to reserve slots at the scheduling level. The latter is an interesting option for structural demand-driven transport services and for supply-driven transport services working with relatively crude timetables, including many freight transport services. This again requires the explicit separation of selection and scheduling slot allocation levels.

Next chapters

Based on the recommendation of this chapter that selection and scheduling slot allocation should be separated, the following chapters focus on selection slot allocation as a separate slot allocation problem. The structure of the selection slot allocation problem is the focus of the next chapter. The stability and flexibility issue that has been introduced in this chapter will be elaborated on in the next chapter to identify the desired validity of selection slot allocation decisions. Furthermore, other characteristics of selection slot allocation are discussed, such as the desired dynamics of the selection slot allocation process.

In this chapter, it has been concluded that finding the optimum level of heterogeneity means finding a balance between capacity and quality-of-service. Chapter 5 discusses the generic problem of finding this balance. More specifically, the relationship between heterogeneity and infrastructure capacity is discussed in chapter 6. Furthermore, this chapter elaborates on the identification of types of infrastructure elements and traffic processes as infrastructure bottlenecks in current practice. Finally, the current levels of

infrastructure capacity scarcity will be identified in this chapter, based on which conclusions may be drawn on the desired levels of slot allocation.

The issue to find a balance between competition and diversity is based on conflicting objectives among and between authorities, carriers, and end users. For instance, the objective of a dominant carrier to consolidate its market share may be in conflict with the objective of the authorities to increase the level-of-competition in the transport market. The question of what are likely objectives of allocation bodies is one of the issues discussed in chapter 7.

4

SELECTION SLOT ALLOCATION SYSTEM DESIGN

In the previous chapters, the problem to be analyzed in this thesis has been introduced. In chapter 2, a conceptual framework has been formulated with definitions of main concepts, while chapter 3 provided an overview of current practice and current problems with slot allocation. Chapters 4 to 7 continue with the specification of the selection slot allocation problem. The objective of these chapters is to analyze the essence of the selection slot allocation problem and to specify the desired characteristics of selection slot allocation systems. This chapter provides a system design of selection slot allocation, based on the demand for selection slot allocation. By investigating the desired characteristics of selection slot allocation systems, this chapter is a bridge between the empirical perspective of chapter 3 and the theoretical perspective of chapters 5 and further. Based on the demand for slot allocation, this chapter discusses the desired characteristics of selection slot allocation with respect to the validity of selection slots (section 4.1), the dynamics of selection slot allocation processes (section 4.2), and the specification of selection slot size (section 4.3). Information from interviews with a number of carriers is used to specify the characteristics of selection slot allocation as desired by carriers. It is concluded that selection slots should be valid for several timetable seasons, and that at an opportunity to reserve selection slots should be offered at least every timetable season. This implies that various selection slot requests may be evaluated simultaneously. Furthermore, either a discrete or a multi-level slot specification is proposed, which means that standard slot periods are specified by the allocation bodies. A straightforward formulation of capacity constraints in the slot allocation decision problem is enabled by avoiding partial overlapping of selection slots. Based on these choices, economic theory will be used to analyze the traffic market in the next

chapter. The identification of traffic supply and demand is the subject of chapter 6. Finally, the choices made in the current chapter and in chapter 6 will be used in chapter 7 to formalize the selection slot allocation decision problem.

4.1 Demand for selection slots

The subject of this section is the demand for slots at the selection level. We first focus on the desired validity of selection slots. The demand for selection slots is largely based on the demand for stability of transport services. Recalling section 2.5, the selection slot allocation level has been specified based on a specific level of transport service planning, i.e. the design of transport service patterns. Furthermore, we have seen in section 3.2 that the application of selection slot allocation, for instance in the shape of framework agreements in the railway sector, is related with the establishment of concessions and long-term transport contracts. The second topic of this section is the question of who should be allowed to request and hold slots (slot holdership). A related issue is the desirability of restrictions on slot usage.

4.1.1 Desired validity of selection slots

The validity of slot allocation decisions is the period that the slot allocation decision cannot be changed without the consent of the slot holders. As we have seen in section 3.2, the validity of scheduling slots is generally the same as the timetable season. However, based on the discussion about stability and flexibility in section 3.3 we may conclude that a longer validity is desired for selection slots. In this section, the desired validity of selection slots given the desired stability (and flexibility) of transport service patterns is investigated.

Stability of transport service patterns

In general, carriers only redesign their transport service networks if there are convincing reasons to change it. Changing a transport service pattern requires an extensive design and evaluation process. Therefore, the Dutch national railway company NS does not evaluate the transport service pattern annually; only adaptations of the infrastructure network may lead to small or larger adaptations of the transport service pattern (Stellingwerff, 2002). For instance, the last major change in the Dutch intercity rail transport network dates from 1994 when a new railway was opened providing a shortcut between Amsterdam Airport Schiphol and various destinations in the Northeastern provinces of the Netherlands. But not only the spatial pattern of transport service is relatively stable. The temporal structure of the transport service network has been stable since 1970, when NS introduced a system of hourly and half-hourly services and fixed connections between lines named *Spoorslag 70*. On the other hand, the annual review of the NS timetable also often results in only minor changes, but this is a direct result of the stability of the transport service pattern.

A possible reason for the stability of transport service patterns is that the current situation may be the best solution. However, there are also other factors explaining this stability. A major economic reason to demand stability is that setting up new transport services requires specific investments. These costs may be recovered given a sufficient period of operation of these transport services. According to carriers in the rail and air transport sectors, this period of operation is much longer than the timetable period, in particular if investments in vehicles are at stake (see Vos, 2002; Stellingwerff, 2002).

The costs of setting up new transport services include a number of aspects, including the following. In the first place, setting up new transport services entails design- and marketing costs. These costs include the ex ante evaluation of the new service and publicity to inform existing and potential customers. For example, when Transavia Airlines introduced direct flights from Amsterdam to Bordeaux, Marseille and Madrid in 2002, they launched an extensive marketing campaign. Secondly, specific investments may be needed when setting up new transport services, such as the procurement of vehicles. To buy a new plane or train generally takes more than a year, while these may stay in operation for several decades. Another example of a required investment is that introducing new destinations in an airline network may require setting up sales offices in the destination country. Finally, the yield will generally be low in the initial period of operation because it generally takes some time for a new service to achieve its full market share. Therefore, the objective of Transavia is to have new services profitable within a year (Vos, 2002).

Another reason for stability of transport service patterns over several timetable periods is that the supply of some transport services has been guaranteed for a number of years, for instance in a government concession or in a private contract. For instance, European governments may tender air services between regional airports and major hubs to ensure that the quality of transport service supply meets the desired standards (Regulation 2408/92, EC, 1992b, art. 4). The same holds for public transport by bus and rail in many European countries. With the publication of the new Transit Act, a formal concession system for public transport has been introduced in the Netherlands (*Wet personenvervoer 2000*, 2000). The maximum concession duration for non-rail public transport is 5 years, which is a compromise between continuity and competition (*Wet personenvervoer 2000*, 1998). Recently, the Transit Act has been changed to extend this concession system to the rail transport sector (*Concessiewet personenvervoer per trein*, 2003). Apart from government contracts and concessions, carriers may conclude private contracts with shippers. In 2002, for instance, chemical manufacturer Bayer concluded a multi-year contract with railway carrier Rail4Chem to cater for shipments of their supplies (Latten, 2002). Both the conclusion of concessions and the conclusions of private transport contracts result in stability of transport service patterns. To facilitate these stable transport service patterns, long-term selection slot validity is required.

Conclusions on selection slot validity

From the discussion in this section, we may conclude that transport service patterns change less often than timetables. Furthermore, carriers demand certainty about the long-term availability of sufficient infrastructure capacity to be able to recover investments and meet contractual obligations. This demand for stability is met given the application of selection slot allocation with a sufficient validity of the selection slots. Given the long-term nature of investments and the typical duration of concessions, a selection slot validity of at least five years is advisable, at least in situations where these issues are dominant. An infinite validity (historic rights) of selection slots is not recommended here, unless there are sufficient other incentives to give up selection slots relatively easily. A reasonable rate of 'freeing' capacity to be reallocated is desired, in order to attain sufficient flexibility to change transport services and to introduce new transport services.

4.1.2 Slot holdership

According to the definition of slot allocation in section 2.4, slots are allocated to slot requests. Slot requests include information about the identity of the slot holder and the intended usage of the slot. The question remains, however, which actors should be allowed to request and hold slots and to which extent slot holders should be allowed to use slots for different types of vehicle paths or to sell slots to or exchange slots with other carriers. Furthermore, it has been noted before that, although it is no slot allocation according to the definition of section 2.4, allocation bodies may also be allowed to reserve selection slots for specific uses, without allocating these slots to a specific carrier.

Changes of slot usage

We first focus on the question of to which extent usage of slots for different types of vehicle paths should be allowed. In the first place, restrictions on slot usage are required because traffic characteristics such as vehicle type and operational speed may influence capacity. Furthermore, since allocation decisions may be based on characteristics of the envisaged slot usage, and to avoid change of usage that are not in accordance with the objectives of the allocation body, slot usage may be limited to a single purpose. Finally, the main argument for long-term validity of selection slots is the desired stability of transport service patterns. When changes of slot usage are due to changes of the transport service pattern, this argument for long-term validity is not valid anymore. Consequently, it is reasonable to reallocate these selection slots in this case, enabling others to issue competing slot requests. On the other hand, stability of utilization of vehicles and staff is an argument to allow carriers to keep selection slots when changing their usage. To conclude, restrictions of selection slot usage may be required to avoid violation of capacity constraints, and may be desired to avoid unwanted changes of slot usage. In

this thesis, it is assumed that slot usage is restricted, allowing the valuation of alternative slot allocations to be based on characteristics of slot usage.

A related issue is to which extent slot holders should be allowed to sell or exchange slots, given that the restrictions on slot usage discussed above are satisfied. Two possible objections against sell or exchange of slots are reviewed here. In the first place, slots may be allocated for free or for a low fee, allowing slot holders selling slots to make an easy profit. On the other hand, it may be argued that this transaction only reflects economic optimization. In the second place, slots allocated to small competitors may be bought by large incumbents, which may have a negative effect on competition. This is especially an issue when one of the reasons to allocate the selection slot under consideration to the small competitor was to promote competition. In this case, it is reasonable to introduce a competition check for slot sale and exchange. However, apart from this possible competition check, this thesis allows for the possibility that slots can be sold and exchanged between slot holders, if only the slots continue to be used as specified in the slot allocation decision. Nonetheless, slot exchanges should be subject to approval from the allocation body, for instance to check whether they are not in conflict with objectives regarding competition or diversity.

Possible slot holders

In section 2.4, it has already been stated that slots will usually be allocated to carriers, because these are responsible for slot usage. In current practice, both airport slots and railway slots are usually held by carriers. However, it is possible to allow the reservation of selection slots to various other parties. For instance, EU Directive 2001/14 (EC, 2001c), which provides the institutional framework for the European railway slot allocation regime, states that framework agreements may be concluded not only with carriers and authorities, but also with other interested parties such as shippers and freight forwarders.

There are a number of reasons to allocate selection slots to other parties than carriers. In the first place, allocation of slots to authorities enables them to promote transport services that meet their objectives. For instance, several local communities in the USA currently hold slots to enhance services between their airports and large congested airports such as Chicago O'Hare Airport (Starkie, 1998). Furthermore, authorities issuing concessions may directly ensure the feasibility of the desired transport service pattern when they reserve the required selection slots themselves, and it is reasonable that when for some reason a concession is revoked and given to another carrier, the associated selection slots are also to be used by the new carrier. Similarly, allocation of selection slots to shippers, tour operators, etc., has the advantage that shippers or tour operators can decide to contract a different carrier when the current carrier does not fully meet its contractual obligations, without running the risk that the desired transport pattern is discontinued because they cannot obtain new selection slots.

To conclude, selection slots are associated with the desired stability of transport service patterns. In many occasions, authorities, end users, or in-betweens such as tour operators are the primary interested parties in maintaining these transport service patterns, and therefore these should be allowed to request and hold slots. In this case, the actor holding the slots is entitled to choose the carrier that will use the slot to produce transport services. However, when slots are requested by carriers related with investments in new transport services, it is reasonable that slot holdership remains with the carrier.

The question of whether slots should be held by carriers or by tour operators is currently an issue in the aviation sector. In the air transport sector, it is common practice that tour operators sell holiday packages to travelers that include flight and accommodation. For this purpose, they contract a carrier to organize the required flights. However, when a tour operator contracts a different carrier for the same transport service in the next season, the associated slots remain with the first carrier. In the past, this has been disputed in court, albeit without success (Vos, 2002). However, when a tour operator has the role of transport service provider, as is the case when the tour operator is the only interface with the travelers while the carrier only organizes and performs the flight, it would be reasonable to give the tour operator slot holdership ('slots follow business'), instead of the carrier.

4.2 Slot allocation dynamics

In this section, the difference between static and dynamic decision-making is introduced. This section first reviews the differences between these two extremes. Next, the dynamics of current slot allocation systems are briefly reviewed. Finally, this section discusses the desired dynamics of different levels of slot allocation, focusing on selection slot allocation.

4.2.1 Static and dynamic slot allocation

The dynamics of the slot allocation decision process is an aspect that has not yet been discussed in this thesis. The decision-making process may have different dynamics for different levels of slot allocation. This thesis distinguishes static and dynamic slot allocation. Figure 4.1 illustrates the main differences between static and dynamic slot allocation decision processes. Static slot allocation decisions are valid for a certain period, for instance a timetable period, and may be based on a comprehensive evaluation of possible allocation alternatives. All slot requests are collected before the final slot allocation decision is made. During this process, conflicting slot requests have to be resolved by, for instance, optimization, negotiation, priority rules, or auctioning. In contrast, dynamic decision processes consist of a sequence of decisions that together determine the allocation of selection slots. All slot requests are evaluated sequentially, ideally directly after being received, while previous allocation decisions determine which slots remain to be allocated. Given this sequential character, a comprehensive simulta-

neous evaluation is not possible, since there is no direct comparison of conflicting slot requests. However, the allocation body may take into account expected future slot requests when deciding on current slot requests.

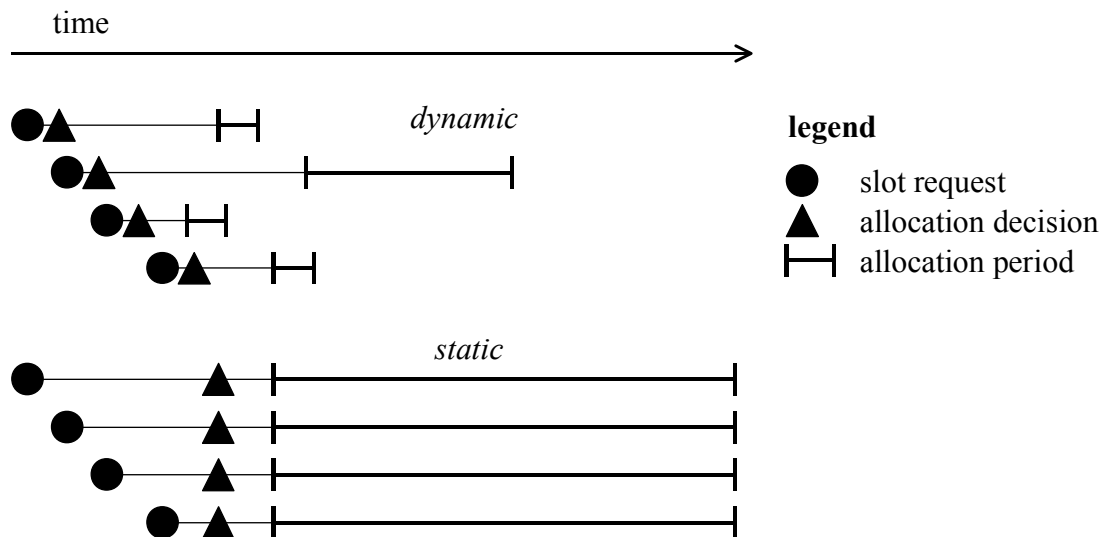


Figure 4.1: Difference in dynamics of static and dynamic slot allocation

The main advantage of static slot allocation is the possibility of comprehensive simultaneous evaluation of all slot requests regarding the next allocation period. Given that slot allocation is applied in situations where traffic supply is lower than traffic demand, the ability to optimize infrastructure usage is desirable. As a result, however, there may be a large delay between issuing a slot request and the allocation decision, which may imply a delayed introduction of new transport services. Dynamic planning increases the flexibility of slot allocation systems by introducing direct response to slot requests. In this case, the decision whether to grant the requested slot or not is based on the existing allocations and possibly also on the expected future requests. There is, however, no explicit comparison of the interests of users with conflicting slot requests, at least if no provisions are made for other carriers to submit competing selection slot requests. The latter situation is disadvantageous for newcomers, because incumbents would usually be the first to submit a new selection slot request when giving up slots they held previously.

Slot allocation dynamics in current practice

Static decision-making is currently applied to the establishment of allocation timetables in the railway and aviation sectors. In contrast, dynamic decision-making is applied to *ad hoc* slot allocation (including ATC slot allocation). Finally, framework agreements apply to a long period and the conclusion of framework agreements will probably be based on a profound evaluation, but nevertheless it seems that framework agreements will be concluded separately from each other. Consequently, although framework agreements have a long validity and may be concluded by carriers for a large volume of traffic simultaneously, the allocation procedure is dynamic rather than static. The next

sub-section proposes for each level of slot allocation a suitable type of decision dynamics.

4.2.2 Proposal for selection slot allocation dynamics

In the current slot allocation regimes described in chapter 3, a static procedure is applied to scheduling slot allocation, while a dynamic procedure is used for *ad hoc* slot allocation. Given the relationship between transport timetable planning and allocation timetable planning, it is natural to apply static slot allocation at the scheduling slot allocation level. Furthermore, dynamic decision-making is probably the best solution for *ad hoc* slot allocation. Although some *ad hoc* transport demand may be known far enough in advance to be included in the seasonal timetable, *ad hoc* transport services will usually be planned at shorter notice. For example, special flights for cup finals can only be scheduled when the finalists are known. This short-notice planning can best be performed dynamically.

The choice of allocation dynamics for selection slot allocation is less straightforward. A semi-static procedure (see figure 4.2) is possible if selection decisions are generally only required at the start of a new timetable season. This is the case in a situation with dominance of scheduled transport services. In the semi-static procedure, selection slot allocation decisions are made every timetable season, but these allocation decisions are valid for several years, or infinitely if historic rights are applied. In this case, slot allocation decisions may be based on an evaluation of alternatives, but selection decisions of earlier periods have to be respected as long as these decisions are still valid.

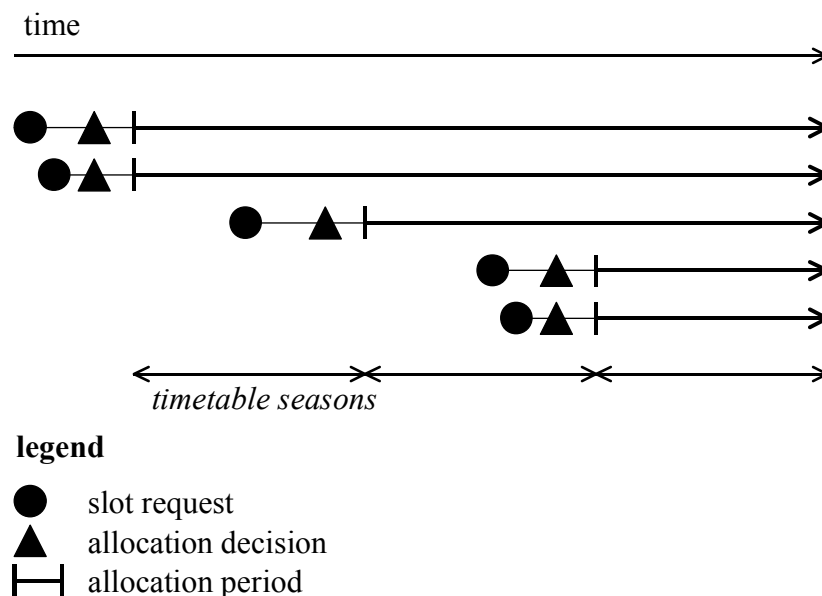


Figure 4.2: Semi-static slot allocation

The best solution in a situation with both supply-driven and demand-driven transport services is probably a combination of semi-static selection and dynamic selection.

Given that new structural supply-driven transport services will usually be introduced in a new timetable season, it is reasonable to make a comprehensive selection slot allocation decision before each timetable season, evaluating all selection slot allocation requests for that season. Furthermore, it should be noted that based on a single concession, a large percentage of all traffic may be affected. This would for instance be the case if a framework agreement were concluded with NS Reizigers, which is the largest railway carrier in the Netherlands being responsible for more than 80% of railway traffic on main railway lines. Additionally to this semi-static procedure, the slot allocation regime may allow the reservation of selection slots throughout each timetable season to facilitate the introduction of new demand-driven transport services, which corresponds with a dynamic slot allocation procedure. Particularly new structural demand-driven transport services may be introduced at any moment within a timetable season. Parallel to concluding a transport contract between a carrier and a shipper, selection slots required to facilitate this transport service may be requested by either the carrier or the shipper. If dynamic selection slot allocation is enabled, they do not have to wait for the start of the new timetable season to obtain the required selection slots.

Because scheduled transport services are dominant in the rail and air sectors, this thesis proposes a semi-static selection slot allocation procedure for these sectors. Alternatively, dynamic slot allocation may be preferred in sectors where demand-driven transport services are dominant, e.g. the road sector. Nonetheless, to enable the evaluation of alternative allocations, a semi-static procedure with frequent evaluations of slot requests is also a good alternative. Therefore, the selection slot allocation problem formalized and analyzed in the remainder of this thesis is based on the assumption that several slot requests are evaluated simultaneously.

To conclude this section, table 4.1 summarizes which types of slot allocation problems apply for which slot allocation stages. In this section, we have concluded that dynamic slot allocation is suitable for *ad hoc* slot allocation decisions. Scheduling slot allocation demands for static decision-making, because this allows an explicit evaluation of the interests of different carriers, resulting in a balanced allocation timetable. Finally, a semi-static procedure is suitable for selection slot allocation problems, at least if a significant percentage of transport services are scheduled in seasonal timetables. Additionally, dynamic selection slot allocation may be allowed, but the remainder of this thesis assumes a semi-static decision procedure.

Table 4.1: Slot allocation planning stages and preferred decision dynamics

planning stages	problem types
selection slot allocation	semi-static (or dynamic)
scheduling slot allocation	static
<i>ad hoc</i> slot allocation	dynamic

4.3 Specification of selection slots

In section 2.4, slot has been defined as 'a time-space domain on an infrastructure element designed to facilitate a specific vehicle path with a certain minimum quality-of-service'. The slot specification problem, which is the main subject of this section, is to specify the borders of this time-space domain. In the first sub-section, three alternative ways to specify slots are reviewed, i.e. overlapping, discrete, and multi-level slot specification. The desirability of slot size differentiation is discussed, as well as the desirability of straightforward slot specification, which may be realized by applying multi-level slot specification. The second sub-section introduces acceptability as the main criterion to determine selection slot size. To illustrate this, the application of a number of specific acceptability criteria to specify selection slots is discussed.

4.3.1 Specification of slot borders

Although selection slot allocation focuses on the selection of slot requests and not on the specification of time and space specific slots, slot size restrictions are also required at this level of slot allocation. As we have seen in section 2.4, slot size is defined by three variables, i.e. slot period, slot length, and bandwidth. Furthermore, the difference between non-exclusive and exclusive slots has been introduced in that section. For selection slot allocation, non-exclusive slot specification is the only reasonable alternative since by definition this level of slot allocation does not focus on the scheduling of vehicle paths.

Spatial and temporal slot dimensions

According to section 2.4, the spatial slot dimensions are the longitudinal and transversal dimensions, corresponding with slot length and slot bandwidth respectively. In principle, the spatial size of the bottleneck under consideration will determine the spatial size of the corresponding slots. However, in some situations spatial slot size is subject to discussion. For instance, assume that selection slot allocation is applied at London Heathrow airport. The question is whether each selection slot should include all four Heathrow terminals. Currently, all domestic flights depart from terminal 1, which enables a different treatment of domestic and international passengers with respect to check-in and customs procedures. Hence, it is reasonable to assume selection slots for domestic services to be restricted to terminal 1. On the other hand, selection slot allocation might be applied to the entire London metropolitan airport system, considering this as a single bottleneck, which would imply that in some cases selection slots may include different airport alternatives such as London Gatwick and London Stansted. Consequently, selection slots considering the same bottleneck may have a different bandwidth or slot length.

A reasonable restriction of temporal slot size is to exclude periods that are not included in the set of potential scheduling or *ad hoc* slots. The size of this period may differ

between slot requests. In principle, slot size may be determined demand-driven, i.e. by including it in the slot requests, but in this case slot size restrictions may be demanded that are not necessary from the point of view of the allocation body. For instance, a carrier might desire a relatively restricted selection slot in order to be ensured of the desired timing of the corresponding scheduling slots. Therefore, we assume supply-driven slot specification, i.e. the allocation body determines slot size.

Discrete, overlapping, or multi-level slot specification

Discrete, overlapping, and multi-level slot specification are three alternative approaches that may be implemented at the selection level, but also at other levels of slot allocation. We first discuss the two extremes. In the discrete case, each slot request and allocated slot should correspond with a *basic slot*. Basic slots are disjoint, i.e. they have no overlap. The time-space domain of each slot should correspond with one of these basic slots. On the other hand, overlapping slot specification means that slots may correspond with any time-space domain within the limits of the bottleneck under consideration. The difference between discrete and overlapping slot specification is illustrated by figure 4.3. In this figure, slot periods are indicated by two-sided arrows. Although figure 4.3 only shows the temporal dimension of slots, the same principle applies to the transversal spatial dimension of slots.

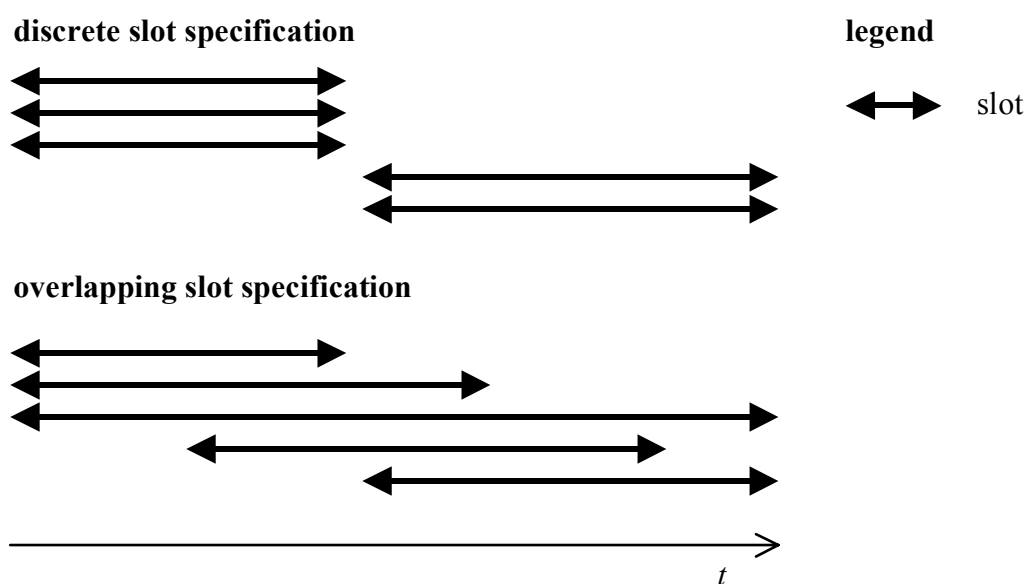


Figure 4.3: Slot periods given discrete and overlapping slot specification

Allowing overlapping slot specification has the advantage that every selection slot can be tailored to suit the slot request it is allocated to. On the other hand, the formulation of capacity constraints is easier given discrete slot specification. Capacity constraints state that the total traffic volume within a period should remain within a certain limit. Given discrete slot specification, this can easily be implemented by letting these periods correspond with the slot periods. However, when slots can have any size, these periods will

not correspond, which means that additional assumptions are required for the specification of capacity constraints.

A solution that allows for diversity in slot size, while maintaining a relatively simple problem specification, is multi-level slot specification. Given multi-level slot specification, selection slots may consist of one or several pre-defined *primary basic slots*. Primary basic slots are the smallest possible units of slots. Slots specified at the same hierarchy level are never overlapping, enabling a complete and consistent specification of capacity constraints at each hierarchy level. Each set of identical slots at a secondary or higher hierarchy level corresponds with a separate *secondary* or *higher-order basic slot*. At each higher level in the hierarchy, basic slots at this level consist of more primary basic slots than basic slots at any lower level. Multi-level slot specification is illustrated by figure 4.4. As in figure 4.3, only the temporal dimension is shown, and slot periods are indicated by two-sided arrows.

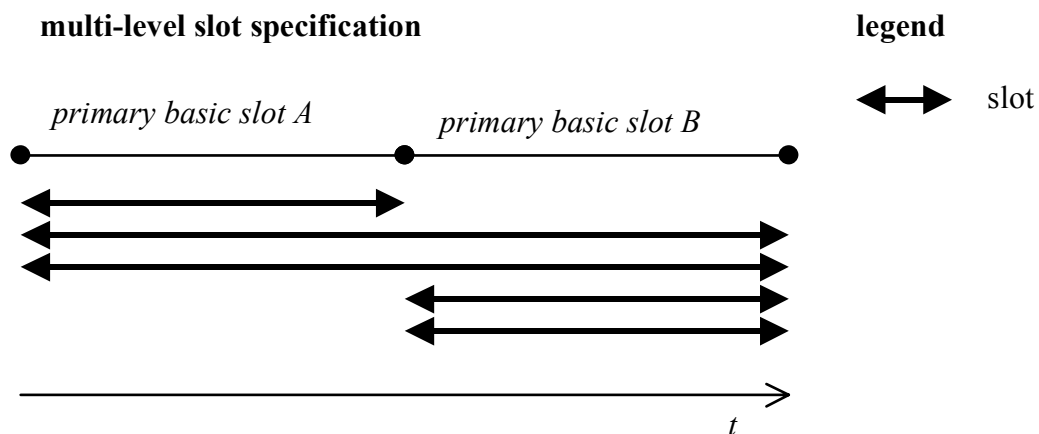


Figure 4.4: Slot periods given multi-level slot specification

Multi-level slot specification results in a straightforward specification of the selection slot allocation decision problem (see section 7.3), contrary to overlapping slot specification. Separate capacity constraints should be formulated for each different level of basic slots. Capacity constraints at higher levels should not only account for slot requests at its own level, but also for slot requests corresponding with lower-level basic slots encompassed by the higher-level basic slot under consideration.

The multi-level principle may apply to both temporal and spatial slot dimensions, as long as it applies to both dimensions simultaneously, i.e. if two slots overlap in time and space, one of them should be entirely encompassed by the other. Figure 4.5 illustrates multi-level slot specification in a situation where higher-level slots have both a larger slot period and a larger bandwidth than primary slots. Assume that selection slot allocation is applied to a station or (air)port consisting of several terminals. Slot 1 corresponds with the arrival of a CityExpress transport service, for which terminal A is the only acceptable alternative because it connects there with other transport services. To ensure

these connections, the slot period is restricted to the arrival bank. Slot 2 corresponds with the arrival of a SunExpress transport service, which does not rely on specific connections. Hence, the second slot has not only a larger slot period, but also a larger bandwidth, since it also includes the alternative terminal B.

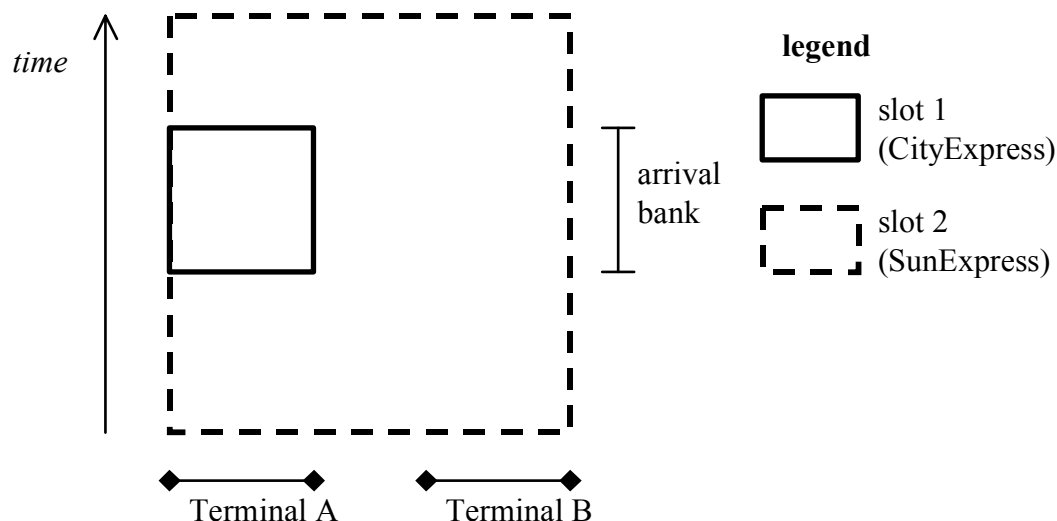


Figure 4.5: Multi-level specification of terminal slots with and without hub connections

To enable a straightforward specification of capacity constraints, this thesis assumes discrete or multi-level slot specification. The next sub-section elaborates on the specification of selection slots. We will see that multi-level slot specification will usually be preferred to discrete slot specification, because in most situations there will be a clear demand for slots of different hierarchy levels.

4.3.2 Acceptability criteria

Selection slot allocation focuses on the selection of slot requests, not on the scheduling of the corresponding vehicle paths. Therefore, the main principle followed here is that the size of selection slots should not be smaller than necessary in order to maintain as much flexibility as possible for scheduling and *ad hoc* slot allocation decisions. However, it is also clear that selection slots should have a limited size. As was argued in the previous sub-section, slot specification is the responsibility of the allocation body.

Acceptability

The criterion used in this thesis to specify selection slots and minimum quality-of-service criteria is acceptability. *Acceptability* means that selection slots should be specified in such a way that carriers can be offered acceptable scheduling slots or *ad hoc* slots, which again implies that acceptable vehicle paths will be the result. Acceptability is defined by the acceptance of alternative vehicle paths by the carrier, as opposed to the alternative that desired trips are cancelled. Acceptability may be influenced by the

criteria imposed by authorities issuing concessions, or by end-users consuming transport services. Therefore, a concession to operate intercity trains, for instance, may include criteria regarding operational speed.

In section 4.1 it was stated that selection slots are mainly required to ensure the stability of transport service patterns, either with or without associated contractual obligations. The question is which possible timetables are acceptable realizations of a transport service pattern. Usually, transport services have a restricted range of departure and arrival times, travel times, and terminal locations in which it can be run profitably. Vehicle paths within this range are accepted by carriers as alternatives to their preferred vehicle path, albeit with a lower quality-of-service. However, vehicle paths outside this range will not be accepted by the carriers. If no acceptable alternative is available, the transport service corresponding with the vehicle path under consideration will not be performed. It is reasonable to require that selection slots are specified in such a way that carriers are ensured of vehicle paths with an acceptable quality-of-service. This includes the restriction of slot sizes to exclude unacceptable periods and locations from selection slots.

This section reviews a few acceptability criteria that are likely determinants of slot size. The assumption of discrete or multi-level slot specification may imply that acceptability may not be ensured for every individual slot request, but nonetheless it may be used as a guideline to specify slots. The following acceptability criteria are discussed in this section:

- frequencies and demand peaks;
- connections between transport services;
- terminal facilities.

Frequencies and demand peaks

Firstly, carriers offering transport services with a certain frequency often desire the departure times to be separated with regular intervals. This minimizes average waiting times and enables efficient utilization of vehicles and staff. For instance, a transport service with a frequency of 4 runs per hour is generally preferred to run in equal intervals of 15 minutes rather than with irregular intervals, for instance, every 5, 15, 25 and 30 minutes past each hour. A reasonable solution would be to specify selection slots of identical size, corresponding with the average intervals between runs. Additionally, a sufficient level of scheduling freedom is required to enable scheduling of runs with equal intervals. Secondly, for transport services without regular intervals it is generally important that the departure times are favorable given the demand pattern. In this case, it is reasonable to let selection slots correspond with the peak demand periods.

Because infrastructure elements are often used by transport services with different frequencies, it is reasonable to expect that different lengths of slot periods will be the

result. Multi-level slot specification is relatively easy when intervals are also chosen from a set of multitudes, for instance 15, 30, 60, and 120 minutes. If not, some adaptations may be required to facilitate a multi-level slot specification. Only the specification of slot periods is affected by the issue of frequencies and demand peaks.

Connections between transport services

Restrictions of departure and arrival times may also be demanded to facilitate connections with other services. Offering opportunities to change between transport services is an important function of many larger railway stations and airports ('hubs'). A hub is the central node in a transport network of a carrier or an alliance of carriers, allowing short connections between flights by synchronizing arrivals and departures of flights. When transport services have relatively low frequencies they usually arrive and depart in 'banks', i.e. concentrated periods of arrival and departure. Between the arrival bank and the departure bank, passengers and freight can be exchanged between transport services, resulting in relatively short changeover times. An additional requirement for short changeover times is that all services use a single terminal.

The changeover function may not be important to all transport services using a hub. For instance, the changeover function is said to be of great importance to the transport network of KLM, which has Schiphol as its main hub. Therefore, the KLM timetable offers good connections between long-distance services and short-distance feeder services. In contrast, the timetables of its subsidiary Transavia are largely determined by the optimization of fleet and staff usage. Transavia Airlines is not interested in providing connections with other air transport services, because it concentrates on leisure travel between the Netherlands and the Mediterranean.

Terminal facilities

Another possible reason for spatial slot size restrictions is the availability of terminal facilities. Carriers often locate certain vehicle and passenger facilities on or near terminals. For instance, air carriers concentrate cleaning and maintenance at their home-base airports, and it is not desirable that carriers run the risk of frequent shifts of their home bases. Furthermore, carriers may have passenger facilities concentrated in a specific terminal of a large airport. These facilities may change location over time, but shifting locations is costly and therefore frequent changes are not desired.

Synopsis

In this section we have seen that slot size restrictions may be related to various acceptability criteria. Slot periods may be related to service frequency, demand peaks, and the connections between transport services. Slot bandwidth, in particular with respect to terminal choice, may be restricted by the desired connections between transport services, and the availability of terminal facilities. Given differences in frequency and with

respect to the importance of connections, a multi-level slot specification will often be required instead of a discrete slot specification. Additionally, the acceptability principle may be used to specify minimum quality-of-service criteria (see section 6.4). Given the relationship between capacity and quality-of-service briefly analyzed in section 2.2, this has implications for the capacity at the selection slot allocation level. This chapter ends with an overview of the main conclusions and assumptions about the selection slot allocation problem.

4.4 Conclusions

In this chapter, the desired characteristics of selection slot allocation systems have been specified from a traffic demand perspective. Issues that have been discussed in this chapter are slot validity, slot holdership, slot allocation dynamics, and the specification of selection slots.

Demand for selection slots

The demand for selection slots is largely due to the demand for stability of transport service patterns. One of the factors explaining the demand for stability of transport service patterns is that carriers invest in transport services. These investments include design costs, marketing costs, and investments in the vehicle fleet. Furthermore, the stability of transport service patterns may be enhanced by the requirements of long-term concessions concluded between authorities and carriers, or of long-term private transport contracts concluded between carriers and end-users. Given this demand for stability, it has been concluded in this chapter that, in situations where investments or concessions are an issue, selection slots should generally have a validity of at least 5 years. The application of infinite validity (historic rights) is also possible, but has the disadvantage of low flexibility and impairing the introduction of new transport services. When requesting selection slots is related to investments in transport services by carriers, it is reasonable that slots are requested and held by carriers. However, when selection slots are requested to facilitate concessions or transport contracts, the best solution is probably to let selection slots be requested and held by the other party, i.e. authorities, shippers, travel agents, etc. In these cases, the transport service patterns are largely determined by the latter actors and these are primarily interested in the stability of transport service patterns.

A function of selection slot allocation is to facilitate the stability of specific transport services. However, the stability argument for long-term selection slot validity is not valid anymore when slot holders desire to use their selection slots for different transport services. Furthermore, allocation decisions may be based on the valuation of the assumed usage of selection slots. Therefore, this thesis assumes that selection slots may only be used for the transport services specified in the allocation decisions. However, apart from a possible competition check it is reasonable to allow selection slots to be

sold or handed over to other carriers, as long as they remain to be used for a similar purpose.

Slot allocation dynamics

In this chapter, three types of slot allocation dynamics have been introduced, i.e. static, dynamic, and semi-static slot allocation. Static slot allocation allows for simultaneous evaluation of all slot requests regarding a certain period, while dynamic slot allocation allows for direct evaluation of slot requests at any moment within a period. Semi-static decision-making is preferred for selection slot allocation for supply-driven transport services. Given semi-static decision-making, selection slot requests regarding new transport services may be submitted every timetable season before the timetables are established. This enables simultaneous evaluation of slot requests, but earlier slot allocation decisions should be respected as long as these are still valid. Additionally, dynamic decision-making is preferred to allow selection slot requests for demand-driven transport services throughout the timetable season.

Specification of selection slots

The specification of selection slots implies the specification of temporal and spatial slot size. Given the focus of selection slot allocation on the question of which vehicle paths should be accommodated, and not on the timing of vehicle paths, the main principle followed here is that selection slots should be large in order to maintain as much flexibility as possible for scheduling and *ad hoc* slot allocation decisions. In this thesis, the acceptability criterion is the leading principle to specify slots. The acceptability criterion is the acceptance of alternative vehicle paths by the carrier, as apposed to the alternative to cancel the desired vehicle path under consideration. However, additional slot size restrictions on slot specification are preferred in this thesis to simplify the selection slot allocation problem. Instead of allowing selection slots to overlap freely, discrete or multi-level slot specification is assumed. In both cases, basic slot periods and bandwidths are specified. Every slot is either identical to a basic slot or encompasses a number of basic slots.

Next chapters

In this chapter, the selection slot allocation problem has been characterized. The next chapter continues to specify the essence of the selection slot allocation problem by specifying and analyzing the main characteristics of traffic markets in situations requiring selection slot allocation. Chapter 6 focuses on traffic supply, and elaborates on the specification of capacity constraints, in relation to minimum quality-of-service criteria. Finally, the specification of the selection slot allocation problem in chapter 7 is based on the characterization of the selection slot allocation problem made in this chapter. The selection slot allocation decision problem assumes a semi-static allocation procedure, and discrete or multi-level slot specification.

5

SPECIFICATION OF TRAFFIC MARKETS

In the previous chapter, the desired characteristics of selection slot allocation have been reviewed, based on the demand for selection slot allocation. This chapter continues with providing a generic overview of the traffic market associated with selection slot allocation. This chapter first reviews current congestion theory, which is a branch of economic theory that is used to describe and analyze traffic bottlenecks. The second section contributes to congestion theory by analyzing the main characteristics of traffic selection markets, i.e. traffic markets in situations where selection slot allocation is applied. Main issues are the specification of periods and the shapes of traffic supply and demand curves. Furthermore, two types of capacity scarcity are introduced: relative and absolute capacity scarcity. In case of relative capacity scarcity an equilibrium between traffic supply and demand is reached, resulting in a traffic volume lower than capacity, while in case of absolute capacity scarcity traffic demand exceeds traffic supply. It appears that selection slot allocation is only desirable in situations of absolute capacity scarcity. In section 5.3, the application of infrastructure tolls is reviewed. Congestion tolls can be applied to reduce the difference between user equilibrium and system optimum in situations of relative capacity scarcity. Furthermore, the possible application of scarcity tolls to bridge the gap between traffic supply and demand in situations of absolute capacity scarcity is analyzed in this section, but it is concluded that optimal scarcity tolls will often leave a remaining gap between traffic supply and demand, due to uncertainty about traffic demand. The economic analysis presented in this chapter assumes a single bottleneck with a single capacity constraint and homogeneous traffic characteristics. An elaborate discussion about traffic supply and the formulation of capacity constraints, relaxing the homogeneity assumption, is left to the next chapter. A complete formulation of the selection slot allocation problem, including multiple capacity constraints and heterogeneous traffic, will be given in chapter 7.

5.1 Congestion theory

Selection slot allocation is applied in situations of absolute capacity scarcity, i.e. in situations where, in absence of additional measures to reduce demand (or to increase supply), traffic demand exceeds traffic supply. In this chapter, economic theory is applied to describe and analyze the balance and imbalance of traffic supply and demand. This section reviews congestion theory, which is the branch of economic theory regarding traffic bottlenecks. Main topics are the specification of traffic supply and demand curves and the difference between system optimum and user equilibrium.

5.1.1 Traffic supply and demand curves

The specification of supply and demand curves is one of the basic tools of economists to describe and analyze markets, including traffic markets. Actually, the first known application of a demand curve was to analyze the consumer value of infrastructure. Dupuit (1844) introduced a demand curve to represent the value to consumers of infrastructure facilities such as roads and canals. He assumed absence of congestion, which means that the supply curve was trivial. Figure 5.1 illustrates the traffic supply and demand curves as assumed by Dupuit. The horizontal axis indicates the quantity of traffic supply and demand q . The generalized costs c can be regarded as an inverse quality-of-service measure.

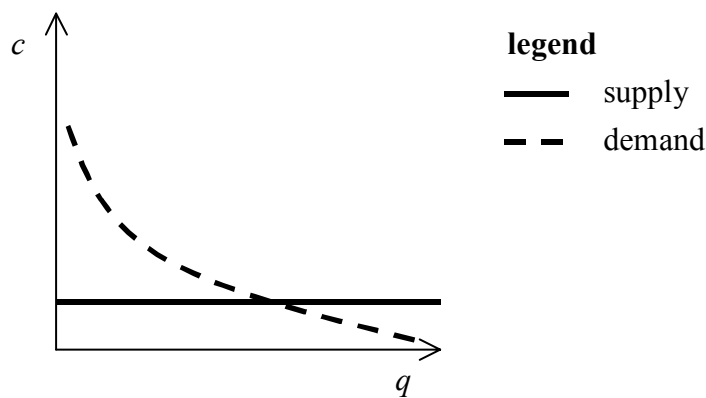


Figure 5.1: Supply and demand curves as assumed by Dupuit (1844)

Later, non-trivial supply curves were introduced, enabling the inclusion of congestion in the analysis; one of the first to apply this to transportation problems was Pigou (1920). An elaborate analysis of highway congestion based on analysis of traffic supply and demand curves has been performed by Walters (1961). His work has been used as a basis for later research on congestion economics, but has been the source of a lengthy debate. Walters used a two-valued supply curve, a direct translation of the speed-flow model of Greenberg (1959). This supply curve has a 'normal' and a 'hypercongested' branch (see figure 5.2). The 'normal' branch describes *flow congestion*, i.e. the level of congestion that occurs in situations where traffic volumes are still lower than the capac-

ity κ ; while the 'hypercongested' branch describes congestion as experienced in traffic queues. Congestion delays are included in the generalized costs c .

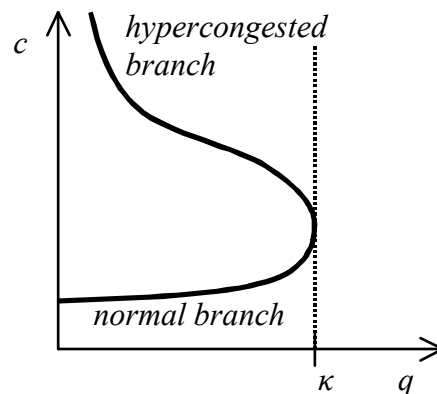


Figure 5.2: (Erroneous) two-valued traffic supply curve including hypercongestion

However, various authors have shown that a backward bending supply curve should not be applied to the analysis of traffic markets (see e.g. May *et al.*, 2000). The 'normal' branch applies to the bottleneck under consideration, but the 'hypercongested' branch applies to the queuing process upstream of a bottleneck, not to the bottleneck itself (Hall & Montgomery, 1993). Based on a review of the existence of 'hypercongestion', May *et al.* (2000) concluded that critical aspects for a correct specification of traffic supply curves are distinguishing demanded traffic flows from performed traffic flows, and correctly interpreting the dynamics of traffic processes.

Economic analysis of supply and demand is based on the assumption of homogeneous markets, which implies that traffic markets should be specified in such a way that all traffic experiences the same costs, including delay costs. The 'classic' solution to this problem is to assume static (steady state) traffic supply and demand (see e.g. Walters, 1961). However, the static model is not very suitable to explain infrastructure capacity scarcity, because traffic demand usually fluctuates considerably. Congestion is typically a dynamic phenomenon, and consequently the level of (hidden) congestion costs due to longer travel times and departure time adjustments varies with time. Furthermore, static models can only handle *flow congestion*, i.e. traffic delays in situations where traffic volumes are still lower than capacity, not the heavy levels of congestion explaining recurrent queuing phenomena on freeways which are common in many urbanized areas throughout the world. Therefore, dynamic analysis of traffic markets is required to analyze infrastructure capacity scarcity and traffic congestion.

Dynamic analysis of traffic markets

To analyze the dynamics of congestion, a continuous-time deterministic queuing model has been introduced by Vickrey (1963, 1969), which has been the basis of a separate branch of transportation economic research (see Arnott *et al.*, 1998, for an overview). The deterministic queuing model used by Vickrey neglects flow congestion, i.e. no

congestion is experienced as long as the traffic flow rate is lower than capacity. Traffic flows in excess of capacity are assumed to be buffered in a queue upstream of the bottleneck (see figure 5.3). This phenomenon is referred to here as *capacitated congestion*. In a situation of capacitated congestion, traffic volumes passing the bottleneck under consideration are equal to its capacity.



legend

- no congestion
- bottleneck with capacitated traffic
- congestion (queuing) upstream of bottleneck

Figure 5.3: Example of bottleneck with capacitated congestion

The deterministic queuing model may be useful for our purposes, but the continuous-time analysis of congestion as proposed by Vickrey and Arnott *et al.* is not suitable to specify homogeneous traffic markets. As an alternative, Henderson (1974) introduced a discrete-time flow congestion model. In his model, queuing is not included, it only considers flow congestion depending on the traffic volume entering the bottleneck in each period, which implies that this model neglects that infrastructure elements have finite capacity. A complete analysis is only possible if both capacity constraints and flow congestion are included. The discrete-time congestion model has been extended by Lerz (1996) to include both flow congestion and capacitated congestion. According to Lerz, both a speed-flow relationship representing flow congestion and infrastructure capacity constraints should be included in his optimal railway capacity allocation model.

5.1.2 System optimum and user equilibrium

Having specified traffic supply and demand curves, these may be used to evaluate to which extent traffic supply and demand equilibrate and to evaluate the differences between this user equilibrium and a system optimum. The difference between a user equilibrium and a system optimum has been introduced by Wardrop (1952). A user equilibrium is attained when all traffic network users optimize their own choices. The user equilibrium corresponds with the intersection of the marginal private cost curve (traffic supply) with the traffic demand curve. Hence, a user equilibrium can only be attained in situations where traffic supply and demand curves intersect, which is the case in figure 5.4. The arguments in favor of the chosen shape of the supply and demand curves are given in the next section. In figure 5.4, the equilibrium traffic volume q^e is lower than the capacity κ . In this equilibrium, infrastructure users experience cost c^e .

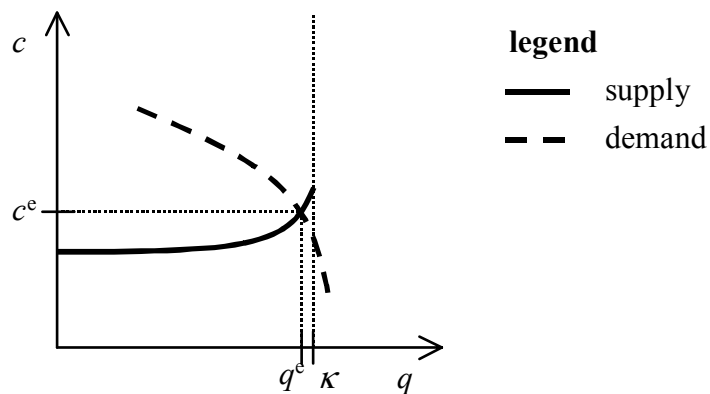


Figure 5.4: User equilibrium

The system optimum is the situation that is optimal from a collective point of view, given the objective that the sum of generalized costs of all users is minimized. According to economic theory, the system optimum corresponds with the intersection of the demand curve with the marginal social cost curve (see figure 5.5). The marginal private costs are the costs experienced by the last single user that enters the system. Additional to these costs, the marginal social costs include the costs incurred by other users due to the entrance of this additional user, which are the external congestion costs. Other external costs may be included in the social cost curve, for instance pollution costs, but these are neglected here.

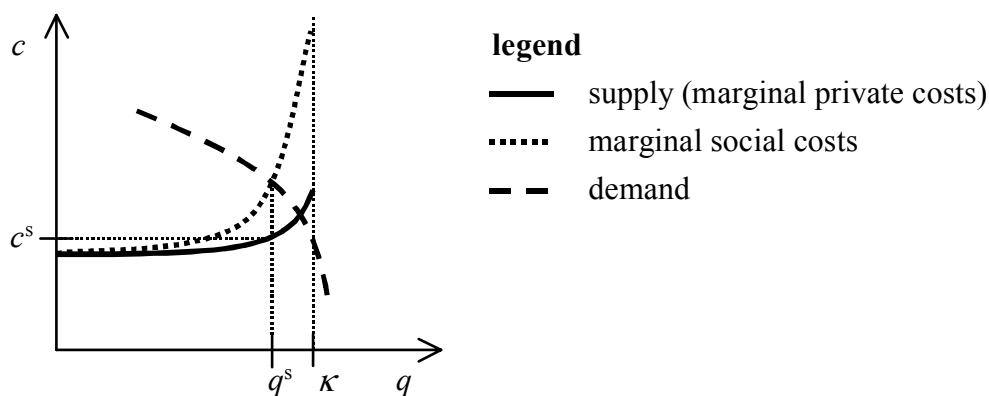


Figure 5.5: System optimum

According to Beckman *et al.* (1956), there is often a significant difference between the user equilibrium and the system optimum. For instance, in figure 5.5, the system optimum traffic volume q^s is clearly lower than the user equilibrium traffic volume q^e , and the system optimum generalized costs c^s are clearly lower than the user equilibrium generalized costs c^e . In situations where the marginal social costs of traffic are higher than the marginal private costs, the optimum traffic volume is lower than the equilibrium traffic volume. The marginal social costs are higher than the marginal private costs in situations with congestion, because every additional vehicle does not only experience congestion delays but also inflicts congestion delays on other traffic. Consequently, the equilibrium traffic volume is too high if congestion costs are not internalized into the

marginal private costs (Beckmann *et al.*, 1956). The latter may be achieved by imposing infrastructure usage fees that depend on expected congestion levels. Section 5.3 elaborates on the application of infrastructure tolls to optimize the balance between traffic supply and demand.

5.2 Main characteristics of traffic markets at the selection level

Having described the state-of-the-art of congestion theory, we now discuss how congestion theory may be used to describe and analyze traffic markets at the selection level. The first step of this section is to specify relatively homogeneous traffic selection markets. The main idea is that each basic slot corresponds with a different traffic selection market. The next step is to specify the shape of traffic supply and demand curves specific for traffic selection markets. Finally, two types of capacity scarcity are discussed, based on the traffic selection market characteristics specified earlier in this section. Relative capacity scarcity is a situation where a user equilibrium is attained, implying a traffic volume lower than capacity. Absolute capacity scarcity is a situation without equilibrium, implying an imbalance between traffic supply and demand. Since such an imbalance is a key characteristic of the selection slot allocation problem, we may assume that selection slot allocation is applied to bottlenecks experiencing absolute capacity scarcity.

