The design of road and air networks for express service providers

Meuffels, W.J.M.

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

The Design of Road and Air Networks for Express Service Providers

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
Tilburg University, op gezag van de rector
magnificus, prof. dr. E.H.L. Aarts, in het
openbaar te verdedigen ten overstaan van een
door het college voor promoties aangewezen
commissie in de aula van de Universiteit op
woensdag 10 juni 2015 om 14:15 uur door

Wilhelmina Johanna Maria Meuffels

geboren op 13 januari 1985 te Sittard.

PROMOTIECOMMISSIE

PROMOTORES:	prof. dr. ir. H.A. (Hein) Fleuren
	prof. dr. ir. E.R. (Edwin) van Dam
OVERIGE LEDEN:	prof. dr. W.E.H. (Wout) Dullaert
	dr. ir. C.M.H. (Cindy) Kuijpers
	dr. ir. ing. M.J.P (René) Peeters
	prof. dr. T. (Tom) van Woensel

The Design of Road and Air Networks for Express Service Providers

Preface (in Dutch)

Het is bijna zeven jaar geleden dat ik begon aan dit proefschrift, en vandaag is het eindelijk zover dat ik dit voorwoord mag schrijven. Een moment om bij stil te staan, vol trots, maar vooral ook een moment om terug te blikken op zeven bijzondere jaren. Bovenal wil ik dit moment ook gebruiken om hen te bedanken die dit proefschrift mogelijk hebben gemaakt.

Het was Hein die zeven jaar geleden bij de afronding van mijn scriptie bij ORTEC aan me vroeg "heb je al eens nagedacht over promotieonderzoek?". Ja dat had ik, en nee, het leek mij helemaal niets om vier jaar lang theoretisch onderzoek te verrichten aan de universiteit. Ik had tijdens mijn scriptietijd de praktijk gezien en had meer interesse in de daadwerkelijke toepassing van besliskunde. "Maar lijkt het je niet leuk om juist in de praktijk te promoveren?". Dat was een optie die ik niet kende, en na enkele vervolggesprekken besloot ik die uitdaging aan te gaan: het leek me bijzonder interessant om wetenschap en praktijk gelijktijdig een stapje verder te kunnen brengen. Er volgden gesprekken met ORTEC en met TNT Express en beiden boden mij de kans om aan dit proefschrift te beginnen.

Het onderwerp van mijn proefschrift werd het ontwerpen van weg- en luchttransport voor expressnetwerken. Dat brengt mij ook bij het eerste team dat ik graag wil bedanken: TNT Express. Voor hun bijdrage aan praktische onderwerpen die verder onderzoek benodigden, voor de vele discussies en inzichten die daaruit volgden, en voor het enthousiasme van mensen waarmee ik heb samengewerkt. En in het bijzonder dank ik Marco Hendriks die dit proefschrift mogelijk heeft gemaakt. De onderzoeken op strategisch en tactisch vlak voor wegtransport zijn nauw verbonden aan de ontwikkeling van de modellen "TRANS" en "DELTA". Ik wil daarvoor de mensen uit de "Strategic Operations Development & Engineering"-groep, geleid door Marco Hendriks, van harte bedanken voor hun bijdrage aan dit onderzoek. Voor het onderzoek naar het ontwerp van luchttransport dank ik de betrokkenen vanuit het "European Air Network", in het bijzonder Bas van Dalfsen. Ook de groep vanuit het "European Road Network" mag niet ontbreken, in het bijzonder hen die betrokken zijn bij het "ROSCO-programma" dat geleid wordt door Camiel van der Velden. Op dit moment bewandelen we samen het pad van strategisch ontwerp tot operationele implementatie: een inspirerend pad met veel bekende maar ook onbekende uitdagingen, dat ik met veel plezier met jullie bewandel.

Uiteraard wil ik ook mijn dank getuigen aan mijn werkgever, ORTEC. Een speciaal woord richt ik tot Lambert van der Bruggen ("CEO" van de ORTEC Consulting Group bij aanvang van dit onderzoek) en Michael van Duijn (huidig "CEO" van de ORTEC Consulting Group). Ook wil ik graag een woord van dank richten tot de verschillende leidinggevenden die mij de afgelopen zeven jaar de mogelijkheid hebben geboden aan dit onderzoek te werken. Arjen van de Wetering, die mij de ruimte heeft gegeven om aan dit proefschrift te starten. John Poppelaars: ik heb mooie herinneringen aan onze samenwerking aan het artikel over de GO-Game en het werk voor de Franz Edelman Award; beide artikelen zijn opgenomen in dit proefschrift. Gregor Brandt was mijn manager in 2013; met veel plezier kijk ik terug naar de inspirerende sessies gedurende het ROSCO project en ik verwacht ook dat we in de toekomst bij ORTEC nog regelmatig zullen samenwerken aan ingewikkelde OR-problemen. Gregor wil ik ook bijzonder bedanken voor zijn belangrijke bijdrage aan mijn carrière, de geboden kansen en het vertrouwen dat ik heb ontvangen. Tot slot mijn huidige leidinggevende: Dave van den Hurck. Begin van dit jaar vertelde je me dat het één van je doelstellingen was dat ik dit jaar mijn proefschrift zou kunnen afronden. Dave, het is gelukt, en ik dank je voor het bewaken van de tijd die ik aan dit proefschrift heb kunnen besteden. Ik kijk met veel plezier uit naar onze verdere samenwerking!

Daarnaast wil ik ook graag alle lieve collega's bij ORTEC bedanken. Samenwerken met jullie is altijd een groot plezier! Mijn onderzoek is onlosmakelijk verbonden tot de projecten bij ORTEC, en om die reden wil ik graag een kort dankwoord richten tot mijn collega's. In het bijzonder richt ik mij tot de mensen uit het "oude DELTA-team": Timon van Dijk, Frank van der Wal en Annelies Woutersen. De projecten waaraan we samen hebben gewerkt waren over het algemeen niet de makkelijkste projecten. Maar ach, extreme situaties scheppen een band! Timon, je bent uitzonderlijk sterk op het OR-vlak. Ik heb graag met je gespard over moeilijke vraagstukken en ben trots op de samenwerking binnen de projecten maar vooral ook op ons gezamenlijk werk dat geleidt heeft tot het artikel over luchttransport. Frank, mijn steun en toeverlaat, op wie ik altijd terug kan vallen. Ik dank je voor onze fijne samenwerking, voor alle keren dat je me hebt laten lachen, maar vooral ook voor je steun tijdens minder leuke momenten. Tot slot wil ik Annelies bedanken: ik weet dat ik bij jou altijd terecht kan zowel zakelijk als privé. Ik hoop nog lang met jullie allen te kunnen samenwerken.