5.2.1 Specification of traffic selection markets

The traffic selection markets analyzed in this chapter have a number of characteristics that enable the specification of homogeneous markets. A traffic selection market considers a single bottleneck with a fixed capacity. Capacity is assumed to be independent of composition-of-traffic. Selection slot allocation is applied to this bottleneck. Slots are specified by the allocation body. Either discrete or multi-level slot specification is applied (recall section 4.3). Assuming discrete slot specification, each traffic selection market considers a single basic slot. Additionally, a separate traffic market may be specified for each higher-level basic slot in case multi-level slot specification is applied. In a situation with multi-level slot specification, however, slot markets can only be analyzed assuming known results of other slot markets considering the same period (at lower or higher levels in the hierarchy). Given multi-level slot specification, slot requests of lower levels are included in higher levels, but not the other way around.

Discrete-time dynamic analysis of traffic markets is possible if periods can be identified within which traffic demand may be assumed to be homogeneous. At the selection slot allocation level, this means that within the slot period of a basic slot, there are no differences in expected departure time adjustments and expected delays. Differences in planned or realized departure times and delays are not relevant at this planning level, since selection slot allocation is traffic planning with a low level of precision with respect to the timing of traffic. Another important assumption is that slot requests in the

traffic market under consideration have no acceptable alternatives outside the period and bandwidth of the basic slot or type of higher-level slot under consideration. When a slot request is not selected, no alternative slot requests will be made, and hence the corresponding planned transport service will have to be cancelled. This means that traffic selection markets should be specified according to the acceptability principle, as proposed in section 4.3.

As defined earlier, selection slot allocation is the selection of slot requests with the purpose of determining which vehicle paths can be accommodated given capacity constraints and acceptability criteria (section 2.4). Consequently, selection slot allocation is applied to situations where the traffic volume corresponding with the set of selection slot requests exceeds the capacity of the bottleneck within the slot period under consideration. This means that traffic selection markets are characterized by a situation where only a part of the set of slot requests can be selected, given that an acceptable quality-of-service has to be realized.

5.2.2 Shapes of traffic supply and demand curves

Given what we have learned about the specification of traffic selection markets, the shapes of traffic supply and demand curves may be derived. Figure 5.6 exemplifies the shapes of supply and demand curves applying to traffic selection markets. As indicated earlier, c denotes generalized costs, q denotes traffic volume, and κ denotes capacity.

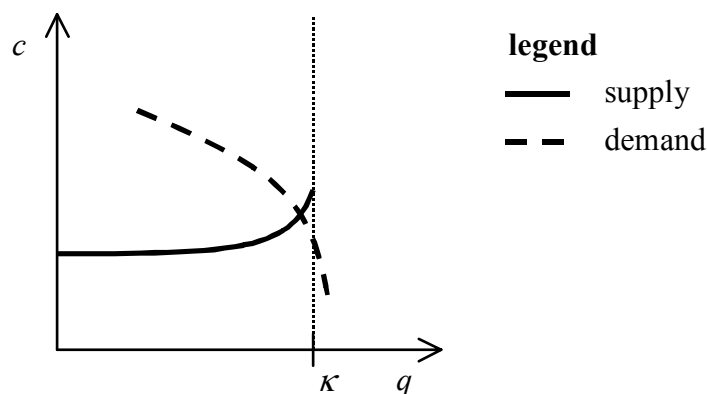


Figure 5.6: Traffic supply and demand curves

Traffic supply curve

The supply curve specifies the quality of traffic supply for traffic volumes between zero traffic flow and capacity κ . By definition, traffic volumes higher than capacity are not possible. Although quality-of-service criteria or external restrictions may restrict capacity (recall section 2.2) and consequently higher traffic volumes may be feasible technically, traffic volumes in excess of the stated available capacity are not allowed. Traffic costs will usually increase as a function of traffic flow for two reasons. In the first place, capacity can often only be reached with a lower speed than the desired or maximum

level. Furthermore, the probability of random delays within the period under consideration increases with increasing traffic volumes. Similarly, the freedom to schedule traffic as desired and the average departure time adjustments that are required in congestion-free timetables will increase with increasing traffic volumes.

The monotonically increasing shape of the traffic supply curve can be motivated by a number of factors. In the first place, we already noted in sub-section 2.2.3 that maximum flow is usually a function of speed with a single maximum. If this optimum speed is lower than the maximum or desired speed, then it can be expected that the maximum speed is a decreasing function of traffic flow. Furthermore, scheduling freedom decreases with increasing traffic volumes, and according to queuing theory, expected queuing delays increase monotonically as a function of traffic volume. In each case, the marginal increase in generalized costs increases with increasing traffic volumes. The experienced delays resulting from the factors discussed here will be limited by a certain factor of the total travel time through the bottleneck, resulting in a traffic supply curve reaching a certain maximum cost when traffic volume reaches capacity, rather than approaching infinity.

Traffic demand curve

Traffic demand is defined as the total traffic demand of a certain bottleneck during a certain period, given the generalized costs in that period. The approach followed here is to derive traffic demand curves from an inverted logistic choice curve. This approach has been suggested by Van Nes (2002) to analyze public transport demand. We consider three alternative curve shapes: linear, concave, and convex, corresponding with different parts of an inverted logistic curve. A concave demand curve corresponds with a situation where with decreasing costs a situation of near saturation is attained, i.e. the effect of decreasing costs on demand decreases with decreasing costs. In contrast, a convex demand curve corresponds with a situation where costs are an important factor reducing demand. In this case, the effect of decreasing costs on demand increases with decreasing costs. Finally, a linear curve may be used to approximate an inverted logistic curve, especially in situations in between the two extreme situations described above.

The choice for a concave demand curve in figure 5.6 is motivated by the assumption that within its usual range, generalized costs will be low enough to deter only a relatively small percentage of the potential traffic demand. However, linear or convex demand curves may also be applied. Which shape is chosen is of minor importance to the analysis of traffic selection market introduced here.

5.2.3 Levels of infrastructure capacity scarcity

In the previous section, the specification of traffic supply and demand curves has been discussed. Given the proposed dynamic analysis of traffic markets, levels of infrastruc-

ture capacity scarcity can be specified. These levels of capacity scarcity will be used in chapter 6 to classify current bottlenecks in different infrastructure sectors. Two levels of infrastructure capacity scarcity are distinguished in this thesis, i.e. relative capacity scarcity and absolute capacity scarcity.

Relative capacity scarcity

Relative capacity scarcity is a situation where all traffic demand can be accommodated, but not with the quality-of-service desired by the carriers. For instance, departure times may be different than desired, or longer travel times may have to be accepted. All traffic requests can be accommodated, but it will be shown in the next section that traffic volume restrictions may still be advisable if the negative effect of additional traffic on the quality-of-service of facilitated traffic outweighs the negative effects of not accommodating this transport service.

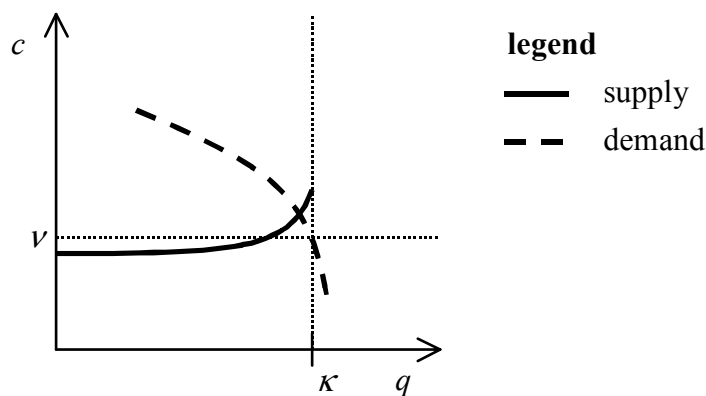


Figure 5.7: Relative capacity scarcity

Figure 5.7 illustrates a situation of relative capacity scarcity. Traffic supply and demand are in equilibrium. The resulting traffic volume q_e is lower than the available capacity κ , which means that all traffic can be facilitated by the bottleneck in the period under study. To distinguish flow congestion from no congestion, an arbitrary level of generalized costs v has to be chosen to distinguish no (or negligible) capacity scarcity from relative capacity scarcity. A reasonable criterion is that the delays and departure time adjustments should exceed the minimum precision of traffic planning, for instance the precision of allocation timetables. This implies that scheduling slot allocation is useful in situations of relative capacity scarcity, but not in situations without capacity scarcity.

Absolute capacity scarcity

Absolute capacity scarcity occurs when a bottleneck cannot accommodate all traffic demand with an acceptable quality-of-service within the period under consideration. The result of this level of capacity scarcity is that carriers cannot offer all transport services with the frequencies they desire. Furthermore, private transport may have to be bundled (e.g. carpooling instead of driver-only car traffic) or replaced by public trans-

port services. For instance, all road links between two city quarters separated by a river are congested for many hours every day. Congestion is so severe that many drivers regard the quality-of-service of these links as unacceptable. If this remains the case, they will choose alternatives such as rail transport or alternative destinations. However, it is also possible to reduce the traffic volumes by stimulating carpooling and a modal shift to bus transport. Another example is an airport with absolute capacity scarcity. If its capacity would be increased, possible results could be more direct connections, higher frequencies, and more direct competition between transport services. Because the increase in transport demand would be less than proportional, it is probable that smaller planes would be used.

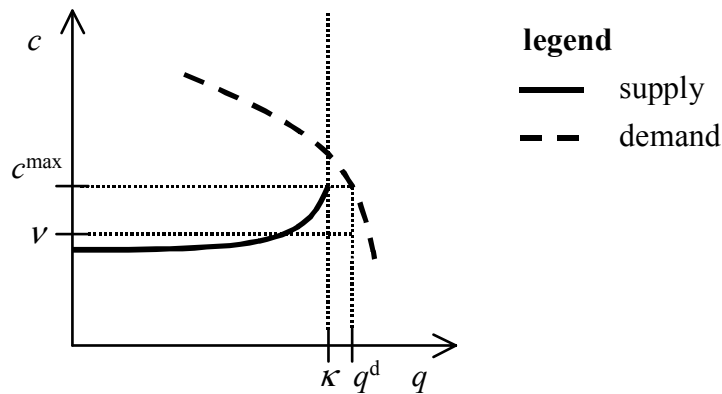


Figure 5.8: Absolute capacity scarcity

The lack of equilibrium between traffic supply and demand described above is typical for absolute capacity scarcity. In figure 5.8, capacity is fully utilized, but the supply and demand curves do not intersect, which means that there is no equilibrium between traffic supply and demand. Given the maximum level of generalized costs c^{\max} , a traffic demand volume q^d remains that is larger than capacity κ . What happens in such a situation depends on the level of traffic control. Selection slot allocation solves this problem by making a selection of traffic demand, reducing the traffic demand volume accommodated by the bottleneck in the period under consideration from q^d to κ .

Capacity scarcity and the selection slot allocation problem

The selection slot allocation problem is characterized by a surplus of traffic demand, and therefore the selection slot allocation problem as assumed in this thesis applies to situations of absolute capacity scarcity. However, situations of relative capacity scarcity may also be transformed 'artificially' to situations of absolute capacity scarcity with the objective of optimizing the balance between traffic supply and demand. In situations with relative capacity scarcity and a significant difference between system optimum and user equilibrium traffic volumes, allocation bodies may decide to lower capacity values to a level corresponding with the system optimum traffic volume. As a consequence, the level of capacity scarcity increases from relative to absolute, and hence the selection slot

allocation problem applies in this case. Figure 5.9 illustrates this artificial capacity value κ' , which is lower than the equilibrium traffic demand.

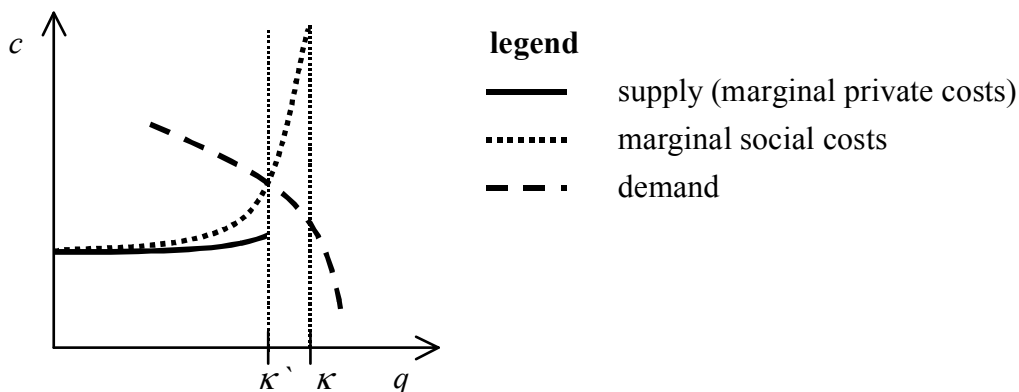


Figure 5.9: Artificial lower capacity

5.3 Infrastructure tolls

In section 5.1, the difference between system optimum and user equilibrium was introduced. It appeared that system optimum traffic volumes at infrastructure bottlenecks are often lower than user equilibrium traffic volumes. In this situation, economic theory suggests the application of congestion tolls, as is argued in sub-section 5.3.1. However, pricing mechanisms cannot only be used to optimize the location of the user equilibrium, but also to attain a user equilibrium in situations of absolute capacity scarcity. In the latter situation, economic theory proposes the application of scarcity tolls, as is discussed in sub-section 5.3.2.

5.3.1 Congestion tolls

The selection slot allocation problem corresponds with situations where traffic demand exceeds traffic supply. In this section, however, the emphasis is on situations without an excess of traffic demand. In these situations, it may be still be desirable to restrict traffic volumes in order to optimize the balance between traffic quality and traffic quantity.

According to congestion theory, shifting the equilibrium traffic volume to the optimum value should be achieved by imposing congestion tolls. Among the first to advocate the imposition of congestion tolls for congested roads and freeways in the literature, based on different theoretical models of road traffic economics, were Walters (1961) and Vickrey (1963). The application of congestion tolls to airports has been studied, amongst others, by Carlin & Park (1970). A method to estimate optimal congestion tolls for railway traffic has been introduced by Harker & Hong (1994). The application of congestion theory to railways in order to derive optimal congestion tolls has been discussed at length in the Ph.D. thesis of Lerz (1996).

Figure 5.10 shows how the congestion toll should be derived according to standard congestion theory. The basic idea is that the equilibrium between supply and demand is shifted towards the optimum traffic volume q^0 by imposing a congestion fee that equals the difference between private and social marginal costs given the optimum traffic volume.

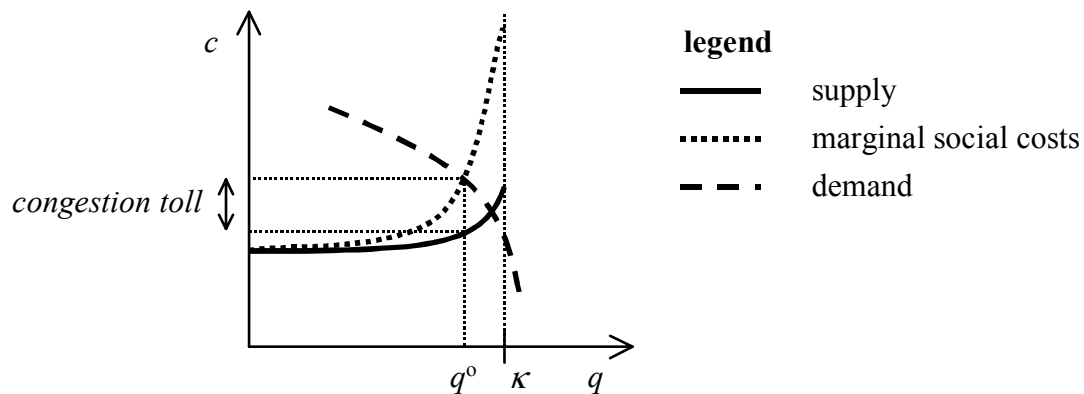


Figure 5.10: Derivation of optimum congestion toll given flow congestion

The application of congestion tolls has the characteristic that it results in a selection of slot requests associated with the highest willingness-to-pay, which is often considered more desirable than more or less random or 'administrative' selection mechanisms. Furthermore, the alternative of introducing an artificial lower capacity (see previous section) is more rigid than congestion tolls, and it can be expected that the given uncertainty about the exact shape of the demand curve, congestion tolls are often a better approximation from the optimum than artificial capacity constraints. However, there may be significant resistance of infrastructure users against the imposition of congestion tolls, especially if its income effects are not compensated. For example, the introduction of congestion tolls on frequently congested roads and freeways in the Netherlands has been blocked by significant political and societal resistance, resulting in the withdrawal of the bill that would have enabled its introduction (V&W, 2000).

5.3.2 Scarcity tolls

In situations with absolute capacity scarcity, traffic demand exceeds traffic supply. Economic theory suggests several market-based (or willingness-to-pay based) methods to bridge the gap between traffic supply and demand. One solution suggested by economists to bridge this gap is the application of a scarcity toll (see figure 5.11). In order to reduce traffic demand from q^d to κ , a scarcity toll has to be applied to shift the demand curve downward until it reaches the demand curve.

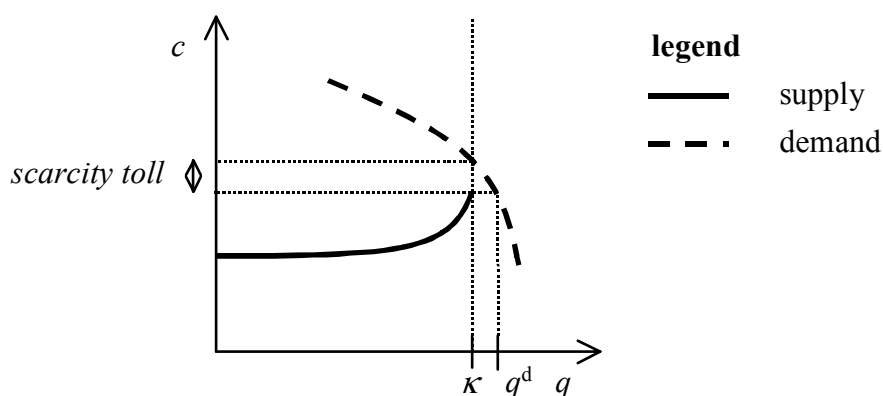


Figure 5.11: Derivation of optimum scarcity toll given capacitated congestion

One of the first to propose the application of a scarcity toll, although applied to transport capacity instead of infrastructure capacity, was Hotelling (1938). The application of this solution to traffic markets is advocated by Starkie (1998) for airports and by Lerz (1996) for railways experiencing absolute capacity scarcity. Hong & Harker (1992) describe how scarcity tolls (slot prices in their terminology) may be derived for airports in a network context with possibly several congested airports. The optimal scarcity tolls in the context of railway traffic have been analyzed by Lerz (1996). Lerz states that if the optimum congestion toll is higher than the scarcity toll (assuming absolute capacity scarcity), the congestion toll should be applied. A reduction of traffic demand to a level lower than the capacity would be the result.

The optimum scarcity toll can be derived given an estimation of the demand curve. However, overestimation of the scarcity toll required to bridge the gap between traffic supply and demand will result in underutilization of the infrastructure, while a conservatively estimated scarcity toll will result in a remaining gap between traffic supply and demand. Assuming that underutilization of the infrastructure is considered undesirable, estimated scarcity tolls should be estimated conservatively, and hence there will often be a remaining gap between traffic supply and demand. In this situation, still a selection slot allocation problem remains, and therefore the application of scarcity tolls is not considered here as a complete solution to selection slot allocation problems.

Auction techniques to measure scarcity tolls

As an alternative to estimating scarcity tolls, it is possible to measure the optimum scarcity toll by applying a second-price auction. In a second-price auction, the winner pays the highest non-winning bid, while in a first-price auction, the winner pays his own bid (see e.g. Krishna, 2002). Second-price auctions are proposed by Lerz (1996) and Nilsson (1999) to measure the required scarcity toll for railway slot allocation, because according to auction theory this principle stimulates bidders to reveal their true willingness-to-pay. However, it can also be shown that carriers may follow the strategy to bid higher than their willingness-to-pay for slots they hardly desire if they have the objective of decreasing the profits of their competitors.

Among the first to analyze the applicability of slot auctions to airports were Rassenti *et al.* (1982) and Brander *et al.* (1989). Main contributions on the application of slot auctions to railways have been made by Lerz (1996) and Brewer & Plott (1996). However, it should be noted that all these contributions consider integration of selection and scheduling slots. Relatively simple slot auctions can be created given a single bottleneck used by traffic with homogeneous characteristics. However, complexity is added when network interactions or heterogeneity of traffic is introduced. One of the objections of Hong & Harker (1992) against airport slot auctions is that airport slots come in pairs, i.e. landings and takeoffs and a slot at the airport of origin is useless if no corresponding slot can be obtained at the airport of destination. As pointed out by Nilsson (1999), this problem may be overcome by introducing a simultaneous auction of all network slots, while an optimization method is used to determine the optimum bid combination. A further complication is heterogeneity of traffic, which is mainly an issue for railway traffic. In this case, the key problem is to determine the optimum feasible set of bids in complex situations. It has been shown by Brännlund *et al.* (1996) that this is a relatively complex problem at the scheduling slot allocation level, in particular when this problem is solved for network problems. Because of this inherent complexity, they propose an approximation algorithm to find a nearly optimal solution in a relatively short time. Furthermore, it will be argued in chapter 7 that assuming linear capacity constraints, the selection slot allocation problem is equivalent to a binary linear programming problem. This problem belongs to the NP-hard class, for which only exponential-time exact solution algorithms are known (see Wolsey, 1998). A further complicating factor for the application of slot auctions is that carriers may have complex preferences with respect to alternative slots, which may be difficult to incorporate in the auction system. Hence, we may conclude that slot auctions are possible, also given complicating factors such as network interactions and interactions between user classes, but it is only a way to measure the value of slots; the problem remains how to derive the optimum selection slot allocation decision.

Discussion about the optimality of market-based selection slot allocation

Despite the theoretical advantages of scarcity tolls (and slot auctions) claimed by various economists, the superiority of these solutions to alternatives such as administrative rationing is disputed. Firstly, no references to empirical evidence supporting the alleged superiority of willingness-to-pay based slot allocation approaches have been found. Furthermore, the assumptions behind these theoretical considerations have been questioned in the literature. Borenstein (1988), for instance, discusses the assumption that a high willingness-to-pay correlates perfectly with a high value of the proposed service. Since infrastructure capacity can be used for services in different markets, he claims that there are no theoretical grounds to expect a high correlation. According to Borenstein, there is little justification for economists to conclude that the market approach is better than bureaucratic mechanisms without empirical evidence.

5.4 Conclusions

In this chapter, the selection slot allocation problem has been studied from an economic point of view. Based on the current state-of-the-art of congestion theory, this chapter has introduced a theoretical framework to analyze the traffic market associated with the selection slot allocation problem. The main contribution of this chapter is the extension of congestion theory with an approach to analyze traffic markets at the selection slot allocation level. This includes a generic approach to specify homogeneous traffic markets corresponding with basic slots and the specification of the shape of traffic supply and demand curves.

Specification of traffic selection markets

Congestion theory may be used to describe and analyze traffic selection markets, i.e. traffic markets at the selection level. A discrete-time dynamic approach is applied, i.e. separate traffic markets are specified for each period. More precisely, the traffic markets correspond with the basic slots. Additionally, for each type of higher-level slot a separate traffic market may be specified if multi-level slot specification is applied.

For each traffic selection market, traffic supply and demand are specified by supply and demand curves. Traffic supply can be described by a monotonically increasing traffic supply curve. This curve is specified on a restricted domain of possible traffic volume values ranging from 0 to capacity κ . The traffic demand curve may have various shapes, but a concave demand curve is suggested. A concave demand curve corresponds with a situation where with decreasing costs a situation of near saturation is attained, i.e. the effect of decreasing costs on demand decreases with decreasing costs.

In this section, two different levels of capacity scarcity have been defined: relative capacity scarcity and absolute capacity scarcity. The main difference between these levels of capacity scarcity is that equilibrium between traffic supply and demand can be attained in the case of relative capacity scarcity, while it is impossible to accommodate all traffic demand in the case of absolute capacity scarcity. Given that selection slot allocation is characterized by an excess of traffic demand, the assumed situation requiring selection slot allocation specified in this chapter corresponds with absolute capacity scarcity. However, situations with relative capacity scarcity may be translated into a situation with absolute capacity scarcity by imposing an artificial capacity constraint with the objective of optimizing the balance between traffic volume and quality-of-service. In this case, the selection slot allocation problem also applies.

Infrastructure tolls

Economic theory provides a powerful framework to analyze traffic markets. Two types of congestion problems are recognized in economic theory. Given relative capacity scarcity, traffic supply and demand are in equilibrium. However, this user equilibrium

may differ significantly from the desired system optimum. In order to attain this optimum, economic theory suggests the application of congestion tolls. Furthermore, given capacitated congestion or absolute capacity scarcity, no equilibrium is reached between traffic supply and demand, resulting in an excess of traffic demand. In this situation, economic theory suggests the application of scarcity tolls. However, it has been concluded in this chapter that to avoid underutilization of the infrastructure, estimated scarcity tolls should only be used to diminish the gap between traffic supply and demand. Alternatively, it may be possible to auction slots, but then the selection slot allocation problem is replaced by a similar problem concentrating on how to find the optimum feasible set of bids. Consequently, both in situations with and without slot auctions, a selection slot allocation problem may be formulated, i.e. the problem to optimally select slot request given capacity constraints.

Next chapters

The selection slot allocation problem will be specified in more details in the next chapters. Chapter 6 elaborates on the specification of traffic supply. The relative simplistic homogeneity assumptions facilitating simple analysis of traffic markets will be relaxed and replaced by more complex capacity constraints. Furthermore, the occurrence of different levels of capacity scarcity in current practice will be reviewed. Chapter 7 elaborates on the structure of traffic demand and the resulting valuation of alternative slot allocation decisions.

6

IDENTIFICATION OF TRAFFIC SUPPLY

In the previous chapter, traffic supply was simplistically represented by a fixed capacity value. However, as we have already seen in section 2.2, capacity may be dependent on composition-of-traffic and on the desired quality-of-service level. Therefore, this chapter introduces a generic framework to identify and model traffic supply in different situations. The first step is the identification of bottleneck types, including the classification of network elements and of types of traffic processes, which is the focus of section 6.1. Our next step is to specify capacity constraints by analyzing the time-space occupancy of primary traffic processes and of traffic service processes, and analyzing the nature of traffic externalities. Our focus is on the type of parameters that should be included in the capacity constraints. Section 6.2 focuses on capacity constraints associated with primary traffic processes, while section 6.3 focuses on traffic service processes and traffic externalities and the associated capacity constraints. Most capacity constraints formulated in these sections appear to be either maximum totals or linear capacity constraints. However, some types of capacity constraints appear to include parameters that are dependent on composition-of-traffic, and hence these are not truly linear. Different types of capacity constraints will result in different analytical formulations of the selection slot allocation problem (see chapter 7). Section 6.4 continues with a discussion on the specification of capacity constraint parameters. First, quality-of-service variables are reviewed, as well as their relationship with capacity constraint parameters. The next subject of section 6.4 is the determination of optimal values of traffic variables that are related to different types of capacity constraints applying to the same bottleneck. An overview of examples of bottlenecks in current practice is provided in section 6.5. It appears that many rail traffic, air traffic, and road traffic bottlenecks are sufficiently capacity-constrained to result in demand for selection slot allocation, at least by structural transport services. However, levels of capacity scarcity in the navigation sector appear to be insufficient to justify the application of selection slot

allocation in that sector. The main contribution of this chapter is the introduction of a generic framework to formulate (mostly linear) capacity constraints, which will be used for the formalization of the selection slot allocation decision problem in chapter 7.

6.1 Identification of bottlenecks

The first step of the identification of traffic supply is to identify the type of bottleneck under consideration. Therefore, this section first focuses on the identification of functional types of traffic network elements. A related issue is the identification of types of capacity constraints. To this end, a classification of capacity constraint types is introduced, which is based on a classification of primary traffic processes, traffic service processes, and environmental processes. Furthermore, the second sub-section introduces a generic modeling approach for capacity constraints.

6.1.1 Types of traffic network elements

Because different types of traffic network elements may result in different types of capacity constraints, a function-based typology of traffic network elements is introduced in this section. In general, network elements are classified as either nodes or links. However, various more detailed classifications of infrastructure elements and traffic network elements may be found in the literature (see e.g. Schoemaker, 2002). These classifications are often based on the transport and traffic processes that are being served. The following typology of traffic network elements is used in this thesis:

- traffic links;
- traffic buffers;
- crossings;
- traffic interchanges;
- parking accommodations;
- terminals.

Both traffic links and traffic buffers are link-type traffic network elements. *Traffic links* are traffic elements with the primary function to facilitate movement of vehicles between nodes. Freeways, railway lines, and air corridors are all examples of traffic links. *Traffic buffers* are locations in the traffic network where vehicles may queue, waiting to enter the next traffic network element. In many situations, the same infrastructure elements may function as both traffic links and buffers. Buffers are modeled here as links and not as nodes, because queuing is a dynamic process. Queued vehicles may be stationary, but this is not a general rule. For instance, airplanes that have no clearance to land are buffered by using a circular holding pattern. Just as air corridors, holding patterns have fixed locations (Nolan, 1994).

Traffic network elements with other functions are modeled as nodes. A *crossing* is an infrastructure element allowing other traffic to cross a link at the same level. Crossings

can only be used by traffic from one link at the same time and crossings cannot be used to change between links. Crossings may be between links of the same traffic network, for instance crossings between railway lines, but crossings may also be between different traffic networks, for instance bridges that have to be opened for vessels that are too high to pass under the bridge, or at grade railway crossings. *Traffic interchanges* are infrastructure elements connecting links. Traffic interchanges generally enable traffic to go in different ways (for instance freeway interchanges). However, locks can also be classified as traffic interchanges, although locks generally connect sequential links only. Another specific type of interchange is a connection with other traffic networks, for instance on-ramps and off-ramps of a freeway network. *Parking accommodations* are needed to store vehicles when they are not used to perform transport services. These facilities are generally located near terminals, for instance railway yards near train stations. Finally, *terminals* function as a connection between traffic and transport networks. At terminals, vehicles stop and open their doors to allow boarding and alighting of passengers or loading and unloading of freight. In transport service networks, terminals function as access points or interchanges between transport services. Examples of terminals are railway stations, bus stops, wharfs, and airport terminals.

Infrastructure elements such as railway stations or airports can be described as single traffic network elements, but also as small traffic networks consisting of various servers of different types. Figure 6.1 compares the specification of a traffic network at an aggregate level and at a detailed level. The box in the upper figure denotes the part of the network that has been elaborated in more detail in the lower figure. The part of the network elaborated in detail is a terminal area. The functional layout of this terminal is typical for a medium-size railway station with two through tracks, two platform tracks, and a reversing track.

Generic methods to describe traffic networks as networks of servers of different types have been introduced by Morlok (1978) and Schwanhäusser (1994). However, this thesis focuses on the aggregate capacity constraints and levels of capacity scarcity of bottlenecks that are the result and therefore the detailed analysis of bottlenecks as server-queue systems is not elaborated on here. Nonetheless, simulation of bottleneck processes based on server-queue network models may be useful to determine the capacity of bottlenecks.

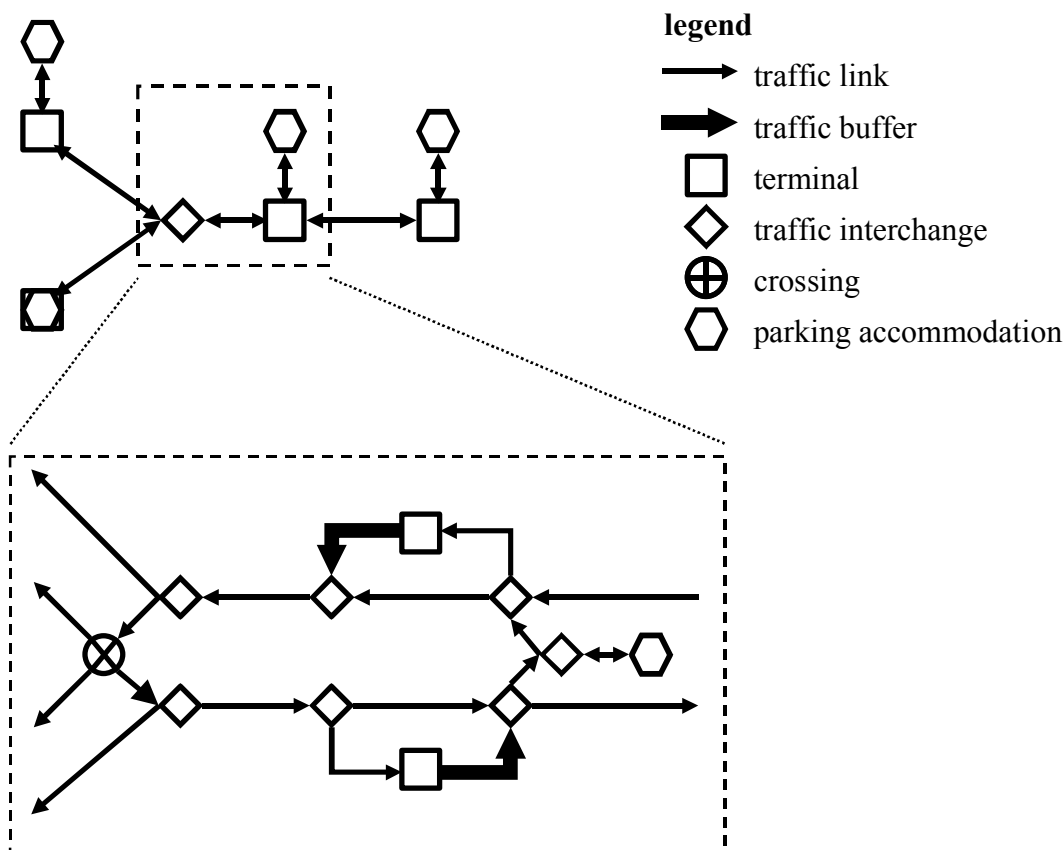


Figure 6.1: Examples of aggregate and detailed descriptions of traffic networks

Infrastructure elements

In section 2.1 it has been pointed out that traffic services do not necessarily correspond with infrastructure elements. Therefore, a functional typology of traffic network elements has been provided here, rather than of physical infrastructure elements. The same type of infrastructure element may serve different traffic functions. For instance, platform tracks at railway stations are often not only used for terminal stops, but also as traffic links, buffers, and parking places.

6.1.2 Capacity constraints

Having identified the type of traffic network element of the bottleneck under consideration, the next step is to assess its traffic supply. In this thesis, traffic supply is modeled by specifying analytical capacity constraints. An alternative would be to check the feasibility of selection slot allocations by performing simulations, which may be preferred in case of complex network interactions. In the literature, various references may be found on how traffic network elements may be modeled as a server-queue system (e.g. Bell, 1980; Wakob, 1985). However, no references have been found to a generic approach to specify analytical capacity constraints corresponding with different types of bottlenecks, reason why this chapter introduces a generic approach to specify capacity constraints corresponding with different types of bottlenecks. These constraints will be

used in the solution approach to the selection slot allocation decision problem described in chapter 7.

In section 2.2, basic maximum flow and maximum load capacity constraints have been introduced. Capacity constraints can be defined as either maximum flow or maximum load, depending on the type of traffic network element under consideration. Capacity has been defined as a maximum number, instead of maximum flow rate or maximum density, which has the advantage that both types of capacity are expressed in the same dimension, i.e. number of vehicles (recall equations 2.1 and 2.2). Whether capacity can best be specified as maximum flow or maximum load depends on the type of traffic network element under consideration. Links, such as freeways and channels, have primarily a flow function and therefore their capacity is usually defined as maximum flow. In contrast, buffers and parking accommodations (e.g. car parks and railway yards) have primarily a storage function and therefore their capacity is determined by their maximum load. Finally, the capacity of terminals can best be defined as maximum flow, assuming that the period under consideration is much larger than average stationing period, but as maximum load when these periods are about equal. Given the perspective of a large period, buffering and stationing time is just a small delay of traffic flows, while given the perspective of a small period, the main issue is whether the traffic network element under consideration is fully loaded or not.

In this thesis, capacity constraints correspond with basic slots. A capacity constraint may specify either a maximum flow or a maximum load. Following the traffic selection market specification proposed in section 5.2, a separate capacity constraint has to be specified for each basic slot b , including higher-level basic slots in case of multi-level slot specification. To simplify notation, however, the index indicating the basic slot (b) under consideration will generally be omitted in the remainder of this thesis.

Homogeneous, linear, and non-linear capacity constraints

Three categories of capacity constraints are distinguished in this thesis: homogeneous capacity constraints, linear capacity constraints, and non-linear capacity constraints. Equation 6.1 represents a homogeneous capacity constraint. In this case, capacity equals maximum traffic volume, irrespective of the composition-of-traffic resulting from the selection slot allocation decision under consideration. A binary decision variable x_i is introduced to denote the selection of slot request i from the set of slot requests R . If slot request i is selected, then $x_i = 1$:

$$\sum_{i \in R} x_i \leq \kappa \quad \forall i : x_i \in \{0, 1\} \quad (6.1)$$

In section 2.2, the dependency of capacity on composition-of-traffic has been discussed. Dependency of capacity on composition-of-traffic is incorporated in linear capacity constraints by allowing different weight parameters α for each slot request i . These parameters are not necessarily different for each slot request; often a limited number of

homogeneous user classes may be distinguished. As was proposed in section 2.2, capacity κ has been defined assuming a given composition-of-traffic as a reference. Equation 6.2 is the generic formula for linear capacity constraints:

$$\sum_{i \in R} \alpha_i \cdot x_i \leq \kappa \quad \forall i: x_i \in \{0,1\} \quad (6.2)$$

It is possible that in some cases a non-linear capacity constraint is required. This implies that a non-linear function has to be applied to determine the capacity occupancy g given selection slot allocation decision \mathbf{x} . The total selection slot allocation decision is denoted by vector variable \mathbf{x} , which dimension is determined by the number of slot requests n :

$$g(\mathbf{x}) \leq \kappa \quad \mathbf{x} \in \{0,1\}^n \quad (6.3)$$

In principle, the capacity of bottlenecks is assumed to be independent of its network context, i.e. bottlenecks are modeled in isolation. If possible, however, limitations from other bottlenecks, such as reduced freedom to schedule vehicle paths, will be included in the quality-of-service criteria. Section 6.4 elaborates in more detail about the relationship between capacity and quality-of-service, including the specification of scheduling freedom criteria that are required in a network context.

Primary and secondary capacity constraints

This thesis distinguishes various types of primary and secondary capacity constraints. Primary capacity constraints are related to primary traffic processes, while secondary capacity constraints are related to traffic service process or traffic externalities restricting traffic supply.

In the next section, various types of primary traffic processes will be reviewed, including the associated types of capacity constraints. The classification of types of primary traffic processes in this thesis is related to the classification of traffic network elements in the previous sub-section:

- following, counter-flowing, and overtaking (at traffic links);
- queuing (in traffic buffers);
- crossing;
- merging (at interchanges);
- parking;
- stopping (at terminals.)

Additional to capacity constraints related to types of primary traffic processes, secondary capacity constraints may be formulated that are related to traffic service processes and traffic externalities. Two types of traffic service processes have been introduced in section 2.2: switching of infrastructure elements and traffic control. Traffic externalities

include noise, pollution, and hazards. Capacity constraints resulting from traffic service processes or traffic externalities are discussed in section 6.3.

The capacity constraints corresponding with primary traffic processes and traffic service processes may be derived by analyzing the available time-space k and the occupancy in time or space by traffic processes and traffic service processes corresponding with slot requests. Whether time or space is the variable under consideration depends on the question of whether a maximum load or maximum flow capacity constraint is formulated. In the next sections, time-space constraints corresponding with primary traffic processes and traffic service processes will be derived. These time-space constraints can be translated into capacity constraints by dividing both sides of the equation with the average time-space occupancy given standard composition-of-traffic. Similarly, environmental constraints may be constructed given the maximum environmental burden and the environmental burden associated with each slot request.

In the case of primary traffic processes, the available time-space k of any basic slot depends on its bandwidth w , slot period m (in case of maximum flow capacity constraints), or slot length l (in case of maximum load capacity constraints). Additionally, a reduction factor φ with a value smaller than 1 may have to be applied to attain a sufficient quality-of-service. For instance, a linear time-space constraint may have the following functional form (with θ_i denoting the time each slot request i occupies a traffic server):

$$\sum_i \theta_i \cdot x_i \leq \varphi \cdot w \cdot m \quad (6.4)$$

6.2 Capacity constraints from primary traffic processes

In this section, various types of primary traffic processes are reviewed. The main aim is to derive the functional form of the associated time-space constraints, including the identification of main parameters and their relationship with quality-of-service variables. As was explained in the previous section, a simple arbitrary linear transformation is sufficient to derive the corresponding capacity constraints.

6.2.1 Following and queuing

Elementary traffic links such as lanes and tracks may be used in one direction uniquely, or alternately in two opposing directions. The primary traffic processes on one-way links are following and overtaking. Assuming that following is the limiting traffic process, the capacity of traffic links is determined by minimum headways. Additionally, overtaking possibilities are restricted by the limited availability of parallel servers such as lanes and tracks and the limited availability of possibilities to switch between lanes and tracks, which will be discussed in the next sub-section.

Minimum headways

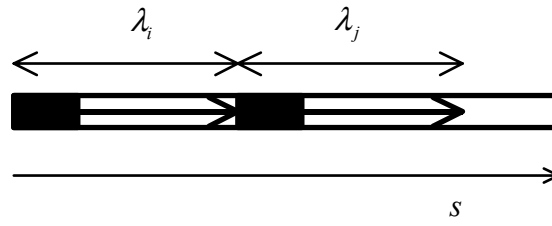
A major factor limiting the maximum performance of infrastructure elements is the distance in time and space that is required for a safe separation of a pair of vehicle paths. The need for safe separation of traffic does not only refer to maintaining safe headways, but also to crossing traffic and counter-flows, which will be discussed in the next sub-section.

Minimum safe headways are speed dependent. The relationship between speed and minimum headway depends on the distance keeping strategy. The specification of distance keeping strategies can be used to design traffic control systems separating traffic, but also to model uncontrolled traffic flows. The following distance keeping strategies are discussed by Minderhoud (1999):

- reaction time distance keeping;
- brick wall distance keeping.

Reaction time distance keeping takes into account the actual speed and maximum deceleration of conveyances ahead. The following behavior of, for instance, car traffic may be modeled as a reaction time strategy (see, for instance, Leutzbach, 1988). *Brick wall* distance keeping is based on the assumption that the vehicle directly in front is either non-moving or may decelerate to a full stop instantly. For instance, the brick wall strategy is currently used as the design standard for rail traffic control systems. Given the current traffic control system, the actual speed of trains is not known by the traffic control system, and hence reaction time distance keeping cannot be applied. Moreover, in case that a train collides with a non-moving object, this strategy avoids a following train colliding with the first train.

Minimum vehicle separation λ_i is assumed to be the sum of vehicle space and net minimum separation. The vehicle space of a vehicle is generally greater than its physical dimensions, since vehicles are always separated with some distance, even when standing still. Net minimum separation is the distance between two vehicle spaces, which is observed as a safe separation between vehicles that is required when traffic is flowing. In figure 6.2, vehicle space is marked black, while the arrow indicates the net minimum separation between vehicles i and j .

**legend**

- vehicle space (λ^0)
- net minimum separation

Figure 6.2: Minimum safe headways

The length λ^0 of the vehicle space is assumed to be independent of speed, and it may vary from less than a meter between cars to several hundreds of meters in rail traffic. The net minimum separation between vehicle paths is speed-dependent, and also depends on assumptions about the maximum decelerations o_i and o_j of the vehicle path under consideration i and preceding vehicle path j respectively. Assuming a brick wall following strategy, $o_j = \infty$. Given reaction-time based following strategies, o_j is assumed to have a finite value. We assume $o_j > o_i$, which corresponds with a prudent reaction time following strategy (Minderhoud, 1999). The vehicle flow under consideration is assumed to be homogeneous and stationary, resulting in equal speeds for all vehicles. Given these assumptions, the minimum headway θ_i of vehicle path i depends on the minimum separation (given zero speed) λ^0 , reaction time τ , speed v , and the assumed maximum decelerations o_i and o_j :

$$\theta_i = \frac{\lambda^0}{v} + \tau + \frac{1}{2} \cdot \left(\frac{v}{o_i} - \frac{v}{o_j} \right) \quad (6.5)$$

Assuming a brick wall or a prudent following strategy, minimum headway is a convex function of speed, which means that it has a minimum for the optimum speed v_{opt} . However, this optimum speed may be higher than maximum allowed speed v_{max} .

Given equation 6.5, we may derive the flow capacity constraint of a link with homogeneous flow and known norm speed v . Norm speed may be based on the optimum speed or the maximum allowed speed, but also on other norm speeds corresponding with given quality-of-service criteria. A time-space constraint similar to equation 6.4 may be derived. The available service time is determined by the length m of the period under consideration and the bandwidth w of the bottleneck. However, in order to allow for a higher reliability level and more flexibility to schedule vehicle paths, a reduction factor φ is introduced. The parameter θ_v corresponds with the minimum headway given norm speed v . The selection of vehicle paths is denoted by the binary variable x_i . The resulting time-space constraint is simply a maximum total:

$$\theta_v \cdot \sum_i x_i \leq \varphi \cdot w \cdot m \quad (6.6)$$

Descent approach types

Minimum headways may not only depend on speed, but also on other factors. An issue with respect to minimum headways that is specific for air traffic is the type of descent approach. It appears that the minimum headways between landing aircraft depend on the type of descent approach that is adopted. At Amsterdam Airport Schiphol, for instance, landing aircraft are normally separated by 2 minutes. As an alternative to the standard descent approaches, continuous descent approaches (CDA) are preferred at night, because they are associated with lower noise levels. The downside of CDA, however, is that minimum separation increases to 4 minutes. These larger separations are required to allow for the higher freedom in descent speed choice (Kershaw *et al.*, 2000). In this situation, equation 6.6 may be applied, with the modification that the minimum headway parameter θ is not dependent on norm speed but on norm descent approach.

Heterogeneity

One of the aspects determining capacity reviewed in section 2.2 is heterogeneity of traffic. In case of heterogeneous traffic flows, minimum headways vary between user classes. Two dimensions of heterogeneity are of particular importance. In the first place, speeds differ between users or user classes, for instance between trucks and cars or between local trains and intercity trains. These speed differences will lead to different safe headways. In the second place, minimum vehicle spaces λ^0 may vary, for instance due to different vehicle lengths. For instance, trucks need more space than cars, and long freight trains need more space than short regional passenger trains. Finally, deceleration capabilities may differ, resulting in different same minimum headways. The formulation of capacity constraints given heterogeneous traffic is enabled by assuming that every vehicle path i has its own minimum headway θ_i . Assuming the absence of interaction effects, the corresponding time-space constraint has the following linear form:

$$\sum_i \theta_i \cdot x_i \leq \varphi \cdot w \cdot m \quad (6.7)$$

Interactions

Apart from the possible dependency of minimum headways on the type of user under consideration, it is also possible that minimum headways depend on the type of user it is following. The capacity gain or capacity loss due to a mixture of traffic types is often called the interaction effect. Bliemer (2001) distinguishes symmetric and asymmetric interaction between vehicle classes. Interactions are symmetric if the effect of vehicles of class A on the minimum headway of vehicles of class B is equal to the effect of vehicles of class B on the minimum headway of class A vehicles.

Interactions between user classes are, for instance, known to occur with respect to planes. The wake turbulence behind large aircraft imposes a potential danger for smaller aircraft. Therefore, the minimum separation of a large aircraft followed by a small

aircraft is larger than the separation between two small aircraft or between two large aircraft (see Nolan, 1994). The interaction between small and large planes is asymmetric, because wake turbulence is not a problem when a large aircraft follows a small aircraft. Another possible example of dependency of headway on the type of vehicle in front is the interaction between trucks and cars. Based on vehicle passage time observations at a two-lane freeway in the Netherlands, Hoogendoorn & Bovy (1999) conclude that truck drivers on average tend to follow person cars more closely than other trucks during congested traffic flow. The interaction between cars and trucks is also assumed to be asymmetric, because block the view of car drivers, cars do not block the view of truck drivers on traffic downstream. (Hoogendoorn & Bovy, 2000). Finally, a positive instead of negative effect of mixture of traffic types may be realized on airport runways. According to Eurocontrol (2002), mixture of arrivals and departures on the same runway may result in a slightly higher capacity than in a situation with arrivals or departures only.

The exact level of interaction within a period depends on the sequence of vehicle paths. The total interaction effect associated with a pair of slot requests is the product of the probability of interaction $p_{i,m}$ between user i and user class m (i.e. the likelihood that user i is following a vehicle path of class m) and the interaction level $\chi_{i,m}$. In the first place, the probability of interaction $p_{i,m}$ depends on the percentage of traffic belonging to class m . Secondly, this probability increases with increasing speed differences, which again depends on the composition-of-traffic. Consequently, the percentage of traffic from each category determines the expected percentage of longer (or shorter) separation associated with specific sequences (e.g. small plane after large plane). The interaction level is positive if the mixture has negative effects on capacity. Apart from the dependency of $p_{i,m}$ on the selection slot allocation decision, the resulting time-space constraint with interaction is linear:

$$\sum_i \left\{ \theta_i \cdot x_i + \sum_m p_{i,m} \cdot \chi_{i,m} \cdot x_i \right\} \leq \varphi \cdot w \cdot m \quad (6.8)$$

6.2.2 Overtaking

Vehicles that are driving on the same lane, track or navigation path cannot overtake each other without changing between traffic servers. In the air, on water, and on freeways, overtaking is possible if the other path or lane has sufficient capacity. Guided traffic systems, such as railways, can only change tracks at switches. On roads and railways with a single lane or track for each direction, overtaking is not possible at all. Assuming that slow traffic cannot speed up, fast traffic has to slow down in situations with overtaking restrictions. Hence, assuming that the minimum speed of each vehicle path is determined by given quality-of-service criteria, overtaking constraints result in lower capacity values. Figure 6.3 illustrates the possible relationship between flow and speed in a situation with two user classes with different speeds on a link with limited overtaking possibilities. The independent variable q_2 is the flow of the 'slow' user class,

while the dependent variable v_l is the speed of the 'fast' user class. The traffic flow of the 'fast' user class is constant.

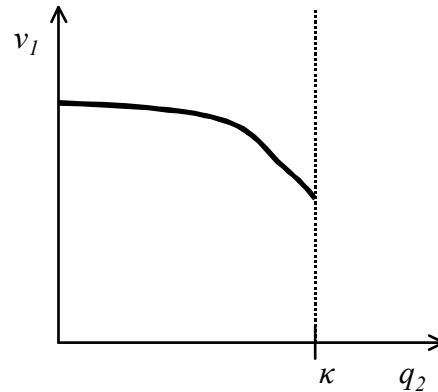


Figure 6.3: Speed-flow interaction of two user classes given overtaking constraints

No overtaking possibilities

A specific situation that is studied here is a situation without overtaking possibilities. This situation applies, for instance, to rail traffic if no overtaking tracks are available, and to landing air traffic given a single available runway. An analysis of the relationship between velocity distribution and capacity for landing air traffic, assuming no overtaking probabilities, has been performed by Blumstein (1959). Blumstein analyzes the landing capacity of runways, assuming a uniform velocity distribution. He shows that significant capacity reductions are possible given large speed differences and long approach paths to the runway, but in practice other bottlenecks such as the minimum spatial separation of aircraft before the landing begins will reduce the effect of speed differences significantly.

We study the following basic situation without overtaking possibilities. A traffic link consisting of a single server is used for traffic in a single direction. For instance, this may be a railway track between stations A and B without passing tracks. It is assumed that stations A and B have sufficient capacity, which means that vehicles may overtake each other there. However, it may also refer to the approach path of landing air traffic, or other types of traffic links without overtaking possibilities. The traffic link under consideration is used by two types of traffic, each with a different speed resulting in different travel times π_i . In figure 6.4, minimum headway is again denoted by θ_i , and traffic server occupancy due to interaction between vehicle paths i and j is denoted by $\chi_{i,j}$.

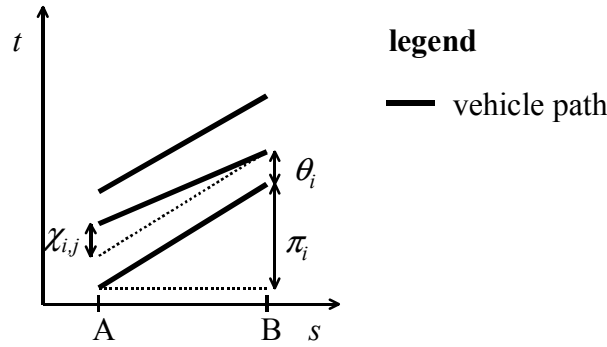


Figure 6.4: Heterogeneous traffic on a link without overtaking possibilities

To assess the effect of overtaking constraints on capacity, we need to make assumptions about the sequence of traffic. Assuming a randomized sequence of vehicle paths, a time-space constraint similar to equation 6.8 may be applied. An extra adaptation may be required to allow for the possibility that the last vehicle path in the previous period is slower than the first vehicle path in the period under consideration, but this possibility is neglected here to enable the separate analysis of subsequent periods. Link capacity is defined as maximum inflow, and consequently it is reasonable to specify interaction as inflow loss due to speed differences. Given this definition, extra occupancy of the traffic server due to speed differences (denoted by $\chi_{i,j}$ in figure 6.4) occurs when the next vehicle path is faster than the previous one. The total interaction effect is determined by the probability $p_{i,m}$ that i is following a vehicle path of class m , and by the interaction parameters $\chi_{i,m}$. In this case, the interaction level equals the positive travel time difference between vehicle path i and of vehicle paths of class m :

$$\chi_{i,m} = [\pi_i - \pi_m]^+ \quad (6.9)$$

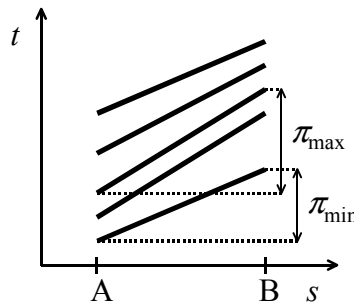


Figure 6.5: Optimized sequence of heterogeneous traffic on link without overtaking

A special case is when the sequence of vehicle paths is optimized in order to maximize capacity. For instance, the sequence of trains is often planned at the timetable planning level. In this case, the sequence of vehicle paths within each period may be chosen as to minimize capacity occupancy within each period. Given such an optimized sequence, only the difference between minimum and maximum travel time (π_{\min} and π_{\max}) has to be added to the sum of minimum headways (see figure 6.5). Although the minimum and maximum travel time may depend on the selection decision, we may consider these as

given parameters for reasons of simplicity. In some situations, the inclusion of at least one slot request with $\pi_i = \pi_{\max}$ and one or more slot requests with $\pi_i = \pi_{\min}$ is beyond discussion. The resulting time-space constraint is linear:

$$\sum_i \theta_i \cdot x_i + \pi_{\max} - \pi_{\min} \leq \varphi \cdot m \quad (6.10)$$

In an optimized sequence of vehicle paths, only the addition of paths slower than the currently slowest path or quicker than the currently quickest path results in extra required service time due to interaction effects. Equation 6.10 is still valid when a single slot request has a higher travel time than π_{\max} or a lower travel time than π_{\min} is considered, if the additional travel time due to the extreme high or low travel time value of this slot request is added to the minimum headway θ . However, equation 6.10 is not suitable to evaluate the selection of more than one slot requests with travel times higher or lower than the assumed range.

Finally, a possible situation is that the majority of traffic has an identical travel time π_{norm} . We assume that the vehicle paths with 'extreme' travel times are scheduled between vehicle paths with 'average' travel times. The following linear time-space constraint corresponds with these assumptions:

$$\sum_i \theta_i \cdot x_i + |\pi_i - \pi_{\text{norm}}| \cdot x_i \leq \varphi \cdot m \quad (6.11)$$

In the situations described here, the addition of vehicle paths with 'extreme' travel times results in capacity loss due to interaction, while the addition of vehicle paths with 'average' characteristics does not lead to extra capacity loss. This principle may be used to optimize the service pattern on capacity-constrained railway links. Koolstra (2000), for instance, proposes to give express trains on capacity-constrained railway links in the urbanized Randstad area in the Netherlands at least twice the frequency of (faster) intercity trains and (slower) local trains.

6.2.3 Counter-flowing, crossing, and merging

In the previous sub-sections, we studied following and overtaking restrictions on traffic links used in one direction. However, some traffic links consist of a single traffic server that is used by traffic from both directions. This situation applies to single-track railway links, and may also apply to runways used in both directions. Crossings are common traffic network elements in all sectors, and traffic flows from different directions may result in an interaction component in the capacity constraint. Finally, traffic flows may change directions at traffic interchanges. Merging traffic from different origins may merge at traffic interchanges, which may give rise to similar interactions between traffic flows from different directions.

Single-track links

Single-track railway links are generally capacity-constrained. Given that traffic from opposite directions can only meet at a limited number of locations, traffic scheduling possibilities are clearly limited. Furthermore, the capacity of a single-track railway link with bi-directional traffic is much less than half the capacity of a double-track railway link, which is due to the large capacity loss of changing the direction of operation ($\chi_{i,j}$ in figure 6.6).

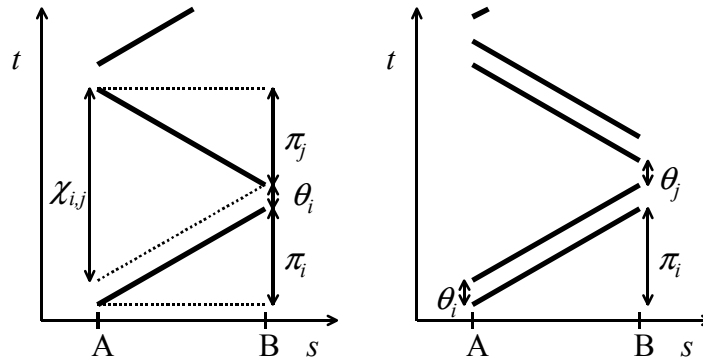


Figure 6.6: Examples of capacitated single-track railway traffic

Another example of a traffic network element that may be regarded as a single-track link is a runway used in both directions. According to Nolan (1994), capacity is lost for each change of direction if runways are used in both directions in the same period, similar to the situation with single-track railways.

Figure 6.6 illustrates that the frequency of changes of direction dominates the capacity of single-track railways, in particular when the travel times π are much larger than the minimum headways θ . Again, the interaction term in equation 6.8 may be used to model the sequence-dependent interaction between subsequent vehicle paths. The probability of interaction $p_{i,m}$ is interpreted as the probability that i is followed by a vehicle path of class m , just as in the previous case. Figure 6.6 illustrates that if vehicle path i is followed by a vehicle path running in the opposite direction, the interaction level $\chi_{i,m}$ is the sum of travel times of both (types of) vehicle paths:

$$\chi_{i,m} = \pi_i + \pi_m \quad (6.12)$$

Crossings and traffic interchanges

A crossing is another example of a traffic element with changes of direction. The minimum headway between two crossing vehicle paths may be larger than the minimum headway between two vehicle paths following each other. For instance, the headway between two crossing vehicle paths can be determined given the 'brick wall' principle introduced earlier. If reaction time based distance keeping is applied to traffic going in the same direction, the headways may be smaller than between crossing vehicle paths. In this case, the capacity of a crossing is lower when frequent changes of direction are

required. Another aspect is that traffic approaching a crossing or a traffic interchange may have to wait briefly before the crossing or interchange is available, even when these are used significantly below capacity. In this case, crossing or merging may occur at lower speeds than optimal, resulting in larger minimum headways and hence in a lower capacity. Finally, additional capacity restrictions may be required to avoid that queuing delays would result in a quality-of-service lower than minimally desired.

The interaction between traffic flows from different directions may again be modeled by including the interaction level in the capacity constraints. Hence, the same type of equation as described by equation 6.8 may be applied. The capacity of railway crossings given different traffic flow levels from both directions has been analyzed numerically by Schaafsma (1987). It appears that the set of feasible combinations of traffic flows q_1 and q_2 maximizing capacity is non-convex, with a shape similar to figure 6.7.

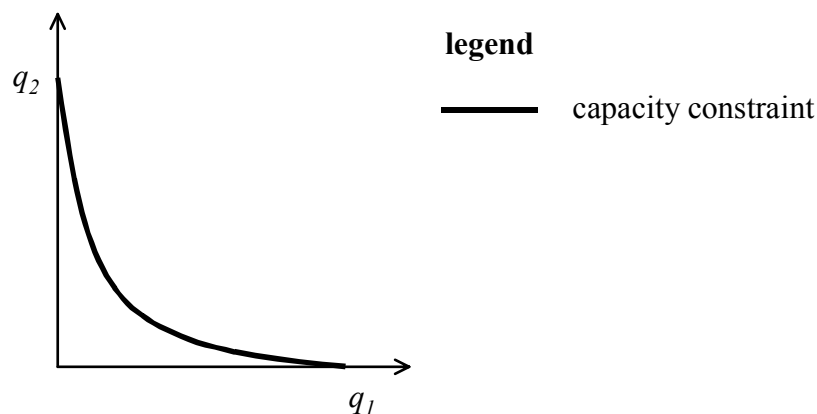


Figure 6.7: Capacity constraint of crossing (or merging) traffic

6.2.4 Queuing, parking, and stopping at terminals

Queuing, parking, and stopping at terminals are traffic processes where maximum load may be binding instead of maximum flow. Queuing is the typical traffic process of buffers, and parking is the typical traffic process of parking accommodations.

Buffers

Buffers are usually applied upstream of a bottleneck. Buffers are required in situations where the arrival rate may be temporarily larger than the maximum possible throughput of the bottleneck downstream. However, buffers have a limited capacity themselves, and hence may also be a bottleneck in specific situations. In section 6.1 it has been stated that the capacity of buffers may best be modeled as maximum load. Modeling buffer capacity as maximum flow would imply modeling these as flow links, neglecting the very nature of buffers. Hence, buffer capacity is specified as maximum load.

A simple approach to model buffer capacity is to calculate the maximum load of buffers given an assumption about maximum density. However, most vehicles in a buffer will

leave it in relatively short notice, and it is not realistic to assume that all traffic will arrive simultaneously. The maximum inflow rate is limited by capacity constraints, and often it is reasonable to assume an even lower maximum inflow rate. The outflow rate is limited by the capacity of the bottleneck downstream. For instance, assuming that the buffer has a maximum load of 100 vehicles, and assuming that maximum outflow is 50% of maximum inflow, 50 vehicles will have left the buffer in the period that these 100 arrived. Therefore, we may define the capacity of a buffer as the maximum number of vehicles that may be accommodated assuming that the inflow rate equals the inflow capacity, and the outflow rate equals the outflow capacity which is limited by the bottleneck downstream.

In this case, the available time-space can be determined by the following geometric series, with l denoting slot length, λ_i denoting required separation distance for each vehicle path i , ρ denoting the ratio between maximum outflow and maximum inflow, and j being an internal variable without physical interpretation:

$$\sum_i \lambda_i \cdot x_i \leq \sum_{j=0}^{\infty} \varphi \cdot l \cdot w \cdot \rho^j \quad (6.13)$$

Equation 6.13 can be simplified to the following linear time-space constraint:

$$\sum_i \lambda_i \cdot x_i \leq \frac{\varphi \cdot l \cdot w}{1 - \rho} \quad (6.14)$$

Please note that the separation distances λ_i depend on the speed within the buffer, where speed again depends on maximum outflow. Assuming reaction-time based or brick wall distance keeping, maximum load decreases as speed increases, and consequently maximum load decreases as outflow increases. However, the effect on capacity of larger separation distances due to a larger outflow rate will be more than compensated by the increasing flow ratio ρ , and hence buffer capacity will increase as the outflow rate increases.

Parking accommodations

The approach outlined here to determine the capacity of buffers may also be applied to parking accommodations if the outflow rate may be assumed to be independent of duration of parking. This assumption, however, is not very realistic. Therefore, the alternative is proposed here to specify slot periods in such a way that these correspond with the period for which a parking place is needed. In this case, the number of available parking places w can be directly adopted in the time-space constraints:

$$\sum_i x_i \leq \varphi \cdot w \quad (6.15)$$

Terminals

Just as parking accommodations, terminals facilitate stationary traffic. The main difference between parking places and terminals is that terminals are specifically used to allow boarding and alighting of passengers and loading and unloading of freight. The best approach to formulate capacity constraints depends on the type of terminal under consideration.

We first assume that the terminal under consideration is a hub, and the slot period corresponds with a bank of connecting transport services. According to the quality-of-service criteria, all transport services should be present in the terminal in this period to allow exchange of freight or passengers. In this case, the available number of platforms w determines the available time-space, resulting in the type of time-space constraint described in equation 6.15. However, in many other situations the duration of stops is much shorter than the period under consideration, implying that a maximum flow formulation is more suitable. Maximum flow at terminals is not only restricted by minimum headways, but also by the duration of stops (see, for instance, Kreuzberger & Vleugel, 1992). In the first place, these stops should be sufficient to allow for boarding and alighting, or loading and unloading. Additionally, opening and closing of doors, security checks, etc. may take some time. Assuming known required stopping time per vehicle path σ_i , the following time-space constraint applies, which is similar to equation 6.7:

$$\sum_i (\theta_i + \sigma_i) \cdot x_i \leq \varphi \cdot w \cdot m \quad (6.16)$$

6.3 Capacity constraints due to secondary processes

In the previous section we focused on capacity constraints imposed by primary traffic processes such as following, overtaking, queuing, and parking. This section continues with the specification of secondary processes, i.e. traffic service processes and traffic externalities, and their impact on capacity constraints. Traffic service processes are related to the operation of infrastructure, including switching between different states of infrastructure elements and the operation of traffic control systems. Traffic externalities include noise emissions, pollution, and hazards resulting from traffic flows.

6.3.1 Switching processes

Switching is the process of changing the states of infrastructure elements. Some types of interchanges have two or more states. Examples of such interchanges are railway switches with the states 'straight' and 'branching', locks with the states 'high level' and 'low level', and bridges with the states 'open' and 'closed'. Furthermore, traffic light controlled road intersections can be modeled as switches with each green phase corre-

sponding with a separate state. The capacity of these traffic network elements depends on the service time required for switching operations.

The required frequency of switching depends on the sequence of traffic. A possible solution to model the loss of service time of interchanges due to switching operations is to use a time-space constraint similar to equation 6.8. In this case, the interaction level $\chi_{i,m}$ is determined by a binary measure $\delta_{i,m}$ indicating the necessity of switching in the case that a m -type vehicle path follows vehicle path i , and the service time lost due to switching operations ζ :

$$\chi_{i,m} = \delta_{i,m} \cdot \zeta \quad (6.17)$$

Alternatively, we may assume that the optimum frequency of switching operations given capacitated traffic is independent of traffic volumes. Given capacitated traffic, locks are usually operated with fully loaded lock chambers, in order to maximize the locking capacity. Furthermore, the frequency of railway bridge openings is usually determined in advance, independent of vessel traffic demand. In these cases, a time-space constraint may be applied which includes switching frequency f as a separate variable. We assume that all parallel servers have the same switching frequency:

$$\sum_i \theta_i \cdot x_i \leq \varphi \cdot w \cdot m - \zeta \cdot w \cdot f \quad (6.18)$$

In equation 6.17, the occupancy time of switching operations was introduced as a single parameter, which means that the occupancy time of switching operations is indeed independent of traffic flow characteristics. This is a reasonable assumption for railway switches and bridges, because the switching process of these is not related to the traffic process. The switching process of locks, however, is somewhat related to the traffic process. The opening and closing time of the gates is not related to the traffic process, but because ships are in the lock chamber while it is being filled, the filling process is somewhat related to the number and type of ships in the lock chamber. Groeneveld (1999) discusses this relationship in detail. However, we assume that switching time is not related to traffic volume.

Additional aspects included in switching time

Included in the switching time parameter ζ is the headway between the switching process and the first vehicle. This headway may be the same as the headway between vehicles given a brick-wall distance keeping strategy. A brick wall approach should be applied for the headway between switching operations and vehicle paths, because switches are fixed points and not moving vehicles with a minimum stopping distance.

A specific aspect of locks is that their operation time is also determined by the time needed by the ships to enter and to exit. Consequently, not only the extra headway between the last ship and the operation of the lock doors restricts the capacity of locks,

but also the fact that ships not only have to enter but also have to exit the lock chamber. The time required for vessels to exit the lock should be included in the switching time parameter.

6.3.2 Traffic control

Another type of traffic service process that may restrict capacity is traffic control. Traffic control systems may have a limited handling capacity, and if the workload of traffic control systems is related to traffic volumes, then capacity restrictions may be the result.

The most notable example of traffic control related traffic restrictions is the handling capacity of Air Traffic Control (ATC). ATC centers have limited staff and the maximum workload of air traffic controllers is also limited. Consequently, ATC handling capacity may be insufficient to handle peaks in traffic demand. Furthermore, each air traffic controller is responsible for his own sector of airspace (Nolan, 1994). To a certain extent, the handling capacity of air traffic control centers may be increased by increasing the number of air traffic controllers and air traffic control sectors. However, in a number of situations the gains of increasing the number of sectors would be outweighed by the increase in coordination needed between the controllers of neighboring sectors and the increase in communications between pilots and air traffic controllers (Magill, 1998).

According to Majumdar *et al.* (2002, p. 382), the capacity of an ATC sector is "*the maximum number of aircraft controlled in that ATC sector in a specified period of time, whilst still permitting an acceptable level of controller workload*". Studies on the optimum ATC slot allocation problem usually assume that the division of airspace into sectors is given, and that the capacity of each sector is a given maximum number of flights entering a ATC sector in each period (e.g. Lindsay *et al.*, 1993; Van den Akker & Nachtigall, 1997). This is equivalent to assuming a time-space constraint for each sector that is similar to equation 6.7, with ξ_i representing a measure of workload for each flight i :

$$\sum_i x_i \cdot \xi \leq \varphi \cdot m \quad (6.19)$$

However, this approach is too simplistic according to Majumdar *et al.* (2002). Based on a simulation study of ATC workload, they conclude that only cruising and ascending flights are directly related with workload. Furthermore, ATC workload appears to depend on interactions between cruising, ascending, and descending flights, i.e. workload is relatively high given a mix of ascending and descending flights. We may regard these as different user classes m . Although it will not be known beforehand whether flight i will be cruising, ascending, or descending in sector a , the probability may be estimated and may be different for different flights. The resulting time-space constraint is similar to equation 6.8, with ξ_i indicating workload (as server occupancy time) associated with

vehicle path i , $\chi_{i,m}$ indicating the expected interaction level between vehicle path i and traffic of user class m , and $p_{i,m}$ indicating the probability that vehicle path i interacts with traffic of user class m :

$$\sum_i \left\{ \xi_i \cdot x_i + \sum_m p_{i,m} \cdot \chi_{i,m} \cdot x_i \right\} \leq \varphi \cdot m \quad (6.20)$$

6.3.3 Traffic externalities

Until now, we have focused on capacity constraints that are directly related to traffic processes and traffic service processes. However, the capacity of traffic network elements may also be determined by external factors. These external factors may include all kinds of restrictions for environmental reasons, including restrictions on hazards, noise, and pollution. This section reviews the functional form of alternative types of capacity constraints.

Environmental capacity constraints are administrative capacity constraints. They may be formulated functionally, e.g. as maximum noise level or maximum hazard level, but also indirectly as maximum traffic level. A relatively straightforward approach to implement environmental capacity constraints is to formulate (un)weighted maximum traffic volumes for each bottleneck. An example of such a constraint is the maximum yearly number of flights that is allowed at Amsterdam Airport Schiphol, which is restricted by noise regulations (Stout, 2001). In this case, a contour has been determined which limits the area in which a certain noise level may be exceeded during day periods. This noise contour corresponds with 35 *Kosten units* (named after Prof. Kosten, chairman of a former governmental noise hindrance working group). *Kosten units* define a cumulative maximum noise level per year (Kershaw *et al.*, 2000). To determine the capacity of this airport, this cumulative noise level is translated into a maximum annual number of flights (Schiphol, 2001). This is irrespective of the plane types that are used, although there are large differences in noise emissions between types of aircraft. Another example of an environmental restriction resulting in a cumulative capacity constraint is the restriction of accident rates on at grade railway crossings in the Netherlands. The probability of (fatal) accidents on at grade railway crossings is assumed to be positively correlated with traffic volume and traffic speed (V&W, 1999b). Finally, noise regulation may specify maximum noise levels that may not be exceeded, without aggregation to annual totals. This type of environmental restriction results in maximum flow capacity constraints. For instance, Dutch noise regulations restrict train frequencies, which in practice can be restrictive for (relatively noisy) heavy freight trains. Similar noise regulations are applied to air traffic in the Netherlands during night periods.

Both flow-based and aggregate environmental capacity constraints may be modeled as linear capacity constraints. We may derive environmental constraints by analyzing the maximum environmental burden η , instead of available time-space. The total level of

environmental burden is assumed to be a sum of individual environmental burden levels β_i :

$$\sum_i \beta_i \cdot x_i \leq \eta \quad (6.21)$$

Please note that the environmental burden levels β_i may be dependent on traffic flow characteristics such as speed. For instance, noise is usually positively correlated with speed. Furthermore, the probability of fatal accidents is often positively correlated with speed. Finally, an issue already discussed in sub-section 6.2.1 is that noise emissions of descending aircraft are related to descent speed. Often, aircraft descend stepwise to their destination airports. Alternatively, and (idle thrust) continuous descent approach may be followed, which means that the aircraft engines may run stationary throughout the descent, resulting in less noise emissions and lower fuel use.

A specific issue is that if we choose norm speeds and descent speeds as to maximize environmental capacity, flow capacity will usually not be maximized. In the first place, the optimum speed will often be different. Minimum noise levels are often attained at low speed levels, while minimum headways usually require high speeds. Furthermore, the optimum descent speed of an aircraft may differ significantly between aircraft types, while minimum headways will be realized given homogeneous descent speeds. Hence, we may conclude that not only a proper balance has to be found between maximizing capacity and maximizing quality-of-service, but also between maximizing capacity values associated with different types of capacity constraints. The relationship between capacity and quality-of-service and the problem how to optimize their balance is one of the topics of the next section.

6.4 Specification of capacity constraint parameters

In the previous sections, various types of time-space constraints (and environmental constraints) have been introduced. As was already stated in section 2.2, these time-space constraints can simply be translated into capacity constraints. This section focuses on the specification of capacity constraint parameter values, given the dependency of capacity on minimum quality-of-service. This section aims to discuss how capacity and quality-of-service are related, without providing exact guidelines how the optimum capacity constraint parameter values should be derived. Furthermore, this section discusses the optimization of quality-of-service variables that are related with two or more capacity constraints. For instance, traffic variables such as speed and descent approach type may correspond with different optimal capacity values for different types of traffic processes.

6.4.1 Types of capacity constraints and corresponding parameters

In sub-section 6.1.2, three types of capacity constraints have been introduced: homogeneous capacity constraints, linear capacity constraints, and non-linear capacity constraints. The primary and secondary time-space constraint types formulated in the previous sections may be directly translated into homogeneous, linear, and non-linear capacity constraints. The only homogeneous capacity constraint is equation 6.6. Examples of linear capacity constraints are equations 6.7, 6.14, and 6.21. Finally, some capacity constraints formulated in the previous sections are non-linear linear, because they include variables such as probability of interaction p (equation 6.8) or switching frequency f (equation 6.18) that are (partly) dependent on composition-of-traffic. Additionally, travel time parameters such as maximum and minimum travel time (π_{\max} and π_{\min}) are conditional on the selection of slot requests with these associated travel times. However, as will be concluded in chapter 7, the representation of these composition-of-traffic dependent variables as parameters in linear capacity constraints is often an acceptable approximation. In many situations only a part of all traffic is evaluated, and hence many interaction values are already known.

Capacity constraint parameters

Capacity constraints may include different types of parameters. In the previous sections, the following possible parameters of linear capacity constraints (α in equation 6.2) have been identified:

- minimum headway θ ,
- minimum vehicle separation λ ;
- minimum stopping time σ ,
- switching time ζ ;
- workload ξ ,
- environmental burden β .

Furthermore, the level of the interaction parameter χ , which has been introduced in equation 6.8, is included in the α parameter of linear capacity constraints. For various situations, the dependency of this interaction parameter on traffic related parameters has been outlined in the previous sections. Apart from the probability of interaction $p_{i,m}$, the level of interaction may depend on the following parameters:

- switching necessity δ ,
- switching time ζ ;
- travel time π ,
- minimum travel time π_{\min} ;
- maximum travel time π_{\max} .

Finally, the available capacity κ on the right hand side of the capacity constraints depends on a number of variables and parameters. The most notable parameter is the reduction factor φ , which is present in most capacity constraints. Furthermore, the fol-

lowing specific parameters influencing the available service time-space can be identified:

- outflow-inflow ratio ρ ;
- switching time ζ ;
- environmental capacity η .

In the next sub-sections, the relationships between quality-of-service variables and capacity constraint parameters, and the determination of parameter values will be discussed.

6.4.2 Quality of service variables

The values of capacity constraint parameters are related to quality-of-service variables. In general, available capacity decreases with increasing required minimum quality-of-service. The relationship between capacity and quality-of-service has been discussed in section 2.2. The three key quality-of-service variables introduced in that section were reliability, travel time, and scheduling freedom. This sub-section elaborates on the relationship between these quality-of-service variables and capacity values. Additionally, the effect of network dependencies on the specification of capacity constraint parameters will be discussed.

Reliability

In section 2.2, it was concluded that available capacity is the relevant type of capacity in the context of slot allocation. Furthermore, the relationship has been specified between available capacity and reliability. It is assumed that the actual capacity per period is an independent random variable, i.e. the stochastic capacity variables of subsequent periods are independent. Recalling equation 2.3, reliability is the probability that the actual capacity K is higher than the available capacity κ :

$$r = P(K \geq \kappa) \tag{2.3}$$

This implies that capacity and reliability are negatively related. A higher reliability implies a higher probability that the actual capacity is greater than the available capacity, which means that the available capacity should be lower. If the mathematical form of the probability function in equation 2.3 is known, it is possible to derive the value of the available capacity κ :

One way to implement this capacity adjustment in the capacity constraints is to adjust the value of the reduction factor φ . This is the most straightforward solution when randomness of capacity is due to infrastructure specific factors. For instance, Bär *et al.* (1988) propose to incorporate the reliability of infrastructure systems into capacity definitions, given that infrastructure elements may be out of service due to random factors, while the out-of-service time of infrastructure elements is also a random vari-

able. However, various other factors determining reliability may be identified, including randomness of minimum headways, randomness of overtaking and crossing conflicts between vehicles, and randomness of emissions, which are often specific for individual vehicle paths or classes of traffic. Their stochastic nature may be implemented directly in the parameters related to these factors, such as minimum headways θ , interaction parameters χ , and environmental burden parameters β . Again, more conservative parameter estimations are required to attain a higher reliability.

To conclude, the capacity value that is applied in the slot definition depends on the chosen reliability level, or *vice versa*. This thesis assumes that it is possible to derive the capacity of a slot given a minimum required reliability level, without explicitly prescribing how this analysis should be performed. However, a few general statements may be made based on general queuing theory. In the first place, it can be inferred given equation 2.3 that a greater level of stochastic variation in traffic processes results in either a lower capacity or a lower reliability. Furthermore, it can be shown that a bottleneck can only operate at its capacity if the probability that no vehicle is queued when a traffic server is ready is 0. Consequently, maintaining a queue is generally needed to use a bottleneck at its full actual capacity, which implies a lower quality of service. For instance, assuming deterministic service times and a Poisson arrival process, it can be shown using the Pollaczek-Khintchine formula that the expected steady-stated queue length of a single-server bottleneck is more than 9 given that the flow rate is 95% of the service rate (see e.g. Hillier & Lieberman, 1990). This means that a high capacity does not only imply a lower reliability but also longer expected travel times.

Travel time

Even in situations where slot allocation is applied, queuing delays may be experienced by traffic on links upstream of bottlenecks. In the first place, we have seen that queuing may be necessary to enhance the available capacity of a bottleneck downstream. Furthermore, bottleneck capacity can only be maximized if the scheduling freedom is low, which may result in scheduled delays.

Travel time is inversely related with speed. In the previous sections we have seen that various types of capacity constraints are related to speed. In the first place, minimum headway is a function of speed. Assuming brick wall or prudent following strategies, this function is convex, which means that a certain finite speed value corresponds with minimum headways and hence maximum capacity. Furthermore, noise emissions are generally positively correlated with speed. Consequently, a lower speed and hence a lower quality-of-service may be required to attain a higher capacity, or a lower capacity has to be accepted to attain a higher quality-of-service.

Minimum and maximum speed or travel time may often have to be restricted by quality-of-service criteria, because travel time is probably a key variable determining the ac-

ceptability of alternative vehicle paths. Larger travel times result in higher transport costs, a lower quality for customers and hence lower revenues for the carrier. Specifying maximum travel time may be especially relevant given heterogeneous traffic. For instance, imposing travel time restrictions may be relevant for fast railway services operating in mixed traffic with slower trains, because else it might appear in later slot allocation stages that only slow slots are available. Furthermore, speed restrictions may result in the specification of minimum travel times. For instance, these speed restrictions may be due to vehicle specific legal or physical speed limits.

Given the feasible speed and travel time values, capacity constraint parameter values can be set in such a way that capacity is maximized. Main examples of parameters related with speed are minimum headways θ , travel time parameters π , interaction parameters χ , and environmental burden parameters β .

Scheduling freedom

As we have seen in the previous sections, capacity may depend on the sequence of vehicle paths, for instance in case of overtaking constraints. In this case, optimization of capacity values requires a reduction in scheduling freedom, because only certain sequences of vehicle paths result in a sufficiently high capacity to accommodate all selected slot requests. However, flexibility of sequence choice may be required to attain a sufficient reliability level. Furthermore, as will be discussed in later in this section, a certain minimum level of scheduling freedom is often required in situations where the bottleneck under consideration is part of a network that includes one or more other bottlenecks. Finally, a high level of scheduling freedom may be required to ensure that acceptable timetables may be designed, for instance from the point of view of efficient vehicle fleet utilization.

Carriers may desire that minimum quality-of-service criteria of selection slots are specified in such a way that an agreed maximum of vehicles is sufficient to operate the corresponding transport services. An optimum spread of departure times of different transport services may be required to optimize the usage of vehicles. For instance, an airplane may be used to fly between Düsseldorf and Stockholm in the morning and between Düsseldorf and Malta in the afternoon of the same day. Furthermore, a sufficient scheduling freedom may be required to ensure good connections between transport services. Offering opportunities to change between transport services is an important function of many larger railway stations and airports ('hubs'). When transport services have relatively low frequencies they usually arrive and depart in 'banks', i.e. concentrated periods of arrival and departure. A sufficient scheduling freedom may be required to ensure that all key connections have an acceptable quality.

The relationship between capacity and scheduling freedom is clearly negative. A relatively large freedom of sequence choice may be ensured by assuming a reasonable

probability of unfavorable sequences, resulting in relatively high interaction levels χ in the capacity constraints for unfavorable combinations of vehicle paths. Even more conservative is to assume a worst-case scenario. Furthermore, a larger scheduling freedom may be ensured by applying stronger (smaller) reduction factors φ . For instance, it may be required that every vehicle path should be scheduled within a certain margin of the desired timing. This approach would imply that the available service time is reduced to this period, assuming the worst-case situation that all vehicle paths have the same desired timing. However, this situation is not very likely and results in an unacceptable reduction of capacity, especially if the scheduling margins around the desired times are small. An alternative is to simulate potential timetables and derive the existence of feasible and acceptable solutions numerically, instead of formulating analytical capacity constraints.

Network dependencies

In a network context, scheduling freedom may be restricted by other bottlenecks in the same network if these are used by the same vehicle paths as the bottleneck under consideration. Although this thesis adopts a single bottleneck approach, we briefly discuss the adaptations of capacity values that may be required in a network context.

In the first place, other bottlenecks may also limit the possible sequences of vehicle paths in a situation with heterogeneous traffic, and an optimal sequence may not be possible for all bottlenecks in a network. This is especially an issue in rail traffic, where sequence changes between bottleneck railway links are desired, but often not possible, or would result in unacceptable travel times. For instance, a selection slot period of one hour has been specified for bottleneck nodes A, B, and C (figure 6.8). Traffic running between these nodes is of two different classes I and II. To avoid capacity loss due to interactions, the presence of traffic of these classes in the bottleneck nodes should be separated in time. However, traffic departing early from node C arrives late in nodes A and B. Consequently, maximum separation of traffic classes is not possible in this example.

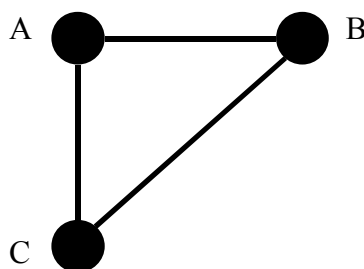


Figure 6.8: Three-node network illustrating network interactions

Another example of restricted scheduling freedom given the network of three bottleneck terminals A, B, and C (figure 6.8) is when the acceptable travel times are relatively large compared with the slot period. We assume that the occupancy of the single termi-

nal servers A, B, and C of departing and arriving transport services is 0.2 selection period. Within the selection period, transport services are desired between each pair of terminals. Consequently, each terminal facilitates 4 transport services, and the total occupancy of each terminal is 0.8. The acceptable travel times AB, BC and CA and *vice versa* are between 0.45 and 0.5 period. Hence, departing services have to leave before $t = 0.35$ and may arrive after $t = 0.65$, implying that in intermediary period no arriving and departing services can be scheduled. Clearly, the remaining 0.7 usable period is insufficient to accommodate the desired terminal occupancy of 0.8.

6.4.3 Determination of optimal capacity constraint parameter values

The previous sub-section focused on the determination of capacity constraint parameter values given the desire to find an optimal balance between capacity and quality-of-service. A related issue is that different types of capacity constraints attain their optimum at different values of a certain traffic variable. In this case, the optimal value of this traffic variable has to be found, and the related capacity constraint parameter values will be based on this value. This sub-section elaborates on the issue of finding the optimal parameter values of related capacity constraints. Related capacity constraints may be different types of capacity constraints applying to the same basic slots, different types of capacity constraints applying to different hierarchy levels, and capacity constraints corresponding with adjacent basic slot periods.

Traffic variables

Capacity constraint parameters may correspond with various *traffic variables*. For instance, the dependency of capacity on maximum speed has been discussed in sub-section 2.2.3. Other examples of traffic variables related with capacity are expected delays, level of mixture of user classes, and heterogeneity of speed. A relatively high capacity may correspond with relatively large expected delays, a relatively low level of mixture of user classes, and a low heterogeneity of speed. Because different types of capacity constraints may attain their optimal value at different values of a common traffic variable, the choice of capacity values of different types of capacity constraints or capacity constraints corresponding with different basic slots may be related.

Optimization of parameter values of different types of capacity constraints

Different types of capacity constraints may be dependent on the same traffic variable. However, maximum capacity values κ^{\max} may correspond with different values of this traffic value. A possible solution is to specify capacity constraints as a function of traffic variables, resulting in a relatively complicated formulation of the selection slot allocation problem. The approach proposed here, however, is to determine the optimum capacity values in advance. In principle, if the traffic variable under consideration is a quality-of-service variable, for instance speed, quality aspects should also be taken into account. However, the latter aspect is neglected here.

We focus on the situation where different types of capacity constraints apply to the same traffic selection market, with different associated optimum values of a common traffic variable. For instance, we have seen in the previous section that the application of continuous descent approaches (CDA) in air traffic results in less noise and hence a higher capacity if environmental capacity constraints are binding, but also in higher minimum headways and hence a lower capacity if flow capacity constraints are binding. Consequently, the specification of capacity constraint parameters is not only a question of finding a balance between capacity and quality-of-service, but also between different types of capacity constraints. Figure 6.9 illustrates possible capacity values given different percentages of landing flights applying a CDA. The relevance of this issue is illustrated by the fact that London Heathrow and Amsterdam Schiphol airports partly apply continuous descent approaches. At Heathrow, CDA is used for about 45% of all landings, while Schiphol applies CDA for night landings only (Kershaw *et al.*, 2000).

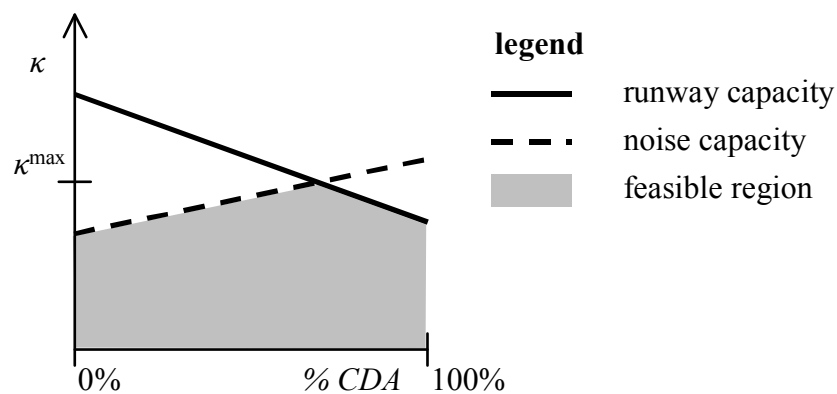


Figure 6.9: Hypothetical dependency of runway capacity and noise capacity on percentage of arriving flights applying CDA

The following approach is proposed here to determine the optimal parameter values of capacity constraints when two or more capacity constraints apply to the same bottleneck. The optimum parameter value is defined as the parameter value corresponding with the highest capacity value.

Capacity usually depends on composition-of-traffic, and therefore the first step is to make an assumption about composition-of-traffic. Secondly, the optimal feasible value of the traffic variable under consideration should be determined for each capacity constraint. Next, each capacity constraint has to be classified as binding or non-binding for its optimal value of the traffic variable under consideration. A capacity constraint is binding a certain value if the associated capacity value is lower than or equal to the associated capacity values of all other capacity constraints. If any of the capacity constraints is binding for its optimal value of the traffic variable, this is also the optimal value for the minimum of all capacity constraints. By definition, the optimum is reached for this value if this capacity constraint is binding, and where other capacity constraints are binding, this may only result in a further reduction of capacity values. Hence, the parameter values associated with this optimum value of the traffic variable should be

applied. It can be shown that, if none of the capacity constraints is binding in its optimum and assuming continuous capacity constraint functions, the optimum value of the minimum of all capacity constraints should either be one of the local optima of the capacity constraints, one of the intersections between capacity constraints, or the minimum or maximum value of traffic variable v . From these set of possible solutions, the value of v corresponding with the highest binding capacity value should be chosen. Therefore, if none of the capacity constraints is binding for its optimal value, the intersections of the capacity constraints as a function of the quality-of-service variable should be investigated as possible solutions, as well as local optima.

The situation where none of the capacity constraints is binding for its optimal value is illustrated by figure 6.10. The capacity constraint attaining the lowest capacity value is binding. Capacity constraint A attains its optimum at $v1$, and capacity constraint B at $v3$. However, capacity constraint B is binding at $v1$, and capacity constraint A at $v3$. The capacity constraints intersect at $v2$ and $v4$, but they have no local optima. Clearly, $v2$ is the optimal value of traffic variable v .

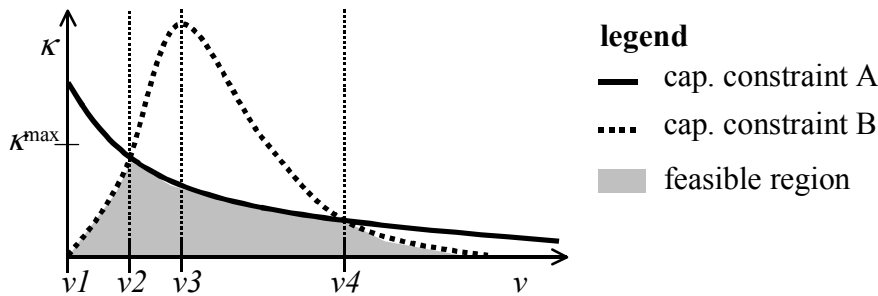


Figure 6.10: Capacity constraints as function of a traffic variable

Different hierarchy levels

A slightly different situation occurs when the related capacity constraints under consideration apply to different hierarchy levels, instead to basic selection slots at the same hierarchy level. For example, the capacity of a bottleneck may be restricted by an environmental capacity constraint that applies to annual totals (upper constraint), and runway capacity constraints applying to periods of one or two hours (lower constraints). However, the procedure that has to be followed to determine the optimal value of the traffic variable under consideration is similar to the situation without differences in hierarchy level. First, the optimum value of the traffic variable under consideration should be determined for the upper capacity constraint, and this value should be applied to determine the parameter values of both the upper and the lower capacity constraints. The next step is to check for which periods the lower capacity constraint is binding. Only for the slot requests in these periods, parameter values will have to be adapted. For these slot requests, the value of the traffic variable under consideration will be optimized to maximize the lower capacity constraints, at least if these are binding for these optimized values. If not, intersections between capacity constraints and local optima can be investigated as possible optimal solutions.

Balancing capacity values of adjacent selection slots

A similar problem as finding the optimal balance between related types of capacity constraints applying to the same selection slot period is to find the optimal balance between capacity values of adjacent selection slot periods. For instance, it is possible to compensate a higher probability of delays in one period by more buffer capacity in a subsequent period. This principle is, for instance, applied by airports such as Amsterdam Airport Schiphol, where 'fire breaks', periods with only a few scheduled arrivals and departures, are planned after peak departure or arrival periods. The determination of the optimal differentiation between slot periods of the balance between capacity and quality-of-service, however, is beyond the scope of this thesis.

A similar problem is the optimization of capacity values of different traffic network elements in a network. The capacity of traffic network elements may depend on the choice of capacity constraint parameters of adjacent traffic network elements. For instance, the optimization of the capacity of railway stations requires that access links can be used as buffers, which may have a negative influence on their flow capacity. However, since this thesis assumes a situation with a single bottleneck, a situation where two adjacent traffic service network elements are bottlenecks is not taken into account. If the railway station is the bottleneck under consideration, capacity constraint parameter values should be chosen in such a way that buffer links are not capacity-constrained. In general, it is advisable to maximize the capacity of traffic network elements where it is needed most, and therefore it is advisable to concentrate buffer and queuing processes in traffic network elements with excess capacity, and not in bottlenecks.

6.5 Examples of bottlenecks in current practice

So far, this chapter has focused on the identification of bottlenecks and the specification of capacity constraints in general. This section focuses on the occurrence of different types of bottlenecks in practice. The following sub-sections each cover a different transport infrastructure sector. For different transportation sectors, main bottlenecks are identified, focusing on the current situation in the Netherlands. For each type of bottleneck, an indication is given of the highest level of capacity scarcity that is attained in current practice. We focus on the question of whether or not the selection slot allocation problem applies, either because traffic supply is lower than traffic demand, or because traffic volume restrictions are desirable to attain a higher quality-of-service.

This overview of bottlenecks and capacity scarcity levels is based on a limited literature survey. An extensive survey of current infrastructure bottlenecks is beyond the scope of this thesis. To get an idea about the maximum level of capacity scarcity, data on maximum intensity to capacity ratios are used, and a few bottlenecks are analyzed as a case study. These case studies have been selected to represent the highest level of capacity scarcity currently attained in each sector.

6.5.1 Rail traffic bottlenecks

In the railway system, almost every infrastructure element may be a bottleneck. Main railway links are generally double-track, which means that one track is available for each direction, and consequently trains with different average speeds cannot overtake each other. Secondary railway links are frequently single-track, which means that traffic from opposite directions can only meet at a limited number of double-track locations. Furthermore, railway nodes and stations may also be bottlenecks, for instance because of crossing train paths, lack of platforms, or lack of reversing tracks.

As far as railways are used by a single carrier, capacity conflicts may be seen as an internal problem of the carrier. However, as was noted in section 3.1, railway networks are often shared by several carriers. Nonetheless, even on main European railway lines the number of competing carriers is often limited to one passenger carrier and a few freight carriers. Bottleneck capacity is often limited (e.g. 20 trains per hour per direction for a double-track railway link with homogeneous traffic), and the excess of traffic demand relative to traffic supply will usually be limited to a few trains per hour.

Single-track railway links

Single-track railway links are generally capacity-constrained, at least at a relative level. Given that traffic from opposite directions can only meet at a limited number of locations, the traffic scheduling possibilities are clearly limited. Furthermore, the capacity of a single-track railway link with bi-directional traffic is much less than half the capacity of a double-track railway link, which is due to the large capacity loss of changing the direction of operation (see sub-section 6.2.3). Hence, there may be situations where single-track railway links experience absolute capacity scarcity.

At some single-track railway links, less transport services are offered than desired by the carrier for capacity scarcity reasons. An example is the regional railway line between Nijmegen and Venlo in the Netherlands. Since 1996, the former half-hourly local service has been replaced by an hourly express service, an hourly local service between Nijmegen and Venray, and an hourly local service between Nijmegen and Boxmeer (see figure 6.11). The local train between Venray and Venlo has been replaced by a bus service, because it is not possible to extend the Nijmegen – Venray local service to Venlo given the current timetable setup. We can see in figure 6.11 that these associated paths would cross somewhere halfway Venray and Venlo. As is indicated below the path diagram, the Venray-Blerick link is single-track, while Blerick-Venlo is double track, and a passing track is available at Venray station, and hence trains can only pass at Venray station or between Blerick and Venlo.

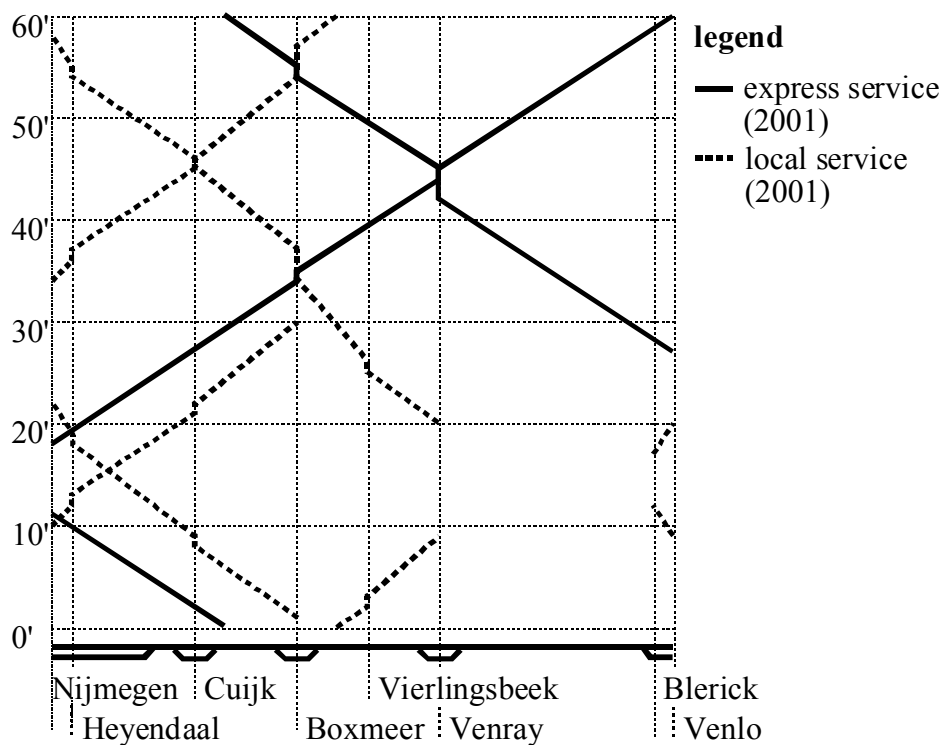


Figure 6.11: Timetable Nijmegen - Venlo (data: NS, 2001)

An alternative timetable setup would probably allow both a through local service between Nijmegen and Venlo and an express service (see figure 6.12). However, this would result in a longer travel times for the express service. Consequently, this is an example of relative capacity scarcity, where a reduction of traffic volume is desirable for quality-of-service reasons.

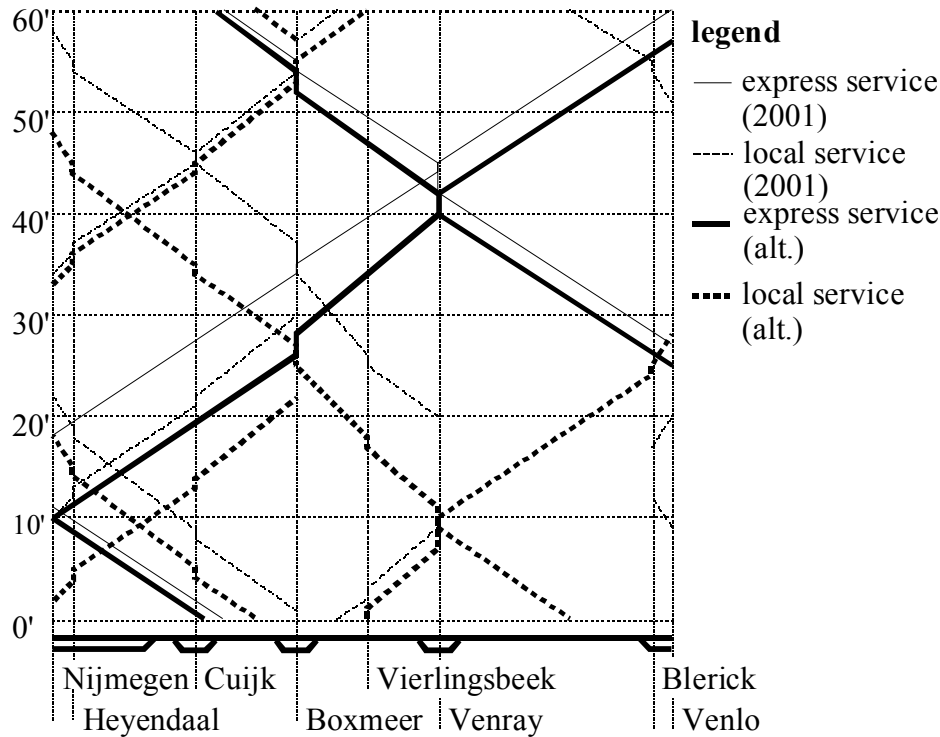


Figure 6.12: Alternative timetable Nijmegen - Venlo

Double-track railway links

Main railways are generally used by different types of transport services with different speed characteristics. Given double track railway links, trains can only overtake each other at a limited number of locations with side tracks, and consequently the scheduling possibilities are limited, especially given heterogeneity with respect to travel times (see figures 2.5 and 6.4). In some situations, traffic links can only allow less traffic than demanded by carriers, at least not without a significantly lower quality-of-service, resulting in less traffic than desired. For instance, only 3 instead of 4 local services per hour are offered on the mainline between Rotterdam and Den Haag during off-peak hours, giving additional room for freight trains (NS, 2005). An example of a railway link that was absolutely capacity constrained in the past is the Rotterdam – Dordrecht railway link. Shortly after the construction of additional tracks on this link in the late 1990's, enabling fast trains to overtake local passenger trains and freight trains, two additional express trains per hour and two additional local trains per hour were introduced in the new timetable.

Figure 6.13 shows which railway links in the Netherlands have been classified by ProRail (2004) as congested, i.e. capacity constrained in our terminology. According to the definition of Directive 2001/14/EC, railway infrastructure is congested if "it is not possible to satisfy requests for infrastructure capacity adequately", or if insufficient capacity is foreseen in the near future (EC, 2001c, art. 22). However, ProRail has not yet established a capacity analysis procedure to determine which infrastructure should be classified as 'congested'.

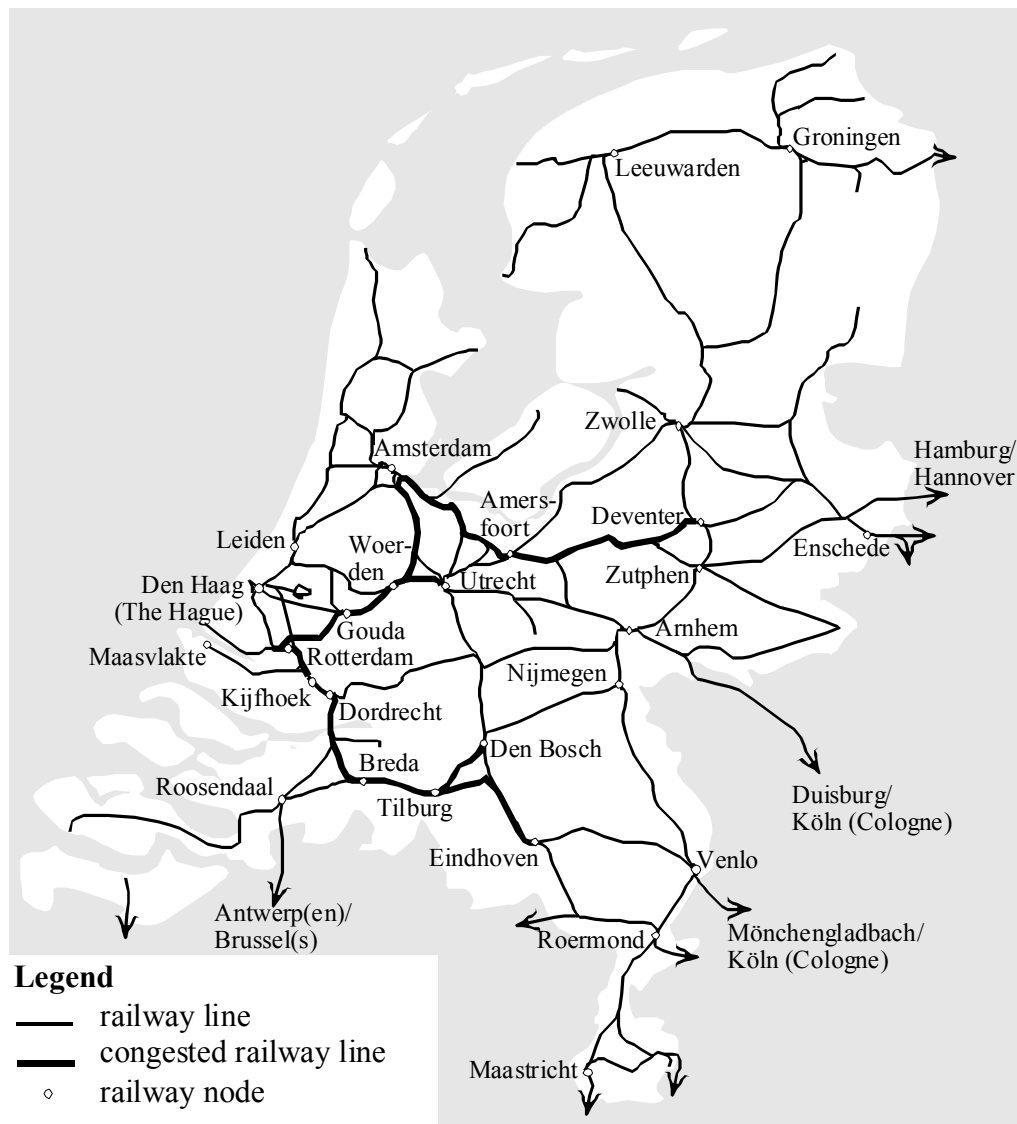


Figure 6.13: Congested and non-congested Dutch Railways (data: ProRail, 2004, appendix 22)

The railway links described as capacity-constrained in figure 6.13 may be either relatively or absolutely capacity-constrained. The level of capacity scarcity may increase in the next years. Railned expected in 2001 that in 2005 there would be several situations where traffic demand of passenger carriers and freight carriers will exceed capacity Railned (2001c). These capacity conflicts cannot be solved by rescheduling, and hence will have to result in either less freight trains or less passenger trains than desired. This situation was expected to occur on the following railway links:

- between Woerden and Utrecht;
- between Utrecht and Arnhem;
- between Den Haag and Rotterdam;
- between Dordrecht and Breda / Roosendaal;
- between Amsterdam and Amersfoort.

Railway nodes

Not only the links between nodes in railway networks but also the nodes themselves may be capacity-constrained. According to Dutch Railways (NS, 1988), increasing the capacity of railway links is commonly useless if the capacity of railway nodes is not increased as well. For instance, the number of platforms may be insufficient, or the number of reversing tracks. Reversing tracks or platform tracks separate from the main tracks are generally needed at the endpoints of railway lines. Although these are generally available where needed, or can be constructed given planned changes of the railway service network, there are examples of transport services that were not possible until a reversing track was built. Furthermore, platform capacity may be a bottleneck at major railway nodes. At major nodes in a passenger transport network, connections are offered between many different services. To offer short connections between all services, all trains should have their terminal stops simultaneously, implying that platforms and access tracks may easily be a bottleneck. However, no examples have been found of current situations where the introduction of new services is inhibited by platform capacity constraints, and hence there is probably little demand to apply selection slot allocation to railway stations.

Another possible bottleneck at railway nodes are interchanges and crossings. In particular, crossing and weaving train paths near main railway stations can be a bottleneck. In the Netherlands, for instance, one of the key measures resulting from the Rail 21 investment plan (NS, 1988) was the construction of a number of flyovers to increase the capacity of railway interchanges. These investments have enabled the introduction in the late 1990's of new express and intercity services, for instance the Rotterdam/Den Haag – Utrecht – Arnhem express service. However, the construction of flyovers to eliminate conflicts between train paths primarily aims at increasing punctuality and decreasing scheduled waiting times (V&W, 2001). A research of railway traffic in the Frankfurt area in Germany has revealed that train delays most often develop at railway nodes (Hermann, 1996). Consequently, railway interchanges may experience absolute capacity scarcity and it is likely that in the recent past, railway interchanges have been absolutely capacity constrained bottlenecks in the Netherlands. However, given the realization of the Rail 21 investment program, it is likely that currently most railway interchanges in the Netherlands are either only relatively capacity-constrained or not capacity-constrained.

Bridges

Bridges that frequently have to be opened for vessel traffic may be a bottleneck for railway traffic. In the first place, they may restrict the scheduling of railway traffic. For example, the opening regime of bridge over the river Oude Maas near Dordrecht in the Netherlands is an important restriction on the scheduling possibilities of train services on the intensively used Rotterdam - Dordrecht line. The loss of capacity for railway traffic may result in a situation of absolute scarcity. For example, plans to increase the

frequencies of suburban services between Purmerend and Amsterdam cannot be implemented given the desired opening frequency of the bridge over the Zaan waterway. Only a structural solution, for instance replacing the current bridge by a higher bridge, can solve this problem (Railned, 2001a).

External capacity constraints

In section 2.2 we noted that external capacity constraints may be imposed, for instance for environmental reasons. In the Netherlands, for instance, the stand-still principle is applied with respect to the safety of railway crossings, implying that any increase in traffic volume should be compensated by less or safer railway crossings, for instance, by closing at grade crossings or replacing crossings by tunnels (V&W, 1999b). Furthermore, noise regulations restrict noise emissions by railway traffic, and consequently may limit rail traffic volumes (Wet geluidhinder, 1979, art. 105; Besluit geluidhinder spoorwegen, 1999). As a result, environmental capacity is in many occasions more restrictive than physical capacity. According to a Railned study on traffic demand in 2005, less freight transport services than desired during night hours may be the result of noise regulation (Railned, 2001c). However, noise regulation may also limit (freight) train frequencies during the day. For instance, only 21 freight trains per day are allowed on the Arnhem - Zutphen railway line in the Netherlands when the Betuweroute, a dedicated freight railway line between the port of Rotterdam and Germany, is finished, which is an environmental capacity constraint (Railned, 2000a).