Dan wil ik graag een dankwoord richten tot de coauteurs van artikelen in dit proefschrift: jullie bijdrage is van grote toegevoegde waarde geweest voor dit proefschrift. Een bijzondere vermelding voor Frans Cruijssen, die intensief betrokken is geweest bij de start van dit proefschrift. Ook promotor Edwin van Dam, ben ik dankbaar voor zijn waardevolle feedback op de stukken in dit werk. Daarnaast richt ik mij tot alle studenten waarmee ik de afgelopen jaren gewerkt hebt, en die een belangrijke bijdrage hebben geleverd aan onderzoek in de expressmarkt. Dank jullie wel: Harm, Theresia, Ilse, Taco, Jan-Willem en Sander!

Wie niet mag ontbreken aan deze lijst is mijn promotor, Hein Fleuren. Hein, ik ben je enorm dankbaar voor de samenwerking in de afgelopen zeven jaar. Ik kijk terug op een bijzonder mooie periode en dat komt voor een groot deel door jou. Bedankt voor je enthousiasme, je inspiratie, voor de waardering die ik van je heb ontvangen, en voor de manier waarop je me bent blijven motiveren. Ik denk met plezier terug aan de vele brainstormsessies bij ORTEC, en de voorbereidingen voor de Franz Edelman Award beschouw ik nog steeds als een hoogtepunt in onze samenwerking. Ook de laatste weken hebben we intensief samengewerkt aan dit proefschrift en dank ik jou (en Dorine) voor de vele uren leeswerk en de waardevolle feedback die verwerkt is in dit proefschrift. Door jou ben ik aan deze reis begonnen, maar bovenal heb ik het met jou ook tot een mooi einde kunnen brengen. Dank je wel daarvoor!

Tot slot dank ik familie en vrienden, voor de fijne afleiding en steun tijdens dit proefschrift, en de warme thuishaard waar ik altijd op terug kan vallen.

Ineke Meuffels
Gouda, November 2014

Contents

Opening remark

This dissertation includes articles that we published in scientific journals. Only minor textual adjustments are applied to these articles for alignment of terminology used in this dissertation. Also note that some specific parts of our published work are used in general discussions, particularly in the introductory sections on express terminology.

The following scientific contributions are part of this dissertation:

- *Heuristics for the Uncapacitated Hub Location and Network Design Problem with a Mixed Vehicle Fleet and Regional Differentiation*[1]
  *V*an Essen J.T.; Meuffels W.J.M.; Aardal K; Fleuren H.A.; Submitted to a scientific journal

- *Enriching the Tactical Network Design of Express Carriers with Fleet Scheduling Characteristics*
  Meuffels, W.J.M.; Fleuren, H.A.; Cruijssen, F.C.A.M.; Van Dam, E.R.; Flexible Services Manufacturing Journal (2010), Vol. 22, Is. 1-2, pp. 3-35

- *Scheduling Movements in the Network of an Express Service Provider*[1]
  Louwerse, I.; Mijnarends, J.; Meuffels, I.; Huisman, D.; Fleuren, H.; Flexible Services Manufacturing Journal (2014), Vol. 25, Is. 4, pp. 565-584

- *Reduced Hub Handling in Multiple Hub Air Network Design for Express Providers by the Introduction of Pre-Sorted Unit Loading Devices*
  Meuffels W.J.M.; Van Dijk T.; Fleuren H.A.; Submitted to a scientific journal

Additionally we included the following scientific contributions as background information in the Appendix of this dissertation:

- *The Design of Express Networks in a Nutshell – Playing the Global Optimisation Game (GO-Game)*
  Meuffels, I.; Fleuren, H.; Poppelaars, J.; Hoornenborg, H.; De Rooij, F.; OR News (2010), Vol. 39, Is. 2, pp. 6-8

- *Supply Chain-Wide Optimization at TNT Express*
  Fleuren H.; Goossens C.; Hendriks M.; Lombard M.-C.; Meuffels I.; Poppelaars J.; Interfaces (2013), Vol. 43, Is. 1, pp. 5-20

[1] This research was conducted in cooperation with a student as part of a Master's thesis trajectory.

Research introduction and motivation

1 General introduction and the definition of research objectives

This chapter serves as a general introduction of the motivation and inspiration of research conducted in this Ph.D. research period (Section 1.1). Subject of this dissertation is the design of road and air networks for express service providers and for that reason we introduce express service providers and the express and parcel market (CEP-market) in which express companies operate in Section 1.2. Details on the express supply chain are provided in Section 1.3. We conclude this chapter with the definition of research objectives in Section 1.4.

1.1 Research inspiration and motivation

TNT Express N.V., one of the world's leading business-to-business express service providers, operates one of the largest express road and air networks in Europe and air and road transportation networks in China, South America, Asia-Pacific, and the Middle East. Express service providers move packages (i.e., parcels, documents, or palletised freight) from a sender to a receiver under various but guaranteed service level agreements. Each service level agreement consists of collecting packages at a sender, transporting them generally via a road and/or air network, and delivering them to a receiver within a specific delivery date and time.

The application of operations research (OR) at TNT Express during the past years has significantly improved decision-making quality and resulted in cost savings of many million euros. It is hard to imagine that about ten years ago, hardly any decision supporting tools were available at TNT Express, but the fact is that only in 2005, TNT Express embarked on its first operations research (OR) project. Triggered by a story on optimisation by Tilburg University professor Hein Fleuren, Marco Hendriks, Director of Strategic Operations & Infrastructure at TNT Express, sensed that quantitative methods should become the key enabler to increase the company's competitiveness. This awareness led to TNT Express' first OR project, which was aimed at optimising Italy's domestic road network. The results were promising: by rescheduling vehicles and reassigning packages, asset utilisation increased and transportation costs decreased by 6.4 percent. This initial success paved the way for the Global Optimisation-Program (GO-Program) and the close working relationship between TNT Express, Tilburg University, and ORTEC, an OR consulting and optimisation software provider that partners with TNT Express on optimisation activities (Fleuren, et al., 2013).

Within that partnership the opportunity emerged to contribute to both practice and science on express network design as part of this Ph.D. research. In fact, the application of OR within TNT Express was still in its infancy at the time that I started my Ph.D. research in 2007. Hence, my appointment was aimed at applying existing knowledge in practice, which was part of my job as consultant within a larger team of consultants at ORTEC, but also to expand existing research on fundamental areas that were not yet covered by academic literature. The latter has resulted in this dissertation that describes several extensions in the field of express network design.