Demand for selection slot allocation

To conclude, situations of absolute capacity scarcity are likely to be present in the Dutch railway network, for instance on the Rotterdam – Den Haag railway line, and more situations of absolute capacity scarcity are expected to arise in the future. Furthermore, traffic volume limitations will be desired for quality-of-service reasons. Consequently, there are many bottlenecks where traffic demand exceeds traffic supply, which is the main characteristic of the selection slot allocation problem. These bottlenecks are mainly single track and double track railway links. Furthermore, it is indicated in Appendix A that railway traffic is dominantly associated with structural supply-driven transport services. As we have seen in section 2.5, selection slot allocation is a desirable slot allocation level for these types of transport services. Consequently, selection slot allocation is a desirable slot allocation level for the railway sector, at least in the Dutch situation with many bottlenecks.

6.5.2 Air traffic bottlenecks

In general, airports are the only bottlenecks for air traffic for which primary capacity constraints are binding. However, air traffic may also be constrained by the handling capacity of air traffic control centers and by environmental regulations. Airports and airspace are shared by many carriers, often with different nationalities, and often these

carriers are competitors. The daily number of flights facilitated by an airport may be up more than 1000, and air traffic control centers may handle similar numbers of flights each day.

Airports

Many airports apply slot allocation because their capacity is constrained, at least at a relative level. According to Vos (2002), various elements of airports may be bottlenecks, for instance runways, airplane parking places, and terminal gates. Furthermore, the handling capacity of passengers and luggage of the passenger terminal itself may be a bottleneck. Figure 6.14 illustrates the types of traffic network elements that constitute an airport, including terminals, parking places, and runways. Additionally, the handling capacity of air traffic control, emission norms, and night closures may limit the capacity of airports. However, we will first focus on the effects of primary capacity constraints.

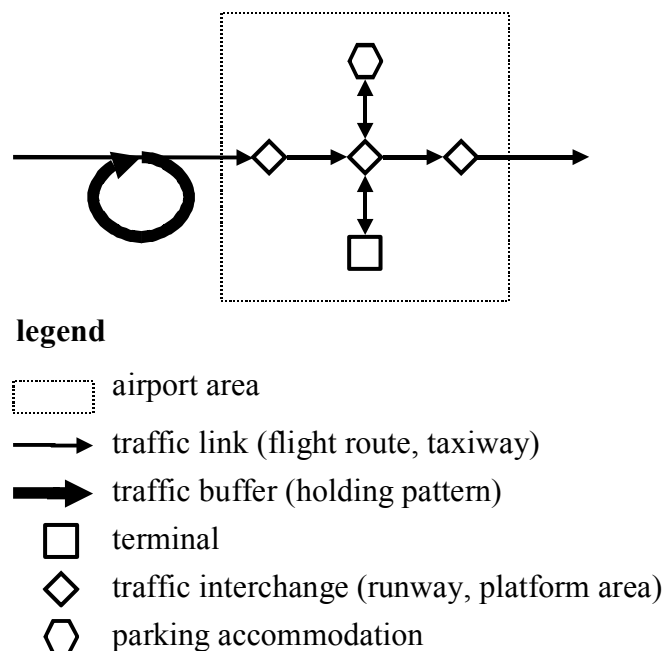


Figure 6.14: functional traffic network model of an airport

Examples of airports where physical capacity constraints are binding are large airports such as London Heathrow, Frankfurt Airport, and the New York airports La Guardia, Newark, and John F. Kennedy. However, some small Mediterranean island airports are also fully coordinated airports because of primary capacity constraints (Vos, 2002). At these fully coordinated airports (see sub-section 3.2.2), capacity scarcity is at least such that during peak hours the supply of slots is smaller than the demand.

Large airports such as London Heathrow, Frankfurt Airport, and Amsterdam Airport Schiphol are hubs (see section 4.3). However, the ideal of synchronous arrivals and departures is generally not feasible due to limited runway capacity. Furthermore, optimal synchronization of flights to provide good connections between air transport ser-

vices is often prohibited by a limited number of gates, similar to the limited platform capacity problem in the railway sector. With respect to Schiphol, for instance, the assumed maximum number of departures and arrivals may frequently be attained during these banks (see Appendix B). Consequently, physical capacity scarcity is limited to these banks.

At London Heathrow, however, primary capacity scarcity is not limited to peak hours. Slot availability data for two weeks in the (usually less congested) winter season reveal that only a few slots are available for *ad hoc* demands every week, and these slots are generally either in the late evenings on week days or in the late afternoon and evening on Saturdays (see Appendix C). Given that it is not reasonable to expect that all flights that cannot be accommodated in the mornings and afternoons can be shifted to the evenings, we may conclude that Heathrow faces absolute capacity scarcity.

Metropolitan airport systems

A different classification of capacity scarcity may be the result if we consider Gatwick and Heathrow as alternative international airports of the London metropolitan area. London Heathrow is severely capacity-constrained, but London Gatwick generally has available slots throughout the whole day (see Appendix C). Both airports offer an extensive network of European and intercontinental connections and therefore we may assume that Gatwick is a reasonable alternative for airline services that cannot obtain a slot at Heathrow. Furthermore, other airports in the London area (City, Luton, and Stansted airports) may be reasonable alternatives for flights with few transfer passengers or freight. Other metropolitan areas, such as Paris in France, Frankfurt in Germany and the Randstad in the Netherlands do not have alternative hubs, but also in these cases smaller airports in the region may be used as an alternative for flights that do not use the hub function of the main airport. For instance, the connections offered by hubs such as Amsterdam Airport Schiphol and Frankfurt International are not important to airlines focusing on leisure markets such as Transavia in the Netherlands and Air Berlin in Germany. Consequently, if there are acceptable non-congested alternatives for congested hubs for sufficient flights, capacity scarcity is classified as relative instead of absolute if both airports are considered as alternatives within the same selection slot for these flights.

However, shifting leisure transport from congested hubs to smaller airports will not be accepted unanimously as a solution to relieve hub congestion. Apart from possible resistance by carriers focusing on leisure markets, increasing traffic levels at smaller airports may be opposed to from an environmental perspective. For instance, in 1999 the Dutch Minister of Housing, Spatial Planning and the Environment stated that a growth of leisure air traffic at regional airports is undesirable. According to the Minister, regional airports should specialize on general aviation and business transport; lei-

sure transport should be concentrated on the Dutch national airport Schiphol (Van Asbeck & Oranje, 1999).

Air traffic control centers

The capacity of air traffic is generally not limited by minimum headways, because conflicts of air traffic can usually be solved by assigning different altitudes to different flights. However, as we have seen in section 2.2, air traffic is restricted by the handling capacity of air traffic control centers. The volume of European air traffic is currently bound by the handling capacity of traffic control centers (Majumdar & Polak, 2001). According to Eurocontrol data, over 80% of European air traffic delays in 1999 were due to inadequate ATC capacity. In that year, more than 30% of all flights had a delay of 15 minutes or more (EP, 2001). However, no indications have been found that these congestion levels are unacceptable to air carriers to an extent that structurally desired flights are cancelled for congestion reasons. Consequently, the capacity scarcity of European airways is relative, not absolute. However, cancellations for congestion reasons do occur in situations where air traffic is restricted by adverse weather conditions, strikes, etc.

Air traffic follows air corridors, using fixed beacons for navigation (Nolan, 1994). In particular the nodes of this network of air corridors may be bottlenecks, given that conflicts between flight paths are likely in situations where several air corridors intersect. According to Magill (1998), less frequent conflicts between vehicle paths could be realized if aircraft could freely choose direct routes instead of following air corridors. Consequently, the capacity of air traffic would increase, or at least the workload of air traffic centers would decrease. Direct routing is technically possible by using area navigation and satellite navigation systems instead of point-to-point navigation using fixed beacons.

External capacity constraints

External capacity constraints may also limit air traffic volumes. A typical example in Europe is Amsterdam Airport Schiphol. The main reason that Schiphol has become a fully coordinated airport (see sub-section 3.2.2) is to limit the total annual number of departures and arrivals to such a level that noise regulations are respected (Besluit slotallocatie, 1997; Stout, 2001). However, runway capacity constraints and airspace capacity constraints also limit the capacity of Schiphol (Kamphuis, 1998). The latter capacity constraints are only binding during peak hours, while noise emissions limit the 24-hour capacity of Schiphol. The fluctuation of departure and arrival rates at Schiphol indicates that the flow capacity of Schiphol is not restrictive at a 24-hour basis, but is restricted to a few 'banks' (see Appendix B).

In 1998, the externally imposed limit on the total annual number of arrivals and departures at Schiphol airport was 360 000, given 420 000 desired flights (Stout, 2001). On

the other hand, all slot requests (regarding the day period) for the winter season 2002-2003 have been accepted (data: ATCN, 2002). However, air traffic levels are generally lower in the winter season, and traffic demand has decreased due to economic recession. Furthermore, less noisy planes, optimized flight routes, and less strict noise criteria have allowed an increase in the total number of flights in the last years. In 2002, the declared capacity of Schiphol had increased to 456 700 arrivals and departures (Schiphol, 2001). Furthermore, a fifth runway has been opened in 2003, which does not only expand the runway capacity but also the environmental capacity of Schiphol, since this new runway allows a shift of flight patterns, allowing more flights given similar noise limits. To conclude, Schiphol experienced absolute capacity scarcity in the past for both the day and night period, and may again experience absolute capacity scarcity in the future if traffic demand continues to grow. Furthermore, runway capacity may be binding during arrival and departure banks, which are characteristic for the hub function of Schiphol, resulting in a large percentage of transfer passengers.

Demand for selection slot allocation

In situations where environmental constraints limit the yearly total number of landings and departures of airports, there is clearly a demand for selection slot allocation, at least if the transport services using it have fixed transport service patterns. The latter is the case, since structural supply-driven transport services are dominant in this sector (see Appendix A). Furthermore, airports may experience absolute capacity scarcity, either only during peak hours or throughout the day, and hence selection slot allocation may also be applied to these periods. Peak hours at hubs are usually associated with arrival and departure banks. However, although capacity scarcity due to limited air traffic control handling capacity appears to be responsible for about 80% of air traffic delays in Europe, no indications have been found that it currently attains such levels that selection slot allocation is desired. Nevertheless, scheduling slot allocation for congested air sectors may be desirable given these high levels of air traffic delays. Furthermore, the need for slot allocation 'in the air' may increase as air traffic intensities continue to increase.

6.5.3 Road traffic bottlenecks

Car traffic congestion is a common phenomenon in many countries. The most commonly experienced type of car traffic congestion is congestion on freeways in urbanized areas. Common freeway bottlenecks are interchanges, and bottleneck links, which frequently coincide with river bridges and tunnels. Furthermore, the connections between freeways penetrating larger cities and urban road networks may be bottlenecks. However, main roads in urban areas may also be congested; main bottlenecks of urban road traffic are road crossings. For instance, congestion of main urban roads frequently occurs in large cities such as Paris and London. Yet another type of urban car traffic bottleneck is parking. City and town centers in Europe often face a large parking demand

during shopping hours, especially at Saturdays. Often, this problem is tackled by a combination of supply measures such as building parking garages, and traffic market measures such as imposing high parking fees during shopping hours.

Car traffic congestion is often restricted to a few peak hours. For instance, according to data of the Dutch Ministry of Transport (AVV, 1997), more than 70 percent of registered queues on Dutch freeways in 1993 occurred between 7 and 9 a.m. and 4 and 6 p.m. Scarcity of parking places, however, is generally not restricted to a few peak hours, but in many occasions it is restricted in time. For instance, inner city parking scarcity may be restricted to Saturdays due to shopping activities, and parking scarcity near beaches may be restricted to sunny days during the summer holidays and summer weekends.

A characteristic of car traffic is that many actors are involved, i.e. many independent car drivers. Furthermore, bottleneck capacities may be relatively large compared with the other sectors. Freeway capacity may be more than 2000 vehicles per lane per hour (TRB, 2000), and urban city centers may have a car park capacity of thousands of vehicles.

Freeways

The main bottlenecks of freeway traffic networks are merges, including merges at freeway interchanges and on-ramps. Freeway capacity is limited by merging and weaving processes in particular (TRB, 2002). Another problem is that congestion may result in blocking back effects, i.e. queues may block vehicle paths that do not pass the bottleneck, for instance because they leave the freeway before the bottleneck (recall figure 3.1). This would have been avoided if sufficient buffer capacity had been offered, for instance by increasing the number of lanes upstream of the bottleneck (Broeren & Westland, 1998).

Studies evaluating the effect of capacity expansion in situations of structural freeway congestion suggest that the total traffic volume may increase with a few percent, while peak hour traffic volumes usually increase much more due to "back to peak" effects. For example, in 1990 the Northern part of the Amsterdam orbital freeway A10 was opened, offering an alternative to the congested Western part of the A10. A comparison of car traffic patterns before and after the opening revealed that car traffic increased with 3%, which is only 1% more than the autonomous growth of car traffic in the period of study (HCG, 1992). This small increase was largely due to a shift from car passengers to car drivers. The increase in car traffic after the reduction of congestion levels, albeit relatively small, suggests that the capacity scarcity was absolute in the previous situation. However, it is also possible that in many situations traffic reductions due to capacity scarcity would not be necessary given an optimized allocation of the available peak hour capacity. Hence, it is not easy to classify road and freeway capacity scarcity as absolute or relative. However, it is clear that congestion may attain such levels that

traffic demand is reduced, for instance due to substitution of transport modes or substitution of destinations.

Parking places

Capacity scarcity of parking places has received much less attention in the scientific literature than roadway congestion. Nonetheless, there are many examples of parking capacity scarcity in practice, especially in urban areas. To reduce parking demand, parking fees are levied in many cities. In the first place, price differentiation may help to optimize the spatial distribution of parking place usage (Young, 2001). Secondly, there may be an absolute shortage of parking places of sufficient quality. Given that near inner city destinations there may not be sufficient parking places to accommodate all traffic demand, and given the high parking fees that are often imposed to reduce parking demand, many potential car drivers to inner city destinations will choose alternative modes such as biking or public transport, or choose alternative destinations. Consequently, we may assume that many inner cities face absolute parking capacity scarcity.

Demand for selection slot allocation

We have seen that road capacity may be absolutely capacity-constrained, especially during peak-hours. Main bottlenecks in freeway networks are merges at intersections, on-ramps, and links with fewer lanes. Furthermore, car park capacity may be capacity-constrained throughout the day. Hence, there are clearly situations of where capacity scarcity attains sufficient levels to demand for selection slot allocation from this point of view. However, as we have seen in section 3.1, slot allocation is currently not applied to road traffic networks, although proposals have been made in the literature to apply slot allocation to parking places or freeways. One of the possible reasons not to apply slot allocation in this sector is the different composition-of-traffic with respect to transport service types compared with rail and air traffic. Road traffic is dominantly associated with private transport, which is much more flexible than supply-driven transport services. Hence, even if slot allocation would be applied in this sector, selection slot allocation would probably only be desired to facilitate (structural) public transport services.

6.5.4 Navigation bottlenecks

In contrast to the other sectors reviewed in this section, few scarcity problems are currently known in the navigation sector. In general, the inland navigation sector has no flow capacity problems. According to Kreutzberger & Vleugel (1992), the peak hour intensity to capacity (I/C) ratios of Dutch canals and rivers in 1990, describing the ratio of flow capacity that is used, were commonly less than 0.1 and never much higher than that. The only potential bottlenecks of inland navigation networks are locks and ports. Typically, even large locks may facilitate just a few ships simultaneously, and only in larger ports the total number of available berths at sea terminals is more than 20. For instance, the average number of sea ships daily arriving at the port of Rotterdam, until

recently the largest port of the world, is about 10 (Havenbedrijf Rotterdam, 2004). Similarly, vessel traffic will not meet congestion problems at open sea, but locks and ports may be bottlenecks. Furthermore, navigation support (by pilots) is required for some coastal areas due to frequent changes in the location of sand plates. Hence, traffic control is a potential bottleneck for vessel traffic, just as appeared to be the case with air traffic.

Locks

Locks may be classified as relatively capacity-constrained bottlenecks if expected waiting times exceed a certain threshold value, which may already be the case due to random delays in situations where traffic volumes are still well below capacity. Lock capacity is determined by a number of aspects of the locking process, including the time required for the locking process (including entry and exit of locks), the number of lock chambers, and the maximum load of lock chambers. A small single-chamber lock may have a capacity of less than 4 ships per hour (Koehler, 1994). However, although the capacity of locks may be low, there are currently no indications of severe congestion of locks, at least not in the Netherlands. According to Kreuzberger & Vleugel, the most intensively used lock in 1990 was the Oranjesluis near Amsterdam with an I/C ratio of 0.66. Higher levels of capacity scarcity have been avoided by timely investments in the navigation network.

Bridges

Although many bridges have an opening regime giving priority to vessels, some bridges have a limited opening regime. For instance, railway bridges usually have fixed opening times. Bridges which have to be opened for regular commercial traffic, such as the Zaan railway bridge in Zaandam and the Margrietkanaal railway bridge in Grou, usually have opening frequencies of at least twice per hour in order to avoid significant delays for vessel traffic (AVV, 2003). In principle, vessel traffic has priority over other traffic, which means that a significant reduction of opening frequencies to allow for more railway or road traffic is often not considered as an acceptable alternative. Consequently, it is unlikely that bridge opening regimes will result in significant capacity scarcity for vessels.

Ports

Ports may be bottlenecks, especially the number of berths available at sea terminals. Container transport networks increasingly have a hub-and-spoke character, which means that large container ships operate between a few hub ports, while smaller short-sea vessels operate between these hubs and feeder ports. In these hubs, terminals with specialized equipment are required (Thomas, 2001). Given that sea traffic is increasingly concentrated on these main ports, and given the cost associated with expanding the number of berths at sea terminals, it is reasonable to expect that sea terminal capac-

ity will be scarce. On the other hand, the competition between ports such as Antwerp, Rotterdam and Hamburg has resulted in much pressure to adjust port capacity in such a way that sea traffic congestion is avoided. According to the Rotterdam port authorities, container terminal capacity in Rotterdam and in competing ports is currently more than sufficient (Schuttevaer, 2001).

Traffic control

Finally, the capacity of ports and waterways may be restricted due traffic control capacity constraints. This may be the case in situations where the assistance of pilots is required for navigation in waterways near ports. For instance, the port of Antwerp is connected with the North Sea by the Westerschelde, an estuary with frequently changing depths. Therefore, sea ships are accompanied by a pilot between the North Sea and Antwerp. Sea ships may be delayed in situations when there are not sufficient pilots available. However, no reports in the literature and the media have been found about structural delays due to a lack of available pilots, except during strikes or in situations of adverse weather conditions inhibiting bringing and collecting pilots from ships on the North Sea.

Demand for selection slot allocation

In this section, no examples have been identified of absolute capacity scarcity. In principle, only random delays will occur near relative bottlenecks such as locks, bridges, and terminals. At these bottlenecks, there may be demand for scheduling slot allocation for supply-driven transport services. For instance, we have already seen in section 3.1 that many container sea transport services run according to a fixed timetable. However, given the low levels of capacity scarcity, there will be no demand for selection slot allocation in this sector.

6.6 Conclusions

In this chapter, we focused on traffic supply. An overview has been given of types of traffic network elements, and the capacity constraints associated with different types of primary traffic processes, traffic service processes, and traffic externalities have been reviewed. Furthermore, an overview has been provided of types of bottlenecks in current practice.

Capacity constraints

Different types of bottlenecks may be distinguished that are associated with different types of capacity constraints. In this chapter, an overview has been given of capacity constraints due to primary traffic processes, traffic services processes, and traffic externalities. It appears that various traffic processes and traffic service processes can be

modeled as linear capacity constraints or even a simple maximum total. This includes homogeneous following processes, buffering, parking, switching, and also some environmental capacity constraints. Given linear capacity constraints, each individual vehicle path has its own capacity parameter. However, in situations with interactions between user classes, additional parameters have to be included with composition-of-traffic dependent values. Hence, these parameter values depend on the selection slot allocation decision that is to be determined. Interactions between user classes frequently appear given heterogeneous following processes, overtaking, crossing, and traffic control processes.

Capacity constraint parameters are often related to one or several quality-of-service variables. There are several examples of situations where a high capacity can only be attained given a relatively low quality-of-service and vice versa. Based on previous discussions on quality-of-service in chapters 2, three important quality-of-service variables have been identified and discussed: reliability, speed/travel time, and scheduling freedom. Strict quality-of-service criteria with respect reliability and a high scheduling freedom result in a relatively low capacity. The same may be the case with speed, depending on the applicable speed-flow relationship.

The determination of quality-of-service and capacity constraint parameter values is the responsibility of allocation bodies, but analyzing the relationship between capacity and quality-of-service may help to make these decisions. Furthermore, different types of capacity constraints may apply to the same bottleneck, and the optimum parameter values of these different constraints may correspond with different values of the same traffic variable. In this case, optimum parameter values may be determined by analyzing which capacity constraint is binding given which values of the traffic variable under consideration. The approach to find this optimum, both in situations with identical and multi-level slot specification, has been presented in this chapter.

The approach to specify capacity constraints introduced in this chapter applies to various types of primary traffic processes, traffic service processes, and traffic externalities. Furthermore, it can be applied irrespective of the sector under consideration. This overview of capacity constraint types is largely based on sector-specific references on the dependency of capacity on various traffic variables. However, the integral generic approach adopted in this chapter is claimed to be new.

It can be concluded from the successful attempts in this chapter to formulate analytical capacity constraints that traffic processes, traffic service processes, and traffic externalities known as possible bottlenecks can be captured in linear capacity constraints, or linear capacity constraints with a composition-of-traffic dependent interaction term. A complication that has only been discussed briefly in this chapter is when the single bottleneck assumption is relaxed to allow for a more complex network context. The solution suggested in section 6.4 is to estimate the required degree of scheduling free-

dom and to optimize capacity constraint parameters in a network context. The issue how to extend the single bottleneck approach to network problems will be elaborated on in section 7.5.

Bottlenecks in current practice

In section 6.5, an overview has been given of bottlenecks in current practice. It appears that there are many differences in traffic bottlenecks between sectors and between locations. In the first place, this chapter shows that there are not only many different causes of infrastructure capacity scarcity in theory, but also in practice. Different types of traffic processes appear to be the main bottlenecks in different sectors, and within sectors there may be geographical differences. Secondly, there are clearly differences in levels of capacity scarcity between sectors and between bottleneck types. Examples of bottleneck types that are often capacity-constrained to such extent that traffic volume restrictions are required or desired are single and double track railway links, airports (large hubs), freeway merges, and urban parking accommodations. Consequently, capacity scarcity in the rail traffic, air traffic, and road traffic sectors may often be sufficient to result in demand for selection slot allocation, at least by structural transport services. However, the selection slot allocation problem appears not to apply to the navigation sector, because of the relatively low level of capacity scarcity in that sector. Nonetheless, scheduling slot allocation may be applied in the latter sector, for instance to optimize the utilization of container terminals.

Next chapter

The specification of capacity constraint types in this chapter is an important step in the formalization of the selection slot allocation problem. The selection slot allocation decision problem is the focus of the next chapter. First, it focuses on traffic demand and on the valuation of alternative slot allocation decisions. Next, the selection slot allocation decision problem is formalized mathematically, and types of selection slot allocation problems are formulated. The classification of problem types is not only dependent on type of objective function, but also on type of capacity constraint. Finally, solution procedures for the selection slot allocation problem will be discussed.

7

SELECTION SLOT ALLOCATION DECISION PROBLEMS

In the previous chapter, the focus was on traffic supply. This chapter continues with the specification of the selection slot allocation decision problem, requiring a shift of focus to traffic demand and to the valuation of alternative slot allocations by the allocation body. The first section of this chapter discusses the main characteristics of rational slot allocation decision-making. A key step of rational decision-making is the specification of objectives, which is the focus of the second section. This thesis adopts a normative perspective, which implies that the decision-maker provides the objectives of slot allocation. Nonetheless, since we deal with transport infrastructure, we may assume that the objectives will be related to general public objectives such as welfare optimization. The (primary) objectives will be formalized in an objective function, but it appears that some secondary objectives, for instance promotion of competition between carriers, may best be implemented as satisficing constraints. It is concluded that in most cases, a linear objective function may be adopted to represent the primary objectives. The optimum selection slot allocation problem (selection problem) with a linear objective function is formalized in the third section. Four problem types are distinguished, ranging from an optimization problem with a linear objective function and a single linear capacity constraint to fixed-point problems where the coefficients of the (linear) objective function or the (linear) capacity constraints are dependent on the selection decision itself. The optimization problems can be classified as binary programming problems, which may be solved by exact optimization algorithms. However, a greedy-type approximation algorithm is preferred here, mainly to avoid some undesirable characteristics of exact solutions. Based on this greedy algorithm, a general solution procedure is formalized in the fourth section that can be applied to most formulated problem in-

stances. An extension to the basic solution procedure is proposed to enable the implementation of satisficing constraints. In section 7.5, the implementation of satisficing constraints regarding minimum claim rights and minimum level-of-competition is discussed in detail. Furthermore, the possibilities to extend the general single-bottleneck approach to network problems are reviewed in this section. The implementation of the selection slot allocation decision approach proposed in this chapter is exemplified by the application to a hypothetical case in chapter 8.

7.1 Rational decision-making applied to selection slot allocation

This chapter focuses on selection slot allocation as a decision problem, i.e. the optimal selection of slot requests. In this chapter the selection slot allocation is simply referred to as the selection problem. The selection problem is studied here from a normative rational perspective. The main arguments for a normative rational perspective have already been given in section 1.2. This section elaborates on the main characteristics of rational decision-making, including a basic outline of rational decision processes. Based on this general outline of rational decision processes, the basic structure of a rational selection slot allocation decision process is provided.

7.1.1 Characteristics of rational decision-making

In this section, the main characteristics of rational decision-making processes are outlined. As was stated in chapter 1, principles of rational decision-making are adopted in this thesis to design a decision-making process for selection problems. This section begins with a brief overview of the main characteristics of rational decision-making. The characteristics of rationality adopted here have been derived from various sources. The interested reader may check these references for a discussion about differences between the ideas adopted by these authors and alternative points of view.

Alternative space and set of feasible alternatives

According to Simon (1955), the alternative space and the set of feasible alternatives is one of basic ingredients of rational choice. The alternative space is the realm of alternatives that is available to and considered by the decision-maker. This may be smaller than the theoretical realm of potential alternatives. The dimensions of the alternative space are determined by the decision variables recognized by the decision-maker. The set of (considered) feasible alternatives is again a sub-set of the set of the alternative space, with feasibility constraints specifying which alternatives are feasible and which are not (see figure 7.1). The alternative space and feasibility constraints should be well-defined, but it is not necessary for each alternative to be specified explicitly.

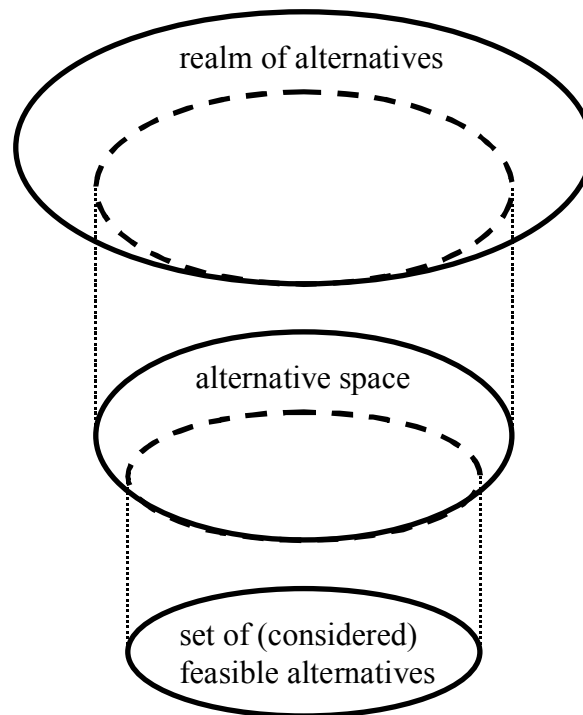


Figure 7.1: Alternative space and set of feasible alternatives

Objectives and preferences

Another element of rational decision-making is the explicit specification of the relationship between the alternative chosen and the degree of satisfaction of the decision-maker (Simon, 1955). The preferences of the decision-maker with respect to alternatives are assumed to reflect this degree of satisfaction. According to Lancaster (1966), the preferences of decision-makers are not directly related to alternatives, but to relevant attributes of alternatives. Hence, instead of attempting to measure the preferences of decision-makers with respect to all alternatives, it is sufficient to establish (analytical) relationships between key characteristics of alternatives and the expected degree of satisfaction of the decision-maker, provided that there is information about the characteristics of alternatives. This relationship is determined by the decision-makers' objectives, and commonly an objective function is used to represent this relationship, at least when operations research is used to support rational decision-making (see e.g. Hillier & Lieberman, 1990). The selection problem specified later in this chapter also adopts an analytical objective function. The specification of objectives and objective functions will be discussed in more detail in the next section.

The characteristics of alternatives may be deterministic and known, but often the characteristics of alternatives will be partially uncertain. According to French (1986), information about alternatives is generally not complete. For instance, probabilities are estimated and not known exactly. According to French, rational decisions may not only be based on 'hard' information, but also on 'soft' or incomplete information, or even on *a priori* assumptions. Nonetheless, empirical information, if available, should largely overrule *a priori* views in rational decision-making. In this thesis, it is assumed that the

objective function regarding slot allocation is largely based on relatively 'hard' information. Uncertainties about the effects of alternative slot allocation decisions may be included in the objective function, for instance adopting the approach for dealing with uncertainty in rational decision making described by French (1986).

Solution procedure

The final step of rational decision-making is the selection of the preferred alternative. A common assumption in classical economics about rational decision-makers ('economic man') is that they always choose the alternative resulting in the highest level of satisfaction to them (Simon, 1955). However, determining which alternative is the best is far from trivial for many real-life problems, and rational decision-makers may consider the potential benefits of finding a better alternative not sufficient to justify the efforts of finding the best alternative. As an alternative to the optimizing approach, Simon proposes satisficing. Satisficing decision procedures do not aim to find the best solution, but a satisfactory solution.

A related aspect is the choice of the method to find an optimal solution in an efficient way. Specification, enumeration, and evaluation of all alternatives is usually not necessary, because alternatives are generated implicitly in the solution procedure. Furthermore, most solution procedures iteratively seek better alternatives until a final (feasible) solution is found. Finally, approximate solution algorithms may be applied as computationally less burdensome alternatives to exact optimization algorithms. Approximation algorithms are not guaranteed to result in the best solution, but many of these can be guaranteed to result in a solution that is close to optimal.

7.1.2 Example of rational slot allocation decision process

Given the characteristics of rational decision-making outlined above, a rational selection slot allocation decision process may be designed. Figure 7.2 outlines the proposed structure of a rational selection slot allocation decision process. This decision process consists of five steps, of which the definition of the alternative space is the first. Given that the decision variable is selection of slot requests, the alternative space corresponds with all possible alternative selection decisions. The set of feasible alternatives is determined next by specifying capacity constraints. The third step is the specification of the objective function, defining the value of each alternative selection slot allocation decision. The next step is to choose and apply a solution procedure to evaluate the alternatives and to find a solution to the slot allocation problem. The final step is to make the slot allocation decision, based on the results of the evaluation procedure.

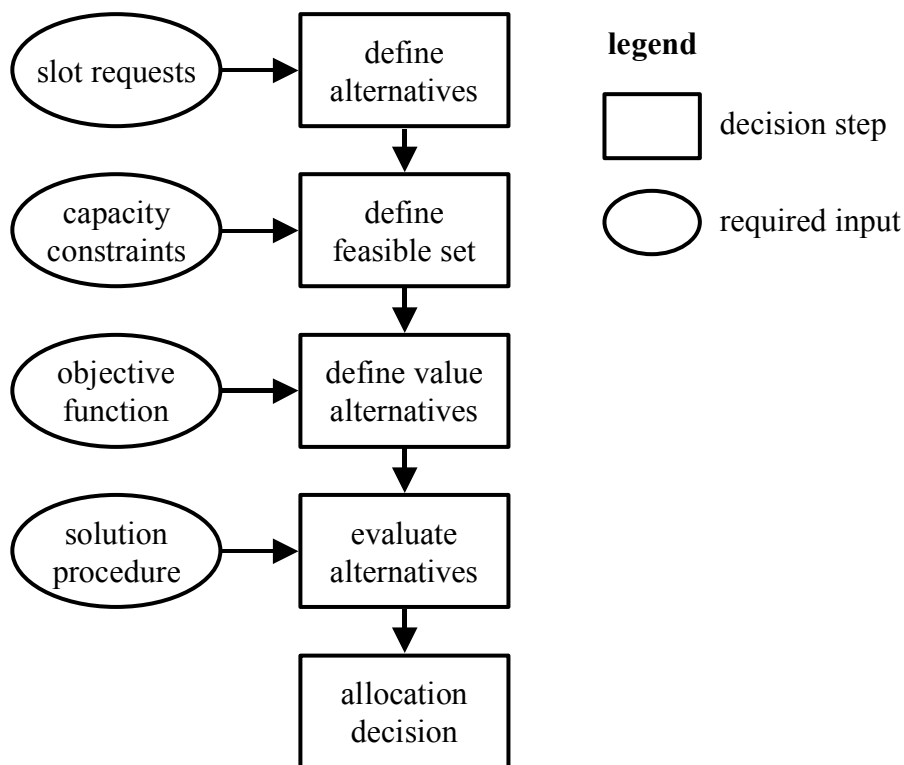


Figure 7.2: Rational selection slot allocation decision process

Alternative space and set of feasible alternatives

In the selection problem, the decision variable is the acceptance or rejection of given slot requests. Every possible combination of selected slot requests is an alternative selection. Which alternatives are feasible is determined by capacity constraints. As we have seen in chapter 6, capacity constraints are not only determined by the nature of traffic supply, but also on quality-of-service (q.o.s.) criteria and the specified slot periods. The specification of the latter two aspects is based on the acceptability principle formulated in section 4.3. This principle implies that the minimum quality-of-service is determined by the acceptance by the carrier of alternative vehicle paths. Vehicle paths with an acceptable quality will be used by the carrier, although the quality may be lower than desired. A non-acceptable quality implies that a vehicle path of that quality will not be used, resulting in the cancellation of the corresponding transport service.

Objective function

A key aspect of rational decision-making is that alternative slot allocation decisions should be valued in order to define which alternative is preferred. Values are assigned to alternative selections given the objectives of the decision-maker. The decision-maker is the allocation body responsible for the selection slot allocation decisions under consideration. In this chapter, an optimization approach is proposed, which implies that an objective function is used to model preferences as a function of the characteristics of alternative selections. The identification and the formulation of the objective function is the main subject of section 7.2.

Solution procedure

Given a set of slot requests, the decision needs to be made which requests are accepted and which are rejected. Given the application of an exact optimization procedure, the feasible alternative with the highest value will be chosen. However, exact optimization may be a complex solution, and therefore other decision procedures may be preferred. In section 7.3, the selection problem is formalized as an optimization problem, and section 7.4 discusses alternative solution procedures that may be applied to the selection problem. Both exact and approximate solution procedures are elaborated in this section. Furthermore, the possibility to include satisficing constraints will be reviewed in section 7.5. Satisficing constraints are applied to select the satisfactory alternatives, before applying the exact or approximate optimization procedure to the set of feasible satisfactory alternatives.

7.2 Specification of slot allocation objectives

As we have seen in the previous section, rational decision-making consists of various steps, including the specification of feasible alternatives, specification of objectives, and specification of the decision procedure. In the context of the selection problem, the specification of feasible alternatives is the specification of traffic supply, which was the main subject of the previous chapter. This section continues with a discussion of objectives, while the formulation of a decision procedure will be discussed in section 7.3. In this section, we first review a few basic assumptions about objective formulation from a normative perspective. Next, a number of likely types of primary and secondary objectives are discussed. Stated objectives of European Union transport policy are used as a basis for this discussion. Finally, we focus on the translation of objectives to objective functions.

7.2.1 Assumptions about objective formulation

As we have seen in chapter 1, this thesis follows a normative perspective. This section starts with reviewing the implications of this normative perspective on the formulation of objectives. The normative perspective implies that, in principle, decision-makers may have any objectives. Nonetheless, a number of possible types of objectives are reviewed in this section. The main reason to review possible objectives is that the specification of a generic solution procedure for the selection problem requires assumptions about the nature of the objective function. Furthermore, a few suggestions are provided in this section how the value of alternative selections may be determined, which is relevant to the design of slot allocation systems. The elaboration of possible objectives of selection slot allocation in this section is based on the assumption that these objectives are formulated from a public perspective, which implies that these are related to the general objective of welfare maximization.

The main implication of a normative, rather than an empirical perspective is that our focus is on how slot allocation decisions should be made, not how they are made. The choice of objective function (or other preference model) used to value alternative slot allocations is a normative decision. In principle, the decision-maker may have any preference function, which determines the value of every alternative. According to the conceptual model of slot allocation outlined in sub-section 2.1.4, the decision-maker is the allocation body, but the slot allocation regime, including the objectives to be pursued and the procedures to be followed, are usually determined by the authorities. Assuming that a public or semi-public body is responsible for transport infrastructure slot allocation, rather than a private company pursuing its own private goals, the objectives with respect to slot allocation will be determined by objectives of public policy, and not by the personal preferences of decision-makers. According to Faludi (1973), objectives of public policy may be a combination of objectives provided by the decision-maker, without being subject to discussion by decision analysts, and empirical revealed or stated preference information. The preferences of all involved actors may be included in the objective function, including the interests of carriers and end users. Additionally, the interests of others, including environmental aspects, may be included in the objective function.

Welfare theory

Welfare theory may be used as a basis for deriving public preferences. The main idea behind welfare theory is that collective (public) preferences may be derived by aggregating individual preferences, using willingness-to-pay as a measure of strength of preference (see e.g. Boadway & Bruce, 1984). Consumer welfare stems from the difference between what consumers are willing to pay for products and services and the actual price they have to pay. Similarly, producer welfare is the difference between revenues and costs.

According to Just *et al.* (1982), the standard willingness-to-pay based method used to aggregate individual preferences is in theory just one alternative approach to aggregate individual preferences, because there is no unique objective way to aggregate individual preferences to social preferences. Furthermore, a complete analysis of all preferences of all relevant individuals is often not possible, if only because preference data are not available for all individuals. Consequently, alternative methods may be used to formulate public preferences, and other preference data may be used than willingness-to-pay. The discussion about alternative methods to aggregate preferences is beyond the scope of this thesis, but in the remainder of this chapter the assumption will be made that the objective function will include the costs and benefits experienced by carriers and end users.

Assumptions about values

Assuming welfare theory as a basis for the derivation of public objective functions, we may assume a number of characteristics of these objective functions. In the first place, the valuation of alternative slot allocation decisions will be based on the resulting transport service patterns; slot allocation decisions have no intrinsic value. Carriers do not directly derive value from the slots they possess, but from the difference between costs and benefits of offering transport services. Similarly, end users are assumed to derive value from the transport opportunities facilitated by the carriers' supply of transport services. Consequently, we assume that the value of alternative selections depends on the resulting transport service pattern, including the logistic aspects of producing this transport service pattern. Additionally, the value of alternative selections may depend on external issues, such as expected environmental effects.

Since the valuation of alternative selections depends on slot usage, a well-defined relationship between allocation decisions and slot usage is required to evaluate selection slot allocation decisions. Recalling the conceptual model of slot allocation outlined in sub-section 2.1.4, slot allocation is an interaction between allocation bodies and carriers. Carriers adapt their supply of transport services given slot allocation decisions, and allocation bodies base their decisions on slot requests by carriers. Furthermore, the assumption has been made in section 4.3 that selection slot requests have no alternatives, and hence no alternative transport service will replace the transport service corresponding with a rejected slot request. Additionally, we assume that slot requests truly represent the planned transport services of carriers, and that if slot requests are accepted, then the corresponding transport services will be realized.

7.2.2 Primary objectives of slot allocation

The assumption that the objectives of allocation bodies are related to welfare maximization is especially true for primary objectives, which are objectives directly related to the value derived from slot usage. Broadly speaking, primary objectives focus on the optimum utilization of infrastructure. Secondary objectives, which will be reviewed in the next sub-section, may also be welfare related, but focus on other issues, such as the effects of slot allocation decisions on competition in the transport market or the environmental effects of resulting traffic flows. Although a comprehensive review of possible objective functions is beyond the scope of this thesis, a number of possible primary objectives and secondary objectives are reviewed in this sub-section.

In sub-section 2.4.1, maximizing infrastructure utilization has been proposed as the primary objective of slot allocation, which is a reasonable primary objective given that slot allocation is applied in situations of absolute infrastructure capacity scarcity (see section 5.2). Transport infrastructure facilities are generally expensive, and one of the objectives of European policy with respect to slot allocation is to attain optimal utilization of infrastructure in order to reduce the costs of transport to society (EC, 2001c).

The main problem, however, is how the value of utilization of slots, given alternative selections, can be determined. Apart from the question of which variables are used as an approximation of value, it may be questioned how these variables are to be measured. The following variables are likely candidates to be used as a proxy for the value associated with alternative selections:

- willingness-to-pay;
- total costs;
- transport volumes.

Willingness-to-pay

Willingness-to-pay of a carrier regarding an alternative selection can be used as a measure of the contribution to welfare of allocating the slot to this carrier. According to Boadway & Bruce (1984), willingness-to-pay is a good measure given the absence of significant externalities. However, Borenstein (1988) argues that willingness-to-pay is not a very good measure to compare slot requests that are related to transport services in different markets (see also section 5.3). For instance, consumer surplus may differ significantly between types of transport services, especially when some services operate in a competitive transport market, while others operate in a (regulated) monopolistic transport market.

The willingness-to-pay can best be measured in an auction procedure. This is equivalent to the application of auction techniques to measure scarcity tolls discussed in subsection 5.3.2. Second-price auctions are recommended in most references on slot auctions, given that profit maximizing bidders are stimulated to reveal their true willingness-to-pay given second-price auctions (Vickrey, 1961).

Total costs

Minimization of total costs or total travel time is frequently used as objective function for transport network optimization studies (Berechman, 1993). Furthermore, Van Nes (2000) concludes that, given a number of assumptions about transport supply and demand, maximizing welfare and minimizing total costs yield similar outcomes for the optimal public transport network design problem. Similarly, minimization of total costs may be used as objective for selection slot allocation. Total costs include transport costs and generalized travel time experienced by end users, as well as the maintenance and operating costs of carriers. Using value-of-time parameters, travel time may be translated into monetary values. Marginal infrastructure maintenance costs may also be included in the analysis.

The estimation of travel time costs of end users is based on the assumption that transport demand is fixed and given; elasticity of transport demand is neglected. This assumption is justified because transport capacity remains unchanged. However, not only

information about transport demand is required to estimate the total costs associated with alternative selections. First, the transport service network resulting from a selection slot allocation decision has to be determined, and then the resulting transport patterns have to be estimated. Furthermore, data on traffic costs of carriers have to be acquired or estimated. The large data requirements and the uncertainty about the correctness of the estimated costs values are a clear disadvantage of this approach, compared with the usage of willingness-to-pay information.

Transport volumes

An alternative to estimating the differences in total costs resulting from alternative selections is to use (expected) transport volumes as a proxy for value of utilization. The implicit assumption is that each unit of transport has the same value. The definition of the unit of transport is a critical issue in this case. In situations with only passenger transport, each individual passenger may be valued equally. In case of mixed transport, weights will have to be attached to passengers and units of (different types of) freight. A relatively 'objective' way to determine equal units of transport is to take total transport capacity as reference. For instance, if the transport capacity of a link is 5000 persons per hour or 500 containers, then 1 container is valued equivalent to 10 persons.

The transport volumes resulting from alternative selection slot allocation decisions may be estimated using a transport model. A simpler alternative is to use estimations of required vehicle capacity for each transport service. However, the latter data have to be supplied by carriers, which may be tempted to overestimate expected transport volumes in order to increase their chances of obtaining slots. Alternatively, historical data about transport volumes and efficiency of capacity usage may be used, as has been proposed by Railion (2001) to be applied to determine distribution of railway capacity among passenger and freight transport services.

7.2.3 Secondary objectives of slot allocation

Secondary objectives of slot allocation are not directly related to the optimal utilization of infrastructure. It appears that the following secondary objectives play a dominant role in the motivation of European slot allocation policy:

- minimal negative external effects of traffic;
- sufficient competition in transport markets;
- sufficient supply of transport services.

As we have seen in sub-section 6.3.3, traffic is often associated with *negative external effects*, and it is a public task to keep these negative effects to reasonable levels. Apart from imposing limits on hazards, hindrance, and pollution, minimization of these negative externalities of traffic may be included in the objective functions. For instance, carriers using relatively silent aircraft may be granted priority over carriers using rela-

tively loud aircraft in the allocation of airport slots. Protecting the quality of the environment is one of the main official objectives of the European Communities:

"The Community shall have as its task ... to promote throughout the Community a harmonious, balanced and sustainable development of economic activities, ... a high level of protection and improvement of the quality of the environment ..."
(Treaty of Rome, 1967, art. 2)

Sufficient competition in the transport market is also an EU objective. In the last decades, EU policies have moved forward from enabling equal market access to actively promoting market access in order to stimulate competition. For instance, Regulation 95/93 states with respect to airports that

"... it is Community policy to facilitate competition and to encourage entrance into the market ... and ... these objectives require strong support for carriers who intend to start operations on intra-Community routes." (Preamble of Regulation 95/93, EC, 1993)

The main idea is that stimulating competition between carriers would result in a diverse supply of transport services with a high quality-to-cost ratio. Direct competition between carriers on the transport market seems to be preferred by the European Union. However, when direct competition between carriers is not economically feasible or not efficient, the preferred alternative is competition between carriers for concessions.

Finally, the objective of having a *sufficient supply of transport services* may be relevant for transport services that are not very profitable to the carriers, but nonetheless desirable from a societal point of view. This objective is often a derivative of other objectives. For instance, a certain minimum supply of connections with peripheral regions may be desired to support regional economic objectives, or a certain minimum supply of suburban transport services may be desired to relieve urban car congestion. The desirability of a sufficient supply of (air) transport services to certain regions is explicitly recognized in Regulation 2408/92:

"It is necessary to make special provisions ... for public service obligations necessary for the maintenance of adequate air services to national regions"
(Preamble of Regulation 2408/92, EC, 1992b).

Conflicting objectives

External slot allocation is only required if several carriers are competing for capacity on the same infrastructure. Consequently, conflicts between the objectives of carriers are inherent to slot allocation. However, the objectives of authorities may also be in conflict, or may be in conflict with the objectives of carriers. A number of these conflicts arising in practice have already been reviewed in section 3.3. Especially secondary objectives may be in conflict with the primary objective. For instance, economies of scope may be achieved if lines are operated by a single carrier, resulting in a higher

welfare level compared with a situation where lines are operated by different carriers. On the other hand, a secondary objective is to promote competition in the transport market, which requires alternative transport services to be operated by different carriers. For instance, having two morning flights between Amsterdam and Lyon separated by less than a half hour, one operated by KLM and one by Air France, is good for competition, but is not likely an example of efficient usage of scarce infrastructure capacity.

7.2.4 Specification of objective functions for the selection problem

Regardless of the chosen type of objective, a relationship has to be established between the transport service pattern corresponding with an alternative selection slot allocation and its net value. Given our choice for an optimizing approach, the objectives identified in the previous sub-sections have to be formalized as objective functions. Table 7.1 gives an overview of the symbols used in this chapter in objective functions and capacity constraints. A full overview of all symbols used in this thesis is provided after the list of references.

Table 7.1: Symbols used in objective functions and capacity constraints

symbol	type	name
h	variable (binary)	indicator of binding constraints
i	index	slot request
j	index	basic slot
k	index	capacity constraint
n	variable (integer)	number of slot requests
R	set	set of slot requests
S	set	set of selected slot requests
u	variable	total value
x	variable (binary)	selection variable
\mathbf{x}	vector (binary)	selection variable vector
α	parameter	weight in capacity constraint
\mathbf{a}	vector	weight in capacity constraint vector
\mathbf{A}	matrix	weight in capacity constraint matrix
κ	parameter	capacity parameter
\mathbf{k}	vector	capacity parameter vector
ω	parameter	value parameter
$\mathbf{\omega}$	vector	value parameter vector

An important issue is whether objective functions are additive, because this determines the type of optimization problem under consideration and hence the type of solution procedure that can be applied. In the case of additive objective functions, total value u is the sum of value parameters ω of selected slot requests, with the set of slot requests being denoted by R . The selection vector \mathbf{x} is the decision variable, indicating for each

slot request i whether it has been accepted ($x_i = 1$) or rejected ($x_i = 0$). In vector notation, total value is the matrix product of the transposed (column)vector $\boldsymbol{\omega}$ of value parameters ω_i and the (column)vector \mathbf{x} of selection variables x_i :

$$u(\mathbf{x}) = \boldsymbol{\omega}' \cdot \mathbf{x} = \sum_{i \in R} \omega_i \cdot x_i \quad \mathbf{x} \in \{0,1\}^n \quad (7.1)$$

However, more complex objective functions may be required to incorporate complement and substitute relationships between the transport services associated with the evaluated slot requests. This section discusses to what extent these relationships exist and whether nonlinear objective functions are required for a realistic representation of values associated with alternative selection slot allocations. In this situation, total value u may still be regarded as a sum of individual values of selected slot requests, but in this case the individual values are not parameters but dependent on the chosen alternative selection \mathbf{x} :

$$u(\mathbf{x}) = \boldsymbol{\omega}(\mathbf{x})' \cdot \mathbf{x} = \sum_{i \in R} \omega_i(\mathbf{x}) \cdot x_i \quad \mathbf{x} \in \{0,1\}^n \quad (7.2)$$

Equations 7.1 and 7.2 show the derivation of total value u . The objective function is to find a selection slot allocation \mathbf{x}^* maximizing total value u . As will be discussed in more detail in the next sub-section, a selection slot allocation decision \mathbf{x} has to be chosen which is part of the set Ω of feasible alternatives. In both equations, matrix multiplication has been applied. Equation 7.3 formalizes the generic objective function:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \Omega}{\text{Argmax}} \{u(\mathbf{x})\} \quad \mathbf{x} \in \{0,1\}^n \quad (7.3)$$

Additivity of production costs and user benefits

Assuming a welfare-based objective, the net value associated with a transport service pattern depends on both the production costs made by carriers and the benefits derived from transport by end users. Traffic dependent infrastructure (maintenance) costs are neglected here.

The costs made by carriers to produce transport services will be largely additive, i.e. the costs of producing a transport service are largely independent of the selection of other transport services to be operated. Each transport service requires energy and the usage of vehicles and staff, irrespective of the operation of other transport services. Significant economies of scale are not likely, given that there are few indications that larger transport networks result in significantly lower overhead costs per unit of production. However, in a number of situations economies of scope or density may result in lower production costs for certain combinations of transport services. Economies of scope may be realized by certain combinations of transport services, if joint production of transport services results in a more efficient usage of vehicles and staff than separate production (Button, 2001). For instance, the costs per run of operating a single-direction transport service from A to B will be more than half the costs per run of operating this service in

both directions. Furthermore, increasing the frequency on a certain line may allow shorter reversing times at the endpoints. Economies of density may occur when higher service frequencies result in lower costs for each run. According to Berechman (1993), empirical studies of economics of density with respect to bus and rail transport reveal a substantial amount of economies of traffic density.

The benefits derived from transport services will only be partly additive. Transport services may be (partial) substitutes, since end users may choose between different destinations, different connections, and different runs as alternatives for the same trip. Different runs of the same transport service line are often acceptable alternatives for a large percentage of transport demand, but similar types of destinations, especially holiday destinations, may also be substitutes for a relatively large percentage of transport demand. If two transport services are (partial) substitutes, the benefits derived from operating both transport services will be lower than the sum of benefits of operating one transport service or the other. This effect will usually compensate the economies of scope and density with regard to production costs.

Another aspect is that sequences of transport services may together constitute a single trip alternative, and consequently combinations of transport services of the same or different carriers may be complementary. However, given reasonable assumptions about the importance of connections offered at main terminals, there is probably no need for an explicit inclusion of complement relationships between pairs of transport services in the objective function. Instead, we may consider the value derived from good connections at a certain terminal to be an attribute of this terminal, neglecting the effect on connectivity of individual transport services.

From the discussion above, we may conclude that the total value of an alternative selection is basically the sum of individual values of slot requests. The most likely potential exception is when slot requests correspond with different runs of the same transport service. In this case, there is a strong case for non-additivity of both production costs and end-user benefits. However, a linear objective function may still be used in this case if a priority of runs is determined in advance. In this case, the value of individual slot requests is based on the assumption that 'higher level' runs have been selected, and 'lower level' runs have not been selected. Given these considerations, a linear objective function is used in the basic formulation of the selection problem presented in section 7.3, and an alternative formulation is proposed using a composition-of-traffic dependent linear approximation of the objective function.

Formalization of secondary objectives

Secondary objectives are not based on welfare maximization but are specific policy objectives. For instance, minimization of environmental burden may be (part of) the objective function. Similar to the linear environmental capacity constraint introduced in

sub-section 6.3.3, we may model environmental objectives as an additive objective function. However, the other two secondary objectives are not formulated in terms of minimization or maximization, but in terms of sufficient levels, and therefore an objective function formulation may not be the best solution. Furthermore, a non-linear objective function would be the result, which is not desirable from a practical point of view. Instead of including aspects such as level-of-competition and minimum supply of transport services in the objective function, this thesis proposes to implement secondary objectives regarding competition as satisficing constraints (see section 7.5).

7.3 Specification of optimum selection problems

In this section, the selection problem is formalized as an optimization problem. In this section, the generic selection problem is specified, but also a number of specific problem types. The main differences between these problem types are the assumptions about the type of objective function and the type of capacity constraints. Given different assumptions about these aspects, a different formulation of the optimum selection problem will be the result. Table 7.1 (sub-section 7.2.4) provides an overview of symbols used in this section.

In the first sub-section, the generic selection problem is specified. The generic selection problem encompasses all 4 problem types. However, satisficing constraints are not yet included in the problem formulation, this will be discussed as an extension to the generic selection problem in section 7.5. In the next sub-sections, four specific problem types are specified, each corresponding with a different level of complexity. Problem type 1 assumes a linear objective function and a single linear capacity constraint. The objective function and capacity constraints are also linear in type-2 problems, but now the number of capacity constraints is greater than one because multi-level slot specification is assumed instead of discrete slot specification. However, capacity constraints are still of the same type. The latter assumption is relaxed in type-3 problems, i.e. two or more capacity constraints (of different types) may apply to the same basic slot. Finally, a fixed-point formulation of the selection problem is introduced for type-4 problems to include a feedback mechanism regarding the assumed parameter values that are dependent on composition-of-traffic.

7.3.1 Specification of the generic selection problem

As defined earlier, selection slot allocation is the selection of slot requests with the purpose of determining which vehicle paths can be accommodated given capacity constraints and acceptability criteria (sub-section 2.4.4). The optimal selection slot allocation problem (selection problem), which is specified in this section, encompasses the core issue of slot allocation, i.e. the question of which slot requests should be selected and which should be discarded in situations where traffic supply exceeds traffic demand. Consequently, the decision variable is the selection vector \mathbf{x} , which indicates for

each slot request i whether it has been accepted ($x_i = 1$) or rejected ($x_i = 0$). The selection problem considers a single bottleneck of predefined longitudinal and transversal size (see sub-section 2.4.3).

Traffic demand

The demand for selection slots is assumed to be known. Traffic demand is denoted by set of slot requests R . Each slot request is identified by a unique number i . Slot requests are issued by prospective slot holders (mainly carriers, see sub-section 4.1.2), and the selection problem is based on these slot requests. Each slot request applies to a specific basic slot. We assume the set of slot requests R to be fixed and given, i.e. slot requests are independent of the selection or non-selection of any other service. For instance, the possibility to submit an alternative slot request if a slot request is rejected is not included in this analysis. Hence, if a slot request is rejected then the corresponding planned transport service will have to be cancelled. This is in accordance with the acceptability principle introduced in section 4.3 as a basis to determine the desired size of selection slots. The number of slot requests is denoted n , and hence the selection vector \mathbf{x} has n elements.

Traffic supply

The selection problem considers a single bottleneck with predefined basic slots. These basic slots have been specified by the allocation body according to the principles outlined in section 4.3. Either discrete or multi-level slot specification may be applied, and multi-level slot specification is assumed in this sub-section in order to obtain a complete overview of the generic selection problem. Basic slots at any level are defined by the primary slot period(s) and the traffic server(s) it includes. Each basic slot is identified by a unique number j .

The selection vector \mathbf{x} is the decision variable, with elements x_i of which the value equals 1 if i is selected and 0 otherwise. The alternative space (see section 7.1) corresponds with the vector space of the selection vector \mathbf{x} , i.e. it is an n -dimensional space of the binary selection variable x . The set of considered feasible alternatives is restricted by capacity constraints. As proposed in section 6.1, a standardized capacity is derived by choosing a slot request or user class as a reference. The generic capacity constraint formula (sub-section 6.1.2) specifies that capacity occupancy g as a function of \mathbf{x} should not exceed capacity κ :

$$g(\mathbf{x}) \leq \kappa \quad \mathbf{x} \in \{0,1\}^n \quad (6.3)$$

However, selection problems usually include more than one capacity constraint. Each basic slot j corresponds with at least one separate capacity constraint. Furthermore, different capacity constraints (of different types) may apply simultaneously to the same basic slot. Each different capacity constraint, regardless whether it corresponds with a

different basic slot j or also with a different type of capacity constraint, is identified by a unique number k .

Basic slots may correspond with any level of a multi-level slot specification. Each selection slot is included in the capacity constraint(s) of its basic slot, and in the capacity constraints of corresponding higher-level basic slots. In this context, a corresponding higher-level basic slot is a basic slot of which the basic slot associated with the slot request under consideration is a subset.

The total set of capacity constraints is formalized in equation 7.4. A capacity vector $\boldsymbol{\kappa}$ is introduced as a vector of all capacity values κ_k , and \mathbf{g} is used instead of g to indicate that the function under consideration results in a vector instead of a single value:

$$\mathbf{g}(\mathbf{x}) \leq \boldsymbol{\kappa} \quad \mathbf{x} \in \{0,1\}^n \quad (7.4)$$

Optimization problem

The final step to formulate the selection problem as an optimization problem is the specification of the objective function. The objective is to determine the optimal selection vector \mathbf{x}^* indicating for each slot request whether or not it is included in the selection. The optimum allocation should yield an optimum value u , given the set of considered feasible alternatives, which is restricted by capacity constraints. The generic optimum selection problem can be formulated as the following binary programming problem:

$$\begin{aligned} &\text{Maximize } u(\mathbf{x}) && \mathbf{x} \in \{0,1\}^n \\ &\text{with} && \\ &\mathbf{g}(\mathbf{x}) \leq \boldsymbol{\kappa} && \end{aligned} \quad (7.5)$$

In this thesis, a shorter notation is used. In equation 7.6, the objective function is notated before the colon, and the specification of the feasible set is notated behind the colon:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{Argmax}} \{u(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq \boldsymbol{\kappa}\} \quad \mathbf{x} \in \{0,1\}^n \quad (7.6)$$

The generic selection problem specified above describes main characteristics of the selection problem, without discussing possible simplifications of the problem. Since the generic selection problem is the least restrictive of all, all assumptions mentioned here also apply to problem types 1 to 4 specified in the next sub-sections.

7.3.2 Type 1: linear objective, single basic slot, single linear constraint

The formulation of the selection problem in this section is based on the assumption that either there are simply no substantive reasons to include interaction terms in the objective function or the capacity constraints, or that the composition-of-traffic is largely

known in advance, which means that the dependency of the objective function or the capacity constraints on composition-of-traffic can be neglected. With respect to the latter argument, we may recall that given the assumption of a semi-static selection slot allocation process as was proposed in section 4.2, a significant percentage of traffic is already determined by earlier selection slot allocation decisions. Furthermore, the application of minimum claim rights (see section 7.5) may further reduce the uncertainty about composition-of-traffic, which reduces the demand for non-linear objective functions or capacity constraints even further. The impact of interactions reduces roughly proportionally with the reduction of percentage of traffic corresponding with slot requests that are currently evaluated. Assuming that the composition-of-traffic is largely known, capacity constraints with interaction may be simplified to linear capacity constraints, and a linear objective function may be used.

In section 4.3, two alternative approaches have been introduced to specify selection slots, i.e. discrete and multi-level slot specification. Assuming discrete slot specification, the optimum selection slot allocation period may be formulated for each period separately. Furthermore, we assume in this case that only a single capacity constraint applies to the bottleneck under consideration. Although we have seen in section 4.3 that a multi-level slot specification will be desirable in many situations, this may not be necessary when, for instance, selection slot allocation applies to peak periods only.

The optimum selection problem applying in this case is formalized by equation 7.7. The capacity of the period under consideration is denoted by κ , while α is a vector indicating the weight of each slot request in the capacity constraint (see sub-section 6.1.2). A linear objective function is formulated similar to equation 7.1, with ω denoting the vector of value parameters. The single-period optimum selection problem with a linear objective function and a single linear capacity constraint can be formalized as the following binary knapsack problem, which is a sub-class of binary programming problems (see Martello & Toth, 1990):

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{Argmax}} \{u = \omega' \cdot \mathbf{x} : \alpha' \cdot \mathbf{x} \leq \kappa\} \quad \mathbf{x} \in \{0,1\}^n \quad (7.7)$$

7.3.3 Type 2: linear objective, multiple basic slots, single linear constraint

In section 4.3, the possibility of a multi-level slot specification has been introduced. In general, a multi-level slot specification is required if considerable differences in demanded selection slot size exist, for instance due to differences in frequency of transport services. Given a multi-level slot specification, several periods need to be analyzed simultaneously, and hence several capacity constraints apply to the same selection problem.

Just as in type-1 problems, both the capacity constraints and the objective function are linear in type 2 selection problems. Furthermore, it is assumed that only one capacity

constraint applies for each period at each hierarchy level, and these capacity constraints are based on the same traffic process, traffic service process, or traffic externalities. Assuming that in many situations only a single type of capacity constraint will be restrictive, this problem type will apply to many situations in practice.

To analyze the selection problem for several basic slots simultaneously, a matrix of weight coefficients $\alpha_{i,j}$ is required instead of a vector, with j identifying different basic slots, as well as a vector of standardized capacity values instead of a single value. The resulting problem can be classified as a set packing problem, which is a subclass of the family of binary programming problems (see e.g. De Vries & Vohra, 2003). In equation 7.8, \mathbf{A} is the matrix of weight coefficients $\alpha_{i,j}$ and $\boldsymbol{\kappa}$ is the vector of standardized capacity values for each basic slot κ_j , with each period (regardless of slot level) being identified by a different index j :

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{Argmax}} \{u = \boldsymbol{\omega}' \cdot \mathbf{x} : \mathbf{A} \cdot \mathbf{x} \leq \boldsymbol{\kappa}\} \quad \mathbf{x} \in \{0,1\}^n \quad (7.8)$$

The weight coefficients $\alpha_{i,j}$ are 0 if capacity constraint t does not apply to slot request i , for instance because they correspond with different slot periods. Slot request i has a positive weight $\alpha_{i,j}$ for each basic slot j it is included in. Since in type 2 selection problems all capacity constraints are all of the same type, the weight coefficients $\alpha_{i,j}$ of any slot request i are identical for every basic slot j it is included in.

7.3.4 Type 3: linear objective, multiple linear constraints

In the third problem type, the assumption that only one type of capacity constraint applies is relaxed. Multiple linear capacity constraints may apply to either situations with discrete or multi-level slot specification. For instance, this problem type applies if primary capacity constraints are combined with environmental capacity constraints, as is currently the case at Amsterdam Airport Schiphol. Furthermore, as we have seen in chapter 6, different types of capacity constraints may apply to railway nodes simultaneously. As will be shown in the next section, solution approaches that can be used for the previous problem type may not apply to the situation where different types of capacity constraints are involved, and therefore two separate problem types are distinguished here.

The formulation of the optimum selection problem is similar to the previous problem type (equation 7.8). Hence, the type-3 selection problem can also be classified as a set packing problem. A difference with problem type 2, of course, is that two or more capacity constraints may apply to the same basic slots, and hence different weight coefficients $\alpha_{i,k}$ may apply to the same slot request i . As was proposed in sub-section 7.3.1, index k is used as a unique indicator of each different combination of basic slot j and capacity constraint. However, because the weight coefficients $\alpha_{i,k}$ have the same inter-

pretation as the weight coefficients $\alpha_{i,j}$, the same symbol \mathbf{A} is used for the matrix of weight coefficients, and hence the optimum selection problem is equal to equation 7.8:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{Argmax}} \{u = \boldsymbol{\omega}' \cdot \mathbf{x} : \mathbf{A} \cdot \mathbf{x} \leq \boldsymbol{\kappa}\} \quad \mathbf{x} \in \{0,1\}^n \quad (7.9)$$

7.3.5 Type 4: non-linear objective or constraints

In the previous problem types, we assumed that the capacity constraints and objective function are linear, and that the parameters which values depend on composition-of-traffic are sufficiently adequate estimates. In most situations, this will be an adequate assumption. As we have seen in section 7.2, linear objective functions will do in most situations, and it has been concluded in section 6.4 that in most situations linear capacity constraints may be applied. However, if necessary it is possible to explicitly take into account the dependency of the value vector \mathbf{u} and the capacity constraint coefficients $\alpha_{i,j}$ in the formulation of the optimum selection problem. For instance, a non-linear objective function may be required to adequately substitute and complement relationships between slot requests (as discussed in section 7.2), or a non-linear capacity constraint may be required to describe heterogeneous traffic on a double-track railway link where weight coefficients α are relatively sensitive to the selection decision, because within each basic slot each slot request corresponds with a different type of transport service.

By making the value vector \mathbf{u} dependent on the solution of the selection problem \mathbf{x}^* , a feedback mechanism is introduced. The resulting selection problem is a fixed-point problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{Argmax}} \{u = \boldsymbol{\omega}(\mathbf{x}^*)' \cdot \mathbf{x} : \mathbf{A} \cdot \mathbf{x} \leq \boldsymbol{\kappa}\} \quad \mathbf{x} \in \{0,1\}^n \quad (7.10)$$

An additional feedback is introduced by making the capacity constraint coefficients also dependent on the solution \mathbf{x}^* :

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{Argmax}} \{u = \boldsymbol{\omega}(\mathbf{x}^*)' \cdot \mathbf{x} : \mathbf{A}(\mathbf{x}^*) \cdot \mathbf{x} \leq \boldsymbol{\kappa}\} \quad \mathbf{x} \in \{0,1\}^n \quad (7.11)$$

The next section elaborates on possible exact solution methods that may be applied to the optimum selection problems formulated here. Furthermore, an iterative solution approach will be suggested to allow for the fixed-point problem formulations of equations 7.10 and 7.11.

7.4 Solution procedures

In the next sub-sections, exact and approximate solution approaches are formulated that can be used to solve different instances of the selection problem specified in the previous section. The solutions produced by exact optimization algorithms are truly optimal,

i.e. there are no better solutions of the specified constrained optimization problem. Sub-section 7.4.1 evaluates which exact optimization method is suitable for each type of the selection slot allocation problem. In contrast with exact optimization, approximate solution algorithms are not guaranteed to yield an optimal solution. Nonetheless, approximate optimization may have certain advantages. Approximate optimization algorithms are often much less computationally burdensome, but may also be preferred for substantive reasons. As will be explained in sub-section 7.4.2, greedy-type approximation algorithms are preferred in this thesis to solve selection slot allocation problems, because of their possible better performance given uncertainty about exact capacity values realized in practice and about future utilization of capacity remaining for allocation after the selection slot allocation decision.

7.4.1 Exact optimization algorithms

This sub-section evaluates which alternative exact solution methods can be used to solve the optimum selection problem types specified in the previous section. One of the aspects discussed in computation time of the alternatives, which may become an issue for larger selection problems. However, since selection problems will typically consider between 10 and 500 requests simultaneously, computation time will not be an important issue in practice (see typical capacity values given in section 6.5). For instance, Amsterdam Airport Schiphol has a declared capacity of less than 110 flights per hour during peak periods (Schiphol, 2004), and the number of slot requests is usually not much higher.

Type 1: linear objective function and single linear capacity constraint

Equation 7.7 specifies the optimum selection problem with only one capacity constraint, which can be classified as a knapsack problem. Exact solutions for the binary knapsack problem are discussed by Martello & Toth (1990). They show that this problem type is NP-hard, which means that these problems cannot be solved in polynomial time, i.e. in a computation time that is bounded by a polynomial function of the size n of the problem. Consequently, the number of evaluations may increase exponentially if the number of slot requests increases, at least in worst-case situations. However, a relatively efficient exact solution algorithm has been proposed by Balas & Zemel (1980) that, according to Martello & Toth, can easily solve knapsack problems with more than 100 000 items within a few seconds, using 1990 state-of-the art hardware.

Types 2 and 3: linear objective function and multiple linear capacity constraints

Equation 7.8 specifies the optimum selection problem with a linear objective function and multiple linear capacity constraints, which was classified as a set packing problem. An overview of exact solution algorithms for binary linear programming problem, including set packing problems, is provided by Wolsey (1998). Balas (1965), for instance, proposes a branch-and-bound type algorithm for binary linear programming

problems. All known exact solution algorithms have an exponential complexity (Wolsey, 1998). Wolsey suggests the usage of approximation algorithms instead of exact optimization to reduce computation time to an acceptable level in the case of 'large' problems. However, since most selection problems in practice will be relatively 'small', an exact optimization algorithm may be applied in most situations.

Type 4: objective function or capacity constraints with interaction terms

Finally, equations 7.10 and 7.11 specify the selection problem as a fixed-point problem. It has been shown that fixed-point problems are equivalent to variational inequality problems (Nagurney, 1993). This type of problem can be solved by iterative procedures, for instance by repeatedly solving the original optimization problem given increasingly precise estimations of composition-of-traffic and the associated parameter values. If in any iteration the assumed composition-of-traffic appears to be inconsistent with the resulting solution of the optimum selection problem, a better estimation may be achieved by adapting the parameters and to recalculate the optimum solution.

In equation 7.10 only the value u depends on the solution \mathbf{x}^* , but in equation 7.11 also the coefficients α of the capacity constraints are made dependent on the solution. Because the feasible set of the problem described by equation 7.11 is also dependent on the solution, this is not a true fixed-point or variational inequality problem. The latter problem is sometimes referred to as a quasi variational inequality problem. This type of problem may be solved by an iterative solution of variational inequality problems (see e.g. Chen, 1999). We will not discuss the solution algorithm of type-4 problems in detail here nor the criteria for convergence of these algorithms and the criteria for uniqueness of the solutions; the interested reader is referred to the literature on fixed-point and variational inequality problems (e.g. Nagurney, 1993).

7.4.2 Approximation algorithm based on efficiency

As an alternative to exact solution algorithms of binary programming problems, various approximation algorithms are proposed in the literature (see e.g. Wolsey, 1998). In the operations research literature, various approximation algorithms have been presented as an alternative to the exact optimization algorithms reviewed in the previous sub-section. The main reason to develop approximation algorithms is that they usually have (much) shorter computation times, especially in the case of 'large' problems. However, given that selection problems will usually consider not more than a few hundred slot request simultaneously, exact optimization algorithms may probably be applied without problem.

Nonetheless, approximation algorithms are proposed in this sub-section for instances of the selection problem. The proposed algorithm for type-1 problems is a greedy algorithm, which solves the continuous relaxation of the binary programming problem,

while the algorithms proposed for the other problem types are extensions to this approach. These algorithms have in common that the objective of maximizing total value is approximated by selecting slot requests with the highest efficiency values. Efficiency is a measure of value per unit of capacity, i.e. it is the contribution to total value u of a slot request (\mathbf{i} in vector notation) resulting in selection \mathbf{x} , divided by its capacity occupancy. Capacity occupancy is a measure of the effect that selecting slot request i has on the capacity remaining for allocation to other slot requests. Given the generic selection problem formulated by equation 7.6, overall capacity occupancy may be calculated based on the contribution of the slot request i under consideration to the capacity occupancy values $g_k(\mathbf{x})$. Relative capacity occupancy values are derived by dividing these with the capacity values κ_k . The overall capacity occupancy is a weighted sum of the effect of the addition of \mathbf{i} on relative occupancy, using z_k as the weight coefficient for each capacity constraint k (see equation 7.12):

$$e_i(\mathbf{x}, \mathbf{i}) = \frac{u(\mathbf{x}) - u(\mathbf{x} - \mathbf{i})}{\sum_k \{z_k(\mathbf{x}) \cdot (g_k(\mathbf{x}) - g_k(\mathbf{x} - \mathbf{i})) \cdot \kappa_k^{-1}\}} \quad (7.12)$$

The calculation of efficiency given different problem types will be elaborated in the remainder of this section. An important issue is the derivation of weight coefficients z_k , which is discussed in the explanation of the extended greedy algorithm for type 3 selection problems.

In the literature, much more sophisticated approximation algorithms may be found than this efficiency approach (see e.g. Wolsey, 1998), but these alternatives usually lack the robustness and clear interpretation of the efficiency approach. The efficiency approach is preferred here, because its performance is less sensitive to the precision of capacity values, which amongst others depends on the choice of period lengths and the choice of minimum quality-of-service. The exact solution of the selection problem is potentially very sensitive to small changes of the capacity values. For instance, several selected slot requests may lose their place in the selection set S given a small decrease or increase of the capacity. Given that capacity values depend on relatively arbitrary choices such as the choice of slot periods, this sensitivity to exact capacity values is undesirable. Therefore, greedy algorithms are proposed in this sub-section as an alternative to exact optimization algorithms. Greedy algorithms are much more robust to chosen parameter values than exact optimization methods, because they evaluate the selection of slot requests sequentially, without previous decisions being dependent on later decisions. Hence, only one slot request may be at stake given a marginal decrease of capacity, which implies that selections determined with a greedy algorithm are less sensitive to adopted capacity values. An additional reason to prefer the efficiency approach is that after the selection slot allocation decision has been made, it may appear that an additional vehicle path may be facilitated, for instance by using the spare capacity of two consecutive periods.

Derivation of greedy algorithm for type-1 problems

The original type-1 selection problem is a binary knapsack problem (equation 7.7). The continuous equivalent is the following linear programming problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{Argmax}} \{u = \boldsymbol{\omega}' \cdot \mathbf{x} : \boldsymbol{\alpha}' \cdot \mathbf{x} \leq \kappa\} \quad \mathbf{x} \in [0,1]^n \quad (7.13)$$

If all weights $\boldsymbol{\alpha}$ are binary and if capacity κ is integer, the LP relaxation provides the exact solution to the original binary problem (Martello & Toth, 1990). In other situations, however, the solution of the linear relaxation may be quite different from its binary counterpart. As is shown by Dantzig (1957), the LP problem represented by equation 7.13 can be solved by selecting all items (slot requests) i in the order of their efficiency e (see equation 7.14). This algorithm may directly be applied to the LP relaxation of type 1 selection problems.

The efficiency principle is the basis of the greedy-type algorithm proposed by Sahni (1975). This algorithm first selects all items that are included completely in the solution of the LP relaxation, and then checks whether any other items can be included completely. To achieve this, each item i should first be ranked according to its efficiency e_i , which is in the case of type 1 selection problems simply its contribution ω_i to the objective function, divided by its weight α_i in the capacity constraint:

$$e_i = \frac{\omega_i}{\alpha_i} \quad \forall i \in \mathbf{R} \quad (7.14)$$

Figure 7.3 visualizes our implementation of Sahni's greedy algorithm for the selection problem. We begin with an empty selection set S and a stack of n slot requests. Then, the efficiency e_i of each slot request i is calculated. Efficiency is used to rank slot requests in descending order. Vector \mathbf{r} is the resulting ordered list of slot request numbers. Next, slot requests are evaluated in descending order of efficiency. Each time, the most efficient non-selected slot request l is added to the selection set ($x_l \leftarrow 1$), and the feasibility of its addition to the selection set is evaluated. If the addition is feasible, it remains in the selection set, else this slot request is discarded ($x_l \leftarrow 0$). This procedure is repeated until all n slot requests have been evaluated. The resulting selection vector \mathbf{x} is the solution \mathbf{x}^* of the selection problem.

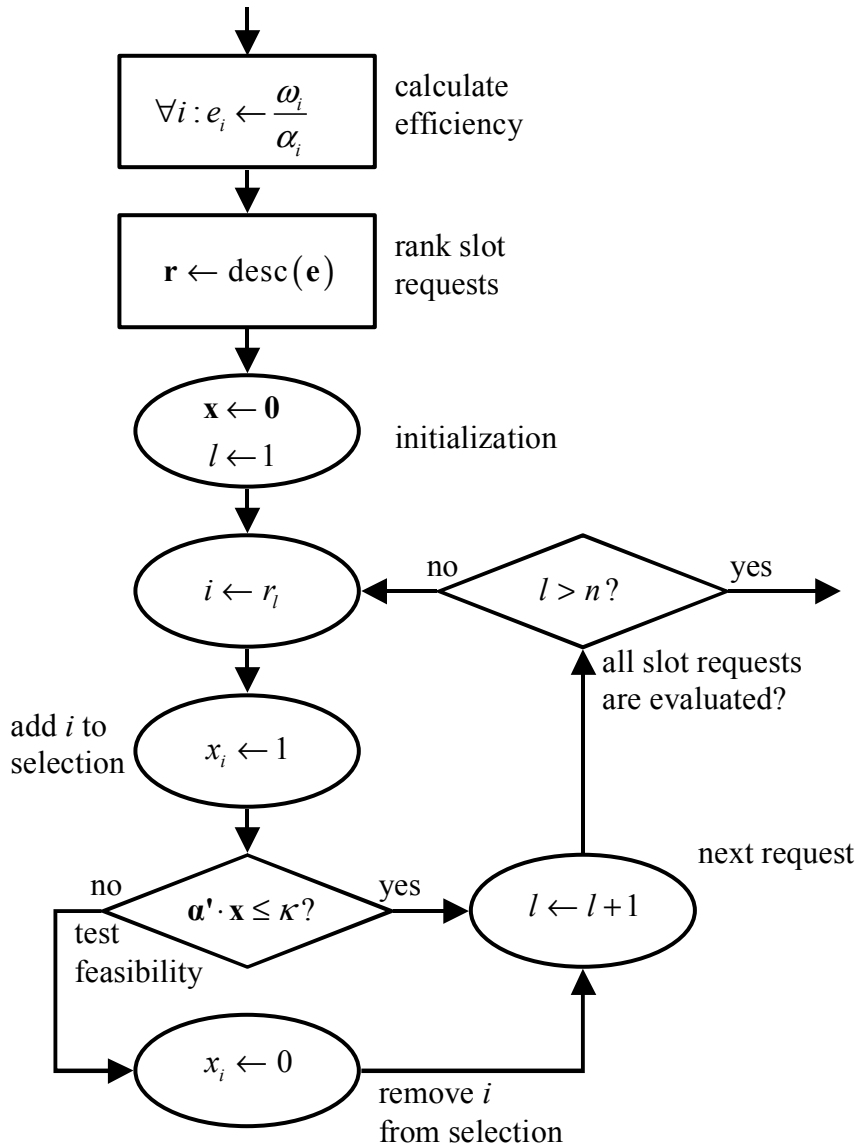


Figure 7.3: Basic greedy algorithm

An interesting question is to which extent the solution proposed by the greedy algorithm differs from the optimal solution of the selection problem. A criterion frequently used in the operations research literature (e.g. Martello & Toth, 1990) is the worst-case performance, which is the relative difference between the least possible value of the greedy solution u_{\min} and the maximum value corresponding with the optimal solution u_{\max} . According to Sahni (1975), the worst-case performance of this approximation algorithm depends on the number of items n :

$$\frac{u_{\min}}{u_{\max}} = \frac{n}{n+1} \quad (7.15)$$

Given only two slot requests, the value of the solution found with the greedy algorithm is at least half the value of the optimum. Given 20 slot requests, the value of the solution is at least 95% of the exact optimum, which means that for but the smallest selection problems, the greedy algorithm is a good approximation of exact optimization algo-

rithms, apart from the substantive arguments in favor of greedy approximation mentioned earlier.

Applicability of greedy algorithm to type-2 problems

With a few small modifications, the greedy algorithm visualized in figure 7.3 can also be applied to type-2 problems. In the first place, the feasibility test should be modified to account for several capacity constraints instead of one. Furthermore, the efficiency e_i corresponding with each slot request i should be calculated using the nonzero $\alpha_{i,j}$ value, which is equal for each basic slot j the slot request under consideration is included in. The greedy algorithm for type-2 problems (see figure 7.4) is similar to the algorithm for type-1 problems, the only difference being the way efficiency is calculated:

$$e_i = \frac{\omega_i}{\max_j \alpha_{i,j}} \quad \forall i \in R \quad (7.16)$$

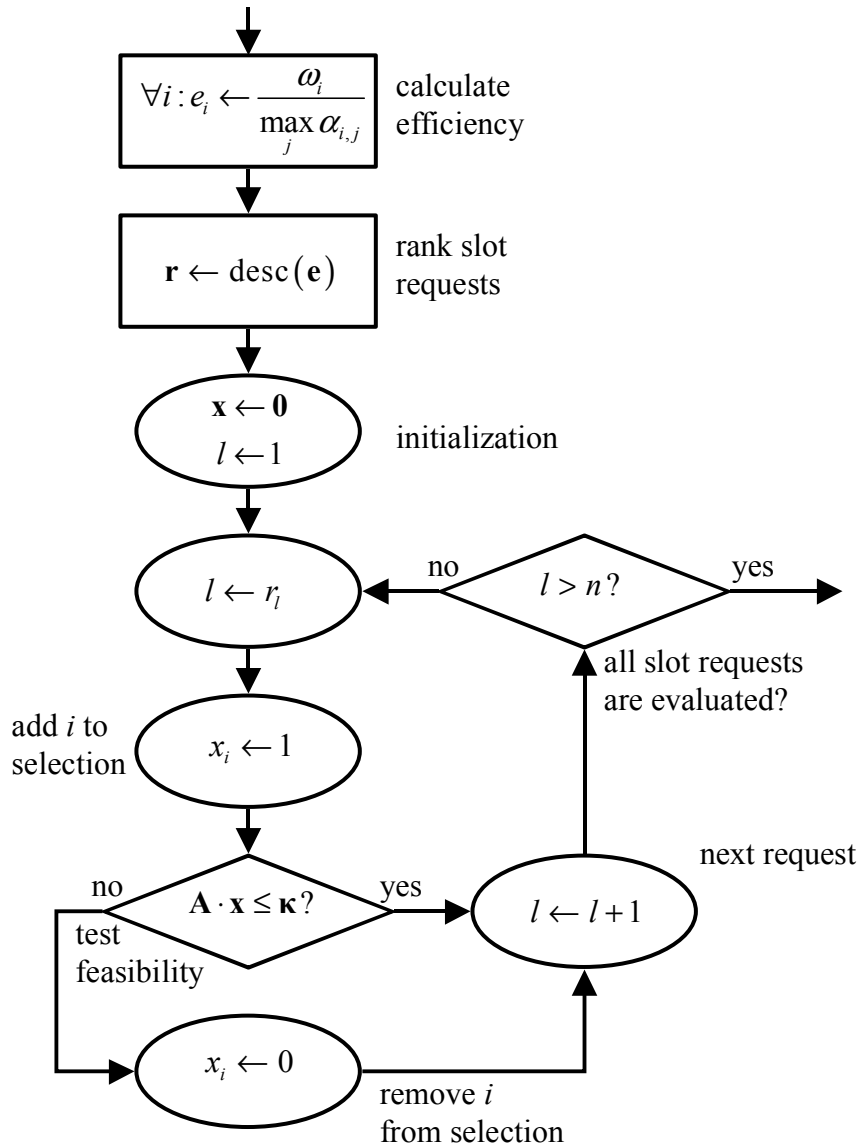


Figure 7.4: Basic greedy algorithm for type-2 problems

We may verify that the algorithm proposed in figure 7.4 indeed results in a maximization of the value per used unit of capacity for each period, just as the efficiency algorithm does for type-1 problems. This argument holds only given multi-level slot specification, not when single slot requests claims different kinds of capacity. Although slot requests included in lower-level capacity constraints are also included in all corresponding higher-level capacity constraints, there is no possibility that one slot request is substituted by different slot requests in different capacity constraints. Therefore, we may directly compare the efficiencies of all slot requests, and the greedy algorithm can be applied to type-2 problems.

Extended greedy algorithm for type-3 problems

In order to be able to deal with type-3 problems, the greedy algorithm introduced in this sub-section has to be extended to allow for the inclusion of two or more capacity constraints. A greedy algorithm to solve the set packing problem has been proposed by Fisher & Wolsey (1982). The only difference with the greedy algorithm for type 1 and type-2 problems is the calculation of efficiency, which involves the weighted addition of all $\alpha_{i,j}$ values of slot request i . Inverse capacity is used to weight the various capacity constraints, which corresponds with equal z values in the generic formula to calculate efficiency (equation 7.12):

$$e_i = \frac{\omega_i}{\sum_k \frac{\alpha_{i,k}}{\kappa_k}} \quad \forall i \in R \quad (7.17)$$

In equation 7.17, capacity constraints are weighted in such a way that the unit of equivalence is the occupancy of an equal percentage of capacity. The main problem, however, with this approach is that all capacity constraints are valued equally, even capacity constraints that are not binding. This is a reasonable approach if all capacity constraints are binding. However, non-binding capacity constraints should have zero weight, since these are of no importance to the selection slot allocation. Therefore, capacity constraints that appear to be non-binding should be excluded from the set of constraints, or should be given a sufficiently low weight to make these binding.

This thesis introduces an extension to the greedy algorithm in which the weight of capacity constraints in the calculation of efficiency is such that only binding constraints influence the efficiency value. Just as in the generic formula for calculating efficiency (equation 7.12), weight coefficients z_k are used to give different weights to different capacity constraints k :

$$e_i = \frac{\omega_i}{\sum_k \frac{z_k \cdot \alpha_{i,k}}{\kappa_k}} \quad \forall i \in R \quad (7.18)$$

The extended greedy algorithm introduced in this thesis for type-3 selection problems is visualized by figure 7.5. Using equation 7.17 to calculate efficiency, the remaining problem is to determine weight coefficients z_k . These weight coefficients should have such values that, after applying the greedy algorithm, all binding capacity constraints in the resulting selection have non-zero z -values, while the z values of non-binding capacity should be 0. Hence, a check for non-binding capacity constraints is required in the solution algorithm. In the extended greedy algorithm, visualized in figure 7.5, a test for non-binding capacity constraints is performed after the feasibility test that is performed after each addition of a new item (slot request) to the selection. Indicator variable h is introduced in figure 7.5 to mark binding capacity constraints. Indicator variable h_j reports for every capacity constraint j whether it is binding ($h_j = 1$) or not ($h_j = 0$). A capacity constraint k is binding if for at least one rejected slot request this capacity constraint was violated:

$$\sum_i \frac{\alpha_{i,k} \cdot x_i}{K_k} > 1 \quad (7.19)$$

When all slot requests have been evaluated, the extended greedy algorithm tests whether all capacity constraints k with non-zero z_k are binding (i.e. $h_j = 1$). The following test is used, using a small threshold value ε enabling the algorithm to converge when the z values of non-binding capacity constraints are close enough to zero:

$$\mathbf{z}' \cdot (\mathbf{1} - \mathbf{h}) \leq \varepsilon \quad (7.20)$$

If equation 7.20 is violated, the weight z of non-binding capacity constraints is reduced for the next iteration of the algorithm. To this end, new weights z are calculated as a weighted average of old z values and the value of h . After recalculating efficiency values, a new selection is determined in a new iteration of the algorithm.

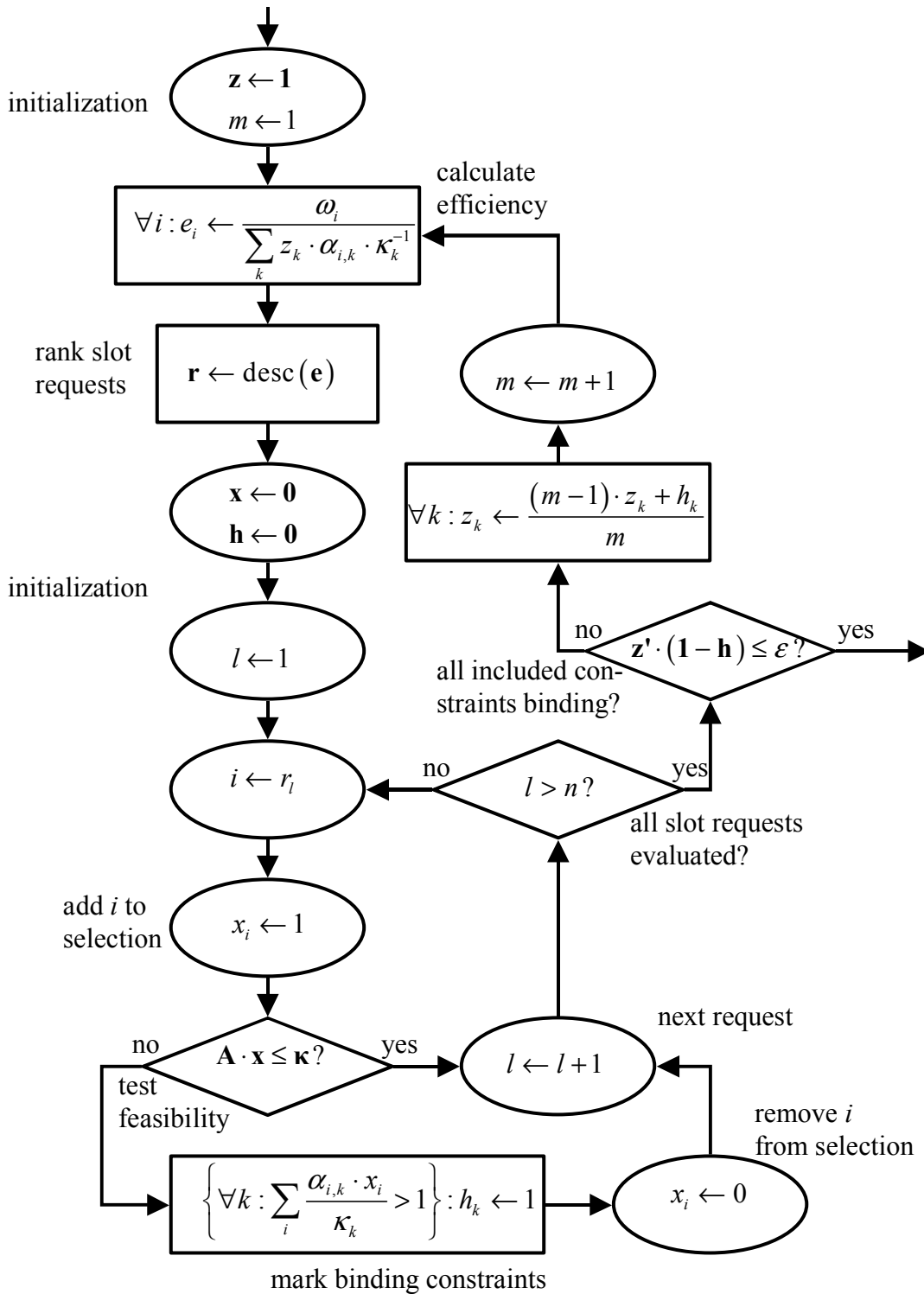


Figure 7.5: Extended greedy algorithm

The weights z in the extended greedy algorithm may be interpreted as shadow prices. In the case of integer programming problems, there is no unique straightforward way to determine the shadow prices corresponding with the optimum (see Wolsey, 1981). An alternative approach to determine shadow prices has been proposed by Rassenti *et al.* (1982), who analyzed the airport slot allocation problem with multiple capacity constraints, which can be classified as a type-3 problem. According to Rassenti *et al.*

(1982), upper and lower bounds of shadow prices may be determined by solving pseudo-dual linear programming problems to determine marginal acceptance and rejection prices, given that the solution to the slot allocation problem has already been determined.

Iterative solution for type-4 problems

Given a type-4 problem, a consistent combination of an assumed selection and the resulting selection has to be found. As was already discussed in the previous sub-section, this may be achieved by repeated application of the solution algorithm. Depending on the problem type, the greedy algorithm or the extended greedy algorithm is used iteratively, until the solution matches the assumptions about this solution.

7.5 Extensions to selection slot allocation decision approach

In the previous section, an approximate optimization approach has been proposed to solve various instances of the selection problem. The specification of traffic supply discussed in chapter 6 and the approximate optimization method outlined in the previous section together result in a selection slot allocation decision approach for single-bottleneck selection problems with a single objective function as the only preference function. However, as will be discussed in this section, this approach may be extended to include satisficing constraints. As was discussed in section 7.2, secondary objectives can best be implemented as additional constraints instead of in the objective function. Furthermore, the single-bottleneck approach may be extended to network problems with several bottlenecks. Both types of extensions are discussed in this section.

7.5.1 Basic slot allocation decision approach

Given the robustness arguments put forward in the previous sub-section, the selection slot allocation decision method proposed in this thesis is not based on exact optimization. Instead, we will assume that a priority list of slot requests may be established, for instance based on the efficiency criterion used in the greedy algorithm. For type 3 and 4 problems, iterative updating of efficiency values given the current solution may be required.

The general selection slot allocation decision method in a situation without satisficing criteria is outlined by figure 7.6. The procedure begins with the identification of slot requests and the specification of the objective function. Next, traffic supply is specified. Capacity constraints are formulated given the slot specification, known traffic demand characteristics such as bottleneck characteristics, and quality-of-service criteria. The procedure continues with specifying a priority list of slot requests, which is (preferably) based on efficiency values. The next step is to check which part of the priority list can be accommodated given the capacity constraints. Based on the feasibility test, the as-

assumptions regarding the efficiency have to be checked, and the priority list may be updated. For instance, capacity constraints that appear to be non-binding should not be included in the calculation of efficiency. The final result of this decision procedure is a proposal for a selection slot allocation.

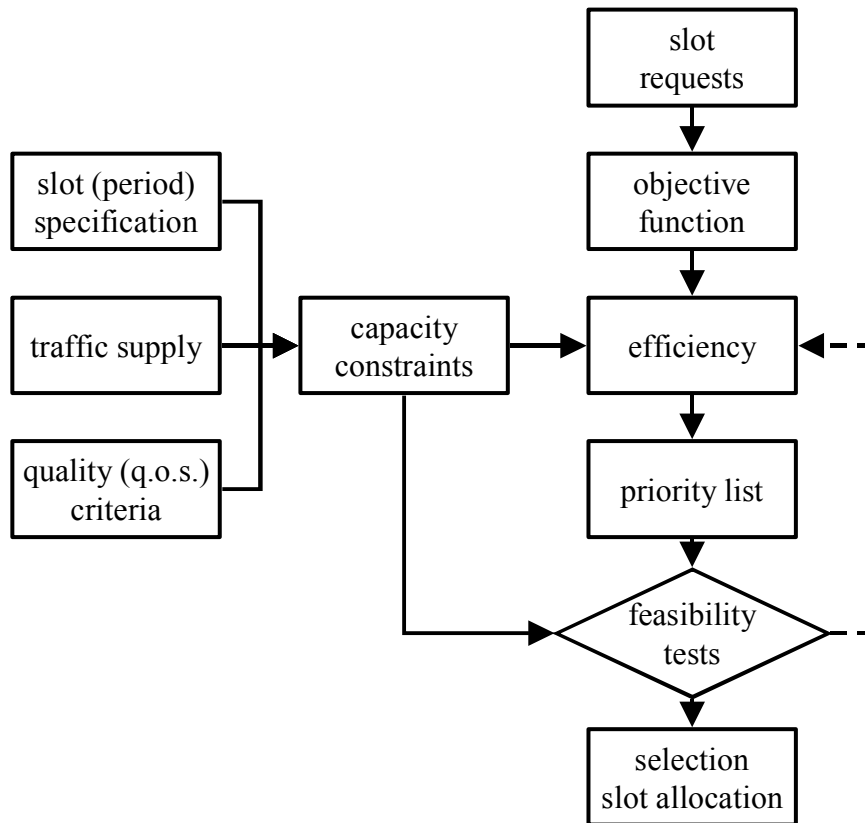


Figure 7.6: Basic selection slot allocation decision procedure

Ties

The selection slot allocation procedure proposed in this thesis is based on sequential evaluation of slot requests. The sequence is based on a priority list, which should be an ordered list of slot requests without ties. However, it is possible that two or more slot requests have an equal estimated efficiency, which means that a tie breaking procedure might be required. It is important that this tie breaking procedure is non-discriminatory, especially when the tie is caused by similar slot requests from competing carriers. A possible solution is to apply additional priority criteria. It seems that only two non-discriminatory solutions will avail if two slot requests only differ with respect to the carrier requesting it. The first solution is to try to refine the estimation of efficiency, for instance by applying an auction procedure. The other solution is to make a random selection.

Solution procedure

The general solution procedure outlined in section 7.3 is based on the formulation of linear capacity constraints. Given the assumptions of problem types 1 to 3, the capacity constraints are indeed linear. In chapter 6, we have seen that in many cases traffic supply can be formalized using linear capacity constraints. In situations where capacity is dependent on composition-of-traffic, linearization of capacity constraints may be possible, but only given a good a priori estimation of composition-of-traffic. To check and update this estimation, the fixed-point solution of problem type 4 may be applied. The necessity to formulate linear capacity constraints may even be avoided altogether by applying a more generic feasibility test. For instance, instead of testing feasibility with linear capacity constraints, available sector-specific traffic (simulation) models may be used.

The general solution procedure described in this section is a simple iterative procedure: slot requests are evaluated sequentially. A similar procedure will be proposed in the next sub-section to find satisfactory basic solutions. However, sequential evaluation, beginning with an empty selection set, is not the only viable approach. It is also possible to start with a full selection set and remove slot requests sequentially. This is probably a faster approach when more than 50% of all requests is expected to be selected. Even more sophisticated may be to start the evaluation process with an 'in-between' alternative that is expected to be close to the final solution.

7.5.2 Extension of solution approach to include satisficing constraints

As we have seen in section 7.1, an alternative to specifying objective functions to be optimized is to specify which solutions are satisfactory. In the context of the selection problem, satisficing constraints specify which types of slot requests should at least be included in the selection. As we have seen in section 7.2, satisficing constraints may be used to implement secondary objectives regarding competition and transport supply. This may be achieved by formulating minimum levels of competition and minimum claim rights. This sub-section reviews how both types of satisficing constraints can be implemented.

Satisficing constraints may be implemented in the solution procedure by introducing an additional step between the establishment of the priority list and the feasibility tests. This additional step is the establishment of a basic selection, which is a selection of slot requests that satisfies the satisficing constraints. An (approximately) optimal basic selection is found by applying the same priority list used in the general solution procedure. This procedure stops when a basic solution has been found, and then the procedure is continued with the sequential feasibility checks of the general solution procedure. We assume that any basic solution is feasible, and hence it is not necessary to incorporate capacity tests in the procedure to establish this basic selection. Figure 7.7 illustrates the selection slot allocation decision method in a situation with satisficing constraints.

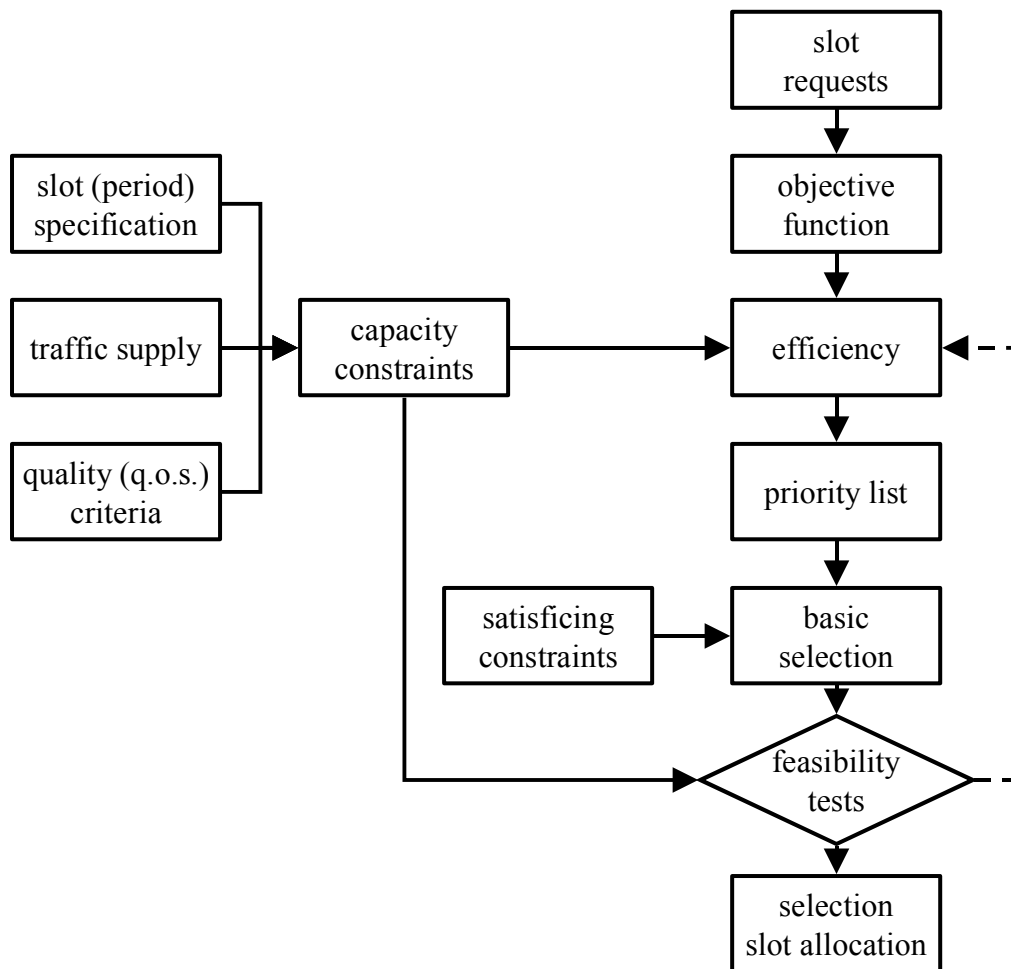


Figure 7.7: Selection slot allocation decision procedure with satisficing criteria

Minimum claim rights

In our review of secondary objectives in section 7.2, we have seen that authorities are often concerned with the minimum supply of transport service in certain transport markets. To ensure this minimum supply, authorities may conclude contracts with carriers and request the required selection slots. However, these slot requests may still be rejected. If the authorities have the opinion that certain frequencies of transport services should be enabled at least, they may apply minimum claim rights. Minimum claim rights are satisficing constraints specifying the minimum number of slot requests within a certain group of slot requests that should at least be selected. For instance, minimum claim rights may specify which transport service frequency should at least be facilitated on which link. An example of minimum claim rights in current practice is the specification of minimum frequencies for each market segment in Dutch regulation regarding railway slot allocation (recall sub-section 3.2.1). Finally, minimum claim rights may be desired to ensure that sufficient transport capacity can be offered for each transport relation, assuming the usage of optimum size vehicles.

Each satisficing constraint relates to a group of slot requests g . A group may correspond with one link or line or with a number of similar links or lines. Groups are disjoint, i.e.

no slot request is member of more than one group. Minimum claim rights state that the number of vehicle paths f within a group of slot requests g as a function of selection slot allocation decision \mathbf{x} , should be at least ψ_g :

$$f_g(\mathbf{x}) \geq \psi_g \quad \forall g \quad (7.21)$$

This total number of vehicle paths is simply a sum of the number of vehicle paths corresponding with each slot request within group g . Because each slot request normally corresponds with one vehicle path, this means that simply the number of slot requests within group g that have been selected should be counted. Equation 7.22 shows how f may be calculated, with $\mu_{i,g}$ as binary variable indicating the membership of slot request i of group g :

$$f_g(\mathbf{x}) = \sum_i \mu_{i,g} \cdot x_i \quad \forall g \quad (7.22)$$

A basic solution satisfying minimum claim rights may be achieved by running the procedure outlined in figure 7.8. Similar to the optimization procedure, the addition of slot requests is evaluated in sequence of their priority, until sufficient slots have been allocated to each group. In this case, priority is determined for each group g separately by multiplying membership vector $\boldsymbol{\mu}_g$ with efficiency value vector \mathbf{e} .

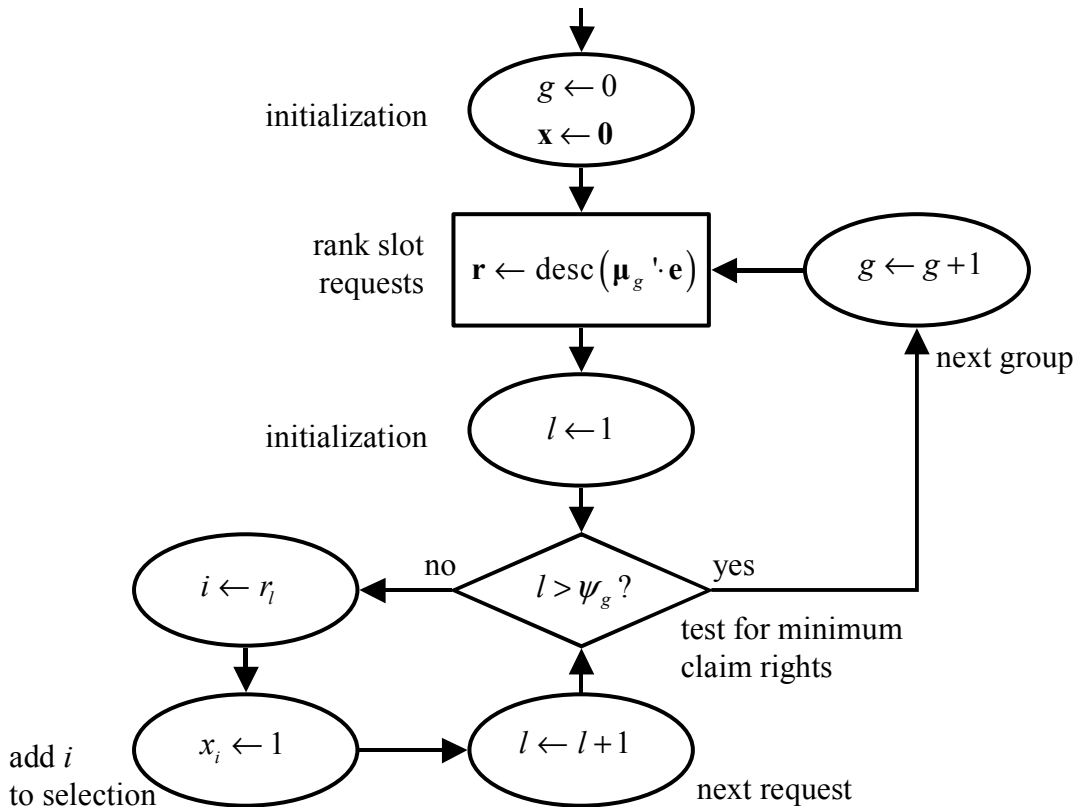


Figure 7.8: Procedure to find basic selection given minimum claim rights

Minimum level-of-competition

Satisficing constraints may also be used to specify a minimum required level-of-competition. For instance, it may be demanded that at least two or three different carriers should be able to obtain slots allowing competing services with reasonable frequencies on main relations in competitive markets such as the air passenger transport market or the rail freight transport market in the EU. The main idea is that given a sufficient number of different carriers offering similar transport services, competition will be the result. Environmental objectives may also be translated into satisficing criteria, but this is similar to the formulation of environmental capacity constraints.

Minimum level-of-competition satisficing constraints are formulated for groups of slot requests. A group of slot requests g may refer to a common origin or destination, period, group of transport network links, etc. As was the case with minimum claim rights, groups are assumed to be disjoint, i.e. without overlap. Minimum level-of-competition constraints state that for a group of slot requests g under consideration, the number of different carriers a that have been granted a slot in group g should be at least γ_g :

$$a_g(\mathbf{x}) \geq \gamma_g \quad \forall g \quad (7.23)$$

Satisficing criteria regarding minimum level-of-competition may be incorporated in the procedure to establish a basic selection by sequentially testing whether the slot request under consideration contributes to a higher level-of-competition, until the required level has been reached. Figure 7.9 illustrates the procedure that may be followed when minimum level-of-competition is defined as a minimum number of different carriers being allocated a slot within a specific group g of slot requests. As in the procedure to find a basic solution satisfying minimum claim rights, $\mu_{i,g}$ is a binary variable indicating the membership of slot request i of group g . The carrier c corresponding with slot request i is identified with t_i . Finally, binary variable h_c is used to identify whether carrier c has been allocated a slot ($h_c = 1$) or not ($h_c = 0$).

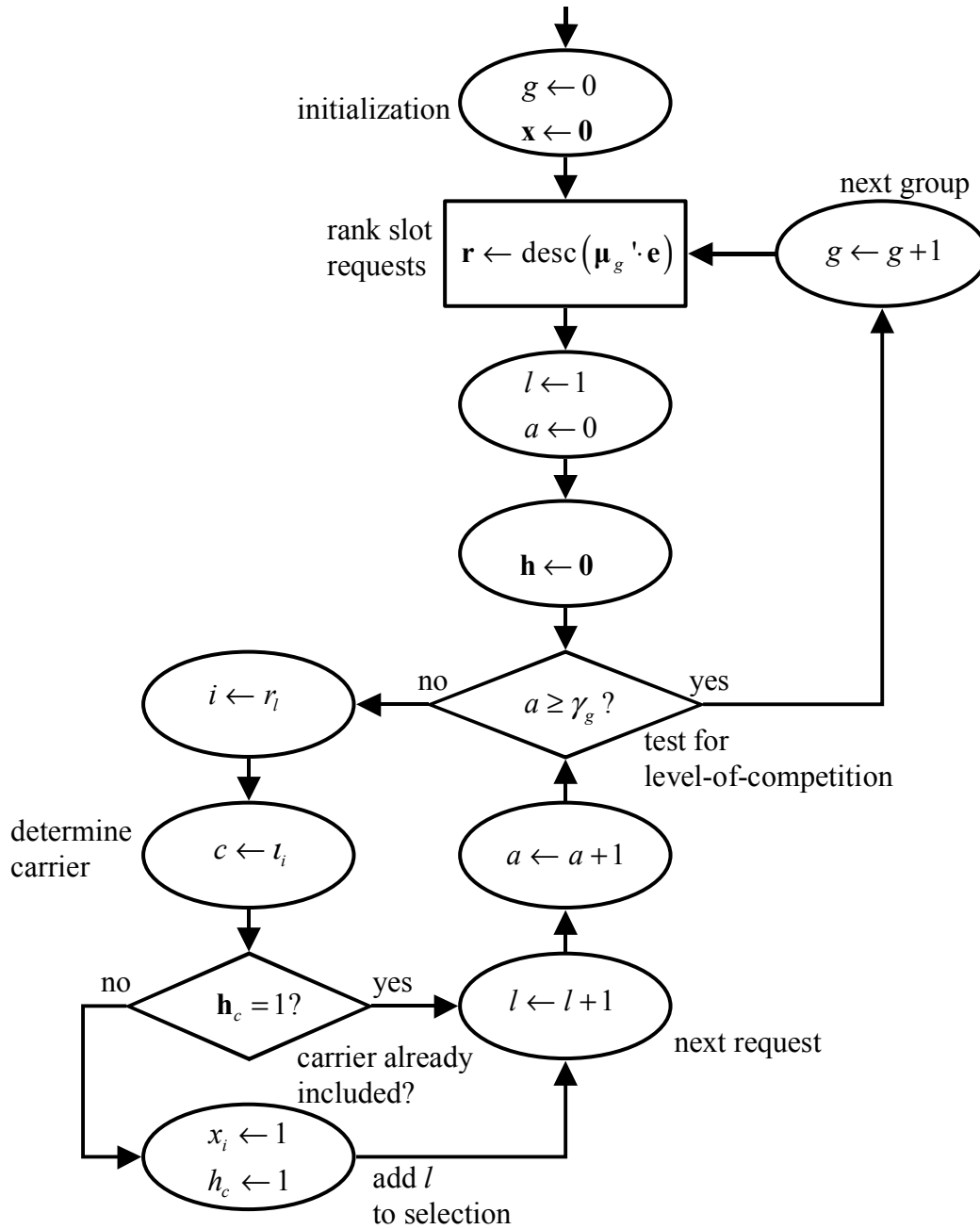


Figure 7.9: Procedure to find basic selection given competition criteria

If both minimum level-of-competition and minimum claim rights are imposed that relate to the same subset of slot requests, both procedures should be run after each other. The procedure regarding competition should be performed first, because in a subset of requests, any accepted slot request helps to increase the total number of vehicle paths within this subset, but only some slot requests help to increase the level-of-competition.