Before discussing our research objectives, we first provide some general background on the courier express and parcel market (CEP-market) in which TNT Express operates, and detail on the supply chain operations of express companies.

1.2 Introduction to the Courier, Express and Parcel-market

The *courier, express and parcel* (CEP) industry concerns the collection, transport and distribution of packages, and can be segmented along the dimensions of service and weight. The service dimensions range from *same day*, *time-certain* to *day-uncertain* deliveries while the weight dimension distinguishes packages by size into *documents* (small and light goods), *parcels*, and *palletised freight* and in the extremes *full loads* and *bulk*. Different service providers operate different types of networks to guarantee specific delivery services, ranging from time-sensitive (air and road) express networks to less expedited sea carriers (see Figure 1, (TNT Express, 2010)).

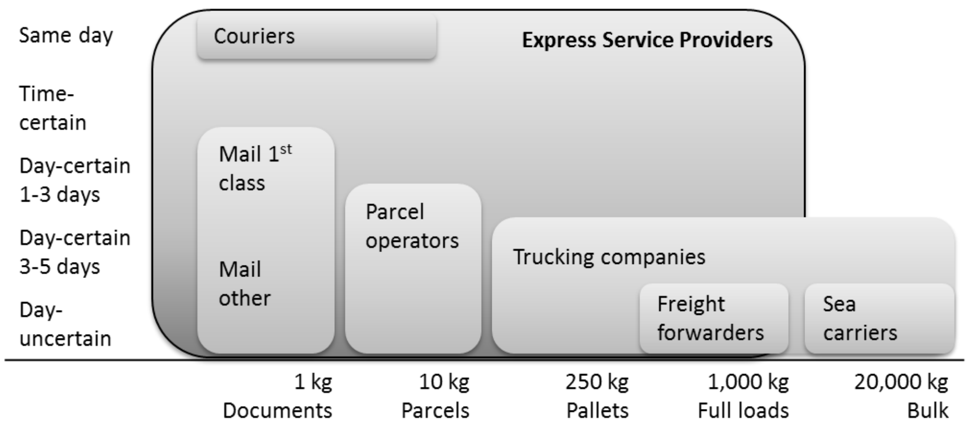


Figure 1: Operators per delivery network by segmentation of time and weight.
Source: TNT Annual Report 2010 with some minor adjustments, (TNT Express, 2010).

Core business of TNT Express lies in the European market, the latter having a total size of 47.2 billion revenues and 5.6 billion shipments a year, based on figures of 2011 provided in a report of A.T. Kearney (A.T. Kearney, Inc, 2012). The market is dominated by the four major global express service providers in the CEP-industry, i.e. DHL, UPS, TNT Express, and FedEx. According to figures presented in the annual report of DHL, the four express service providers already account for 88 percent of the total market share. Note that TNT express is classified as third largest in the European CEP-market (see also Figure 2).

To retain their leading position, express service providers have to respond to the trends that shape the market. The good news is that the market still continues to grow, but the growth pattern is shifting. We see an emerging growth in international transport, while so far domestic transport was clearly dominating (91% of the shipments in 2011); additionally it is expected that Europe's emerging markets (Poland, Russia, and Turkey) will show the strongest relative growth in the near future. Despite the continued growth of shipments in recent years, the increase in revenues stayed behind, setting margins under pressure. According to the report of A.T. Kearney (A.T. Kearney, Inc, 2012), the moderate increase in revenues is caused by the increase in cheaper standard services and lighter shipments and also because service providers not fully forward price increases on to customers. Another trend that is observed in the market is the acceleration of e-commerce and the increasing demand for last-mile solutions particularly caused by an increasing demand in the business-to-consumer segment. All in all vital reasons for express service providers to focus on cost cutting strategies that do not harm service, while investing in alternative and innovative solution methods to stay ahead of the competition. This dissertation is a consequence of that aspiration.

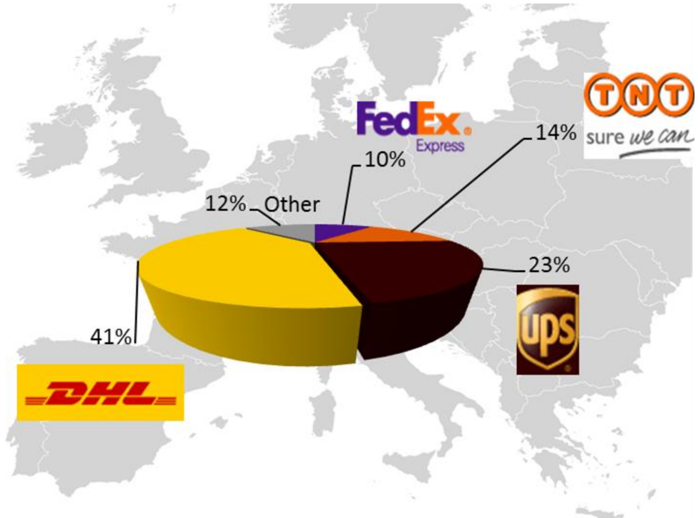


Figure 2: Market shares European international express market in 2011.
Source: DHL annual report 2012, (DHL, 2012)*.*

1.3 The express supply chain

Express service providers in the CEP-industry provide door-to-door services. Packages that are collected at customers are transported to *depots*, which are local sorting centres. Also the delivery of packages is organised by the depots. This process of collection and delivery of packages at the depots is referred to as the *pickup and delivery process (PUD)* (or *collection and distribution process*), while the transport between the depots is known as the *network process* (or *line-haul* process). Research in this dissertation is dedicated to the network process, and does not consider the pickup and delivery process.

The depot at which packages arrive after pickup at the customer is called the *origin depot* or *origin* of the package; the depot from where the package is delivered to the receiver is called the *destination depot* or *destination* of the package. *Cut-off times* (i.e. due times) separate the PUD process from the network process: at the *pickup cut-off time*, the packages must be available at the origin depot for the network process; at the *delivery cut-off time*, the packages must be available at the destination depot for the delivery process. In this, the PUD process encompasses the processing time at the depots, (i.e. at the pickup cut-off time all packages are processed, while the depot processing occurs after the delivery cut-off time).