7.5.3 Extension of solution approach to network problems

Until now, we have focused on single bottleneck selection problems, ignoring specific issues related to networks with several bottlenecks in the basic solution approach. How-

ever, the approach for single-bottleneck selection problems proposed in this thesis may in many cases be extended to network problems. The key question is to which extent network problems are fundamentally different from single bottleneck problems. As we have seen in section 6.4, network interdependencies may complicate the specification of traffic supply. Another difference is that not only capacity constraints corresponding with different periods and traffic processes may apply, but also different capacity constraints corresponding with different bottlenecks.

In section 6.4, a few solutions have been proposed to cope with network interdependencies. In the first place, it has been suggested that a sufficient degree of scheduling freedom can be attained by applying a stronger reduction factor φ . This means that a larger part of capacity is not allocated as to increase scheduling freedom. Secondly, interdependencies of capacity values between adjacent infrastructure elements may be coped with by optimizing the capacity of the 'true' bottleneck. Furthermore, a feedback mechanism as suggested for type 4 single bottleneck problems may be used to update the initial assumptions about bottleneck use as to optimize the capacity values of related bottlenecks. Finally, a (sector-specific) network traffic model may be used for the feasibility tests, instead of linear capacity constraints. In that case, linearization of traffic supply is only required to estimate efficiency, but not for the feasibility tests.

Formalization of optimum selection problem for networks

The selection problem of a network with two or more bottlenecks may be formulated as a comprehensive problem. Each bottleneck corresponds with at least one capacity constraint for each period. If the travel times in the network are relatively large compared with the periods, all periods may have to be analyzed simultaneously, even in the case of discrete slot specification, because vehicle paths departing from one bottleneck in a certain period may arrive at another bottleneck in a later period. Consequently, many capacity constraints may be included in the definition of traffic supply.

Assuming that capacity constraints are linear and independent and assuming a linear objective function, the problem is the same as assumed in single-bottleneck problem type 3. Hence, the optimum selection problem for networks is expressed as in equation 7.9:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{Argmax}} \{u = \boldsymbol{\omega}' \cdot \mathbf{x} : \mathbf{A} \cdot \mathbf{x} \leq \boldsymbol{\kappa}\} \quad \mathbf{x} \in \{0,1\}^n \quad (7.24)$$

Given that the problem formulation is equivalent, the same solution procedure as described for type-3 problems may be used. Furthermore, non-linearity of objectives or constraints may be treated as proposed for type-4 problems, i.e. by solving increasingly good linear approximations of the original problem.

7.6 Conclusions

In this chapter the selection slot allocation decision problem has been specified and the solution procedure of the selection problem has been discussed, which are two essential steps to meet the objectives of this thesis. The first issue covered in this chapter was the specification of selection slot allocation as a rational decision-making process. An essential step of rational decision-making is the specification of objectives, and therefore this subject has been discussed extensively in this chapter. Furthermore, the selection problem has been formalized as an optimization problem, and a general solution approach based on efficiency has been proposed.

Objectives

The specification of objectives is a key step of rational decision-making. Rational decisions are based on preferences with respect to alternative solutions, and these preferences are based on objectives. There are different ways to specify objectives, including objective functions and satisficing constraints. Given the decision approach outlined in this chapter, an objective function is used to represent the primary objectives, but satisficing constraints may be applied additionally.

This thesis adopts a normative perspective, which means that the formulation of objectives is up to the decision-maker. This thesis assumes that these objectives are related to general public interests, including the interests of carriers and end users. Given this assumption, the mathematical characteristics of the objective function have been discussed. It appears that interactions between slot requests may result in non-linearity of the objective function, but these are likely to be relatively small. Given a reasonable estimation of composition-of-traffic, a linear objective function can be used in most situations. It may be noted that selection slot allocation decisions will often be made in a situation where many selection slots have already been allocated previously, and hence the main gross characteristics of composition-of-traffic will be known beforehand. The application of a linear objective function is preferred, because it results in a simpler solution procedure than non-linear objective functions.

Optimum selection slot allocation decision problem: formulation and solution approach

Given this general approach, four selection problem types have been formulated in this chapter. Type 4 is most generic, while type 1 is most specific. Given the argument that a linear objective function and linear capacity constraints may be applied in most practical situations, almost all selection problems arising in practice may be formulated as a type 3 selection slot allocation decision problem, i.e. a binary linear programming problem. Problem types 1 and 2 are sub-categories of problem type 3.

Although exact optimization algorithms may be applied to solve selection problem types 1 to 3, the application of a greedy approximation algorithm is preferred in this

thesis. Exact optimization only gives better results than greedy approximation if spare capacity can be minimized, assuming that unutilized capacity has no positive benefits to the objective function. Furthermore, the solution of exact optimization may be very sensitive to chosen capacity specification, while the latter depends on various independent variables, included desired quality-of-service level. The main advantage of the greedy algorithms proposed in this chapter is their robustness and their straightforward interpretation. For each slot request, its efficiency determines its ranking. Efficiency is calculated by dividing its value associated with a slot request by the percentage of (binding) capacity it occupies. The algorithm evaluates the feasibility of adding slot requests to the selection set in sequence of their ranking until all slot requests have been evaluated.

While the greedy algorithm for type-1 and type-2 selection problems has been directly adopted from the literature, an extension to this algorithm has been developed in this thesis to be applied to type-3 selection problems. If two or more different capacity constraints apply to the same basic slots, these capacity constraints have to be weighted in the calculation of efficiency. This weight is determined by checking to which extent different capacity constraints are binding. The weight of capacity constraints in the calculation of efficiency is adapted iteratively in the extended greedy algorithm, until correct weight coefficients have been found.

The solution procedure to the selection problem proposed in this chapter includes the possibility to specify satisficing constraints. Two types of satisficing constraints have been discussed explicitly. Minimum claim rights can be applied to ensure that a reasonable minimum quality of transport services can be offered on certain network links. Minimum claim rights may be based on the required transport capacity, but also on accessibility criteria with regard to peripheral regions. Furthermore, minimum level-of-competition satisficing constraints are probably the best way to implement the (secondary) objective of attaining a sufficient level-of-competition between carriers on the transport market.

Next chapter

The next chapter focuses on the practical implications of selection slot allocation. Chapter 8 exemplifies the formulation and solution of selection slot allocation problems by a hypothetical case study. The selection slot allocation problem of a metropolitan airport is formulated and the type of selection problem is identified. Furthermore, the selection problem is solved using the extended greedy algorithm that has been proposed in the current chapter. Finally, the next chapter discusses a number of issues with respect to the practical implementation of selection slot allocation as proposed in this thesis.

8

PRACTICAL IMPLICATIONS OF SELECTION SLOT ALLOCATION

In this thesis, the desired setup of a selection slot allocation system has been given, including an approach to solve the corresponding decision problem, adopting the efficiency approach outlined in the previous chapter. As a final step before the conclusions and recommendations, this chapter discusses the implementation of selection slot allocation in practice. This includes a hypothetical case study and a discussion about a number of issues with respect to the practical implementation of selection slot allocation as proposed in this thesis. The case study analyzes how selection slot allocation can be implemented at a hypothetical airport (section 8.1). First, the selection slot allocation problem is specified, and then the selection problem is solved for an example period. It appears that even the selection problem of a fairly complicated bottleneck, featuring terminal and runway capacity constraints and environmental constraints, can be solved relatively easily as a type-3 problem, using the efficiency heuristic proposed in the previous chapter. The second section reviews the consequences of implementation of selection slot allocation, compared with current practice. This includes an overview of the main changes resulting from implementation of selection slot allocation as proposed in this thesis, compared with current practice in the European Union. Furthermore, the issue of limited legal competence of allocation bodies is discussed, focusing on the problem of cross-border traffic. Finally, the main consequences of the implementation of selection slot allocation are evaluated, based on the main issues introduced in section 3.3. It is concluded that the framework introduced in this thesis may help to attain a proper balance between stability and flexibility, and between diversity and competition. Furthermore, it offers a simple but adequate approach to deal with heterogeneity.

8.1 Example: Metropolitan Airport

To illustrate selection slot allocation in practice, we review a hypothetical case study, i.e. the selection problem of Metropolitan Airport. First, the capacity scarcity problems of Metropolitan Airport are described. Because of the absolute capacity scarcity problems, selection slot allocation is applied. In the second sub-section, an overview is given of traffic supply and demand, regarding a specific period. The selection slot allocation is formalized as a type-3 slot allocation problem. A solution to this problem is derived using the extended greedy algorithm introduced in section 7.4.

8.1.1 Introduction

Metropolitan Airport is the main airport of a large city. It is situated close to the urban area on a peninsula, which limits its expansion possibilities (figure 8.1). However, its proximity to the city center gives this airport a clear advantage over peripheral airports.

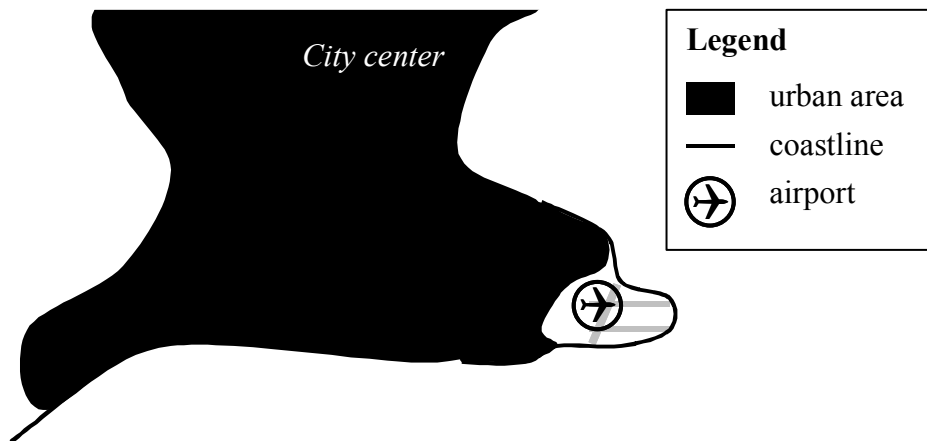


Figure 8.1: Location of Metropolitan Airport

A schematic map of Metropolitan Airport is provided by figure 8.2. Metropolitan Airport has three runways (A, B, and C), two passenger terminals (T1 and T2) and a cargo terminal (Tc). T1 is an older terminal accommodating international transport services, while T2 is a relatively new passenger terminal for domestic transport services. A subway station of the Metropolitan Railway Network is located not far from both terminals, which means that these terminals have a good connection with the city center.

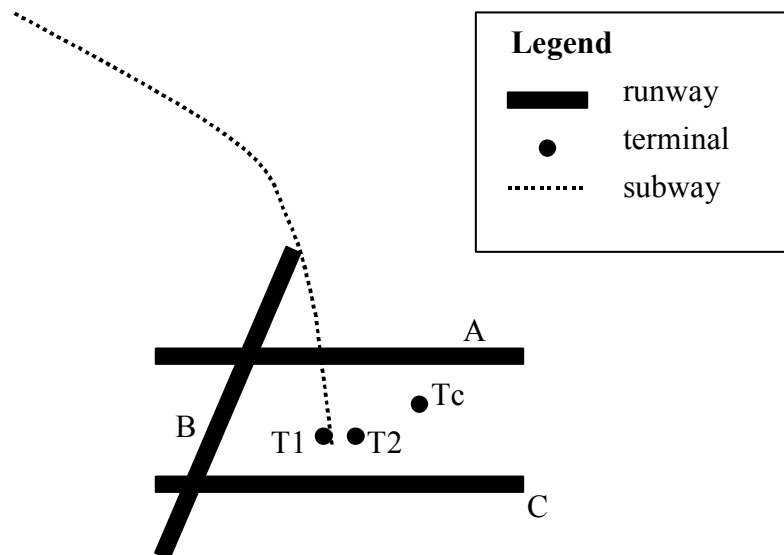


Figure 8.2: Layout of Metropolitan Airport

Metropolitan Airport faces structural capacity scarcity problems. Depending on the period under consideration, the following traffic network elements may be bottlenecks. In the first place, only two runways are available simultaneously, which limits the airport's ability to accommodate peaks in departure and arrival rates. Another potential bottleneck is the capacity of the terminals. The terminals have a limited number of gates, which is mainly a potential bottleneck for Terminal 1. Finally, noise limitations have been imposed by the metropolitan authorities, which considerably limits airport capacity during late evening and night hours.

Transport services facilitated by Metropolitan Airport

Metropolitan Airport facilitates a number of different transport networks that are provided by different carriers. Metropolitan Airport is a home-base of flag carrier InterAir. InterAir offers an extensive network of international and domestic passenger transport services, and Metropolitan Airport is the hub of this network. Domestic transport services are also offered by its competitor CityLink. InterFreight is the branch of InterAir that is responsible for cargo transport. It offers mainly point-to-point transport services. PostExpress is a mail company that has chosen Metropolitan Airport as a regional collection and distribution center in its worldwide network. SunExpress offers point-to-point passenger transport services to leisure destinations, including both 'line' and 'charter' flights. Finally, selection slots have been requested by foreign carriers offering continental and intercontinental passenger transport services. These passenger transport services are not dependent on connections with other flights at Metropolitan Airport, but either convey passengers to another hub or directly to their destination. Table 8.1 summarizes some characteristics of the transport service networks reviewed here, i.e. carrier, type of transport network (passengers or freight), scale, suitable terminals, and the question of whether Metropolitan Airport functions as a hub for this transport service network.

Table 8.1: Main characteristics of transport service networks

carrier	type of transport	scale	suitable terminals	hub
InterAir	passengers	international	T1	yes
InterAir	passengers	domestic	T2	yes
CityLink	passengers	domestic	T2	yes
InterFreight	cargo	international	Tc	no
SunExpress	passengers	international	T1	no
others (foreign)	passengers	international	T1	no

Selection slot allocation at Metropolitan airport

The level of capacity scarcity at Metropolitan Airport is classified as absolute, and therefore selection slot allocation is applied. The Metropolitan Airport Traffic Authority (MATA) is the allocation body; i.e. MATA receives slot requests and decides on their acceptance or rejection. A semi-static selection slot allocation procedure is applied, and selection slots can be requested before every timetable season. Each year consists of a winter and a summer season. Selection slots that are requested for the first time, i.e. that correspond with new transport services, have a validity of 8 years. After each period, extensions may be requested, which have a validity of 4 years. Selection slots are usually requested and held by carriers, although some selection slots are held by regional authorities or by tour operators.

8.1.2 Specification of traffic supply and demand

Selection slot allocation is applied to Metropolitan Airport in accordance with the ideas outlined in chapter 4. In this sub-section we review the specification of selection slots, based on the acceptability criterion, which was the main subject of section 4.3. Next, an overview is given of potential bottlenecks within the airport, i.e. runways, terminals, and externalities. Based on this overview, capacity constraints are specified. The service time available for allocation, i.e. the service time that is not already occupied by selection slots allocated in previous periods, is determined next. Finally, all slot requests are listed, including their main variables determining capacity consumption, and the value associated with each slot request.

Slot specification

As we have seen in the previous sub-section, each terminal is equipped for a different class of transport services. Because terminal capacity is one of the bottlenecks of Metropolitan Airport, it is relevant to specify the spatial slot size of selection slots explicitly, i.e. to specify which terminals are included. Selection slots of international passenger transport services are restricted to Terminal 1, because Terminal 2 has no customs facilities. Similarly, domestic flights are restricted to Terminal 2, in order to avoid that

domestic passengers use tax-free shopping facilities in Terminal 2. Finally, cargo flights can only use the cargo terminal.

The length of the slot period also differs between selection slots, depending on the type of transport service they are used for. As we have seen in section 4.3, the acceptability of alternative arrival or departure times is determined by factors such as the timing of transport demand, frequency of transport services, and the importance of connections with other transport services. The freedom to specify selection slots is restricted by requiring multi-level slot specification. If connections are important, it is reasonable to restrict the slot period to a single bank of arriving or departing flights, which is a period of 1½ hour (primary periods). Good connections between transport services are only essential for most InterAir flights. InterAir has a system of three pairs of arrival and departure banks, which are spread equally throughout the day.

Table 8.2 illustrates the three-level hierarchy of slot periods proposed for Metropolitan Airport. The primary periods correspond with the arrival and departure banks of InterAir. With respect to the other transport services, we may distinguish between services with a frequency of 2 flights or less per day and higher frequency services with a frequency of up to 5 flights per day. The secondary periods for the latter services have a length of 3 hours, while lower frequency services will be allocated tertiary periods with a length of 6 hours each, unless a more specific timing is essential for these transport services.

Table 8.2: Selection slot periods at Metropolitan Airport

<i>time</i>	<i>slot period</i>		
	primary	secondary	tertiary
1:00 – 4:00		early night	night
4:00 – 7:00		late night	
7:00 – 8:30	arrival bank 1	early morning	morning
8:30 – 10:00	departure bank 1		
10:00 – 11:30		late morning	
11:30 – 13:00	arrival bank 2		
13:00 – 14:30	departure bank 2	early afternoon	afternoon
16:00 – 17:30	arrival bank 3	late afternoon	
17:30 – 19:00	departure bank 3		
19:00 – 22:00		early evening	evening
22:00 – 1:00		late evening	

Given the level of capacity scarcity experienced at Metropolitan Airport, selection slot allocation is desirable for most periods. However, for reasons of simplicity we will focus on the late afternoon period (16:00-19:00) in the remainder of this example. For

this period, the maximum efficiency selection slot allocation will be determined in subsection 8.1.3.

Specification of capacity constraints

Each combination of period and terminal corresponds with a separate primary basic slot, and at least one capacity constraint has to be formulated for each basic slot. Furthermore, two different types of capacity constraints apply to the selection problem of Metropolitan Airport. In the discussion about slot specification, we have seen that terminal traffic processes may be restrictive. Secondly, it was noted in the introduction that runway capacity may be restrictive, and finally environmental restrictions may apply, which are different for day and night periods.

In the past, the level of scarcity of available gates at Terminal 1 has been relieved by building Terminal 2 for domestic traffic. Nonetheless, growing traffic demand has resulted in scarcity of available gates at Terminal 1 during peak periods. Terminal 2 and the cargo terminal have sufficient capacity at the selection slot allocation level. Terminal 1 has 35 gates, which means that 35 planes can be served simultaneously by the terminal for boarding and alighting. Assuming that it is not necessary that all aircraft can be present at the terminal simultaneously, more than these 35 planes may be accommodated in each primary period. However, because at most gates passenger bridges are used for transport of passengers between aircraft and gate, aircraft usually occupy a gate for the entire period required for boarding and alighting, refueling, and other servicing activities. The time required for these activities, which is slightly more than the turnaround time of the corresponding aircraft, depends mainly on type of aircraft, but also on type of transport service. For instance, more than one long-distance aircraft using a gate in a 90-minute period is not realistic. However, it is assumed that two short-distance aircraft may use the same gate within a 90-minute period. Therefore, MATA (the allocation body) distinguishes four aircraft types, each with a different associated gate occupancy time. Table 8.3 summarizes the number of aircraft that may be accommodated by each gate in a 90-minute period.

Table 8.3: Maximum number of aircraft per gate per 90 min. period

<i>aircraft type</i>	<i>aircraft characteristics</i>		<i>aircraft per gate per period</i>	
	<i>typical aircraft</i>	<i>typical range</i>	<i>arrivals</i>	<i>departures</i>
A	Boeing 767	long distance	1	1
B	Boeing 757	medium distance	2	1
C	Airbus 320	medium/short distance	2	1
D	Beechcraft 200	short distance	2	2

Another bottleneck to airport traffic is the limited availability of runways for landings and take-offs. In 'standard' operation, only runways A and C can be used simultaneously. As is shown by Eurocontrol (2002), mixing arrivals and departures on both run-

ways is probably the best solution for an airport with two operational runways, unless there is a relatively large percentage of heavy aircraft. The capacity of two parallel runways is less than twice the capacity of a single runway, because of the reduced range of flight paths that may be followed in approach or after departure.

In a number of situations, runway B is used. Firstly, a strong sidewind on the parallel runways may require runway B to be used instead. Furthermore, runway B is used as the preferred runway for arrivals and departures between 22:00 hours in the evening and 7:00 in the morning to reduce noise hindrance during night hours. In this period, capacity is lowered in order to enable exclusive usage of runway B given normal weather conditions. Finally, the application of CDA (Continuous Descent Approach, see section 6.2) is propagated for arriving aircraft that have to fly over urban area, which results in a lower capacity than technically possible.

To determine the available capacity in the late afternoon period, the chosen reference situation is that runways A and C are in use, with CDA being applied to inbound air traffic. We conservatively assume that minimum separations between arrivals are restrictive for the arrival period, and minimum separations between departures for the departure period. The (potential) additional capacity resulting from a mix of arrivals and departures is not used in the selection slot allocation process. Given the application of CDA, the assumed headway between arrivals is 4 minutes. This leaves sufficient room for additional departures in the arrival period. Including buffer time, the assumed separation between each departure is 2.5 minutes for large aircraft (type A/B), and 1 minute for small aircraft (type C/D). Sufficient capacity for additional arriving aircraft, for instance aircraft from earlier or later periods that are delayed or ahead of time, is ensured by leaving some buffer capacity in both periods for additional arrivals.

Service time available for allocation

Given the semi-static nature of the selection slot allocation procedure, capacity is only partly available for allocation in this timetable season. About three-quarters of capacity is occupied by selection slots that have been allocated in previous seasons. Table 8.4 provides an overview of selection slots from previous seasons that occupy capacity at Terminal 1. Because four aircraft types are distinguished in the terminal capacity constraint, the selection slots from previous seasons are listed by desired aircraft type. Additionally, table 8.4 indicates whether selection slots have been allocated a primary or a secondary period. Primary arrival slots correspond to the arrival 3 period, primary departure slots correspond to the departure 3 period, and secondary arrival and departure slots correspond to the late afternoon period.

Table 8.4: Selection slots from previous seasons, listed by aircraft type (Terminal 1 only)

<i>aircraft type</i>	<i>slot period</i>	<i>number of selection slots</i>	
		<i>arrival</i>	<i>departure</i>
A	primary	4	5
	secondary	3	3
B	primary	2	0
	secondary	1	1
C	primary	12	13
	secondary	5	4
D	primary	6	9
	secondary	1	1

Slot requests regarding any terminal are included in the runway capacity constraints. Table 8.5 overviews the selection slots from previous seasons occupying runway capacity. Because aircraft type (small or large) is an important aspect in the runway capacity constraints, selection slots are distinguished by assumed aircraft type, and again slots are classified as either primary or secondary. Large aircraft are assumed to correspond with classes A and B, and small aircraft with classes C and D.

Table 8.5: Selection slots from previous seasons, listed by aircraft type (all terminals and runways)

<i>aircraft type</i>	<i>slot period</i>	<i>number of selection slots</i>	
		<i>arrival</i>	<i>departure</i>
large (A/B)	primary	6	5
	secondary	4	4
small (C/D)	primary	22	26
	secondary	8	9

The next step is to determine the available service time for each capacity constraint. As we have seen in chapter 6, the available service time is generally the product of length of slot period, bandwidth, and a reduction factor. The primary periods arrival 3 and departure 3 have a length of 90 minutes, and the secondary late afternoon period has a length of 180 minutes. The bandwidth of Terminal 1 is 35, and the bandwidth of the runways is 2. However, to increase reliability and flexibility, for instance with regard to additional flights, delays, and maintenance activities, a reduction factor is applied to all capacity constraints. The assumed bandwidth of Terminal 1 is reduced to 32, and the assumed availability of runways is reduced with 15 minutes for every hour. The relatively large reduction of runway availability is required to allow for additional arrivals that are delayed or ahead of time, or additional flights with high priority such as ambulance flights.

The available service time of each combination of basic slot and capacity constraint is listed in table 8.6. The available service time is calculated by multiplying (reduced) bandwidth with (reduced) period for each basic slot and each capacity constraint. Furthermore, an overview is given of the service time occupied by selection slots allocated in previous seasons, and the resulting service time that is available for allocation for the next season. Occupied service times are calculated by multiplying the numbers of selection slots from previous seasons (see tables 8.4 and 8.5) with the assumed minimum headways (for runways) or the assumed occupation periods (for terminals).

Table 8.6: Available service time per basic slot

<i>traffic network</i>		<i>service time (min.)</i>		
<i>element</i>	<i>period</i>	<i>total</i>	<i>occupied</i>	<i>net available</i>
Terminal 1	arrival 3	2880	1260	1620
	departure 3	2880	2025	855
	late afternoon	5760	4635	1125
runway	arrival 3	135	112	23
	departure 3	135	38.5	96.5
	late afternoon	270	217.5	52.5

Taking into account the differences in required service time per vehicle path between terminals and runways, we may conclude from table 8.6 that in the arrival period, runway capacity will be binding, while terminal capacity will be binding in the departure period. For the late afternoon period, it is not immediately clear from this overview whether either one of these or both capacity constraints are binding.

Identification and valuation of new slot requests

We now turn our attention to the slot requests that have been received by the Metropolitan Airport Traffic Authority (MATA). In order to be able to determine the capacity constraint parameter values, slot requests should include information about a number of attributes of their usage. Relevant attributes are, for instance, the slot period the request refers to, and desired aircraft type (A, B, C, or D). Furthermore, the objective function has to be determined as a function of the selection slot allocation decision, which means that values are assigned to slot requests.

The main objective of MATA is to optimize the utilization of the airport with respect to passenger and freight transport volumes. To estimate passenger and freight transport volumes resulting from alternative selections, MATA uses the expected transport capacity as a proxy, based on reports by carriers on expected aircraft types for each slot request. Additionally, information about realized occupancy rates in previous season for each carrier and type of transport service is used to estimate transport volumes corresponding with each slot request.

Table 8.7 lists the slot requests that have been received for the next timetable season. This overview includes for each slot request the expected value, which corresponds with the expected transport volume in passengers. Because none of the slot requests corresponding with this period is a freight transport service, conversion from freight to passenger volumes is not necessary.

Table 8.7: Slot requests

slot request	carrier	type	period	class	terminal	value
1	InterAir	arr.	arrival 3	A	T1	425
2	InterAir	arr.	arrival 3	A	T1	425
3	InterAir	arr.	arrival 3	C	T1	175
4	InterAir	dep.	departure 3	C	T1	175
5	InterAir	arr.	arrival 3	C	T2	105
6	InterAir	dep.	departure 3	C	T2	105
7	InterAir	arr.	arrival 3	D	T1	45
8	InterAir	dep.	departure 3	D	T1	45
9	InterAir	arr.	arrival 3	D	T2	40
10	InterAir	dep.	departure 3	D	T2	40
11	CityLink	arr.	late afternoon	C	T2	110
12	CityLink	dep.	late afternoon	C	T2	110
13	CityLink	arr.	late afternoon	D	T2	65
14	CityLink	dep.	late afternoon	D	T2	65
15	SunExpress	arr.	late afternoon	C	T1	160
16	SunExpress	dept.	late afternoon	C	T1	160
17	BudgetAir	arr.	late afternoon	C	T1	190
18	BudgetAir	dept.	late afternoon	C	T1	190
19	InterFlug	arr.	late afternoon	C	T1	120
20	InterFlug	dept.	late afternoon	C	T1	120
21	InterFlug	arr.	departure 3	C	T1	145
22	TTA	dept.	arrival 3	D	T1	100
23	TWA	dept.	departure 3	A	T1	425
24	Royal Air	arr.	arrival 3	A	T1	415
25	Royal Air	dept.	departure 3	A	T1	415
26	IslandLink	arr.	arrival 3	D	T1	35
27	IslandLink	dept.	departure 3	D	T1	35

8.1.3 Solving the selection slot allocation decision problem

In this sub-section, the selection slot allocation decision problem described in the previous sub-sections is formalized as a constrained optimization problem. Furthermore, given the information from the previous sub-section, a solution to this problem is found

using the extended greedy algorithm introduced in sub-section 7.4.2. It appears that the problem described here is solved after only two iterations of this algorithm.

Formalization of objective function and capacity constraints

The objective function is to maximize expected transport volumes. It is simply assumed that the total transport volume is the sum of expected transport volumes corresponding with each slot request. Hence, a linear objective function as introduced by equation 7.1 can be used, with u representing total value, ω slot request values (as listed in table 8.7), and x_i the decision variable describing the selection of each slot request i :

$$u(\mathbf{x}) = \sum_i \omega_i \cdot x_i \quad \forall i \quad (8.1)$$

As we have seen in the previous sub-section, three periods are distinguished, i.e. two primary periods and a secondary period. Furthermore, two capacity constraints have been formulated, i.e. regarding terminal capacity (of Terminal 1) and regarding runway capacity. Consequently, six capacity constraints k may be identified. Table 8.8 assigns a unique number k to each combination of basic slot (period) and capacity constraint type.

Table 8.8: Identification of capacity constraints

<i>basic slot period</i> [j]	<i>capacity constraint</i> [k]	
	Terminal 1	runways
arrival 3	11	21
departure 3	12	22
late afternoon	13	23

Equation 8.2 describes the structure of each capacity constraint k . In the capacity constraint, κ_k denotes the available capacity, which is assumed to be equal to the available time-space as listed in table 8.6. Furthermore, $\alpha_{i,k}$ denotes the weight of each slot request i in capacity constraint k , i.e. the assumed required service times as described in the previous sub-section:

$$\sum_i \alpha_{i,k} \cdot x_i \leq \kappa_k \quad \forall t \quad (8.2)$$

Solution procedure

The selection problem described in this section features a linear objective function, three basic periods and a higher-level period, and two different types of linear capacity constraints. Hence, it can be classified as a type-3 problem (see section 7.3). This type of problem can be solved with the extended greedy algorithm established in sub-section 7.4.2. Since we have seen that the terminal capacity constraint is likely not binding in the first primary period, and the runway capacity constraint is likely not binding in the second primary period, these capacity constraints will be assigned an initial shadow price z of 0, while all other capacity constraints will be assigned an initial shadow price

z of 1. Given the values of z , ω , β , and κ , an initial estimation of efficiency values e can be made by applying equation 7.18:

$$e_i = \frac{\omega_i}{\sum_k \frac{z_k \cdot \alpha_{i,k}}{\kappa_k}} \quad \forall i \in R \quad (7.18)$$

The derivation of initial efficiency values is illustrated by table 8.9. In this table, the arrival 3 period is referred to as period 1, departure 3 is period 2, and the late afternoon period is period 3.

Table 8.9: Derivation of initial efficiency values

slot request	period	value	terminal occupancy			runway occupancy			efficiency
			α_{11}	α_{12}	α_{13}	α_{21}	α_{22}	α_{23}	
1	1	425	90	0	90	4	0	4	1332
2	1	425	90	0	90	4	0	4	1332
3	1	175	45	0	45	4	0	4	615
4	2	175	0	90	90	0	1	1	955
5	1	105	0	0	0	4	0	4	420
6	2	105	0	0	0	0	1	1	5513
7	1	45	45	0	45	4	0	4	158
8	2	45	0	45	45	0	1	1	445
9	1	40	0	0	0	4	0	4	160
10	2	40	0	0	0	0	1	1	2100
11	3	110	0	0	0	0	0	4	1444
12	3	110	0	0	0	0	0	1	5775
13	3	65	0	0	0	0	0	4	853
14	3	65	0	0	0	0	0	1	3413
15	3	160	0	0	45	0	0	4	1446
16	3	160	0	0	90	0	0	1	1818
17	3	190	0	0	45	0	0	4	1717
18	3	190	0	0	90	0	0	1	2159
19	3	120	0	0	45	0	0	4	1084
20	3	120	0	0	90	0	0	1	1363
21	2	145	0	45	45	0	4	4	916
22	1	100	45	0	45	1	0	1	1031
23	1	425	90	0	90	2.5	0	2.5	1887
24	2	415	0	45	45	0	4	4	2622
25	1	260	90	0	90	2.5	0	2.5	1154
26	1	35	45	0	45	4	0	4	123
27	2	35	0	45	45	0	1	1	346

The next step is to rank the slot requests, depending on their efficiency values. Following the procedure described in figure 7.5, the addition of each slot request to the selection set is evaluated sequentially, in order of its rank. As is shown in table 8.10, only slot requests 3, 5, 7, 9, and 26 are not included in the proposed selection.

Table 8.10: Resulting selection

slot request	iteration 1			iteration 2		
	efficiency	rank	selection	efficiency	rank	selection
1	1332	13	1	5578	7	1
2	1332	13	1	5578	7	1
3	615	21	0	2297	16	1
4	955	18	1	9188	2	1
5	420	23	0	1378	23	0
6	5513	2	1	5513	9	1
7	158	26	0	591	25	0
8	445	22	1	2363	15	1
9	160	25	0	525	26	0
10	2100	6	1	2100	17	1
11	1444	11	1	1444	22	1
12	5775	1	1	5775	6	1
13	853	20	1	853	24	0
14	3413	3	1	3413	13	1
15	1446	10	1	2100	17	1
16	1818	8	1	8400	4	1
17	1717	9	1	2494	14	1
18	2159	5	1	9975	1	1
19	1084	16	1	1575	21	1
20	1363	12	1	6300	5	1
21	916	19	1	1903	19	1
22	1031	17	1	5250	12	1
23	1887	7	1	8925	3	1
24	2622	4	1	5447	11	1
25	1154	15	1	5460	10	1
26	123	27	0	459	27	0
27	346	24	1	1838	20	1

As is illustrated by table 8.11, the only binding capacity constraint after the first iteration is the runway capacity constraint of period 3 (i.e. $k = 23$). For this capacity constraint, the remaining available service time is only 0.5 minutes, while the non-selected slot requests require 4 minutes each. The remaining available service times for other capacity constraints are well above the values required to facilitate additional slot requests. Hence, the procedure has to be repeated with updated z values. After the second

iteration, the slot requests excluded from the selection are numbers 5, 7, 9, 13, and 26. The runway capacity constraint is still binding, which implies that the algorithm has converged after 2 iterations.

Table 8.11: Remaining available service time after each iteration

<i>capacity constraint</i>		<i>remaining available service time</i>	
<i>type</i>	<i>period</i>	<i>iteration 1</i>	<i>iteration 2</i>
Terminal 1	arrival 3	1215	1170
	departure 3	9	5
	late afternoon	585	585
runways	arrival 3	83.5	83.5
	departure 3	45	0
	late afternoon	0.5	0.5

Alternative specifications of the slot allocation problem

The slot requests that have not been selected, given the decision procedure outlined above, are all arrivals. In practice, selecting slots for departing flights, without selecting corresponding slots for arriving flights, may cause logistic problems, in particular for carriers who have their home-base elsewhere. An alternative is to consider pairs of arrivals and departures as single slot requests, and to treat the late afternoon period under consideration as a single period, without distinguishing between arrival and departure banks. Table 8.12 gives an overview of the resulting selection given this alternative approach, which has been obtained after two iterations of the extended greedy algorithm.

Table 8.12: Alternative selection

slot request	value	terminal occupancy	runway occupancy	rank efficiency	selection
1	425	90	4	3	1
2	425	90	4	3	1
3/4	350	135	5	7	1
5/6	210	0	5	11	1
7/8	90	90	5	14	0
9/10	80	0	5	15	0
11/12	220	0	5	10	1
13/14	130	0	5	13	0
15/16	320	135	5	8	1
17/18	380	135	5	6	1
19/20	240	135	5	9	1
21	145	45	4	12	1
22	100	45	1	5	1
23	425	90	2,5	1	1
24/25	830	135	6,5	2	1
26/27	70	90	5	16	0

8.2 Consequences of implementation of selection slot allocation

Having discussed the implementation of a selection slot allocation regime as proposed in this thesis to a hypothetical case, the possibilities to implement this in real practice are discussed in this section. We first focus on the main changes compared with current practice resulting from the implementation of the ideas outlined in this thesis, i.e. the separation of selection and scheduling in the air sector and the extension of framework agreements to a full selection slot allocation system in the rail sector. Next, the issue of cross-border traffic is discussed. In the case of cross-border traffic, two or more allocation bodies maybe responsible for the same slot requests. Furthermore, each country may choose a different slot allocation regime. In these cases, collaboration between allocation bodies is required, and deviations from the slot allocation system outlined in this thesis may be necessary. Finally, we will discuss the main consequences of the implementation of the ideas presented in this thesis with respect to the issues introduced in section 3.3, i.e. stability and flexibility, diversity and competition, and heterogeneity.

8.2.1 Main changes given implementation of selection slot allocation

Implementing the selection slot allocation system proposed in this thesis implies changes with respect to the current slot allocation regimes in the rail and air sectors. Especially the separation of selection and scheduling slot allocation has major consequences for the setup of slot allocation systems. The European railway sector currently

introduces a system of 'framework agreements', which is similar to the idea of selection slot allocation, but introducing selection slot allocation would imply a profound change of the airport slot allocation regime.

Railway selection slot allocation

There are a number of similarities between selection slot allocation as proposed in this thesis and the system of framework agreements that is currently being introduced in the railway sector. As we have seen in section 3.2, framework agreements have the function to ensure the continuity of transport services for a period of several years, without fixing their timetable for this period. The main difference, however, is that framework agreements will be concluded bilaterally between the allocation body and the carrier, authority, or other interested party requesting this framework agreement. An explicit evaluation of competing selection slot requests, as has been proposed in this thesis, has not been included in the new European slot allocation regime as described by Directive 2001/14/EC (EC, 2001c). It is even not clear to which extent framework agreements may be used to solve problems of absolute capacity scarcity, given the provision that *'framework agreements shall not be such as to preclude the use of the relevant infrastructure by other applicants or services'* (EC, 2001c, art. 17). If a regular situation of absolute capacity scarcity, i.e. resulting from a capacity conflict between regular transport services, cannot be solved at the level of framework agreements, these capacity conflicts will emerge in the annual timetable allocation process.

Solving absolute capacity scarcity problems at the selection slot allocation level (rather than in the annual timetable process) is desirable to ensure the stability of transport service patterns desired by carriers and other interested parties. A possible alternative is the usage of minimum frequency criteria (which is current practice in the Netherlands), but this may only solve all capacity conflicts if a complete *a priori* selection slot allocation is made. The latter implies that the allocation body (or, for instance, the Minister of Transport) produces a selection slot allocation that is not based on slot requests but on the desired transport service pattern as perceived by the decision-maker. This implies a shift of responsibility for transport network design towards either the allocation body or the authorities, which is against the common idea that the supply of transport services should primarily be determined by carriers and end users. Therefore, we may conclude that in situations with absolute capacity scarcity, it is desirable to extend the system of framework agreements to a full selection slot allocation system. A full selection slot allocation system may be implemented in the railway sector by introducing a two-stage annual slot allocation procedure. In the first stage, new selection slot requests (requests for framework agreements) are evaluated, and in the second stage the new scheduling allocation is determined.

Another consequence of the implementation of selection slot allocation as proposed in this thesis is that it introduces a clear framework to evaluate alternative selections, in the

context of heterogeneous railway usage. Given assumptions about composition-of-traffic, capacity occupancy levels can be determined for each slot request. This means that differences between slot requests with respect to value and capacity occupancy can be evaluated explicitly. Furthermore, the framework introduced in this thesis is able to analyze different types of capacity constraints simultaneously. For instance, the capacity occupancy of one slot request may depend on a primary capacity constraint, while the capacity occupancy of another slot request regarding the same basic slot may depend on an environmental capacity constraint. This cannot be achieved by simply applying priority criteria, which is current practice in the Dutch railway sector.

Airport selection slot allocation

As we have seen in chapter 3, the principle of historic rights with respect to slots has been chosen for airports. No distinction is made between selection and scheduling slot allocation, and hence the principle of historic rights also applies at the level of scheduling slots. However, this thesis proposes to distinguish between multi-year selection slots on one hand in order to achieve stability, and seasonal scheduling slots without historic prejudice on the other hand in order to achieve flexibility. Implementing this system would mean a radical change with respect to the IATA slot allocation system, which is the basis of the EU slot allocation system as well. Instead of one type of airport slots, two types would need to be distinguished, both of which may have a limited validity. Selection slots would have a validity of at least 5 years, while scheduling slots would still have a validity of a single timetable season (i.e. a half year).

As is the case in the rail sector, priorities are often applied in current practice to solve capacity conflicts at the selection level. This thesis, however, offers a systematic framework to evaluate alternative allocations. Implementing the ideas from this thesis means that the current system based on priorities is replaced by an evaluation based on efficiency. This means that data are required on the value of slot requests, given the objectives of the allocation body. Furthermore, airport capacity should be modeled adequately with linear capacity constraints. This means that heterogeneity of traffic is included in the evaluation, while in current practice often a simple capacity constraint is assumed, i.e. without differences in capacity consumption. Finally, the framework introduced in this thesis enables taking into account different types of capacity constraints simultaneously, e.g. runway capacity constraints and environmental capacity constraints. For instance, the priority assigned to a slot request corresponding with a noisy type of aircraft may depend on runway capacity occupation, environmental capacity occupation, or both, depending on which capacity constraints are binding.

8.2.2 Legal competence and slot allocation

The implementation of selection slot allocation as proposed in this thesis requires legal changes. For instance, implementation in the Dutch railway sector requires the current

decree regarding railway slot allocation to be changed, because the current decree demands a different approach. Furthermore, legal experts should determine to which extend the implementation of selection slot allocation is permitted by current European and national legislation. The implementation of selection slot allocation to airports is clearly not in line with current European legislation, and this means that implementation in the EU requires a profound change of European legislation with respect to airport slot allocation.

Another issue is that different allocation bodies may be responsible for different parts of the traffic network. Hence, slot requests regarding cross-border traffic may be evaluated by two or more allocation bodies, which may apply different allocation procedures. In this sub-section, we review a few problems that may emerge with respect to cross-border traffic.

If two or more allocation bodies are responsible for the same slot request, it is possible that they disagree about the acceptance or rejection of this slot request. A possible solution in this case is that the allocation bodies agree about a joint decision procedure, which seeks the most efficient solution for the total network, for instance using the approach proposed in section 7.5 for network problems. Alternatively, the allocation bodies may negotiate about cross-border slot requests, and mediation may be used to support this negotiation process.

A related issue is the limited jurisdiction of national authorities as the Dutch government, and of international authorities such as the EU. The previous suggestions to decide about cross-border slot requests may not work in situations where, for instance, selection slots allocation is applied in one country, and historic rights are applied in another country. This problem, for instance, would arise if the European Union implemented selection slot allocation for airports, and other countries would continue to apply the IATA slot allocation system. In this case, the most likely solution is that new EU slot allocation system will be applied to EU 'domestic' flights only, while historic rights will still be applied to other flights. Unilateral abolition of the IATA system will probably not be accepted by other countries.

8.2.3 Main consequences of implementation of selection slot allocation

The implementation of selection slot allocation in practice has a number of consequences. This sub-section reviews the consequences with respect to the main issues introduced in section 3.3. These issues are stability and flexibility, competition and diversity, and heterogeneity.

In the first place, the separation of selection and scheduling slot allocation implies a choice for a relatively high level of stability at the selection level, and a relatively high level of flexibility at the scheduling level. Introducing selection slot allocation as pro-

posed in this thesis would result in more flexibility than in a situation with historic rights, and more stability than in a situation with only minimum frequencies. We may expect that the implementation of selection slot allocation would result in an improved balance between stability and flexibility in both sectors. Flexibility, in particular at the scheduling level, is increased with respect to airport slot allocation by abolishing historic rights at this level, without significant loss in stability. Stability is increased in the rail sector by explicitly solving absolute capacity scarcity problems at the selection slot allocation level, while framework agreements are not suitable to solve these capacity conflicts due to their bilateral nature.

Another issue that has been reviewed in section 3.3 is the balance between competition and diversity. The priority criteria currently used in the rail and air sectors include criteria with respect to diversity and competition. For instance, in certain situations entrants have priority in both sectors, and new destinations may be given priority over other destinations in the airport slot allocation procedure. In the framework introduced in this thesis, diversity may be included in the valuation of alternative selections. The proposed method to implement criteria with respect to competition is the inclusion of satisficing constraints. This approach can be used to ensure a sufficient potential for competition by allocating slots used for competing transport services to different carriers. However, authorities and allocation bodies should realize that strict satisficing constraints with respect to competition would result in a lower level of diversity. Attaining a good balance between diversity and competition is the responsibility of decision-makers, i.e. authorities and allocation bodies.

Finally, the approach to model traffic supply described in chapter 6 has consequences for the heterogeneity issue. The priority criteria used in current practice are not suitable to deal with differences in capacity occupancy of slot requests in different situations. In chapter 6, a generic approach has been introduced to model traffic supply with linear capacity constraints. Even in situations where, in principle, capacity constraints are non-linear, it appears to be possible to make a reasonable linear approximation, if only a reasonable estimation of composition-of-traffic can be made. Given this approach, the capacity occupancy of each slot request may be estimated, and the efficiency approach introduced in chapter 7 provides a generic framework to assess the value per unit of capacity associated with each slot request. The implementation of the efficiency approach offers a simple but powerful framework to compare slot requests given heterogeneous traffic.

8.3 Synopsis

In this chapter, a hypothetical case has been reviewed to illustrate the selection slot allocation approach proposed in this thesis. We focused on the allocation of selection slots in the morning period. The example problem of a metropolitan airport shows that the approach proposed in this chapter can handle selection problems with a realistic

level of complexity. Despite the relative complexity of the case, with environmental regulation limiting runway capacity, two types of (linear) capacity constraints are sufficient, i.e. with respect to runway and terminal capacity. Furthermore, two levels of selection slots have been identified. The solution procedure converged in only two iterations, which may be interpreted as an indication of the robustness of this solution approach.

The implementation of the ideas outlined in this thesis about the introduction of separate selection slot allocation has consequences for current practice. In the railway sector, the system of framework agreements that is currently being introduced in Europe should be extended with a system to evaluate competing slot requests. In the aviation sector, a separate selection slot allocation level should be introduced for airports, and according to the ideas outlined in this thesis historic rights should be abolished, especially at the scheduling level. Finally, the implementation of the framework outlined in this thesis would imply that the simple priority criteria of current practice would be replaced with a more sophisticated approach. Contrary to the current approach with priority criteria, the selection slot allocation decision approach introduced in this thesis takes into account heterogeneity of traffic and enables simultaneous analysis of different types of capacity constraints.

In order to implement the desired changes of slot allocation regimes, legal changes are required. The current slot allocation regimes in Europe are largely prescribed by European and national legislation, and therefore the implementation of selection slot allocation in Europe as proposed in this thesis will require adaptation of legislation at both the European and the national levels. However, each authority, e.g. the European Union, has a limited jurisdiction, and hence selection slot allocation cannot be introduced unilaterally to cross-border traffic.

In section 3.3, three main issues have been introduced, and in this chapter the consequences of the implementation of the ideas outlined in this thesis with respect to these issues have been reviewed. As was already concluded in chapter 3, the separation of selection and scheduling slot allocation is desired to attain an optimal balance between stability and flexibility of transport service patterns. Therefore, we may expect that the implementation of selection slot allocation would result in an improved balance between stability and flexibility of transport service patterns in both the rail and the aviation sector. Furthermore, the framework to analyze the selection problem introduced in this thesis may be used to optimize diversity, given a proper objective function, while maintaining a sufficient level-of-competition by formulating satisficing constraints. Finally, the application of linear capacity constraints as proposed in chapter 6 is a simple but adequate approach to deal with heterogeneity.

9

CONCLUSIONS AND RECOMMENDATIONS

The final chapter of this thesis summarizes the conclusions and contributions, and provides recommendations for the implementation of the ideas outlined in this thesis and for further research. First, the contents of this thesis are summarized, which gives the reader a brief overview of the problem studied in this thesis. A summary of the main conclusions of this thesis is provided by the second section. The third section gives a brief overview of the main contributions of this thesis to transportation science. In the fourth section, recommendations with respect to the implementation in practice of the ideas developed in this thesis are given, as well as recommendations for further research on (selection) slot allocation.

9.1 Short summary

In this thesis, transport infrastructure slot allocation has been studied, focusing on selection slot allocation, i.e. on longer-term slot allocation decisions determining the traffic patterns served by infrastructure bottlenecks, rather than timetable-related slot allocation problems. The allocation of infrastructure capacity among carriers is a major issue in various transport infrastructure sectors, and therefore a theoretical framework on slot allocation would be desirable to support rational decision-making on slot allocation. The current state-of-the-art of slot allocation research does not provide such a theoretical framework, and therefore a theoretical framework to analyze slot allocation problems has been developed in this thesis.

The first step in the development of a theoretical framework to analyze slot allocation problems has been the specification of a conceptual framework, which includes the definition of key concepts such as capacity. Capacity has been defined as being depend-

ent on conditions such as composition of traffic and traffic context as well as assumptions about the desired balance between capacity and quality-of-service. The next step was to review the current application of slot allocation in the railway and aviation sectors, and the potential application of slot allocation in the road and navigation sectors. Slot allocation is currently applied in the railway and aviation sectors, and slot allocation may potentially be applied in other sectors.

This thesis introduces the important distinction between selection and scheduling slot allocation. In both the railway and aviation sectors, the tradition has been to integrate selection and scheduling slot allocation. This thesis, however, considers selection slot allocation as a separate slot allocation level. Separating selection and scheduling slot allocation enables the application to each level of different rules with respect to slot validity, valuation of alternative slot requests, etc.

The desired characteristics of selection slot allocation have been formulated in this thesis by analyzing the main desires of carriers and other interested parties such as shippers and authorities. It has been concluded that selection slots should be valid for a significantly longer period than scheduling slots, and a semi-static slot allocation procedure has been proposed. Furthermore, the acceptability principle has been introduced as a basis to specify desired slot size. However, the specification of standard basic slots by the allocation body (at different levels to attain differentiation of slot size) is desirable.

A semi-static slot allocation procedure implies that selection slot allocation decisions may be based on an explicit evaluation of selection slot requests. The selection problem may be analyzed using congestion theory, resulting in a generic specification of traffic supply and demand. The next step is to specify traffic supply in more detail by specifying capacity constraints. Examining various types of primary traffic processes, traffic service processes, and traffic externalities, capacity constraints have been formulated, which may be applied to different types of bottlenecks. Three categories of capacity constraints have been distinguished, i.e. homogeneous capacity constraints, linear capacity constraints, and non-linear capacity constraints. The next step is the specification of objectives. The (primary) objective of slot allocation may usually be specified as a linear objective function.

Depending on the type and number of capacity constraints, various instances of selection slot allocation decision problems may be formulated. The corresponding optimization problems may be solved using an exact solution algorithm, but for various reasons this thesis proposes a greedy approximation instead. Besides a standard greedy algorithm for selection problems with a single type of capacity constraint, an extended greedy algorithm has been developed to solve problems with two or more different types of capacity constraints. The latter algorithm has been tested for a hypothetical case study.

9.2 Main conclusions of this thesis

Having briefly reviewed the contents of this thesis, our next step is to summarize the conclusions. Three main conclusions are discussed in this section. The first main conclusion is that selection and scheduling should be considered as separate slot allocation levels having a hierarchical relationship. Selection slot allocation is of primary importance and scheduling slot allocation is only of secondary importance, because selection decisions determine which traffic is facilitated and which is not. The second main conclusion is that the validity of slots is a compromise between stability and flexibility. To ensure a sufficient level of stability, a validity of at least 5 years seems reasonable for selection slots. To ensure a reasonable level of flexibility, infinite validity of selection slots (historic rights) is not desirable, and at least every timetable season the opportunity should be offered to reserve selection slots. The final main conclusion is that the objectives and constraints of the selection problem can be modeled as linear functions, and the resulting binary linear programming problem can best be solved with the greedy efficiency algorithm presented in this thesis. This efficiency algorithm does not provide an exact solution of the binary linear programming problem, but its results are more robust and are easier to interpret than exact solution approaches.

Slot allocation levels

In this thesis, three levels of slot allocation have been distinguished, i.e. selection, scheduling, and *ad hoc* slot allocation. Which slot allocation levels are useful to implement depends on the types of transport services to be facilitated. Selection slot allocation is mainly desirable for structural transport services, i.e. transport services that are provided with regular intervals over a longer period of time. In both the railway and aviation sectors, scheduling slot allocation is currently the dominant level of slot allocation, while selection slot allocation is relatively underdeveloped in these sectors.

For a number of reasons, it is desirable to separate selection slot allocation from scheduling slot allocation. Firstly, it appears that discussions about the design of slot allocation regimes would be much clearer if the differences between selection and scheduling decisions were recognized explicitly. Another reason to separate these levels of slot allocation is that scheduling slot allocation is not required for demand-driven structural transport services. Introducing selection slot allocation as a separate planning stage avoids that the carriers offering this type of transport service have to plan these services in advance with much more detail than desired. Finally, separating selection and scheduling slot allocation is required to attain a good balance between stability and flexibility. The desired stability of transport service patterns is ensured by making the validity of selection slots significantly longer than the usual timetable period. Allowing relatively frequent optimization of timetables is required to attain a sufficient level of flexibility. The demand for selection slots is largely due to the demand for stability of transport service patterns. By separating selection and scheduling, a different balance between stability and flexibility may be found for selection and scheduling slot allocation. Con-

sequently, selection and scheduling should be formulated as separate problems in a hierarchical relationship, instead of a single integral problem.

Desired characteristics of selection slot allocation

Given the demand for stability, we have concluded that selection slots should generally have a validity of at least 5 years, while scheduling slots usually have a validity of only a half year or a full year. The application of historic rights, which is current practice in air traffic, implies the application of infinite validity, which has the disadvantage of low flexibility and impairing the introduction of new transport services. However, a dynamic or semi-static selection slot allocation procedure is desired to enable flexible planning of new transport services. Given semi-static decision-making, selection slot requests regarding new transport services may be submitted every timetable season before the timetables are established. This enables simultaneous evaluation of slot requests, where earlier slot allocation decisions should be respected as long as these are still valid. Additionally, dynamic decision-making is preferred to allow selection slot requests for demand-driven transport services throughout the timetable season.

Selection problem

Based on economic theory regarding transportation and congestion, two types of capacity scarcity have been identified, i.e. relative and absolute capacity scarcity. Selection slot allocation is a solution for traffic selection markets facing absolute capacity scarcity. Absolute capacity scarcity is a situation where no natural equilibrium between traffic supply and demand can be reached because traffic supply is still well below traffic demand when traffic supply reaches capacity. A possible solution is to apply scarcity tolls to bridge the gap between traffic supply and demand. However, given imprecise knowledge of traffic demand, conservatively estimated scarcity tolls will usually not be sufficient to bridge the gap between traffic supply and demand. A better solution is to apply selection slot allocation in this case.

Although traffic processes are often complex and non-linear, traffic supply can largely be modeled by linear capacity constraints at the level of detail required to analyze selection slot allocation. Non-linear capacity constraints that may be identified in theory may usually be reformulated as linear capacity constraints, given that realized composition-of-traffic will not be very different from the assumed composition-of-traffic. Since selection slot allocation decisions will often be made in a situation where many selection slots have already been allocated previously, the main gross characteristics of composition-of-traffic indeed will be known beforehand. Given the latter assumption, it is also reasonable that a linear objective function can be formulated that adequately represents the (primary) objective of the decision-makers, i.e. the allocation bodies and the authorities who are responsible for designing slot allocation regimes.

Given the argument that a linear objective function and linear capacity constraints may be applied in most practical situations, almost all selection problems arising in practice may be formulated as a binary linear programming problem. Although this problem may be solved with an exact solution algorithm, an efficiency-based greedy approximation algorithm is preferred for the selection problem because the latter approach is more robust with respect to small changes in the capacity value. The main advantage of this greedy algorithm is its straightforward interpretation: for each slot request, its efficiency determines its ranking. Efficiency is a measure of the amount of contribution of each slot request to the objective value, divided by the amount of (binding) capacity that is 'occupied' by this slot request. The greedy algorithm evaluates the feasibility of adding slot request to the selection set in sequence of their ranking, until all slot requests have been evaluated. This algorithm may easily be extended to include satisficing constraints with respect to minimum claim rights and level-of-competition. The example problem of a metropolitan airport shows that for selection problems the approach proposed in this thesis can adequately handle problems with a realistic level of complexity.

9.3 Main contributions of this thesis

Having reviewed the main conclusions of this thesis, we now focus on the main contributions of this thesis to the state-of-the-art of transportation science. These main contributions have already been previewed in sub-section 1.3.2, but are repeated here to give a complete overview.

In the first place, this thesis contributes to the establishment of a sector-independent theoretical framework to analyze slot allocation problems, which is in accordance with the main objective of this thesis. A conceptual framework with respect to slot allocation has been developed, which can be used to describe and analyze slot allocation systems in a systematic way. This conceptual framework has been applied to analyze the current (European) practice with slot allocation in the air and rail sectors. Based on this analysis, three main issues with respect to the design of slot allocation systems have been formulated, which appear to cover the dominant current discussions about desired improvements of slot allocation systems. The theoretical framework established in this thesis may be used to design slot allocation systems, by finding a balanced solution to these issues. For instance, separating selection and scheduling slot allocation, as is proposed in this thesis, enables both a high level of stability with respect to long-term decisions, and a high level of flexibility with respect to short-term slot allocation decisions.

In the second place, this thesis contributes to the state of the art of (analytical) transportation research. A unified approach to formulate capacity constraints of different types of bottlenecks has been introduced, which is valid for all kinds of transportation sectors. This multiple user-class approach can deal with heterogeneous traffic, and it can be applied to formalize transportation problems such as the optimal selection slot alloca-

tion decision problem, which is another main contribution of this thesis. The most likely instances of the selection problem have been specified as 'standard' optimization problems, which implies that any exact (or approximate) solution algorithm described in the literature may be applied to these problems. However, it has been argued in this thesis that a greedy approximation algorithm is most suitable for this type of problem. We have seen that for situations with only a single capacity constraint, an existing basic greedy algorithm can be used. Furthermore, an extended greedy algorithm has been developed in this thesis that can be applied in situations where two or more different capacity constraints apply to the same bottleneck.

Finally, this thesis contributes to discussions about slot allocation systems in practice. In the first place, the theoretical framework developed in this thesis may contribute to a transparent discussion about objectives, alternatives, complications, etc. In current discussions, the objectives are not always clear, and the arguments in favor of the chosen approach are often fragmented. A first contribution of this thesis is distinguishing different slot allocation problems at different levels. Explicitly distinguishing selection from scheduling slot allocation, as proposed in this thesis, enables decision-makers to make different choices with respect to slot validity, priorities, etc., at different levels of slot allocation. The selection of traffic in situations where traffic supply is insufficient to accommodate all traffic demand with a reasonable quality-of-service is identified here as the essence of the slot allocation problem. Furthermore, this thesis has introduced a systematic approach to evaluate alternative selection slot allocation decisions, which may be used as a guideline to redesign slot allocation systems. The main strength of this approach is that it is relatively simple and easy to interpret, while it is also suitable for bottlenecks that are often considered as relatively complex. Contrary to the simple priority rules applied in current practice, it offers a systematic approach to deal with heterogeneity of traffic and with different types of capacity constraints applying to the same bottleneck.

9.4 Recommendations

One of the main contributions of this thesis is that it helps to structure discussions about the application of slot allocation in practice by providing a theoretical framework regarding slot allocation and by introducing a systematic approach to evaluate alternative selection slot allocation decisions. Furthermore, the implementation of the ideas about slot allocation outlined in this thesis has been discussed in section 8.2. In this section a number of recommendations are given with respect to the implementation of selection slot allocation in practice. Finally, the main opportunities and desires for further research are analyzed in the second part of this section.

9.4.1 Recommendations about slot allocation in practice

In his thesis a number of ideas about desired adaptations of slot allocation systems have been outlined. Our recommendations focus on three issues, i.e. separation of selection and scheduling slot allocation, optimization of infrastructure utilization, and slot pricing.

Separation of selection and scheduling slot allocation

One of the main ideas outlined in this thesis is that selection and scheduling slot allocation should be treated separately. In the first place, selection and scheduling slot allocation should be separate decision stages. Furthermore, in discussions about slot allocation regimes, selection and scheduling should be treated as separate problems. If not, arguments about the desired characteristics of slot allocation regimes may become confused, in particular discussions regarding the desired balance between stability and flexibility, and discussions with respect to the value to be attached to alternative slot allocations.

In the railway sector, the recently introduced system of framework agreements can best be extended to a full system of selection slot allocation. The current system of bilateral conclusion of framework agreements between carriers (or other interested parties) and allocation bodies corresponds with a dynamic allocation procedure, without explicit evaluation of conflicting interests of different railway infrastructure capacity users. In situations of absolute capacity scarcity there should be a fair and 'objective' procedure to solve capacity conflicts, which is currently not the case. In the case of absolute capacity scarcity, it is likely that dynamic selection slot allocation will result in a situation where incumbents can easily conclude new framework agreements when their current agreements expire, which is advantageous to incumbents but disadvantageous to newcomers. Furthermore, it is desirable that, in principle, all capacity conflicts resulting from absolute capacity scarcity are solved at the selection slot allocation level, which increases the possibility that negotiation between carriers will be sufficient to solve capacity conflicts at the scheduling slot allocation level.

With respect to slot allocation in the aviation sector, the primary recommendation of this thesis is to abolish the principle of historic rights, at least at the scheduling level. Furthermore, to attain a better balance between stability and flexibility, separation of selection and scheduling slot allocation is desirable. Given the tradition of slot allocation based on international cooperation between carriers in the IATA, it is possible to apply self-regulation to slot allocation at the scheduling level. With respect to selection slot allocation, however, independent allocation bodies applying fair and objective allocation criteria are required, given that selection slot allocation decisions have a significant impact on the diversity of supply of air transport services and on the level of competition between air transport services.

Optimization of infrastructure utilization

Another recommendation is to use a systematic procedure such as the optimal selection slot allocation decision approach introduced in this thesis to optimize the utilization of infrastructure capacity. In current practice, historic rights (air) and minimum claim rights (rail) are the dominant mechanisms to solve absolute capacity scarcity problems. In both cases, carriers will probably optimize the utilization of their slots internally, but there are no incentives to free capacity if another carrier can use it more efficiently. Therefore, a systematic procedure should be applied to evaluate alternative slot allocation decisions, at least at the selection level, in order to optimize infrastructure utilization.

The analysis of current bottlenecks in this thesis provides a number of suggestions how the utilization of infrastructure capacity may be enhanced, assuming that this primary objective dominates over secondary objectives such as attaining a desired level of competition in the transport market. The first example is the duplication of transport services such as analyzed in Appendix D. It appears that (in aviation) two carriers may offer competing services with similar departure times between the same pairs of airports. A higher utilization value may probably be achieved by offering more different transport services. Another example is that more homogenization of railway traffic may be possible, for instance by slowing down fast trains, or speeding up slow trains by applying improved rolling stock.

Assuming that better utilization of infrastructure capacity will often be possible, proposed infrastructure capacity expansion schemes should always be evaluated against the alternative of optimized utilization of existing capacity. For instance, a hub airport may argue that expansion is necessary in order to improve its hub position. However, an alternative way to create more room for hub-and-spoke flights is to shift charter flights to another airport.

Slot pricing

We have seen in chapter 5 that pricing may be used to reduce traffic demand in situations of capacity scarcity. According to many economists, pricing is usually the best mechanism to attain efficient allocation of scarce resources such as infrastructure capacity. It has been argued in chapter 5 that auction techniques may be applied to determine the optimal level of pricing. This means that pricing is used to determine the value of slot requests, which can be used as input for the slot allocation decision procedure proposed in this thesis. However, policy makers should be aware that the selection mechanism resulting from applying this market mechanism to the traffic market does not necessarily correspond with their objectives. For instance, the allocation of capacity among different types of uses (i.e. different types of transport services) may be significantly different from the allocation desired from a public point of view. Another example is that government objectives regarding competition in transport markets may be

frustrated if large incumbent carriers appear to be able to block competition by buying up (virtually) all selection slots. Therefore, we may conclude that applying pricing mechanisms to determine selection slot values is an option that should be carefully considered by policy makers and evaluated against alternative ways to determine slot values. When applying selection slot auctions, it is advisable to ensure a reasonable level of competition and a reasonable diversity of transport services by applying satisficing criteria as proposed in chapter 7 of this thesis. Finally, apart from the question of whether pricing should be used to determine the value of selection slots, it is reasonable to introduce at least a reservation fee in order to cover the costs of the allocation body.

9.4.2 Recommendations for further research

This section provides suggestions for further research regarding the selection slot allocation decision approach outlined in this thesis, new slot allocation systems, and the evaluation of infrastructure capacity expansion schemes against the alternative of improved bottleneck utilization.

Selection slot allocation decision approach

With respect to a number of aspects of the selection slot allocation decision approach proposed in this thesis, further research is desired. In the first place, further research is desired with respect to the formulation of objective functions. This includes research about the linearity of objective functions and the desirability to introduce the possibility to submit related slot requests as 'packages' that together have to be accepted or rejected. Secondly, more research is desired about the formulation of capacity constraints. The correctness of the capacity constraint formulations proposed in chapter 6 may be tested using traffic data from the corresponding types of bottlenecks, or using simulation studies. Finally, an interesting extension to the selection slot allocation decision approach introduced in this thesis would be to include the desired balance between capacity and quality-of-service in the analysis, instead of assuming a fixed capacity value.

Another subject for further research is the extended greedy algorithm developed in this thesis to solve type-3 selection slot allocation problems. More information is desired about the characteristics of this solution approach, in particular the uniqueness of the solutions (i.e. are there different solutions possible corresponding with different z values) and the convergence of the algorithm (i.e. can convergence be guaranteed in all circumstances). Furthermore, the application of the extended greedy algorithm to network problems (sub-section 7.5.3) may be evaluated in further research. For instance, the possibilities suggested in this thesis to extend the single-bottleneck approach adopted in this thesis to network problems may be tested for a number of realistic problem cases.

The selection problem specified in this thesis has been tested in a hypothetical case study. More convincing, however, would be experimentally testing the ideas outlined in this thesis. For instance, a simulation game may be applied to test the functioning of a proposed slot allocation regime in a simulation environment, ideally with roles being played by representatives of real-life actors such as carriers and allocation bodies (see e.g. Mayer & Veeneman, 2003; Nilsson, 1999).

New slot allocation systems

In chapter 3 of this thesis, a number of new applications of slot allocation have been discussed. In the aviation sector, for instance, slot allocation at the scheduling and selection level is only applied to airports. However, we have seen in chapter 6 that the capacity of Air Traffic Control (ATC) centers may be a bottleneck in aviation. If this problem increases in the future, extension of the current ATC slot allocation system, which only allows reservation of slots shortly before departure, may be considered. Structural delays due to ground holding may be avoided by extending the scheduling slot allocation system for airports to a full scheduling slot allocation system that also solves capacity conflicts at the ATC level. Further research is required to evaluate the possibilities, pros and cons of this extended application of slot allocation in the aviation sector.

Similarly, the application of slot allocation in the road and navigation sectors may be evaluated in further research. For instance, proposals for freeway and parking slot allocation systems have been developed, but their pros and cons have still to be evaluated against the currently available alternatives. Furthermore, the technical aspects of these slot allocation systems will have to be elaborated in more detail. Similarly, more research about the current and potential applications of slot allocation in the navigation sector may be desirable, for instance to gain more insight in the functioning of port capacity allocation.

Improved bottleneck utilization versus capacity expansion

In this thesis, the optimization of infrastructure bottleneck usage has been discussed. An interesting question is to what extent the utilization of infrastructure bottlenecks may be improved compared with current practice. A related issue demanding further research are the consequences of improved bottleneck utilization, for instance by applying the ideas outlined in this thesis, for capacity expansion schemes. It is possible that in some situations improved bottleneck utilization is preferred to (expensive) investments to expand bottleneck capacity, in particular if sufficient buffer capacity is (made) available near the bottleneck under consideration. In this respect, an important question is whether the potential gain of improved capacity utilization is significant enough to be considered a good alternative to capacity expansion plans. It is a question of great societal relevance, given the costs and environmental impacts associated with many capacity expansion plans.

REFERENCES

- ACL (2000), *Procedures for Controlling Ad hoc Operations at Heathrow Airport*, version 6. Airport Coordination Ltd., London, UK. <http://www.acl-uk.org>
- Arnott, R., A. De Palma & R. Lindsey (1990), Departure Time and Route Choice for the Morning Commute, *Transportation Research B* **24/3**, pp. 209-228.
- Arnott, R., A. De Palma & R. Lindsey (1998), Recent Developments in the Bottleneck Model. In: K.J. Button & E.T. Verhoef (eds.), *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*. Edward Elgar, Cheltenham, UK.
- AVV (1997), *Verkeersgegevens jaarrapport 1996*. Ministerie van Verkeer en Waterstaat, Directoraat-Generaal Rijkswaterstaat, Adviesdienst Verkeer en Vervoer, Rotterdam, NL [in Dutch].
- AVV (2003), *Bedieningstijden van sluizen en bruggen*, edition of 17 Nov. 2003. Ministerie van Verkeer en Waterstaat, Directoraat-Generaal Rijkswaterstaat, Adviesdienst Verkeer en Vervoer, Rotterdam, NL [in Dutch]. <http://www.avv-rws.nl>
- AVV & TNO Inro (2001), *STIS referentiemodel voor de scheepvaart: tussenrapportage FS/FO, versie 1.0*. Ministerie van Verkeer en Waterstaat, Directoraat-Generaal Rijkswaterstaat, Adviesdienst Verkeer en Vervoer, Rotterdam, NL [in Dutch].
- Balas, E. (1965), An additive algorithm for solving linear programs with 0-1 variables, *Operations Research* **13**, pp. 517-546.
- Balas, E. & E. Zemel (1980), An algorithm for large zero-one knapsack problems, *Operations Research* **28**, pp. 1130-1154.
- Bär, M., K. Fischer & G. Hertel (1988), *Leistungsfähigkeit - Qualität - Zuverlässigkeit: theoretische Grundlagen für komplexe Systeme*. Transpress, Berlin, D [in German].
- Beasley, J.E. *et al.* (2000), Scheduling Aircraft Landings: the Static Case, *Transportation Science* **34/2**, pp. 180-196.
- Beckmann, M, C.B. McGuire & C.B. Winsten (1956), *Studies in the Economics of Transportation*. Yale University Press, New Haven (CT), USA.
- Bell, M.C. (1980), *Measures of Queuing Performance for a Traffic network*. University of Newcastle upon Tyne, Newcastle, UK.
- Berechman, J. (1993), *Public Transit Economics and Deregulation Policy*. Elsevier Science, Amsterdam, NL.
- Besluit geluidhinder spoorwegen (1993), *Staatsblad* 396, lastly amended by decree of 17. December 1999 [in Dutch].

Besluit slotallocatie (1997), *Staatsblad* 635 [in Dutch].

Bliemer, M.C.J. (2000), *Analytical Dynamic Traffic Assignment with Interacting User-classes: Theoretical Advances and Applications using a Variational Inequality Approach*, PhD Thesis, TU Delft. Delft University Press, Delft, NL.

Blumstein, A. (1959), The Landing Capacity of a Runway, *Operations Research* **7**, pp. 752-763.

Boadway, R.W. & N. Bruce (1984), *Welfare Economics*. Basil Blackwell, New York, USA.

Borenstein, S. (1988), On the Efficiency of Competitive Markets for Operating Licenses, *The Quarterly Journal of Economics* **104/2**, pp. 357-385.

Bouman, P.A., A.A. Bruijn & S.C. Kieft (2001), Intercity stopt ook te Vleuten West: alternatieve bedieningsmodellen door de bril van de reiziger. In: *Colloquium Vervoersplanologisch Speurwerk 2001: Wie doet wat?*, deel 2, pp. 869-888. CVS, Delft, NL [in Dutch].

Brander, J.R.G., B.A. Cook & J.E. Rowcroft (1989), Entry, Exclusion, and Expulsion in a Single Hub Airport System, *Transportation Research Record* 1214 (Modeling and Analysis of Airport and Aircraft Operations), pp. 27-36. Transport Research Board, Washington DC, USA.

Brännlund, U., P.O. Lindberg, A. Nöu & J.-E. Nilsson (1998), Railway Timetabling using Lagrangian Relaxation. *Transportation Science* **32/4**, pp. 358-369.

Brewer, P.J. & C.R. Plott (1996), A Binary Conflict Ascending Price (BICAP) Mechanism for the Decentralized Allocation of the Right to use Railroad Tracks, *International Journal of Industrial Organization* **14**, pp. 857-886.

Broeren, P.T.W & D. Westland (1998), Traffic Flow Control in Congested Motorway Networks using Buffers, *Transportation Research Record* 1612 (Highway Geometric Design Issues), pp. 34-41. Transport Research Board, Washington DC, USA.

Button, K.J. (2001), Economics of Transport Networks. In: K.J. Button & D.A. Hensher (eds.), *Handbook of Transport Systems and Traffic Control*, pp. 61-75. Pergamon/Elsevier Science, Oxford, UK.

Cambridge International Dictionary of English (1995). Cambridge University Press, Cambridge, UK.

Carlin, A. & R.E. Park (1970), Marginal Cost Pricing of Airport Runway Capacity, *American Economic Review* **60**, pp. 310-319.

Ceder, A. (2001), Public Transport Scheduling. In: K.J. Button & D.A. Hensher (eds.), *Handbook of Transport Systems and Traffic Control*, pp. 539-558. Pergamon/Elsevier Science, Oxford, UK.

- Chen, H.-K. (1999), *Dynamic Travel Choice Models: a Variational Inequality Approach*. Springer, Berlin, D.
- Concessiewet personenvervoer per trein (2000), Memorie van toelichting, *Kamerstukken II 1999/00*, dossier 27 216, no. 3 [in Dutch].
- Concessiewet personenvervoer per trein (2003), Wet van 23 april 2003 [NL], *Staatsblad* 265.
- Dahlgren, J. (1998), High Occupancy Vehicle Lanes: Not Always More Effective than General Purpose Lanes. *Transportation Research A* **32/2**, pp. 99-114.
- Danzig, G.B. (1957), Discrete Variable Extremum Problems. *Operations Research* **5**, pp. 266-277.
- De Jong, F. (2002), Interview with Mr. De Jong, Railion Benelux, March 14, 2002.
- De Ruijter, R., Th.H.J. Muller & P. Hakkesteegt (1990), *Optimalisatie afwikkeling treinverkeer: knelpunten- en capaciteitsanalyse traject Rotterdam - Den Haag v.v.* Technische Universiteit Delft, Faculteit der Civiele Techniek, Vakgroep Verkeer, Delft, NL [in Dutch].
- De Vries, S. & R. Vohra (2003), Combinatorial Auctions: a Survey, *INFORMS Journal on Computing* **15/3**, pp. 284-309.
- De Wit, J. & H. Van Gent (1996), *Economie en transport*. Lemma, Utrecht, NL [in Dutch].
- Department for Transport (2001), *The Future of Aviation: Consultation on Air Transport Policy*. Department for Transport, London, UK. <http://www.dft.gov.uk>
- Donohue, G.L. (2001), Air Traffic Control Systems. In: K.J. Button & D.A. Hensher (eds.), *Handbook of Transport Systems and Traffic Control*, pp. 571-578. Pergamon/Elsevier Science, Oxford, UK.
- Dupuit, J. (1844), De la mesure d'utilité des travaux publics, *Annales des Ponts et Chaussées* **14/8** [in French].²
- EC (1991), Council Directive 91/440/EEC of 29 July 1991 on the development of the community's railways, *Official Journal* L237, pp. 25-28.
- EC (1992a), Council Regulation (EEC) 2407/92 of 23 July 1992 on licensing of air carriers, *Official Journal* L240, pp. 1-7.

² English translation by R.H. Barback (1952): On the Measurement of the Utility of Public Works, *International Economic Papers* **2**, pp. 83-110. Reprinted in American Economic Association (1969), *Readings in Welfare Economics*, pp. 255-283. George Allen and Unwin, London, UK.

- EC (1992b), Council Regulation (EEC) 2408/92 of 23 July 1992 on access for Community air carriers to intra-Community air routes, *Official Journal* L240, pp. 8-14.
- EC (1993), Council Regulation (EEC) 95/93 of 18 January 1993 on common rules for the allocation of slots at Community airports, *Official Journal* L14, pp. 1-6.
- EC (1995), Council Directive 95/18/EC of 19 June 1995 on the licensing of railway undertakings, *Official Journal* L143, pp. 70-74.
- EC (2001a), Directive 2001/12/EC of the European Parliament and of the Council of 26 February 2001 amending Council Directive 91/440/EEC on the development of the Community's railways, *Official Journal* L75, pp. 1-7.
- EC (2001b), Directive 2001/13/EC of the European Parliament and of the Council of 26 February 2001 amending Council Directive 9518/EC on the licensing of railway undertakings, *Official Journal* L75, pp. 26-28.
- EC (2001c), 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, *Official Journal* L75, pp. 29-46.
- EC (2004), Council Regulation (EC) 793/2004 of the European Parliament and of the Council of 21 April 2004 amending Council Regulation (EEC) 95/93 on common rules for the allocation of slots at Community airports, *Official Journal* L138, pp. 20-60.
- EP (2001), *European Parliament Fact Sheets*. European Parliament, Directorate-General for Research, Luxembourg, L. <http://www.europarl.eu.int>
- Eurocontrol (2002), *Guidelines on Runway Capacity Enhancement*, edition 1.0. Eurocontrol, Brussels, B. <http://www.eurocontrol.int>
- European Commission (2001), Proposal for a Regulation of the European Parliament and of the Council amending Council Regulation (EEC) No 95/93 of 18 January 1993 on common rules for the allocation of slots at Community airports. *Official Journal* C 270 E, pp. 131-140.
- Faludi, A. (1973), *Planning Theory*. Pergamon, Oxford, UK.
- Fisher, M. & L. Wolsey (1982), On the Greedy Heuristic for Continuous Covering and Packing Problems, *SIAM Journal on Algebraic and Discrete Methods* 3/4, pp. 584-591.
- French, S. (1986), *Decision Theory: an Introduction to the Mathematics of Rationality*. Ellis Horwood, Chichester, UK.
- French, S. (1989), *Readings in Decision Analysis*. Chapman and Hall, London, UK.
- Greenberg, H. (1959), An Analysis of Traffic Flow, *Operations Research* 7, pp. 79-85
- Grether, D., M. Isaac, & C. Plott (1981), The Allocation of Landing Rights by Unanimity among Competitors, *American Economic Review* 71, pp. 166-171.

- Groenveld, R. (1999), *Capacities of Inland Waterways*. Delft University of Technology, Faculty of Civil Engineering and Geosciences, Hydraulic Engineering Division, Delft, NL.
- Hägerstrand, T. (1970), What about people in regional science? *Papers of the Regional Science Association* **23**, pp. 7-21.
- Hall, F.L. & F.O. Montgomery (1993), The Investigation of an Alternative Interpretation of the Speed-Flow Relationship for UK Motorways, *Traffic Engineering and Control* **34**, pp. 420-425.
- Harker, P.T. & S. Hong (1994), Pricing of Track Time in Railroad Operations: an Internal Market Approach, *Transportation Research B* **28/3**, pp. 197-212.
- Havenbedrijf Rotterdam (2004), *Haven in cijfers 2003*. Havenbedrijf Rotterdam, Rotterdam, NL. <http://www.portofrotterdam.com> [in Dutch].
- HCG (1992), *Studie effecten openstelling ringweg Amsterdam: eindrapport onderzoek latente vraag*. Hague Consulting Group, Den Haag, NL [in Dutch].
- Henderson, J.V. (1974), Road Congestion: a Reconsideration of Pricing Theory, *Journal of Urban Economics* **1**, pp. 346-355.
- Hermann, (1996), *Untersuchung zur Verspätungsentwicklung von Fernreisezügen auf der Datengrundlage der Rechnerunterstützten Zugüberwachung Frankfurt am Main*, Ph.D. thesis. Technische Universität Darmstadt, Darmstadt, D [in German].
- Hertel, G. (1994), *Leistungsfähigkeit und Leistungsverhalten von Eisenbahnbetriebsanlagen: Modelle und Berechnungsmöglichkeiten*. In: W. Schwanhäusser & S. Rielke (eds.), *Kapazitätsreserven der Verkehrsträger*, p. 120-156. Deutsche Verkehrswissenschaftliche Gesellschaft, Bergisch Gladbach, D [in German].
- Hillier, F.S. & G.J. Lieberman (1990), *Introduction to Operations Research*, 5th edition. McGraw-Hill, New York, USA.
- Hong, S. & P.T. Harker (1992), Air Traffic Network Equilibrium: toward Frequency, Price and Slot Priority Analysis, *Transportation Research B* **26/4**, pp. 307-323.
- Hoogendoorn, S.P. & P.H.L. Bovy (1999), Estimation and Analysis of Multilane Multi-class Headway Distribution Model, In: *Traffic Management, Safety and Intelligent Transport Systems, Proceedings of Seminar D held at the European Transport Conference*, pp. 287-301. PTRC, Hammersmith Grove, London, UK.
- Hoogendoorn, S.P. & P.H.L. Bovy (2000), A Closer Examination of Overtaking Prohibitions for Trucks, In: E. Schnieder & U. Becker (eds.), *9th IFAC Symposium Control in Transportation Systems 2000, Proceedings Vol. 1*, pp. 256-263. IFAC, Braunschweig, D.
- Horonjeff, R. & F.X. McKelvey (1994), *Planning and Design of Airports*, 4th edition. McGraw-Hill, New York, USA.

Hotelling, H. (1938), The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates, *Econometrica* **6**, pp. 242-269.³

IATA (2002), *Worldwide Scheduling Guidelines*, 7th edition. International Air Transport Association, Montreal, CA. <http://www.iata.org>

Imhof, D. (2001), Interview with Mr. Imhof, Railned, June 19, 2001.

Interimbesluit capaciteitstoewijzing spoorwegen (2000), *Staatsblad* 474 [in Dutch].

Jain, R. & J. M. Smith (1997), Modeling Vehicular Traffic Flow using M/G/C/C State Dependent Queuing Models, *Transportation Science* **31**/4, pp. 324-336.

Just, R.E., D.L. Hueth & A. Schmitz (1982), *Applied Welfare Economics and Public Policy*. Prentice-Hall, Englewood Cliffs (NY), USA.

Kamphuis, H.J. (1998), Air Traffic and Airport Management in the Netherlands. In: *Proceedings of the Symposium "Planning Traffic Networks for a Dynamic Transport Market - Rail Traffic Compared to Air Traffic"*. Railned, Utrecht, NL.

Kershaw, A.D., D.P. Rhodes & N.A. Smith (2000), The Influence of ATC in Approach Noise Abatement, 3rd USA/Europe Air Traffic management R&D Seminar. <http://atm-seminar-2000.eurocontrol.fr>

Koehler, R. (1994), Die Kapazität von Wasserstraßen und Möglichkeiten zu ihrer Beeinflussung. In: W. Schwanhäusser & S. Rielke (eds.), *Kapazitätsreserven der Verkehrsträger*. Deutsche Verkehrswissenschaftliche Gesellschaft, Bergisch Gladbach, D [in German].

Koolstra, K. (1999), Slotreservering op autosnelwegen: een alternatief voor rekeningrijden? In: *Colloquium Vervoerplanologisch Speurwerk 1999: Nederland is af!*, deel 1, pp. 421-437. CVS, Delft, NL [in Dutch].

Koolstra, K. (2000), Delta Metro Expres: betere benutting spoorcapaciteit door hoogfrequent sneltreinverkeer. In: *Colloquium Vervoerplanologisch Speurwerk 2000: wie betaalt bepaalt!*, deel 2, pp. 733-752. CVS, Delft, NL [in Dutch].

Koolstra (2001a), Capaciteitsmanagement in de spoorwegsector. In: E.H. ten Heuvelhof, H.D. Stout & K. Koolstra (eds), *Capaciteitsmanagement: beslissen over capaciteit van infrastructures*, pp. 95-115. Lemma, Utrecht, NL [in Dutch].

Koolstra (2001b), Slotallocatie op het spoor en op luchthavens, een vergelijkende analyse. In: *Colloquium Vervoersplanologisch Speurwerk 2001: Wie doet wat?*, deel 2, pp. 741-760. CVS, Delft, NL [in Dutch].

³ Reprinted in American Economic Association (1969), *Readings in Welfare Economics*, pp. 284-308. George Allen and Unwin, London, UK.

- Kreutzberger, E. & J.M. Vleugel (1992), *Capaciteit en benutting van infrastructuur: capaciteitsbegrippen en infrastructuurgebruik in de binnenvaart en het lucht- rail- en wegvervoer*. Delfse Universitaire Pers, Delft, NL [in Dutch].
- Krishna, V. (2002), *Auction Theory*. Academic Press, San Diego, USA.
- Lafortune, S. *et al.* (1993), Dynamic System-Optimal Traffic Assignment Using a State Space Model, *Transportation Research B* **27/6**, pp. 451-472.
- Lancaster, K.J. (1966), A New Approach to Consumer Theory, *Journal of Political Economy* **74/2**, pp. 132-157.
- Latten, R. (2002), *Spoorwegen 2003*. De Alk, Alkmaar, NL [in Dutch].
- Lerz, S. (1996), *Congestion Theory and Railway Traffic*, Ph.D. Thesis, Rijksuniversiteit Groningen. Labyrint, Capelle a/d IJssel, NL.
- Leutzbach, W. (1988), *Introduction to the Theory of Traffic Flow*. Springer, Heidelberg, D.
- Lindsay, K.S., E.A. Boyd & R. Burlingame (1993), Traffic Flow Management Modeling with the Time Assignment Model, *Air Traffic Control Quarterly* **3/1**, pp. 255-276.
- Magill, S.A.N. (1998), The Effect of Direct Routing on ATC Capacity, *2nd USA/Europe Air Traffic Management R&D Seminar*. <http://atm-seminar-98.eurocontrol.fr/>
- Majumdar, A. & J. Polak (2001), Estimating Capacity of Europe's Airspace using a Simulation Model of Air Traffic Controller Workload, *Transportation Research Record* 1744 (Issues in Aviation: Airports, Capacity, and Air Traffic Control and Management), pp. 30-43, Transportation Research Board, Washington DC, USA.
- Majumdar, A., W. Ochieng & J. Polak (2002), Estimation of European Airspace Capacity from a Model of Controller Workload, *Journal of Navigation* **55/3**, pp. 381-403.
- Martello, S. & P. Toth (1990), *Knapsack Problems: Algorithms and Computer Implementations*. Wiley, Chichester, UK.
- May, A.D., S.P. Shepherd & J.J. Bates (2000), Supply Curves for Urban Road Networks, *Journal of Transport Economics and Policy*, **34/3**, pp. 261-290.
- Mayer, I.S. & W. Veeneman (eds.), *Games in a World of Infrastructures: Simulation-Games for Research, Learning and Intervention*. Eburon, Delft, NL.
- McGowan, F. & P. Seabright (1989), Deregulating European Airlines, *Economic Policy*, **7/10**, pp. 283-344.
- Milner, J.M. (1995), *Dynamic Slot Allocation with Airline Participation*, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge (MA), USA.

- Minderhoud, M.M., P.H.L. Bovy & Th.J.H. Schoemaker (1995), *Een dynamisch parkeermanagementsysteem voor stadscentra*. Technische Universiteit Delft, Vakgroep Infrastructuur, Sectie Verkeerskunde, Delft, NL [in Dutch].
- Minderhoud, M.M., H.Botma & P.H.L. Bovy (1997), Assessment of Roadway Capacity Estimation Methods, *Transportation Research Record 1572* (Highway Capacity Issues and Analysis), pp. 59-65. Transportation Research Board, Washington DC, USA.
- Minderhoud, M.M. (1999), *Supported Driving: Impacts on Motorway Traffic Flow*, Ph.D. Thesis, Delft University of Technology. Delft University Press, Delft, NL.
- Mohring, H. & M. Harwitz (1962), *Highway Benefits*. Northwestern University Press, Evanston (IL), USA.
- Morlok, E.K. (1978), *Introduction to Transportation Engineering and Planning*. McGraw-Hill, New York, USA.
- Nagurney, A. (1993), *Network Economics: a variational inequality approach*. Kluwer, Dordrecht, NL.
- Nilsson, J.-E. (1999), Allocation of Track Capacity: Experimental Evidence on the use of Priority Auctioning in the Railway Industry, *International Journal of Industrial Organization* 17, pp. 1139-1162.
- Nolan, M.S. (1994), *Fundamentals of Air Traffic Control*, 2nd edition. Wadsworth, Belmont (CA), USA.
- Nordic Task Force on Airline Competition (2002). *Competitive Airlines: towards a more vigorous competition policy in relation to the air travel market*. Nordic Competition Authorities, Copenhagen/Helsinki/Oslo/Stockholm. <http://www.kkv.se>
- NS (1988), *Rail 21: sporen naar een nieuwe eeuw*. Nederlandse Spoorwegen, Utrecht, NL [in Dutch].
- NS (2001), *Spoorboekje*, edition 2001-2002. NS Reizigers, Utrecht, NL [in Dutch].
- NS (2005), *Spoorboekje 2005*. NS Reizigers, Utrecht, NL [in Dutch].
- Persaud, B. & V. Hurdle (1991), Freeway Capacity: Definition and Measurement Issues. In: U. Brannolte (ed.), *Highway Capacity and Level of Service*, Proceedings of the International Symposium on Highway Capacity, pp. 289-307. Balkema, Rotterdam, NL.
- Pigou, A.C. (1920), *The Economics of Welfare*. McMillan, London, UK.
- ProRail (2004), *Network Statement 2005*, version 1.0. ProRail, Utrecht, NL.
- Railion (2000), Geef Railion de ruimte, *Railion Issue 1*. <http://www.railion.nl> [in Dutch].
- Railion (2001), Ieder zijn rail, *Railion Issue 3*. <http://www.railion.nl> [in Dutch].

- Railned (2000a), Milieucapaciteit vaak beperkter dan fysieke capaciteit, *Railned 14* [in Dutch].
- Railned (2000b), *Studie ToeRGoed: Toekomst van het vervoer van reizigers en goederen per spoor*, eindrapport. Railned, Utrecht, NL [in Dutch].
- Railned (2000c), Meer treinen maar een lagere kwaliteit, *Railned 15* [in Dutch].
- Railned (2001a), Eerste fase Regionet, *Railned 16*, p. 9 [in Dutch].
- Railned (2001b), *Network Statement 2003*, version 1.1. Railned, Utrecht, NL [in Dutch]. <http://www.prorail.nl>.
- Railned (2001c), *ToeRGoed 2005: toekomst reizigers- en goederenvervoer bij afsluiten prestatiecontract*, eindrapport. Railned, Utrecht, NL [in Dutch].
- Rassenti, S.J., V.L. Smith & R.L. Buffin (1982), A Combinatorial Auction Mechanism for Airport Time Slot Allocation, *Bell Journal of Economics* **13/2**, pp. 402-417.
- Regeling vliegplannen (1998), *Staatscourant* 195, p. 11.
- Sahni, S. (1975), Approximate algorithms for the 0-1 knapsack problem, *Journal of the ACM* **22**, 115-124.
- Schaafsma, A.A.M. (1987), *Beschouwingen over capaciteit van een spoorwegennet*. Nederlandse Spoorwegen, Utrecht, NL [in Dutch].
- Schaafsma, A.A.M. (1997), Liever het ontleedmes dan de botte bijl: radicale scheiding van infrastructuur en exploitatie niet doeltreffend voor concurrentie en innovatie in de spoorwegbranche, *Tijdschrift vervoerswetenschap* **33/3**, pp. 237-258 [in Dutch].
- Schaafsma, A.A.M. (2001), *Dynamisch Railverkeersmanagement: besturingsconcept voor railverkeer op basis van het lagenmodel verkeer en vervoer*, Ph.D. Thesis, Delft University of Technology. Delft University Press, Delft, NL [in Dutch].
- Schiphol (2001), *Gebruiksplan Schiphol 2002 (november 2001 - oktober 2002)*. Amsterdam Airport Schiphol, Schiphol, NL. [In Dutch].
- Schiphol (2004), *Capacity Declaration Amsterdam Airport Schiphol, summer 2004*. Schiphol Group, Schiphol, NL. <http://www.slotcoordination.nl>
- Schipper, Y. (2001), *Environmental Costs and Liberalization in European Air Transport: a Welfare Economic Analysis*. Elgar, Cheltenham, UK.
- Schoemaker, Th.J.H. (2002), *Samenhang in vervoer- en verkeerssystemen*. Coutinho, Bussum, NL [in Dutch].
- Schoemaker, Th.J.H., K. Koolstra & P.H.L. Bovy (1998), Traffic in the 21st Century: a Scenario Analysis of the Traffic Market in 2030. In: M.P.C. Weijnen & E.F. ten Heuvelhof (eds.), *The Infrastructure Playing Field in 2030*, Proceedings of the First Annual

Symposium, Delft Interfaculty Research Centre Design and Management of Infrastructures, pp. 175-194. Delft University Press, Delft, NL.

Schuttevaer (2001), Nieuwe Duitse containerhaven onbereikbaar voor binnenvaart. *Weekblad Schuttevaer*, 15 april 2001 [in Dutch].

Schwanhäusser, W. (1994), Leistungsfähigkeit und Kapazität. In: W. Schwanhäusser & S. Rielke (eds.), *Kapazitätsreserven der Verkehrsträger*. Deutsche Verkehrswissenschaftliche Gesellschaft, Bergisch Gladbach, D [in German].

Simon, H.A. (1955), A Behavioral Model of Rational Choice, *The Quarterly Journal of Economics* **69**/1, pp. 99-118. ⁴

Spoorwegwet (1875), Wet van 9 april 1875 [NL], *Staatsblad* 67.

Spoorwegwet (2003), Wet van 23 april 2003 [NL], *Staatsblad* 264.

Starkie, D. (1998), Allocating Airport Slots: a Role for the Market? *Journal of Air Transport Management* **4**/2, pp. 111-116.

Stellingwerff, L. (2002), Interview with Mr. Stellingwerff, NS Commerce, Department of Product Management, March 21, 2002.

Stout, H.D. (2001), Capaciteitsmanagement van luchthaven Schiphol. In: E.H. ten Heuvelhof, H.D. Stout & K. Koolstra (eds.), *Capaciteitsmanagement: beslissen over capaciteit van infrastructures*, pp. 169-195. Lemma, Utrecht, NL [in Dutch].

Szpigel, B. (1973), Optimal Train Scheduling on a Single Line Railway, *Operations Research* **72**, pp. 344-351.

Ten Heuvelhof, E.H. & M. Kuit (2001), Capaciteitsmanagement: verkenning, uitwerking en operationalisering. In: E.H. ten Heuvelhof, H.D. Stout & K. Koolstra (eds.), *Capaciteitsmanagement: beslissen over capaciteit van infrastructures*, pp. 95-115. Lemma, Utrecht, NL [in Dutch].

Ten Heuvelhof, E.H., H.D. Stout & K. Koolstra (eds.) (2001), *Capaciteitsmanagement: beslissen over capaciteit van infrastructures*. Lemma, Utrecht, NL [in Dutch].

Thomas, B. (2001), Seaport Terminal Management. In: K.J. Button & D.A. Hensher (eds.), *Handbook of Transport Systems and Traffic Control*, pp. 559-570. Pergamon/Elsevier Science, Oxford, UK.

TRB (2000), *Highway Capacity Manual*, Transport Research Board, Washington DC, USA.

Treaty of Rome (1967), Treaty establishing the European Community, lastly amended 1 May 1999.

⁴ Reprinted in H.A. Simon (1957): *Models of Man*, pp. 241-260. Wiley, New York, NY, USA.

- Uebe, G. (1970), *Optimale Fahrpläne*, Lecture Notes in Operations Research and Mathematical Systems **20**. Springer, Berlin, D [in German].
- UIC (1991), *Transformation of the International Railway System as Part of New European Policy*. Union Internationale des Chemins de Fer, Paris, F.
- V&W & VROM, Ministeries van (1988), Tweede Struatuurschema Verkeer en Vervoer, deel A: beleidsvoornemen, *Kamerstukken II 1988/89*, dossier 20 922, no. 2 [in Dutch].
- V&W, Ministerie van (1997), Beleidskader toelating tot het spoorvervoer, *Kamerstukken II 1996/97*, dossier 25 228, no. 1 [in Dutch].
- V&W, Ministerie van (1998a), Aanwijzing Schiphol tot volledig gecoördineerd luchtvaartterrein, besluit van 13 januari 1998, no. DGRLD/ JBZ/L98.210002, *Staatscourant* 238, p.8.
- V&W, Ministerie van (1998b), Aanwijzing op grond van artikel 3 Besluit slotallocatie, besluit van 25 november 1998, no. DGRLD/ JBZ.L98.210729, *Staatscourant* 15, p. 14.
- V&W, Ministerie van (1999a), *De derde eeuw spoor*. Ministerie van Verkeer en Waterstaat, Directoraat-Generaal Personenvervoer [in Dutch].
- V&W, Ministerie van (1999b), Kadernota railveiligheid, *Kamerstukken II 1998/99*, dossier 26 699, no. 2 [in Dutch].
- V&W, Ministerie van (2000), Wet op het rekeningrijden, brief houdende intrekking van het wetsvoorstel, *Kamerstukken II 2000/01*, dossier 25 816, no. 13 [in Dutch].
- V&W, Ministerie van (2001), *MIT projectenboek: stand van zaken 2002*. SDU, Den Haag, NL [in Dutch].
- Van Asbeck, G. & J. Oranje (1999), Vakantiecharters in regio liggen onder vuur, *NRC Handelsblad*, 13 March 1999 [in Dutch].
- Van Binsbergen, A.J., Th.J.H. Schoemaker & C.D. van Goeverden (1995), *Ondergrondse Transportmogelijkheden Doorgrond: een Systematische Verkenning*. TRAIL Onderzoekschool, Delft, NL [in Dutch].
- Van Binsbergen, A.J & J.G.S.N. Visser (2001), *Innovation Steps Towards Efficient Goods Distribution Systems for Urban Areas*, Ph.D. Thesis, Delft University of Technology. Delft University Press, Delft, NL.
- Van den Akker, J.M. & K. Nachtigall (1997), *Slot Allocation by Column Generation*, National Aerospace Laboratory Technical Report no. 97286. NLR, Amsterdam, NL. <http://www.nlr.nl>
- Van Hoorn, H. (2002), Interview with Mr. Van Hoorn, Tulip Air, March 5, 2002.

- Van Nes, R. (2000), *Optimal Stop and Line Spacing for Urban Public Transport Networks*, TRAIL Studies in Transportation Science S2000/01. Delft University Press, Delft, NL.
- Van Nes, R. (2002), *Design of Multimodal Transport Networks: a Hierarchical Approach*, Ph.D. Thesis, Delft University of Technology. Delft University Press, Delft, NL.
- Veenendaal, A.J. (1998), *De ijzeren weg in een land vol water: beknopte geschiedenis van de spoorwegen in Nederland 1834-1958*. De Bataafsche Leeuw, Amsterdam, NL [in Dutch].
- Vickrey, W.S. (1961), Counter-speculation, Auctions and Competitive Sealed Tenders, *Journal of Finance* **16**/1, pp. 8-13.
- Vickrey, W.S. (1963), Pricing in Urban and Suburban Transport, *American Economic Review (Papers and Proceedings)* **53**, pp. 452-465.
- Vickrey, W.S. (1969), Congestion Theory and Transport Investment, *American Economic Review (Papers and Proceedings)* **59**, pp. 251-261.
- Vos, H. (2002), Interview with Mr. Vos, Transavia Airlines, March 14, 2002.
- Wakob, H. (1985), *Ableitung eines generellen Wartemodells zur Ermittlung der planmässigen Wartezeiten im Eisenbahnbetrieb unter besonderer Berücksichtigung der Aspekte Leistungsfähigkeit und Anlagenbelastung*. Rheinisch-Westfälische Technische Hochschule Aachen, Verkehrswissenschaftliches Institut, Aachen, D [in German].
- Walters, A.A. (1961), The Theory and Measurement of Private and Social Cost of Highway Congestion, *Econometrica* **29**/4, pp. 676-697.
- Wardrop, J.G. (1952), Some Theoretical Aspects of Road Traffic Research. *Proceedings of the Institution of Civil Engineers* **1**/2, pp. 325-378.
- Watson, R. (2001), Railway scheduling. In: K.J. Button & D.A. Hensher (eds.), *Handbook of Transport Systems and Traffic Control*, pp. 527-538. Pergamon/Elsevier Science, Oxford, UK.
- Wet Geluidhinder (1979), Wet van 16 februari 1979, *Staatsblad* 99.
- Wet Luchtvaart (1999), Beschikking van 30 juni 1999 houdende plaatsing van de tekst van de Wet luchtvaart, *Staatsblad* 274.
- Wet personenvervoer 2000 (1998), Memorie van toelichting, *Kamerstukken II 1998/99*, dossier 26 456, no. 3 [in Dutch].
- Wet personenvervoer 2000 (2000), Wet van 6 juli 2000, *Staatsblad* 314.
- Wolsey, L.A. (1998), *Integer Programming*. Wiley, New York, USA.

Wong, J.-T. (1997), Basic Concepts for a System for Advance Booking for Highway Use, *Transport Policy* 4/2, pp. 109-114.

Young, W. (2001), Parking and Traffic Control. In: K.J. Button & D.A. Hensher (eds.), *Handbook of Transport Systems and Traffic Control*, pp. 375-386. Pergamon/Elsevier Science, Oxford, UK.

Zigterman, L., R. van Bloemendaal, F. Behr (1999), Treinen kunnen elkaar nu al sneller opvolgen, *OV Magazine* 5/6, p. 6-9 [in Dutch].

LIST OF SYMBOLS

Numbers of vehicles or slots are indicated by #; value measures are indicated by €.

variables

a level of competition [#]
 b buffer (between slots) [s]
 c costs [€]
 d delay [s]
 e efficiency [€]
 f switching frequency, transport service frequency [1/s]
 g capacity occupancy [-]
 h indicator (binary)
 j number of periods [#]
 k available time-space [s]
 l length (of slot) [m]
 m duration of (slot) period [s]
 n number of slots / slot requests [#]
 p probability of interaction [-]
 q traffic volume [#]
 r reliability [-]
 s space [m]
 t time [s]
 u total value [€]
 v speed [m/s]
 w bandwidth (of slot) [-]
 x selection (binary) [#]
 z constraint weight coefficient [-]

stochastic variables

K capacity [#]

parameters

α weight in capacity constraint [-]
 β environmental burden [#]
 γ minimum level of competition [#]
 δ switching necessity [binary]
 ε convergence threshold [-]

ζ switching time [s]
 η environmental capacity [#]
 θ minimum headway [s]
 ι carrier identity
 κ available capacity [#]
 κ^s available scheduling slots [#]
 λ minimum vehicle separation [m]
 μ membership [binary]
 ν cost threshold level [€]
 ξ workload [s]
 o maximum deceleration [m/s²]
 π travel time [s]
 ρ outflow inflow ratio [-]
 σ minimum stopping time [t]
 τ reaction time [s]
 v speed [m/s]
 φ reduction factor [-]
 χ interaction [s]
 ψ minimum claim right [#]
 ω value parameter [€]

vectors and matrices

\mathbf{e} efficiency vector
 \mathbf{g} capacity occupation vector
 \mathbf{h} vector of binary indicators
 \mathbf{i} vehicle path identification vector
 \mathbf{r} ordered list of slot requests
 \mathbf{x} selection vector
 \mathbf{z} constraint weight vector
 $\boldsymbol{\alpha}$ weight vector of capacity constraint
 \mathbf{A} weight matrix of capacity constraint
 $\boldsymbol{\kappa}$ capacity vector
 $\boldsymbol{\mu}$ membership vector
 $\boldsymbol{\omega}$ value parameter vector

(continued on next page)

indices

- a* area
- c* carrier
- g* group of slot requests
- i* slot request, vehicle path
- j* basic slot, preceding vehicle path
- k* capacity constraint
- l* iteration
- m* user class, secondary iteration
- p* period
- r* traffic server
- s* location
- t* time
- v* speed class

sets

- R set of slot requests
- S allocated selection slots

functions

- E expectation
- p probability density
- P probability

GLOSSARY

<i>ad hoc</i> slot allocation	<i>see slot allocation, ad hoc</i> –
bandwidth	number of parallel traffic servers (within a slot)
bottleneck	traffic network element experiencing capacity scarcity
capacity	maximum load or flow
infrastructure –	maximum traffic load or traffic flow that can be served by infrastructure elements or combinations of infrastructure elements
transport –	maximum transport volume (i.e. persons, goods) that may be conveyed on traffic network elements
capacity allocation	allocation of traffic to an infrastructure element and a restricted period, in advance or in real-time, in such a way that capacity constraints are satisfied
capacity constraint	inequality function describing the relationship between bottleneck usage and capacity
capacity management	application of measures to control the allocation of infrastructure capacity to alternative users and uses
capacity occupancy	measure of the amount of capacity
capacity scarcity	situation where, given current levels of traffic supply and demand, traffic costs (significantly) higher than in the situation that the traffic volume approaches zero
absolute –	level of capacity scarcity where traffic demand exceeds traffic supply
relative –	level of capacity scarcity where traffic supply and demand are in equilibrium
hub	central node in a transport network of a carrier or an alliance of carriers, allowing short connections between flights
infrastructure	set of physical elements that are used to provide traffic services
infrastructure capacity	<i>see capacity, infrastructure</i> –
infrastructure element	physical element that is used to provide traffic services
infrastructure server	basic infrastructure element that can serve one vehicle simultaneously
quality-of-service	measure of the quality of traffic served by traffic network elements
run	single transport service belonging to a series of periodically repeated transport services
scheduling	<i>see slot allocation, scheduling</i> –
selection	<i>see slot allocation, selection</i> –
slot	time-space domain on an infrastructure element designed to facilitate a vehicle path

elementary –	slot defined on a single infrastructure element with the smallest spatial dimension
network –	sequence of elementary slots facilitating an entire vehicle path
slot allocation	allocation of slots to vehicle paths, in advance, in such a way that capacity constraints are satisfied
<i>ad hoc</i> –	the allocation of time and space specific slots, given slot requests, with the purpose of determining the feasibility of vehicle paths in short notice
scheduling –	the allocation of time and space specific slots, given slot requests, with the purpose of determining which transport service timetables are feasible given infrastructure capacity constraints
selection –	the selection of slot requests with the purpose of determining which vehicle paths can be accommodated and which have to be rejected
slot margin	temporal size of a slot period
slot period	period corresponding with a slot, specifying the period in which the corresponding vehicle path is accommodated
slot request	request to be allocated a slot which is usually submitted by a carrier to the allocation body
timetable	table specifying scheduled place and time of vehicle paths within a certain margin
allocation -	timetable describing which scheduling slots have been allocated to which vehicle paths
transport service -	timetable describing the supply of transport services
time-space domain	a closed subset of time-space, which is characterized by a common administrator, a common user, etc.
traffic	transportation and parking of vehicles
traffic market	the interaction of traffic supply and demand, resulting in traffic flows
traffic network	network providing traffic services
traffic network element	physical or virtual element providing traffic services
traffic server	smallest possible unit of a traffic network element
traffic service	provision of functions required for transportation and parking of vehicles
transport	transportation of passengers or freight
transport capacity	<i>see capacity, transport</i> –
transport market	the interaction of transport supply and demand, resulting in transport flows
transport service	provision of functions required for transportation of persons or freight
vehicle	transport means (includes also boats, trains, planes, etc.)
vehicle path	trajectory in time-space that is followed by a vehicle

***Appendix A* MAIN TRANSPORT SERVICE TYPES PER SECTOR**

In section 2.5, a number of transport service types have been specified. This appendix aims to give an overview of the occurrence of these transport service types per sector. Using data from various sources, usually applying to the Dutch situation, an indication is given of the percentages of transport services that are public or private, structural or *ad hoc*, scheduled or demand-driven, or passengers or freight. The given percentages are not exact data; they should only be used to get an idea which types of transport services are dominant.

Rail transport services

Railway networks are usually not open to private transport services. Furthermore, many railway links are used by both freight and person transport services, although some railways are dedicated for freight trains and others are dedicated for passenger trains. Large parts on the Dutch railway network are used for both freight and person transport services. According to the largest Dutch freight carrier Railion (2000), about 20% of railway capacity is reserved for freight transport services. However, according to Railion almost 50% of these paths cannot be used. Therefore, the percentage of freight transport services will be less than 15%.

Most passenger transport services are structural scheduled services, running according to a seasonal timetable. Additional *ad hoc* transport services may be operated for special events, for instance transport services between a main railway station and a stadium on the occasion of sports or music events. These events occur usually less than once per week, and usually result in only a few extra trains before and after these events. Therefore, we may conclude that only a small percentage of railway passenger traffic (less than 1%) will correspond with *ad hoc* transport services. Even these *ad hoc* trains are scheduled; the only non-scheduled transport services are trains that are chartered for specific events. The number of non-scheduled transport services will be negligible.

According to the Railion website (<http://www.railion.nl>), the three main types of freight transport services are shuttles, unit-cargo services, and charter services. Unit-cargo services offer transport services between many origins and many destinations. Unit-cargo networks consist of different parts, i.e. the collection and distribution networks connecting the origins and destinations with main shunting nodes, which are linked by a connection network connecting main shunting nodes. Unit-cargo services are largely supply-driven, scheduled, and structural. However, the composition of unit-cargo trains is demand-driven. In contrast, shuttle services are completely supply-driven. Shuttle trains are container trains of a fixed length. Shuttle services are structural and supply-

driven and operate according to a timetable. Finally, charter trains are demand-driven, i.e. their operation is tailored to the demands of the shippers and receivers. However, many charter trains are structural, i.e. they operate with approximately a given weekly or monthly frequency. No figures have been found about which percentage of freight transport services are charter services, and which percentage of charter services are *ad hoc*. However, unit-cargo and shuttle services are clearly dominant in the European situation, and given that the majority of traffic corresponds with passenger transport, the overall percentage of *ad hoc* transport services will be negligible.

Table A.1: Rail transport services

public / private	
structural / ad hoc	
scheduled / demand driven (public only)	
passengers / freight	

Table A.1 gives an indication of the occurrence of various transport service types, based on the arguments given above. Of each pair of notions (public / private, structural / *ad hoc*, scheduled / demand driven, and passengers / freight,) a crude indication is given whether almost all transport services are of the first type (arrow at 100%), the majority is of the first type (arrow between 100% and 50%), etc. According to table A.1, almost all rail transport services are public and structural, while demand-driven and freight transport services constitute a significant minority against a majority of scheduled passenger transport services.

Air transport

Many airports are used for both passenger and freight transport. For instance, between January and November 2002 371 135 flights were performed from Amsterdam Airport Schiphol between January and November 2002, of which 10 968 were cargo flights, and 14 757 general aviation flights (data: Schiphol Group). Consequently, the majority of flights from this mainport are passenger transport services. Consequently, the share of cargo flights is less than 3% and the share of general aviation flights is less than 4%. General aviation includes both passenger and freight transport. A significant percentage of general aviation corresponds with private transport services (air taxi services).