The *service* of a package defines guaranteed delivery date and time, like a next day 09:00 hour delivery (time-definite) or a two-day service that requires delivery before business closure at the second day (day-certain). By the definition of cut-off times, service translates in a transhipment time from origin depot to destination depot within these cut-off times. Packages shipped between the same origin and destination depot within the same cut-off times can be transported together and are said to follow the same (*network*) *service*. These packages are referred to as the *flow* of an *origin-destination-service pair (ods-pair)*. Typically in express systems, depot to depot flows are too low to justify *direct* transport between the depots and consolidation occurs at *hubs*, being large consolidation facilities. The transportation *route* for an ods-pair defines the sequence of hubs to be visited by a package with that particular service, from the origin to the destination, with scheduled times of arrival and departure at the hubs. A *path* is the simplification of a route, denoting only the sequence of hubs and excluding timing information.

The most common transportation modes in the express business are road and air, and express service providers can achieve service fulfilment by the definition of timetables for the (majority of) operations of their vehicles and aircrafts. Package routes and paths are based on these schedules of vehicles and aircrafts. We now introduce some specific (but similar) terminology for the transportation tasks in road and air. When we describe the sequence of locations visited by a vehicle (and driver), including the times at which each location is visited, we talk about a *tour*. A *movement* connects two successive locations of a tour with no intermediary loading/unloading stops. Characteristics of a movement are its departure and arrival times as well as the corresponding vehicle type. An *empty* movement is a repositioning of a vehicle which is not carrying packages. A sector describes the existence of one or more movement(s) between two consecutive locations. Similar, we distinguish *flights*, *flightlegs*, and *legs* in the air network, where a flight denotes the sequence of airports visited by the airplane, the *flightleg* denotes a particular transport between two successive airports and its simplification that excludes timing information is known as a *leg*. An overview of the express supply chain is provided in Figure 3, and the scope of this dissertation is highlighted in the figure.

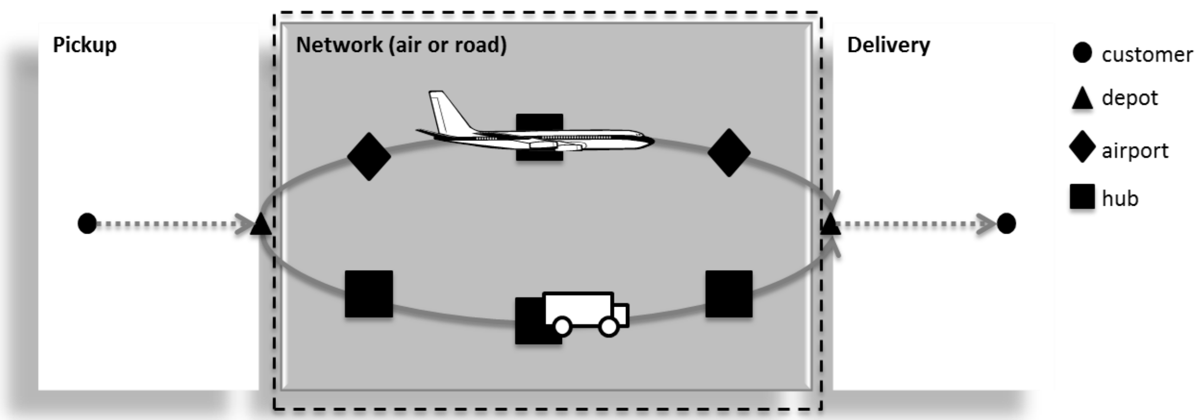


Figure 3: The figure shows the air and road supply chains; the highlighted area concerns the line-haul network composing the research scope of this dissertation.

1.4 Research objectives

The overall objective of this research is to develop OR models that support a cost efficient operation of an express line-haul network with guaranteed service level agreements. As this research was part of an overall program on supply chain-wide optimisation at TNT Express, the objective of this study was not only to provide insights regarding the design of such networks in general, but also to focus on the applicability of those models in the context of daily practice at TNT Express. This has contributed to the selection of subjects that needed further research, as well as the conditions that models had to meet on computational complexity, problem sizes, and practical conditions to be met. We now discuss the individual research objectives that are addressed in this dissertation for road and air network design.

1.4.1 Research objective I: hub location choice in express road networks

The first cost reductions at TNT Express were obtained by redesigning domestic road networks within the current depot and hub operations. The resulting savings were significant and triggered ideas for further optimisation at the infrastructural level. Particularly the question whether hub locations were positioned at the right place in express road networks was raised and asked for further research.

Clearly, if no hub locations would be available, vehicles should be operated between each pair of depots that offers an express service, causing inefficient transports. Also operating too few hubs results in inefficient transport leaving room for further cost savings. On the other hand, operating too many hub locations would cause high housing cost and if packages are handled at too many hub locations the advantage in line-haul cost reductions disappears due to an increasing handling cost. So in order to make the right decisions in hubs to operate, the trade-off between line-haul cost reductions and hub cost increases has to be made carefully.

An additional level of complexity that should be kept in mind when deciding on hub locations is the period of decisions made. Building a hub location might already take up for two to three years, and a hub may easily be operated for ten to fifteen years. This brings a level of uncertainty that should be handled in the best possible way. A possible hub configuration might be verified amongst future package volumes, but also fluctuations in transport cost or hub housing and handling cost should be studied when making decisions on hubs to operate. The solution method should at least support evaluation of such scenarios.

The observed challenges when making infrastructural decisions has resulted in the definition of the first research objective addressed in this dissertation: *Develop a method that supports the decision on hubs to be operated in an express road network such that total cost is minimised under tight service requirements.* Note that the corresponding research questions will be defined during conceptualisation of the problem in Chapter 2.

1.4.2 Research objective II: package routing and fleet scheduling in express road networks

The second research objective was instigated by the deployment of changes that resulted from the opening and/or closing of hub locations in road networks. So we first had to think about what was needed to accelerate the implementation of network changes. Firstly, at the moment that a new hub location becomes available for use, it has to be decided which packages will be routed via this hub and fleet decisions need to be taken to support the adjustments in routings. Note that adapting routes does not only concern the newly opened hub but also causes changes in package flows at other hub locations in the network. Clearly, it is a time-consuming job if the impact in the total network has to be reviewed in a manual way, which was the case at TNT Express at that time. Hence, more advanced planning tools were needed to evaluate the effects of the strategic changes in a short time span.

One may remark that the earlier decisions on the hub infrastructure also involved decisions regarding flows routed via hub locations. So why would we reinvestigate these outcomes and not just implement routes accordingly? There are actually two reasons to reconsider the outcomes of routing decisions at this stage. Firstly, hubs are in general built and/or closed in a sequential way, so that on the path towards the final hub configuration, hubs might be used in a slightly different intermediate network set-up. Secondly, note that the building process of a hub takes about two to three years, so that package flows already may have changed. It is also possible that the existing road infrastructure has been improved meanwhile, so that other routing opportunities arise. For that reason, package routings might be refined and fleet schedules need to be adjusted accordingly. Actually, also when the final hub set-up is available for use, routes and fleet schedules will be refined at regular times in the planning process.

The above motivated the second research objective in this dissertation: *Develop a method that designs the set of movements and supports (refined) routing decisions to achieve service commitment at minimum cost in an express road network.* Note that the corresponding research questions will be defined during conceptualisation of the problem in Chapter 2.

1.4.3 Research objective III: package routing and flight scheduling in multiple hub express air networks

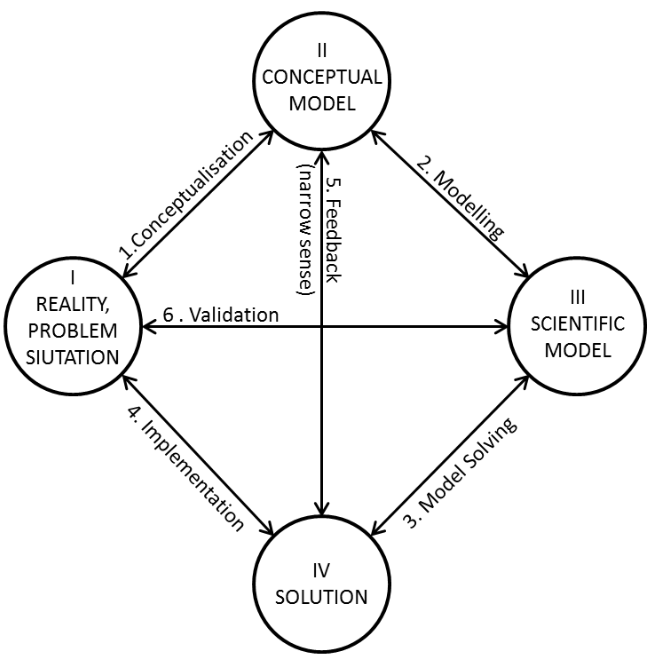
The research performed in this dissertation for the design of road networks also gave reason to think about optimisation opportunities in the design of air express networks. Particularly at the time that TNT Express was observing pressure on hub handling capacities in their European air network as a result of increasing number of lighter shipments, we were asked to support decision making by developing a solution method that can evaluate different set-ups of European air operations.

At the time that we were approached for research in this field, TNT Express operated their European air network with a single hub in Liège. Decisions had to be taken in order to be able to commit to the services offered to their customers as existing hub handling capabilities were insufficient. So we were asked to think about alternative ways of working that would enable TNT Express to offer best services to their customers at lowest possible cost. We were asked to investigate if there existed possible solutions to reduce the desired handling capacities without the necessity of heavily investments. If investment had to be made anyway, it had to be investigated if these had to be made at their current hub, or whether it was more efficient to invest in a second hub in Europe. The operation of a second hub in Europe would also result in contingency advantages, and was considered as a topic that needed further consideration.

That resulted in the definition of our third research objective, which is stated as: *Develop a method that supports routing decisions and designs flight schedules at minimum cost for multiple hubs in express air networks.* Note that the corresponding research questions will be defined during conceptualisation of the problem in Chapter 2.

1.5 Concluding remarks

This chapter introduced the operations of express service providers and the research objectives in the field of express network design that are addressed in this dissertation. Now that we are familiar with the problem situation, the next step is to design the conceptual model. Chapter 2 is used to make a conceptualisation of the problem situation by definition of the scope, assumptions and design variables; also existing research is used to refine the conceptualised problem situation. The conceptualisation phase is closed by the definition of research questions that belong to each of the research objectives. The third and fourth phases in our research are derivation of the scientific model and application of this scientific model on data instances to provide a solution to the problem situation. For each of the research objectives, the scientific model and solution is provided in individual chapters in this dissertation. We finalise this dissertation by stating our feedback to improve the conceptualisation of the problem situation and/or discuss the implementation reality. Note that the four-phase approach that starts from definition of the problem situation, followed by a conceptualisation phase, and the phases to derive the scientific model and solution is a common research method to view the problem in a systematic way. For more details about this research method we refer to (Mitroff, et al., 1974) and an illustration of the approach is provided in Figure 4.



DISSERTATION OUTLOOK:
SYSTEMS VIEW OF PROBLEM-SOLVING

ELEMENTS	
I REALITY, PROBLEM SITUATION	Chapter 1
II CONCEPTUAL MODEL	Chapter 2
III SCIENTIFIC MODEL	Chapters 3 – 6*
IV SOLUTION	Chapters 3 – 6*

PATHS	
1. Conceptualisation	Chapter 2
2. Modelling	Chapters 3 – 6*
3. Model Solving	Chapters 3 – 6*
4. Implementation	Chapters 7 and 8**
5. Feedback	Chapters 7 and 8**
6. Validation	Chapters 7 and 8**

* *The Chapters 3 – 6 concern individual research objectives.*
** *The Chapters 3 – 6 contain already a first indication of the paths of implementation, feedback, and validation in their concluding sections. However, a more general review of these is provided in Chapters 7, and 8.*

Figure 4: A systems view of problem-solving (Source: (Mitroff, et al., 1974)) and resulting dissertation outlook.

2 Conceptualisation of research objectives

The first challenge in conceptualisation of a problem situation is to make decisions on scope. It is important to define the right variables while making assumptions that support problem-solving but remain valid during implementation of the results. Scope, assumptions and design variables strongly relate to the planning horizon for which decisions have to be taken. As a first step in specification of the conceptual model, we therefore start this chapter by specification of the different network planning levels and discuss the resulting scope, design variables and assumptions per research objective in Section 2.1. Afterwards we review existing research available within literature and make use of these earlier studies to draw points of attention for research performed in this thesis (Section 2.2). We use these points of attention to refine the conceptual models when stating our research questions in Section 2.3.