A large percentage of air passenger and freight transport services are scheduled line services. According to the 2002 Schiphol Group statistics, 340 411 flights were scheduled line-services, which is about 92%. The 8% non-scheduled passenger services may include some structural charter flights, which means that the percentage of structural services is even greater.

Line services and structural charter services are scheduled services. Furthermore, some *ad hoc* services, for instance extra flights to sports finals, are scheduled. Consequently, the dominance of scheduled services is even greater than the dominance of structural services. Specific data for passenger and freight transport services have not been found. Freight transport features a certain percentage of *ad hoc* and demand driven services, e.g. air taxi services for specific types of cargo. However, normal freight and mail services are structural scheduled services on fixed routes. Consequently, we may assume similar percentages of scheduled freight and scheduled passenger transport services.

Table A.2 shows that the majority of air transport services are public, structural, scheduled, and passenger transport services.

Table A.2: Air transport services (main airports)

public / private	
structural / ad hoc	
scheduled / demand driven (public only)	
passengers / freight	

Road transport (freeways)

When driving on freeways, we may observe that the percentage of trucks may differ between freeways and between times of day, but usually it is significantly less than 50%. We may also observe that ordinary passenger cars are dominant, suggesting that most transport services are private. Public transport services such as line buses and taxis correspond with only a small percentage of traffic, significantly smaller than freight transport. With respect to freight transport, the percentage of private transport services is lower, i.e. about 30% in 1994 in the Netherlands according to data of the Dutch Statistical Bureau CBS (see De Wit & Van Gent, 1996).

Statistics about travel purposes may be used to get an idea about the percentage of passenger transport that is structural. According to same 1994 CBS statistics, Dutch citizens spent on average about 20 minutes each day on commuting trips, and about 55 minutes on other purpose trips, which will generally have an *ad hoc* nature. However, during peak hours, which is the most relevant situation for capacity management studies, the percentage of commuter travel will be much higher, i.e. more than 50%. No information has been found about the percentages of structural and *ad hoc* freight transport services.

Finally, most public freight transport services will be demand-driven rather than scheduled, while a significant percentage of public passenger transport services will corresponding with line buses, i.e. scheduled transport services. Given that public transport is much more common for freight transport than for passenger transport, the result will be that the majority of public transport services will be demand-driven.

Table A.3 shows that the majority of transport services on freeways is private. With respect to the other variables, the distribution is much less one-sided than in the previous two sectors. This means that the usage of freeways is relatively heterogeneous with respect to transport service types.

Table A.3: Road transport services (freeways)

public / private	
structural / ad hoc	
scheduled / demand driven (public only)	
passengers / freight	

Sea transport

The last sector discussed in this appendix is navigation, focusing on sea transport calling at seaports such as the port of Rotterdam in the Netherlands. Almost all sea transport calling at these mainports is freight transport; the volume of passenger transport is negligible.

According to statistics of the Dutch Statistical Bureau CBS, 42 thousand sea ships were freighted in the Netherlands in 2001, of which 17 thousand scheduled line services, 9 thousand tanker ships and 16 thousand tramp ships (CBS, Maandcijfers internationale

handelsvaart over zee, 2001). According to De Wit & Van Gent (1996), about 40% tanker traffic is private. Line and tramp services are professional, and hence about 90% of sea transport services are public. Furthermore, line services are structural and we may assume that tanker services are also structural, while tramp is *ad hoc* transport. Consequently, more than 60% of all transport services is structural. Finally, we may assume that only line services are scheduled, while tanker services are largely demand driven. Hence, about 40% of ships calling Dutch seaports are scheduled transport services.

Table A.4: Sea water transport services calling at sea ports

public / private	 100% 50% 0%
structural / ad hoc	 100% 50% 0%
scheduled / demand driven (public only)	 100% 50% 0%
passengers / freight	 100% 50% 0%

Appendix B IDENTIFICATION OF BANKS OF AMSTERDAM AIRPORT SCHIPHOL

In this appendix, banks of Amsterdam Airport Schiphol are identified using the timetable of the 2002-2003 winter season. Figure B.1 gives an overview of departure and arrival rates for each half hour of flights to/from all European origins and destinations except London Heathrow. This overview is based on the 2002-2003 winter timetable for Fridays.

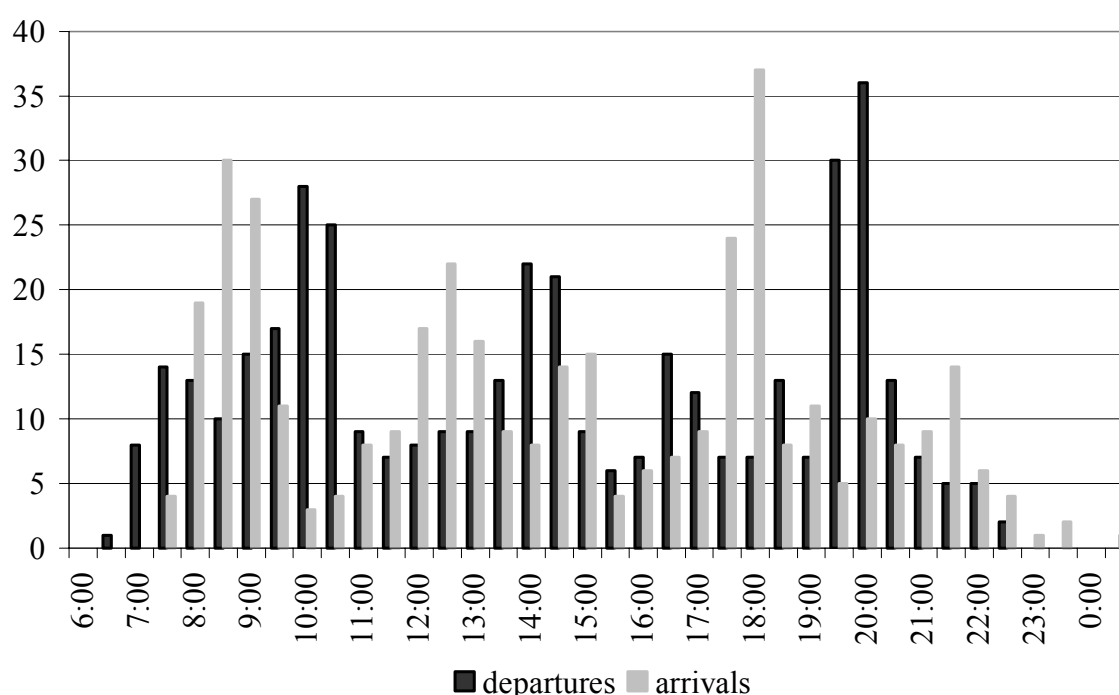


Figure B.1: Arrivals from and departures to European destinations at Amsterdam, 2002-2003 winter timetable for Fridays

Based on figure B.1, the following banks may be identified (see table B.1). Bank 0 is not a true bank, but a relatively broad peak of early departures that have little connections with arriving flights. Banks 1 to 4 are true banks, consisting of an arrival and a departure peak. Between some banks, a relatively quiet period of 1 to 1½ hours may be discerned. However, the relatively large banks 0 and 1 have some overlap. Furthermore, the arrival and departure peak periods overlap for the first two banks, while about 1 hour is needed to change, including the usual reliability margin for late arrivals. Hence, not all pairs of flights from an arrival bank and the corresponding departure bank can be considered as potential connections.

Table B.1: Banks of Amsterdam Airport Schiphol hub

bank	arrival period	departure period
0		7:00 - 9:00
1	8:00 - 10:00	9:30 - 11:30
2	12:00 - 13:30	13:00 - 14:30
3	14:30 - 15:30	16:00 - 17:00
4	17:30 - 18:30	19:00 - 22:30

Appendix C SLOT AVAILABILITY DATA FOR LONDON AIRPORTS

The following tables provide an indication of spare capacity at London Heathrow and London Gatwick airports during the, usually less congested, winter season. Table C.1 and C.2 show data of two subsequent weeks regarding London Heathrow, and table C.3 and C.4 provide the same information regarding London Gatwick.

Table C.1: Slots at London Heathrow between Jan. 13. - Jan. 19. 2003 available for ad hoc demands on Jan. 17. 2003 (Source: Airport Coordination Ltd., 2003)

Time	departures							arrivals						
	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su
0600	0	1	1	0	0	0	0	0	0	2	2	1	6	13
0700	0	0	2	1	0	0	3	0	0	0	0	1	4	7
0800	0	0	0	0	0	0	0	1	0	1	2	0	0	0
0900	0	1	0	0	0	0	0	0	1	0	0	0	0	0
1000	0	2	0	2	1	0	0	0	1	0	0	0	0	0
1100	0	0	0	1	1	0	1	0	1	0	0	1	1	0
1200	0	0	1	0	0	0	0	0	1	0	0	0	1	0
1300	1	3	0	3	3	0	4	0	1	0	1	1	0	0
1400	0	0	0	1	0	0	1	1	2	0	1	0	0	0
1500	2	0	2	1	1	5	0	1	0	0	1	0	0	0
1600	0	2	0	0	0	3	0	2	0	0	0	0	3	1
1700	2	1	2	2	1	7	2	2	0	1	0	0	1	0
1800	1	1	0	1	0	4	0	1	0	1	0	0	6	0
1900	3	1	0	1	1	7	0	0	1	0	0	0	8	0
2000	1	0	1	1	0	3	0	3	2	0	0	0	9	0
2100	4	2	2	3	2	17	1	18	13	15	12	10	13	11
2200	3	6	5	4	5	10	2	2	4	0	3	3	4	4

Table C.2: Slots at London Heathrow between Jan. 20. - Jan. 26. 2003 available for ad hoc demands on Jan. 17. 2003 (Source: Airport Coordination Ltd., 2003)

Time	departures							arrivals						
	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su
0600	0	0	0	0	0	1	0	0	0	0	1	0	5	12
0700	0	0	0	0	0	0	6	0	0	0	0	0	4	9
0800	0	0	0	0	0	0	2	0	0	0	0	0	0	2
0900	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	0	0	0	0	0	0	0	1	0	0	0	0	0	0
1100	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1200	0	0	1	0	0	0	0	0	0	1	0	0	1	0
1300	1	0	0	2	3	0	3	1	1	1	0	0	0	0
1400	0	0	0	1	0	0	1	1	1	1	1	0	1	0
1500	1	0	0	1	1	4	1	0	0	0	0	0	1	1
1600	0	2	0	1	0	2	0	0	0	0	0	0	3	0
1700	1	1	2	2	1	6	2	0	0	0	0	0	2	0
1800	0	1	0	0	0	5	1	0	1	0	0	0	5	0
1900	0	0	0	0	0	8	0	0	1	0	0	0	9	0
2000	0	0	1	1	0	4	0	0	0	0	0	0	7	0
2100	4	3	1	1	2	16	0	19	13	14	13	12	14	12
2200	5	5	3	5	5	11	2	3	5	0	4	4	5	3

Table C.3: Slots at London Gatwick between Jan. 13. - Jan. 19. 2003 available for ad hoc demands on Jan. 17. 2003 (Source: Airport Coordination Ltd., 2003)

Time	departures							arrivals						
	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su
0600	12	12	16	12	12	11	14	5	13	8	11	10	5	13
0700	6	7	8	7	5	4	9	4	4	5	5	4	7	3
0800	3	4	5	3	2	5	10	7	6	6	6	7	3	3
0900	5	7	2	4	8	8	10	6	4	6	1	1	2	5
1000	8	10	9	8	10	9	10	3	6	3	2	5	5	5
1100	9	10	10	10	13	7	6	5	6	5	7	6	8	7
1200	11	7	8	7	8	5	1	6	11	6	11	8	8	10
1300	3	7	5	4	6	3	7	8	5	5	7	7	7	1
1400	8	11	6	8	6	6	1	9	12	13	9	9	6	7
1500	9	9	10	7	7	6	9	11	13	12	11	9	11	4
1600	13	13	19	15	15	15	9	11	11	12	9	9	13	8
1700	10	10	8	7	11	6	5	19	18	21	17	21	18	16
1800	2	2	1	0	0	14	2	13	11	12	9	12	13	11
1900	6	6	8	11	10	14	8	3	3	3	1	2	12	1
2000	8	10	8	8	5	14	9	18	17	17	18	16	24	17
2100	15	16	18	18	11	15	13	26	25	26	26	24	26	24
2200	13	18	16	15	9	18	15	23	23	23	23	21	23	24

Table C.4: Slots at London Gatwick between Jan. 20. - Jan. 26. 2003 available for ad hoc demands on Jan. 17. 2003 (Source: Airport Coordination Ltd., 2003)³

Time	departures							arrivals						
	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su
0600	14	10	15	12	11	10	14	8	13	9	11	10	3	12
0700	7	8	7	7	6	4	10	4	5	5	5	3	7	0
0800	2	5	5	4	3	5	9	6	5	5	7	7	3	2
0900	4	7	2	6	8	6	4	6	4	6	2	2	1	7
1000	10	8	9	9	10	5	3	2	6	3	2	4	2	4
1100	9	10	10	11	12	6	5	1	7	4	7	8	7	6
1200	7	8	9	7	8	4	0	8	12	6	13	9	8	11
1300	4	7	6	5	6	3	8	8	5	7	9	7	7	2
1400	7	9	6	7	5	6	1	9	11	11	9	8	5	8
1500	9	8	10	7	8	5	9	13	12	12	10	11	10	5
1600	13	15	19	14	15	14	9	12	11	13	9	10	10	9
1700	10	9	9	11	11	5	6	19	17	21	17	21	18	14
1800	1	2	0	1	2	14	2	12	11	12	10	12	14	10
1900	9	6	8	10	9	14	10	2	3	2	2	3	13	1
2000	7	10	8	7	9	11	7	18	17	18	19	17	24	17
2100	14	16	18	17	14	14	12	26	26	25	24	24	26	24
2200	13	18	16	15	11	16	15	23	23	23	23	21	23	25

Appendix D DUPLICATION OF TRANSPORT SERVICES

In this appendix, an overview is given of transportation services departing from Amsterdam Airport Schiphol with the same destination and similar departure times (i.e. within an hour) but operated by different carriers, i.e. parallel transport services. Table D.1 gives an overview of all parallel morning departures from Amsterdam Airport Schiphol to EU destinations, given the 2003 winter timetable.

Table D.1: Pairs of departures from Amsterdam Airport Schiphol with same EU destination within an hour, Fridays, departures 6-11 h., January - March 2003

destination	dept. time	carrier	plane size	dept. time	carrier	plane size
Barcelona	6:00	Transavia	M	6:35	Iberia	M
Birmingham	9:40	British Airways	M	10:00	Air UK	S
Copenhagen	7:45	KLM	M	7:55	SAS	M
Copenhagen	9:40	SAS	XS	10:25	KLM	M
Frankfurt	8:45	KLM	S	8:50	Lufthansa	M
Geneva	8:00	KLM	M	8:25	EasyJet	M
Hamburg	8:35	Eurowings	XS	9:30	KLM	S
London Gatwick	10:00	British Airways	M	10:15	EasyJet	M
London Heathrow	7:15	British Midland	L	7:20	KLM	M
London Heathrow	7:30	British Airways	M	8:25	KLM	L
London Heathrow	9:40	British Midland	L	10:25	KLM	M
Lyon	9:10	KLM	S	9:20	Air France	XS
Madrid	6:55	KLM	M	7:30	Transavia	M
Madrid	8:35	Iberia	M	8:50	KLM	M
Manchester	9:20	Air UK	S	10:20	British Airways	XS
Munich	7:50	KLM	M	8:35	Lufthansa	XS
Oslo	10:10	KLM	M	10:35	SAS	S
Paris CdG	7:40	Air France	M	8:05	KLM	M
Paris CdG	9:00	Air France	M	9:45	KLM	M
Rome Fiumicino	9:05	KLM	M	9:40	Al Italia	M
Stockholm Arl.	10:00	KLM	M	10:25	SAS	M
Vienna	7:50	Austrian Air	M	8:05	KLM	M
Warsaw	10:00	KLM	M	10:35	LOT	S
Zurich	7:00	CrossAir	S	7:55	KLM	M
Zurich	9:35	CrossAir	S	9:55	KLM	M

Out of 144 morning departures, 25 pairs of 'duplicate' flights have been identified, of which only 2 are currently operated with aircraft larger than medium size. The classification of aircraft types is based on seat capacity and is illustrated by table D.2.

Table D.2: Classification of aircraft types

Plane size	Aircraft types
XS	Canadair RJ, Dash 8, Embraer RJ 120/135/145, Fokker 50
S	Airbus 318, Avro RJ 100, Boeing 737-500/600, Fokker 70/100
M	Airbus 319/320, Boeing 737-200/300/400/700/800/900, MD 80 series
L	Airbus 321, Boeing 767

Appendix E **OVERVIEW OF INTERVIEWS**

A number of people associated with carriers and allocation bodies have been interviewed as part of this thesis research. The table below gives an overview of all interviewed people and their affiliations.

Table E.1: Interviewed people and their affiliations

respondent	organization	date
Mr. F. De Jong	Railion Benelux	March 14 th 2002
Mr. D. Imhof	Railned (ProRail)	June 19 th 2001
Mr. L. Stellingwerff	NS Commerce	March 21 st 2002
Mr. H. Van Hoorn	Tulip Air	March 5 th 2002
Mr. H. Vos	Transavia Airlines	March 14 th 2002

SUMMARY

Introduction

Due to the structural nature of transport infrastructure capacity scarcity, the allocation of infrastructure capacity is a major issue in various transport infrastructure sectors. In the absence of traffic control measures, capacity scarcity manifests itself as congestion. An effective way to prevent structural traffic congestion is slot allocation. The key characteristic of slot allocation systems is that infrastructure users have to reserve a 'slot' on the network before departure. The main purpose of slot allocation is to solve infrastructure capacity scarcity problems in planning, and not in real-time on the network. Hence, slot allocation helps to avoid congestion on the network and enables the optimization of infrastructure utilization.

Slot allocation is currently applied in the railway and aviation sectors. The allocation of infrastructure capacity to different carriers is an increasingly important policy in Europe. A theoretical framework on slot allocation would be desirable to support rational decision-making on slot allocation. However, the current state-of-the-art of slot allocation research does not provide such a theoretical framework. Consequently, the main objective of this thesis is to establish a theoretical framework to analyze slot allocation problems. This thesis focuses on selection slot allocation, i.e. on longer-term slot allocation decisions determining the traffic patterns served by infrastructure bottlenecks, rather than timetable-related slot allocation problems, which have received much more attention in the scientific literature. Furthermore, this thesis focuses on single-bottleneck problems, rather than network problems, enabling a thorough analysis of the essence of (selection) slot allocation problems. The perspective of this thesis is normative, rational, and substantive, which means that we focus on the question of which slot allocation decisions should be made, rather than the question of how slot allocation is performed in current practice, or the question of which slot allocation decision procedures should be applied.

Conceptual framework

The first step in the development of a theoretical framework to analyze slot allocation problems is the specification of a conceptual framework. As a starting point, a transportation system model is specified that consists of three layers (activities, transport services, and traffic services) and two markets in between these layers (transport market and traffic market). This model is used to position the slot allocation problem, i.e. at the traffic market, and is the basis of a conceptual model of slot allocation. Another part of this conceptual framework is the definition of main concepts such as capacity. Capacity is the maximum planned traffic volume on a traffic network element, either simultaneously or per unit of time, which is defined as being dependent on assumptions about

composition of traffic and traffic context as well as the desired balance between capacity and quality-of-service. Another main concept discussed in the conceptual framework is slot allocation, which is the allocation of slots (time-space domains) to slot requests in such a way that capacity constraints are satisfied. Furthermore, three different levels of slot allocation are specified, i.e. selection, scheduling, and *ad hoc* slot allocation.

Current practice with slot allocation

The next step is to review the current application of slot allocation in the railway and aviation sectors and the potential application of slot allocation in the road and navigation sectors. In the railway sector, the slot allocation regime applies to virtually the entire infrastructure network, except (mainly metropolitan and suburban) networks that are exclusively used by a single carrier. In the aviation sector, however, slot allocation at higher levels than ATC slot allocation is applied to congested airports only, not to *en route* air traffic. Potentially, selection and scheduling slot allocation may also be applied to other types of bottlenecks (airways, ports, congested freeways), but an evaluation of the pros and cons of introducing slot allocation to these types of bottlenecks is beyond the scope of this thesis.

The current European slot allocation regimes in the rail and aviation sectors appear to consist of several decision levels that are similar to the decision levels proposed in this thesis. In both sectors, scheduling slot allocation is currently the dominant level of slot allocation, while selection slot allocation is relatively underdeveloped in these sectors, and the tradition has been to integrate selection and scheduling slot allocation. This thesis, however, considers selection slot allocation as a separate slot allocation level. Separating selection and scheduling slot allocation enables the application to each level of different rules with respect to slot validity, valuation of alternative slot requests, etc.

Based on the evaluation of current European slot allocation regimes, three issues are identified in this thesis. A large amount of discussions in policy documents and scientific papers about the desirability of adaptations in current slot allocation regimes are related to these issues. The first issue is stability *versus* flexibility. Stability is achieved given a relatively long validity of slots, while flexibility is achieved when slot allocation decisions are reconsidered frequently. In general, carriers will have a stronger desire for stability at the selection slot allocation level and a relatively stronger desire for flexibility at the scheduling slot allocation level. The second issue discussed in this thesis is competition *versus* diversity. A high level-of-competition is attained when two or more carriers offer parallel alternative transport services, while a high level of diversity given limited infrastructure capacity implies that as much different transport services are offered as possible, avoiding parallelism. The final issue is heterogeneity, i.e. the question of whether the level of heterogeneity corresponding with the characteristics of vehicle paths as desired by carriers should be accepted. Alternatively, homogenization

of traffic may be actively promoted by allocation bodies, for instance by slowing down fast traffic.

Desired characteristics of selection slot allocation systems

The desired characteristics of selection slot allocation are formulated in this thesis by analyzing the main desires of carriers and other interested parties such as shippers and authorities. With respect to slot validity it is concluded that selection slots should be valid for a significantly longer period than scheduling slots, because stability is more important for decisions of carriers with respect to the selection of transport services to be offered, and flexibility is more important for decisions with respect to the scheduling of transport services. Furthermore, given flexibility at the selection level, a semi-static slot allocation procedure is proposed, which implies that selection slots are allocated every (timetable) season, but these are valid much longer than a single season. By restricting selection slot allocation decisions to each season, evaluation of slot requests is enabled.

Another important issue is the specification of slot size. Given the focus on selection of slot requests rather than scheduling of slots, selection slots should be as large as possible. Nonetheless, it is argued in this thesis that alternatives included in the selection slots should be acceptable for the slot holders, i.e. no alternatives should be included in the selection slots that would result in the withdrawal of the corresponding transport service. However, the specification of selection slots is the responsibility of the allocation body, and for practical reasons it is advisable to specify standard basic slots. Larger slots may be introduced by specifying higher-level slots, consisting of multiple basic slots.

Specification of traffic supply and demand

The problem how to balance supply and demand of selection slots may be analyzed using congestion theory, which is a branch of economic theory. This thesis contributes to congestion theory by analyzing the main characteristics of traffic selection markets. It appears that selection slot allocation is desirable in situations of absolute capacity scarcity, i.e. in situations where traffic demand exceeds traffic supply without being able to reach equilibrium.

The assumption of homogeneous traffic characteristics, which is required to analyze traffic selection markets, is relaxed in our discussion about the identification of capacity constraints. Examining various types of primary traffic processes, traffic service processes, and traffic externalities, capacity constraints are formulated. These capacity constraints may be applied to different types of bottlenecks, including bottlenecks where external constraints are binding. Of these different types of bottlenecks, examples from current practice are provided in this thesis.

Apart from the different types of bottlenecks, it appears that three categories of capacity constraints may be distinguished, i.e. homogeneous capacity constraints, linear capacity constraints, and non-linear capacity constraints. Linear capacity constraints are most common, while non-linear constraints may be translated into linear capacity constraints given a reasonable good estimation of composition of traffic.

Specification of selection slot allocation decision problems

The specification of traffic supply - by specifying capacity constraints - is a key step in the specification of selection slot allocation decision problems. Another step is the specification of objectives. The (primary) objective of slot allocation may usually be specified as a linear objective function. Various instances of selection slot allocation decision problems may be formulated, depending on the type of capacity constraint(s), the number of capacity constraints (single or multiple), and the number of levels of basic slots that have been specified (single or multiple). The optimization problem corresponding with selection slot allocation decision problems with linear capacity constraints and a linear objective function is a binary linear programming problem. This type of problem may be solved using an exact solution algorithm, but also with an efficiency-based approximation algorithm. The basic idea of this greedy algorithm is that the selection of slot requests is evaluated sequentially, the order of evaluation being determined by the efficiency associated with each slot request. Efficiency is a measure of value per unit of capacity that is occupied.

Exact optimization is only better than the efficiency-based approximation algorithm given the assumption that capacity values are exact and that having some extra scheduling freedom or reliability margin does not benefit to the objective function. Capacity values depend on various factors such as the desired quality-of-service and the specification of slot periods, and the omission of benefits derived from higher reliability or higher scheduling freedom from the objective function is a simplification. Consequently, efficiency-based approximation is preferred in this thesis.

Besides a standard greedy algorithm for selection problems with a single kind of capacity constraint, an extended greedy algorithm is developed in this thesis to solve problems with two or more different types of capacity constraints. A further extension to this solution approach is the implementation of satisficing constraints. Satisficing constraints may be used to represent secondary objectives that may be regarded as having simply two values, i.e. satisfactory or non-satisfactory. This approach is suggested for secondary objectives regarding minimum supply levels of transport services and regarding minimum level-of-competition. These secondary objectives may be imposed by authorities to enhance the accessibility of specific regions and to enhance the competitiveness of transport services on key relations.

Practical implications of selection slot allocation

The proposed selection slot allocation decision approach is illustrated in this thesis with a hypothetical case study. This case study regards the selection slot allocation problem of a hypothetical airport. It appears that even the selection problem of a fairly complicated bottleneck can be solved relatively easily, adopting the efficiency approach described in this thesis.

After analyzing this case study, the consequences of implementation of selection slot allocation are reviewed, focusing on the three main issues formulated earlier in this thesis. It is concluded that the framework introduced in this thesis may help to attain a proper balance between stability and flexibility, and between diversity and competition. Finally, it offers a simple but adequate approach to deal with heterogeneity.

Conclusions and recommendations

The main findings of this thesis may be summarized as follows. Our first main conclusion is that selection and scheduling should be considered as separate slot allocation levels having a hierarchical relationship, selection being the primary slot allocation problem. Secondly, the validity of slots is a compromise between stability and flexibility. Finally, the objectives and constraints of the selection problem can be modeled as linear functions, and the resulting binary linear programming problem can best be solved with a greedy efficiency algorithm, as outlined in this thesis.

This thesis contributes in a number of ways to the state-of-the-art of transportation science. In the first place, this thesis contributes to the establishment of a generic, sector-independent theoretical framework to analyze slot allocation problems. This theoretical framework may be used to design slot allocation systems. Secondly, a unified approach to formulate capacity constraints of different types of bottlenecks, independent of transportation sector has been introduced. This may be applied to selection slot allocation problems, but also to other transportation problems. Furthermore, the selection slot allocation problem has been identified and formalized in this thesis. Finally, this thesis contributes to discussions about slot allocation systems in practice by providing a framework for transparent discussions about objectives, alternatives, etc.

In this thesis, a number of ideas have been outlined that have consequences for the application of slot allocation in practice. One of the main ideas outlined in this thesis is that selection and scheduling slot allocation should be treated separately, which means that these should be separate decision stages. This implies that the recently introduced system of framework agreements in the railway sector can best be extended to a full system of selection slot allocation, and in the aviation sector the principle of historic rights to airport slots at the scheduling level can best be replaced with a system of long-term validity of selection slots. Furthermore, in discussions about slot allocation regimes, selection and scheduling should also be treated as separate problems in order to

avoid confusion of arguments about the desired characteristics of slot allocation regimes.

Finally, a number of recommendations for further research are provided in this thesis. Further research is mainly desired with respect to the selection slot allocation decision approach outlined in this thesis. For instance, experimentally testing the ideas outlined in this thesis with a simulation game would be desirable. Another subject for further research emerging from this thesis is the desirability and technical consequences of new slot allocation systems, e.g. allocation of urban parking slots, or allocation of *en route* air traffic slots at higher levels than ATC slot allocation. Finally, a recommendation for further research is to investigate the question of to which extent the improvement of infrastructure bottleneck usage, for instance by improving the capacity allocation system, is an alternative to capacity expansion schemes.

SAMENVATTING

Inleiding

Schaarste aan infrastructuurcapaciteit is een structureel probleem. Hierdoor is de toedeling van infrastructuurcapaciteit een belangrijk thema in verschillende infrastructuursectoren. Tenzij er bepaalde vormen van verkeersmanagement worden toegepast om dit te voorkomen, zal capaciteitsschaarste zich manifesteren in de vorm van congestie. Slotallocatie is een effectieve vorm van verkeersmanagement ter voorkoming van structurele verkeerscongestie. Het belangrijkste kenmerk van slotallocatie systemen is dat infrastructuurgebruikers verplicht zijn om vooraf een slot te reserveren, teneinde toestemming te krijgen om het netwerk te gebruiken. Hierdoor kan slotallocatie bijdragen aan het voorkomen van congestie op het netwerk en aan het optimaliseren van de benutting van infrastructuur.

Momenteel wordt slotallocatie toegepast in de spoor- en luchtvaartsector. De toedeling van infrastructuurcapaciteit aan verschillende vervoerders is in toenemende mate een beleidsissue in Europa. Teneinde rationele beslissingen over slotallocatie te ondersteunen zou een theoretisch kader met betrekking tot slotallocatie wenselijk zijn. Een dergelijk theoretisch raamwerk is echter nog niet beschikbaar; het tot dusverre beschikbare onderzoek met betrekking tot slotallocatie is hiervoor nog onvoldoende. Als doelstelling van dit proefschrift is daarom gekozen een theoretisch kader op te stellen waarmee slotallocatieproblemen kunnen worden geanalyseerd. Dit proefschrift richt zich daarbij met name op selectie-slotallocatie. Selectie-slotallocatie betreft toedelingsbeslissingen op langere termijn die bepalend zijn voor de vraag welk verkeer wel en welk verkeer niet wordt afgewikkeld door een infrastructureel knelpunt, in tegenstelling tot slotallocatieproblemen gerelateerd aan dienstregelingen. De laatste categorie allocatieproblemen heeft veel meer aandacht gekregen in de huidige wetenschappelijke literatuur. Verder richt dit proefschrift zich op afzonderlijke infrastructuurknelpunten in plaats van netwerkproblemen, hetgeen een diepgaande analyse van de essentie van slotallocatieproblemen mogelijk maakt. In dit proefschrift is een normatieve, rationele, en inhoudelijke benadering gekozen, hetgeen betekent dat de nadruk ligt op de vraag wat voor soort slotallocatiebeslissingen het beste zouden zijn. De vraag hoe slotallocatie momenteel plaatsvindt is slechts van secundair belang, evenals de vraag welke slotallocatieprocedures het beste kunnen worden toegepast.

Conceptueel kader

De eerste stap in de ontwikkeling van een theoretisch kader om slotallocatieproblemen te analyseren is het opstellen van een conceptueel kader. Als uitgangspunt hiervoor is gekozen voor een lagenmodel van het verkeers- en vervoersysteem. Het model bestaat uit drie lagen: activiteiten, vervoerdiensten en verkeersdiensten. Tussen de lagen

worden twee markten onderscheiden: vervoermarkt en verkeersmarkt. Dit model wordt gebruikt om het slotallocatieprobleem te positioneren, namelijk op het niveau van de verkeersmarkt. Een belangrijk onderdeel van het conceptueel kader is de omschrijving van belangrijke begrippen zoals capaciteit. Capaciteit is het maximale verkeersvolume dat tegelijkertijd of per tijdseenheid kan worden toegelaten op een element van een verkeersnetwerk, gegeven aannames over de samenstelling van het verkeer, verkeersomstandigheden en de gewenste balans tussen kwaliteit en capaciteit. Een ander belangrijk concept in dit proefschrift is slotallocatie, hetgeen gedefinieerd wordt als de zodanige toedeling van slots (tijd-ruimte domeinen) aan slotaanvragen dat capaciteitsbeperkingen worden gerespecteerd. Verder worden in het conceptuele kader drie niveaus van slotallocatie geïntroduceerd, namelijk selectie-slotallocatie, dienstregeling-slotallocatie en *ad hoc* slotallocatie.

Huidige praktijk

De volgende stap van dit proefschrift is het onderzoeken van de huidige praktijk met betrekking tot slotallocatie in de spoor- en luchtvaartsector, alsmede het onderzoeken van potentiële nieuwe toepassingen van slotallocatie op de weg en in de binnen- en zeevaart. In de (Europese) spoorsector wordt slotallocatie toegepast op bijna het gehele netwerk, met uitzondering van stedelijke en suburbane spoorwegnetwerken die gescheiden zijn van nationale spoorwegnetten. In de luchtvaartsector wordt slotallocatie op dienstregelingniveau echter alleen toegepast op luchthavens. Wel is er een systeem van slotallocatie op operationeel niveau dat juist met name van toepassing is op de vlucht zelf. In potentie kan slotallocatie op selectie- of dienstregelingniveau ook worden toegepast op andere soorten infrastructuurknoelpunten, zoals havens, snelwegen en luchtvaartroutes. Een onderzoek naar de wenselijkheid van dergelijke slotallocatiesystemen is echter buiten het bereik van dit proefschrift.

De huidige slotallocatieregimes in de Europese Unie met betrekking tot de sectoren spoor en luchtvaart blijken te bestaan uit een aantal beslissingsniveaus. Deze beslissingsniveaus komen redelijk overeen met de beslissingsniveaus die zijn omschreven in het conceptueel kader. In beide sectoren is dienstregeling-slotallocatie dominant, terwijl slotallocatie op selectieniveau nog nauwelijks ontwikkeld is. Bovendien is het in beide sectoren gebruikelijk om deze niveaus te integreren. Dit proefschrift beschouwt echter nadrukkelijk selectie als een apart niveau van slotallocatie. Het scheiden van slotallocatie op selectieniveau en dienstregelingniveau is wenselijk om op beide niveaus optimale keuzes te kunnen maken met betrekking tot de geldigheid van slots, de waardering van alternatieve slotaanvragen, enz.

In dit proefschrift zijn drie centrale thema's geformuleerd welke een centrale rol spelen met betrekking tot gewenste verbeteringen van de huidige Europese slotallocatieregimes. Het eerste thema is stabiliteit *versus* flexibiliteit. Stabiliteit vraagt om een relatieve lange geldigheid van slots, terwijl flexibiliteit juist vraagt om

regelmatig heroverwogen van slotallocatiebeslissingen. Het tweede thema is concurrentie *versus* diversiteit. Concurrentie tussen vervoerdiensten vereist dat twee of meer vervoerders vergelijkbare vervoerdiensten aanbieden, terwijl een grote diversiteit gegeven beperkte capaciteit vereist dat zo veel mogelijk verschillende vervoerdiensten worden aangeboden, waarbij parallelliteit juist vermeden moet worden. Het derde thema is heterogeniteit, waarbij het gaat om de vraag in hoeverre de mate van heterogeniteit in gewenste karakteristieken van voertuigpaden moet worden gehanteerd, of dat homogenisering van verkeersstromen moet worden nagestreefd teneinde een hogere capaciteit te bereiken. Homogenisering kan actief worden bevorderd door de toewijzende instantie, bijvoorbeeld door verkeer met een hogere gewenste snelheid dan gemiddeld langzamer te laten rijden.

Gewenste kenmerken van slotallocatiesystemen

Op basis van een analyse van de wensen van vervoerders en andere bij slotallocatie betrokken partijen zijn in dit proefschrift de gewenste kenmerken van slotallocatiesystemen afgeleid. Eén van deze kenmerken is de geldigheid van slots. Het blijkt dat een beduidend langere geldigheid gewenst is voor slots op selectieniveau dan voor slots op dienstregelingniveau, aangezien stabiliteit vooral van belang is met betrekking tot de vraag welke vervoerdiensten kunnen worden aangeboden, terwijl meer flexibiliteit gewenst met betrekking tot het aanpassen van dienstregelingen. Om ook op selectieniveau voldoende flexibiliteit te waarborgen wordt in dit proefschrift een semi-statische slotallocatieprocedure voorgesteld, hetgeen inhoudt dat ieder (dienstregeling) seizoen nieuwe selectie-slots kunnen worden aangevraagd, welke een geldigheid hebben van meerdere seizoenen. Door slechts éénmaal per seizoen selectie-slots toe te delen wordt een afweging tussen verschillende slotaanvragen mogelijk gemaakt.

Een ander belangrijk kenmerk is de grootte van een slot. Aangezien het bij selectie-slots gaat om het selecteren van slotaanvragen en niet om het afstemmen van dienstregelingen ligt het voor de hand om slots zo groot mogelijk te definiëren, dus met een grote tijdmargin. Desalniettemin is het redelijk om de omvang van selectie-slots te beperken tot acceptabele alternatieven, hetgeen betekent dat alternatieven onacceptabel zijn voor de vervoerder zo veel mogelijk buiten de marges van het slot worden gehouden. 'Onacceptabel voor de vervoerder' wil zeggen dat deze zou besluiten de betreffende vervoerdienst niet door te laten gaan. Het vaststellen van selectie-slots is echter de verantwoordelijkheid van de toedelende instantie en uit praktisch oogpunt is het aan te raden om standaard-slots te definiëren. Deze standaard-slots hebben een vaste omvang, waarbij het mogelijk is om grotere standaard-slots te definiëren als combinatie van twee of meer primaire standaard-slots.

Verkeersvraag en verkeersaanbod

Congestietheorie is de tak van de economische wetenschap die zich bezighoudt met de balans tussen verkeersvraag en –aanbod. Dit proefschrift beschrijft hoe congestietheorie

toegepast kan worden op de interactie tussen vraag en aanbod van selectie-slots. Op basis hiervan worden twee verschillende niveaus van capaciteitsschaarste onderscheiden, waarbij absolute capaciteitsschaarste de situatie is waarin er geen evenwicht tussen verkeersvraag en –aanbod bereikt kan worden, waardoor de verkeersvraag groter is dan het verkeersaanbod. Het blijkt dat selectie-slotallocatie wenselijk is in situaties van absolute capaciteitsschaarste.

Om verkeersmarkten te kunnen analyseren met behulp van congestietheorie is aangenomen dat het verkeer homogeen van samenstelling is. Voor een adequate beschrijving van het verkeersaanbod moet echter ook rekening worden gehouden met de heterogeniteit van verkeer. Dit proefschrift introduceert een methode om het verkeersaanbod te beschrijven in de vorm van capaciteitsbeperkingen. Deze methode kan worden toegepast op diverse typen primaire en secundaire verkeersprocessen die zullen resulteren in verschillende typen capaciteitsbeperkingen. Hierdoor kunnen verschillende typen infrastructurele knelpunten gemodelleerd worden, waaronder knelpunten waar externe (milieu) beperkingen maatgevend zijn. Van deze verschillende typen knelpunten worden in dit proefschrift voorbeelden gegeven uit verschillende sectoren.

Naast verschillende typen infrastructurele knelpunten onderscheidt dit proefschrift ook verschillende soorten capaciteitsbeperkingen, namelijk homogeen, lineair, en niet-lineair. Lineaire capaciteitsbeperkingen komen het meest voor en bovendien is het gegeven een redelijke aanname over de samenstelling van verkeer meestal mogelijk om niet-lineaire capaciteitsbeperkingen te benaderen door een lineaire capaciteitsbeperking.

Beschrijving van het selectie-slotallocatie beslissingsprobleem

Het vastleggen van het verkeersaanbod in capaciteitsbeperkingen is een belangrijke stap in de beschrijving van het selectie-slotallocatie beslissingsprobleem, hier verder kortweg het selectieprobleem genoemd. De volgende stap is het vastleggen van de doelstellingen. Het blijkt dat het meest waarschijnlijke type (primaire) doelstelling kan worden benaderd door een lineaire doelfunctie. Verschillende typen selectieproblemen kunnen worden geformuleerd, afhankelijk van het aantal en type capaciteitsbeperking(en) en het aantal niveaus waarop standaard-slots zijn gespecificeerd. Het optimaliseringsprobleem dat overeenkomt met (eventueel meerdere) lineaire capaciteitsbeperkingen en een lineaire doelfunctie is een binair lineair programmeringsprobleem. Voor dit type probleem is een exact oplossingsalgoritme beschikbaar, maar ook een benaderingsalgoritme gebaseerd op efficiëntie komt in aanmerking. Het basisidee achter dit zogenaamde *gulzige* algoritme is dat slotaanvragen één voor één worden beoordeeld, in volgorde van hun efficiëntie. Efficiëntie is hier een maat voor de bijdrage van een slotaanvraag aan de doelfunctie per eenheid capaciteit die hierdoor wordt 'verbruikt'. In dit proefschrift gaat de voorkeur uit naar dit type benaderingsalgoritme, aangezien exacte optimalisering alleen beter scoort in situaties waar capaciteitswaarden exact zijn en waar een extra betrouwbaarheidsmarge niet

bijdraagt aan de doelstelling. In het selectieprobleem zijn capaciteitswaarden echter afhankelijk van relatief arbitraire keuzes, zoals het gewenste kwaliteitsniveau en de gekozen marges van de standaard-slots en het niet opnemen van betrouwbaarheid in de doelstellingsfunctie van het selectieprobleem is slechts een simplificatie.

In dit proefschrift is een zogenaamd *greedy* (gulzig) algoritme gekozen voor selectieproblemen met slechts één type capaciteitsbeperking. Daarnaast is een uitbreiding op dit algoritme ontwikkeld om selectieproblemen met twee of meer typen capaciteitsbeperkingen te kunnen oplossen. De efficiency-benadering is nog verder uitgebreid om binaire doelstellingen mee te nemen. Dit betreft secundaire doelstellingen die (bij benadering) slechts twee waarden kennen, namelijk voldoende en onvoldoende. Doelstellingen die hiervoor in aanmerking komen zijn onder andere het bereiken van een gewenst aanbodniveau van bepaalde vervoerdiensten en het bereiken van een zeker niveau van concurrentie tussen vervoerdiensten. Deze secundaire doelstellingen kunnen worden opgelegd door respectievelijk overheden die de bereikbaarheid van bepaalde regio's willen vergroten en overheden die de concurrentie tussen vervoerdiensten willen stimuleren.

Gevolgen van het invoeren van selectie-slotallocatie voor de praktijk

In dit proefschrift is een hypothetische casus opgenomen om de voorgestelde benadering voor het selectieprobleem te illustreren. De betreffende casus betreft het selectieprobleem van een denkbeeldige luchthaven. Het blijkt uit deze casus dat zelfs het selectieprobleem van een relatief complex infrastructureel knelpunt relatief eenvoudig kan worden opgelost met de voorgestelde efficiencybenadering.

In dit proefschrift worden verder een aantal aspecten besproken van de invoering van selectie-slotallocatie in de praktijk. Hierbij wordt met name gekeken naar de gevolgen voor de drie centrale thema's die zijn geformuleerd naar aanleiding van het overzicht van de huidige praktijk van slotallocatie. De conclusie is dat de door dit proefwerk geïntroduceerde ideeën met betrekking tot slotallocatie kunnen helpen om een goede balans te bereiken tussen stabiliteit en flexibiliteit, en tussen diversiteit en concurrentie. Tenslotte biedt het geïntroduceerde theoretische kader met betrekking tot slotallocatie een simpele maar adequate benadering om om te gaan met heterogeniteit van verkeer.

Conclusies en aanbevelingen

De belangrijkste bevindingen van dit proefschrift kunnen als volgt worden samengevat. De eerste conclusie is dat selectie-slotallocatie en dienstregeling-slotallocatie beschouwd moeten worden als verschillende niveaus van slotallocatie in een hiërarchische relatie, waarbij selectie het primaire slotallocatieprobleem is. In de tweede plaats is de geldigheid van slots een compromis tussen stabiliteit en flexibiliteit. Tenslotte kunnen zowel de doelstelling als de beperkingen van het selectieprobleem worden gemodelleerd met lineaire functies. Het resulterende binaire lineaire

programmeringsprobleem kan het beste worden opgelost met een *greedy* algoritme gebaseerd op efficiëntie dat is beschreven in dit proefschrift.

Dit proefschrift draagt op een aantal manieren bij aan de huidige verkeers- en vervoerswetenschap. In de eerste plaats draagt het bij aan het totstandkomen van een theoretisch kader om slotallocatieproblemen te analyseren. Dit theoretisch kader kan worden gebruikt om slotallocatiesystemen te ontwerpen. In de tweede plaats introduceert dit proefschrift een universele benadering om capaciteitsbeperkingen op te stellen voor verschillende typen infrastructurele knelpunten uit verschillende transportsectoren. Deze benadering kan niet alleen worden toegepast voor slotallocatieproblemen, maar ook voor andere verkeersproblemen. Daarbij is op basis van deze aanpak het selectie-slotallocatieprobleem geformaliseerd in dit proefschrift. Tenslotte draagt dit proefschrift bij aan het stroomlijnen van discussies over slotallocatie in de praktijk door het bieden van een kader voor transparante discussies over doelstellingen, alternatieven, enz.

De in dit proefschrift geformuleerde ideeën met betrekking tot slotallocatie hebben belangrijke gevolgen indien geïmplementeerd in de praktijk. Eén van deze ideeën is dat selectie-slotallocatie en dienstregeling-slotallocatie moeten worden behandeld als verschillende problemen overeenkomend met verschillende beslissingsstadia. Dit betekent dat het recent geïntroduceerde systeem van kaderovereenkomsten in de spoorwegsector het beste uitgebreid kan worden naar een volledig systeem van selectie-slotallocatie. Bovendien kan met betrekking tot slotallocatie op luchthavens het principe van historische rechten op dienstregelingniveau het beste worden vervangen door een systeem van selectieslots met een lange geldigheid, gecombineerd met slots op dienstregelingniveau met slechts een geldigheid van één dienstregelingseizoen. Tenslotte is het ook wenselijk dat slotallocatie op selectieniveau en op dienstregelingniveau gescheiden wordt in discussies over slotallocatieregimes, teneinde discussies omtrent de gewenste kenmerken van slotallocatieregimes helder te houden.

Dit proefschrift bevat tenslotte enkele aanbevelingen voor nader onderzoek. In de eerste plaats is nader onderzoek gewenst met betrekking tot de benadering en oplossing van het selectie-slotallocatieprobleem die is voorgesteld in dit proefschrift. Het gaat hierbij bijvoorbeeld om het experimenteel testen van de voorgestelde aanpak van het slotallocatieprobleem in een spelsimulatie. Een ander onderwerp voor nader onderzoek is de wenselijkheid en de technische haalbaarheid van nieuwe slotallocatiesystemen, bijvoorbeeld de toedeling van (auto)parkeerslots, of de allocatie van *en route* luchtverkeersslots op dienstregeling- of selectieniveau. Tenslotte is een interessante vraag voor nader onderzoek in hoeverre de verbetering van de benutting van infrastructurele knelpunten, bijvoorbeeld door het verbeteren van toedeling van capaciteit, een alternatief is voor het vergroten van de infrastructuurcapaciteit.

ABOUT THE AUTHOR

Kaspar Koolstra was born in Amsterdam on November 27th 1974. In 1997 he graduated from Groningen University, receiving a masters degree in Environmental and Infrastructure Planning. In the same year he joined the Transport and Planning department of Delft University of Technology as a Ph.D. student. His Ph.D. research project was part of the interfaculty research center on Design and Management of Infrastructures. Kaspar's research focused on the allocation of transport infrastructure capacity, but he also participated in the DeltaNet project, a proposal for a new multi-modal suburban public transport system for the *DeltaMetropolis*, i.e. the urban core of the Netherlands. Based on his research, he presented a number of papers at national as well as international conferences. Furthermore, Kaspar was one of the initiators of a research project on capacity management in different infrastructure sectors, carried out by an interdisciplinary team of researchers from Delft University of Technology. This project resulted in 2001 in a book on the current state-of-practice of infrastructure capacity management in the Netherlands. Kaspar also contributed the teaching activities of the Department of Transport and Planning. Currently he works as an *ad interim* teacher of geography.

TRAIL Thesis Series

A series of The Netherlands TRAIL Research School for theses on transport, infrastructure and logistics.

Nat, C.G.J.M., van der, *A Knowledge-based Concept Exploration Model for Submarine Design*, T99/1, March 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Westrenen, F.C., van, *The Maritime Pilot at Work: Evaluation and Use of a Time-to-boundary Model of Mental Workload in Human-machine Systems*, T99/2, May 1999, TRAIL Thesis Series, Eburon, The Netherlands

Veenstra, A.W., *Quantitative Analysis of Shipping Markets*, T99/3, April 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Minderhoud, M.M., *Supported Driving: Impacts on Motorway Traffic Flow*, T99/4, July 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Hoogendoorn, S.P., *Multiclass Continuum Modelling of Multilane Traffic Flow*, T99/5, September 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Hoedemaeker, M., *Driving with Intelligent Vehicles: Driving Behaviour with Adaptive Cruise Control and the Acceptance by Individual Drivers*, T99/6, November 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Marchau, V.A.W.J., *Technology Assessment of Automated Vehicle Guidance - Prospects for Automated Driving Implementation*, T2000/1, January 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Subiono, *On Classes of Min-max-plus Systems and their Applications*, T2000/2, June 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Meer, J.R., van, *Operational Control of Internal Transport*, T2000/5, September 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Bliemer, M.C.J., *Analytical Dynamic Traffic Assignment with Interacting User-Classes: Theoretical Advances and Applications using a Variational Inequality Approach*, T2001/1, January 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Muילerman, G.J., *Time-based logistics: An analysis of the relevance, causes and impacts*, T2001/2, April 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Roodbergen, K.J., *Layout and Routing Methods for Warehouses*, T2001/3, May 2001, TRAIL Thesis Series, The Netherlands

Willems, J.K.C.A.S., *Bundeling van infrastructuur, theoretische en praktische waarde van een ruimtelijk inrichtingsconcept*, T2001/4, June 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Binsbergen, A.J., van, J.G.S.N. Visser, *Innovation Steps towards Efficient Goods Distribution Systems for Urban Areas*, T2001/5, May 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Rosmuller, N., *Safety analysis of Transport Corridors*, T2001/6, June 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Schaafsma, A., *Dynamisch Railverkeersmanagement, besturingsconcept voor railverkeer op basis van het Lagenmodel Verkeer en Vervoer*, T2001/7, October 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Bockstael-Blok, W., *Chains and Networks in Multimodal Passenger Transport. Exploring a design approach*, T2001/8, December 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Wolters, M.J.J., *The Business of Modularity and the Modularity of Business*, T2002/1, February 2002, TRAIL Thesis Series, The Netherlands

Vis, F.A., *Planning and Control Concepts for Material Handling Systems*, T2002/2, May 2002, TRAIL Thesis Series, The Netherlands

Koppius, O.R., *Information Architecture and Electronic Market Performance*, T2002/3, May 2002, TRAIL Thesis Series, The Netherlands

Veeneman, W.W., *Mind the Gap; Bridging Theories and Practice for the Organisation of Metropolitan Public Transport*, T2002/4, June 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Van Nes, R., *Design of multimodal transport networks, a hierarchical approach*, T2002/5, September 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Pol, P.M.J., *A Renaissance of Stations, Railways and Cities, Economic Effects, Development Strategies and Organisational Issues of European High-Speed-Train Stations*, T2002/6, October 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Runhaar, H., *Freight transport: at any price? Effects of transport costs on book and newspaper supply chains in the Netherlands*, T2002/7, December 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Spek, S.C., van der, *Connectors. The Way beyond Transferring*, T2003/1, February 2003, TRAIL Thesis Series, Delft University Press, The Netherlands

Lindeijer, D.G., *Controlling Automated Traffic Agents*, T2003/2, February 2003, TRAIL Thesis Series, Eburon, The Netherlands

Riet, O.A.W.T., van de, *Policy Analysis in Multi-Actor Policy Settings. Navigating Between Negotiated Nonsense and Useless Knowledge*, T2003/3, March 2003, TRAIL Thesis Series, Eburon, The Netherlands

Reeven, P.A., van, *Competition in Scheduled Transport*, T2003/4, April 2003, TRAIL Thesis Series, Eburon, The Netherlands

- Peeters, L.W.P., *Cyclic Railway Timetable Optimization*, T2003/5, June 2003, TRAIL Thesis Series, The Netherlands
- Soto Y Koelemeijer, G., *On the behaviour of classes of min-max-plus systems*, T2003/6, September 2003, TRAIL Thesis Series, The Netherlands
- Lindveld, Ch.D.R., *Dynamic O-D matrix estimation: a behavioural approach*, T2003/7, September 2003, TRAIL Thesis Series, Eburon, The Netherlands
- Weerdt, de M.M., *Plan Merging in Multi-Agent Systems*, T2003/8, December 2003, TRAIL Thesis Series, The Netherlands
- Langen, de P.W., *The Performance of Seaport Clusters*, T2004/1, January 2004, TRAIL Thesis Series, The Netherlands
- Hegyi, A., *Model Predictive Control for Integrating Traffic Control Measures*, T2004/2, February 2004, TRAIL Thesis Series, The Netherlands
- Lint, van, J.W.C., *Reliable Travel Time Prediction for Freeways*, T2004/3, June 2004, TRAIL Thesis Series, The Netherlands
- Tabibi, M., *Design and Control of Automated Truck Traffic at Motorway Ramps*, T2004/4, July 2004, TRAIL Thesis Series, The Netherlands
- Verduijn, T. M., *Dynamism in Supply Networks: Actor switching in a turbulent business environment*, T2004/5, September 2004, TRAIL Thesis Series, The Netherlands
- Daamen, W., *Modelling Passenger Flows in Public Transport Facilities*, T2004/6, September 2004, TRAIL Thesis Series, The Netherlands
- Zoeteman, A., *Railway Design and Maintenance from a Life-Cycle Cost Perspective: A Decision-Support Approach*, T2004/7, November 2004, TRAIL Thesis Series, The Netherlands
- Bos, D.M., *Changing Seats: A Behavioural Analysis of P&R Use*, T2004/8, November 2004, TRAIL Thesis Series, The Netherlands
- Versteegt, C., *Holonic Control For Large Scale Automated Logistic Systems*, T2004/9, December 2004, TRAIL Thesis Series, The Netherlands
- Wees, K.A.P.C. van, *Intelligente voertuigen, veiligheidsregulering en aansprakelijkheid. Een onderzoek naar juridische aspecten van Advanced Driver Assistance Systems in het wegverkeer*, T2004/10, December 2004, TRAIL Thesis Series, The Netherlands
- Tampère, C.M.J., *Human-Kinetic Multiclass Traffic Flow Theory and Modelling: With Application to Advanced Driver Assistance Systems in Congestion*, T2004/11, December 2004, TRAIL Thesis Series, The Netherlands
- Rooij, R.M., *The Mobile City. The planning and design of the Network City from a mobility point of view*, T2005/1, February 2005, TRAIL Thesis Series, The Netherlands

Le-Anh, T., *Intelligent Control of Vehicle-Based Internal Transport Systems*, T2005/2, April 2005, TRAIL Thesis Series, The Netherlands

Zuidgeest, M.H.P., *Sustainable Urban Transport Development: a Dynamic Optimization Approach*, T2005/3, April 2005, TRAIL Thesis Series, The Netherlands

Hoogendoorn-Lanser, S., *Modelling Travel Behaviour in Multimodal Networks*, T2005/4, May 2005, TRAIL Thesis Series, The Netherlands

Dekker, S., *Port Investment – Towards an integrated planning of port capacity*, T2005/5, June 2005, TRAIL Thesis Series, The Netherlands

Koolstra, K., *Transport Infrastructure Slot Allocation*, T2005/6, June 2005, TRAIL Thesis Series, The Netherlands