2.1 Conceptualisation that suits the planning hierarchy

In the previous chapter we gave a description of the express supply chain and the scope of this dissertation being the network process. The organisation of such a network appears to be rather complex and requires decisions at various levels ranging from strategic (long-term) planning via the tactical (medium-term) planning level to the operational (short-term) planning level. There are several papers that discuss differences in planning levels; we present now the definitions made by Crainic (2002) in a survey overview on long-haul freight transportation:

Strategic planning - "The strategic (long-term) planning involves the highest level of management and requires large capital investments over long-term horizons. Strategic decisions determine general development policies and broadly shape the operating strategies of the system. These include the design of the physical network and its evolution, the location of major facilities, the acquisition of major resources such as motive power units, and the definition of broad service and tariff policies".

Tactical planning - "The tactical (medium-term) planning aims to determine, over a medium-term horizon, an efficient allocation and utilisation of resources to achieve the best possible performance of the whole system. Typical tactical decisions concern the design of the service network and may include issues related to the determination of the routes and types of service to operate, service schedules, vehicle and traffic routing, and repositioning of the fleet for use in the next planning period".

Operational planning - "The operational (short-term) planning is performed by local management, yard masters and dispatchers, for example, in a highly dynamic environment where the time factor plays an important role and detailed representations of vehicles, facilities and activities are essential. Important operational decisions concern: the implementation and adjustment of schedules for services, crews; the dynamic allocation of scarce resources".

In the next sections, we discuss the network planning level of each research objective and use this classification in order to define scope, variables and assumptions per topic.

2.1.1 Strategic network design: hub location choice in express road networks

Our first research objective concerns supported decision making on hubs to be operated in an express road network. As stated earlier, the decision to use a hub location is a decision that will remain for ten to fifteen years, which is clearly long-term. Also the investment to operate a hub, which consists of several components (i.e. yard, building, equipment …) accounts for million euros per hub site. So both the term of the investment as well as the amount explain why this problem situation is classified as strategic.

When it is decided to open a hub at a site, one has to be sure that cost in the network reduces for a number of years. Efforts can be put in forecasting of future package flows such that detailed fleet operations can be designed, and all other kind of details can be gathered for the final design of the hub in the network. However, in the end, there will always remain a certain degree of uncertainty that is impossible to predict. In general, the decision to use a hub at a site is never a coin flip and is taken only as clear cost savings can be proven when viewing the solution from different points of view. The exact use of the hub in terms of needed workforce, in terms of package routings, and fleet schedules is in general taken at the time that the hub actually becomes available for use. At the strategic planning stage, decisions on paths for ods-pairs, and sectors to indicate where transportations need to be operated, without going into the details of timetabling, should suffice for infrastructural decisions. Also a general indication of the investment for the opening/closing of hub locations should suffice for the design of the network.

That means that we are now able to define the scope of the first research objective as well as defining decisions considered out of scope. For the sake of completeness we also state the general assumptions that follow from the scope of line-haul network design in express networks.

Scope:

- the decision on the number and location of hubs;
- the selection of paths for each ods-pair, and;
- the decision on sectors that are operated.

Out of scope:

- the design of routes for each ods-pair;
- the decision on movements, tours, drivers, fleet size;
- detailed layout, investment, workforce decisions for hubs.

Assumptions: fixed hub processing times are given, depot locations are known and fixed, flow and service information is given, all cost information is available.

2.1.2 Tactical network design: package routing and fleet scheduling in express road networks

Our second research objective concerns supported decision making on routings and fleet schedules operated in the express road network. Although day-to-day routings and fleet schedules may show deviations, the main part of the network is operated in the same way each day. This is because express service providers make use of a timetable that specifies the operation of vehicles and resulting package routes. At the operational level, planning starts with this standard timetable and minor adjustments are made to reflect the actual situation of the operational day. Examples that result in an adjusted operational planning are expected increases in package flows that might result from marketing campaigns at a customer but also expected reductions in package flows as a result of public holidays and necessary adjustments because of roadblocks. It is clear that the basic timetable has to be reviewed at regular times (e.g. quarterly) to assure that it remains actual and to take care of seasonality effects. Due to these regular reviews, the effort to translate the timetable in an operational plan is low. In this research objective, we study the creation of the basic timetable, which is a problem with a tactical planning nature.

There are two topics that we would like to address as these have been set out of scope during the research performed in this dissertation. Extension of the conceptual model regarding these topics would add value, but also strongly adds a level of complexity to the problem at hand. The first topic to be discussed is the design of hub operations, particularly workforce related operations. At the tactical planning level hub locations are known but we also consider hub handling capacities and processing times as a given. Clearly, changes in vehicle departures/arrivals at hub locations might have influence on the needed handling capacities or the resulted handling times; this topic might be addressed in further research on the integral line-haul and hub location problem.

The second topic that we have excluded from the conceptual model is the generation of tours as well as the assignment of drivers to tours. One may argue that the creation of a timetable is strongly related to decision making regarding fleet sizes and hence the timetabling problem cannot be considered in isolation. The reason to exclude the topic in this research is mainly for the reason of simplification, as the tour generation and driver assignment problem is a very complex problem in itself. Particularly as express service providers often make use of subcontractors to outsource part of the set of movements it is a problem with a high level of uncertainty as prices may vary and be negotiable over time. Due to the latter, it is also seen in practice that this problem is solved by first creating the movements where tour generation follows afterwards; some adjustments may be made to the sets of movements and tours as a result of negotiation with subcontractors. In order to assure that timetabling choices show the right tendencies to estimate the result after tour generation, some notion of tours is considered in this research by desiring a balanced set of movements at the end of the planning horizon, i.e. the number of vehicles that start at a location at the beginning of the planning horizon equals the number of vehicles at the same location at the end of the planning horizon.

That means that we are now able to define the scope of the second research objective as well as defining decisions considered out of scope. For the sake of completeness we also state the general assumptions that follow from the scope of line-haul network design in express networks.

Scope:
- the design of routes for each ods-pair;
- the decision of movements;
- balance of incoming and outgoing movements at all locations.

Out of scope:
- the decision on tours, and fleet size;
- decisions on driver assignment to tours;
- detailed layout, investment, workforce decisions for hubs.

Assumptions: fixed hub processing times are given, hub locations and capacities are known and fixed, depot locations are known and fixed, flow and service information is given, all cost information is available.

2.1.3 Tactical network design: package routing and flight scheduling in multiple hub express air networks

The question on the number and location of hubs in air network design is, similar as in road network design, a strategic decision. What is different in air network design compared to road network design is the dominance of the fleet cost: aircraft cost are dominant in air network design, where hub cost and line-haul cost are both significant in the design of road networks. Additionally, continental flights are in general fully operated by the express service provider itself as few outsourcing opportunities are available that can commit to the service standards. As a result, decisions on fleet size in air network design can be considered as a strategic/tactical problem that needs to be considered jointly in the decision on which hubs to operate.

However, one may note that the hub location choice in road network design is more complex than the hub location problem in air network design. Firstly, because road networks in general are operated via multiple hub locations while continental air network design in general makes use of a single hub location, as is confirmed by the fact that the big four express service providers all operate single hub networks in Europe (see Figure 5). Although we are considering multiple hub air networks, we argue that a minimal cost continental air network makes use of a few hub locations. Secondly, almost no restrictions are tied to the exact site of hubs in road networks; the possible hub sites for air operations are limited by the availability of airports and even lower as a result of night regulations. Hence, the strategic hub location problem boils down to enumeration of possible hub configurations while optimising package routings and flight schedules and the conceptual model is valid enough when it can support the evaluation of a given hub configuration.

When considering the planning hierarchy for air network design, one may note that there is in fact no clear distinction between strategic and tactical planning levels in air network design as the hub location choice and fleet sizing choice is made jointly, and almost none outsourcing opportunities exist to capture additional aircraft capacity. At the tactical planning level, only small changes are applied to the schedules of the aircrafts basically by swapping different-size aircrafts between flights when package volumes shift. For ease of the comparison to road network design we have chosen to classify the problem that we consider during the third research as tactical.

Figure 5: The four largest express service providers operate a single hub European air network (FedEx: Charles de Gaulle, TNT Express: Liège, UPS: Cologne, DHL: Leipzig)[2].

Similar as in tactical road network design, we would like to mention that we have excluded design decisions regarding the hubs used in the network. When creating the scientific model the hub locations are given and we consider hub handling capacities and processing times as given. Additionally, crew scheduling decisions are considered out of scope. The research objective might be extended on these topics in the future.

There is one additional topic to discuss regarding this research topic. When we first considered multiple hub air network design, we could easily conclude via a back-of-the-envelope calculation that on-time delivery comes in danger when packages have to be sorted more than once. On the other hand, as package flows at individual airports are low, operating flights to two hubs is not cost efficient. This had led to the idea of pre-sorted unit loading devices (ULDs), being containers used to load packages in single units that can be loaded at once in an aircraft. When pre-sorting is used, packages can visit multiple hubs while only being sorted once. The idea of pre-sorted ULDs would provide a solution to the operation of a multiple hub network as well as it results in reduced handling opportunities in a single hub network. Note that the availability of ULDs is considered as unlimited. One may also question if we need to consider repositioning of ULDs so that the scheme can be repeated each day. This is however no issue, as aircrafts are always fully loaded with ULDs even when there is no load to fill (some of the) ULDs of an aircraft.

---

[2] Sources:
FedEx - http://www.fedex.com/cn_english/services/euroone/routemap.html
TNT Express - http://www.tnt.com/express/en_lu/site/home/about_us/about/facts_figures.html
UPS - http://www.ups.com/content/nl/nl/about/facts/europe.html
DHL – http://www.dpdhl.com/en/about_us/at_a_glance/publications.html (Report: *The development of Deutsche Post DHL)*

That means that we are now able to define the scope of the third research objective as well as defining decisions considered out of scope. For the sake of completeness we also state the general assumptions that follow from the scope of line-haul network design in express networks.

Scope:
- the design of routes for each ods-pair;
- the decision on flightlegs and flights; the decisions on ULD routing, sorting and loading.

Out of scope:
- decisions on crew assignment to flights;
- detailed workforce decisions at hubs.

Assumptions: fixed hub processing times are given, hub locations and capacities are known and fixed, airport locations are known and fixed, depot locations are known and fixed, flow and service information is given, all cost information is available.

2.1.4 Illustration of strategic and tactical road network design

In this section, we illustrate the concept of strategic and tactical network design in Figure 6 and discuss typical decisions made in these fields in the sections below. Note that we used the network configuration of the *GO Game*, which is discussed in detail in Appendix A, but used simplified data for the ease of the illustrations and discussions.

Starting point for network design

In Figure 6a, the starting point for network design is provided. The figure denotes the depot locations, ten in total, and corresponding flows are provided in the table below. In total, 3,200 packages need to be transported via the network.

There are a few remarks that can be made already regarding the flow of packages. Firstly, note that there exist imbalances in flows: some depots send/receive more packages than other depots. Also for an individual depot, imbalances can occur in the flows that it receives or distributes. Depot *f* for example is the origin depot of 595 packages while it only receives 390 packages for delivery to customers. For that reason, depot *f* is sometimes referred to as a *net sender* while a depot that receives more packages than the number of packages that originate at the depot is called a *net receiver*, like depot *c*. Also note that it is possible that no service is offered between a pair of depots. In this example, depot *g* offers no services to depots *f, h, i,* and *j*.

Additionally, note that the time for transport is provided by the cut-off times which in this illustration are given at 20:00 in the evening and 07:00 in the morning. In this illustration, all depots have the same cut-off time, but in practice cut-off times may vary among the depots. Also note that all packages in this example have the same *express* service and are to be delivered overnight. In practice, a different service might be offered between the same pairs of depots.

Strategic network design

The aim of strategic network design is the selection of the number and location of hubs. However, in order to make the decision for a network that performs best in terms of cost efficiency, we need to make the right trade-offs between hub cost, in terms of handling cost per package and fixed cost to operate a hub (e.g. rental cost), versus line-haul cost for the transportation.

Although in theory hubs can be operated at any site, we observe that strategic network design in practice generally starts by indicating a number of potential hub locations. Typically, potential hub locations are chosen at the sites of existing depots, in order to take advantage of the combined operation of a depot and hub. Sometimes, potential hub locations are chosen at sites that have no depot function (yet), particularly when such a potential hub site is located near highway intersections.

There is a small remark that can be made regarding the selection of hub locations, when approaching it from a modelling point of view. Clearly, the complexity of the problem increases when more potential hub sites are to be considered. In order to reduce computational effort, educated decisions can be made regarding the selection of potential hub sites. For example, sites at the borders of the network are often not selected as potential hub sites, as operating a hub at such a location increases in general transport cost. For example, consider the very small network that needs to be operated between depots *a, b,* and *c* and suppose that depot *a* would be chosen as a hub location. Clearly, when flows from depot *b* to depot *c* would use hub *a* in this situation, the flows first need to be transport completely to the north and are afterwards transported back to the south again to be delivered to depot *c*. Obviously, less kilometres need to be driven if a hub would be operated at depot site *c* instead of depot site *a* from a line-haul perspective. In this illustration, we therefore excluded depots *a, b,* and *j* as potential hub locations.

A possible outcome of strategic network design in this example is provided in Figure 6c. Let us discuss some details of the outcomes. So first note that the selected hub configuration is apparently a network with two hubs, at potential sites *c* and *f*. This resulted from selecting a path for each package flow that visits either none, one or two hubs in this situation. As the example shows, the path between two depots does not need to be symmetrical: the path from depot *a* to depot *g* makes use of hubs *c* and hubs *f* while the path from depot *g* to depot *a* visits hub *c* only. Also note that a depot can be served by other hubs for flows that originate at the depot or destine at the depot. In this example, depot *g* only sends packages to hub *c* but receives packages via hub *f*. Recall that depot *g* does not offer services to depots *f, h, i,* and *j*, which explains why it connects to hub *c* only for packages that originate at this depot.

This hub configuration should support the transport of all packages in the network in the most cost efficient way. Based on the selected paths, it is also known from where to where transports are operated, i.e. the sectors in the network are known and the total flow of packages that traverses via each such sector can be calculated. Based on this, line-haul cost can be estimated and the way to do this depends on the chosen approach. Similar, we know based on the paths which amount of packages is handled at each hub so the cost to operate the hub can be estimated as well.

Tactical network design

As soon as hubs become operational, tactical design questions need to be answered. In fact, this stage reviews the strategic outcomes regarding paths and sectors at a more detailed level while considering the hub configuration as a given (Figure 6c).

The outcome of the tactical network design is illustrated in Figure 6e. The table on the left denotes the selected routes while the table on the right indicates the movements. Note that in the strategic network design, we already decided that flows from depot a to depot b might be routed via hub c but at this time we also know that the packages leave depot a at 20:00 and arrive at hub c at 22:00. The packages are sorted at the hub and leave the hub at 02:00 while arriving at the destination depot in b at 06:00 hour. On the other hand, note that there did not yet exist a sector between depot a and depot g at the strategic network design level. Previously, the chosen path from depot a to depot g visited both hub locations. However, note that the movement from hub c to hub f arrives at hub f at 04:00 while the movement from hub f to depot g already has to leave at 04:00 in order to meet the delivery cut-off time. Clearly there is no time to move packages from the interhub-movement to the hub-depot movement and this is resolved by the inclusion of a direct transport from depot a to depot g instead.

Furthermore there is something to mention regarding the arrival/departure pattern of movements at hubs and depots. When more movements arrive/depart at the same time at a hub or depot the workload to handle these movements increases. A spread of movements is hence preferred when possible. In this example, the direct movement from depot a to depot d has some slack considering the available time for transport and the driving time, and therefore leaves depot a at 21:00 hour while the cut-off time of the depot is 20:00 hour. Also note that in practice it is possible that many movements are scheduled between a pair of locations, e.g. between large metropoles; generally, the departure and arrival times of these movements differ.

Remarks

There are a few remarks that we would like to address for the sake of completeness. First one may question the robustness of outcomes at the strategic stage of the network design. Hub location decisions are made for years, so we need to be absolutely sure about the decisions that we take here. How do we get confidence in the results? This question is in practice answered by running a number of scenarios at the strategic level. One can think of strategic scenarios in which other sets of package flows are used (e.g. future volume scenarios), in which cost parameters are varied (e.g. what is the influence of fuel cost), scenarios with other cut-off times or service days (e.g. what is the impact of a two-day service), among many other scenarios that vary the dynamics of the network. Additionally, for the final outcome of the strategic network design, we sometimes use tactical network design models for validation of the results to assure that timing aspects will not change conclusions drawn. Also, it might occur that an existing hub infrastructure exists that deviates significantly from the modelled outcomes. In that situation, scenarios are often used to validate whether the effort to adopt the hub configuration weighs against the expected savings that can be obtained from an adjusted operation.

Figure 6: step-wise illustration of strategic and tactical network design.

A: **starting point for network design are flows for which transport needs to be organised and the corresponding service levels that are offered to the customer.**

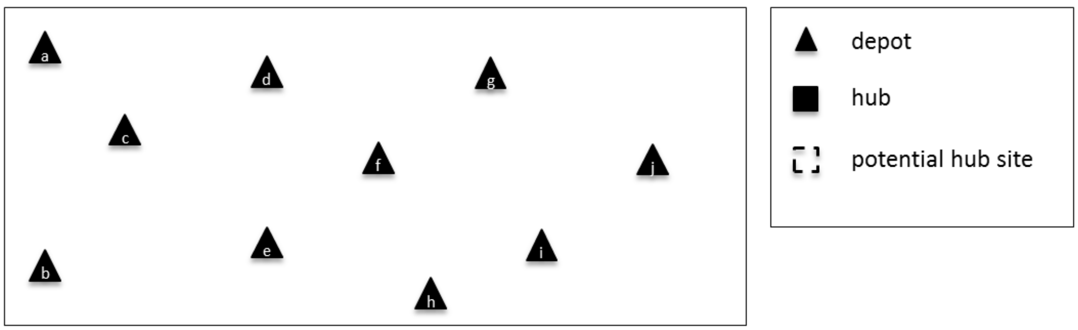
B: **for strategic network design, potential hub locations are identified first.**

C: **the outcome of strategic network design is the chosen hub configuration, and a draft of the network setup via the selection of paths and resulting sectors.**

D: **the strategic network design is a starting point for tactical network design; hub locations are given at this stage, and a review of paths and sectors is done at a more detailed level by the inclusion of times.**

E: **the outcome of tactical network design is the selection of routes and movements.**

Starting point for network design



Flows from depot to depot (express service)

from\to	a	b	c	d	e	f	g	h	i	j
a		20	15	30	80	50	20	10	10	5
b	40		5	20	20	15	80	80	60	120
c	20	40		80	20	100	10	10	5	5
d	60	10	90		10	80	5	20	20	30
e	10	100	10	100		40	50	10	80	10
f	15	100	150	10	180		20	100	10	10
g	80	5	5	10	20	0		0	0	0
h	10	10	20	20	30	30	15		50	30
i	15	20	10	20	10	15	80	10		60
j	10	40	80	30	80	60	10	10	5	

Remarks:
express flows only

pickup cut-off time: 20:00
delivery cut-off time: 07:00

(all next day deliveries)

Figure A: starting point for network design are flows for which transport needs to be organised and the corresponding service levels that are offered to the customer.

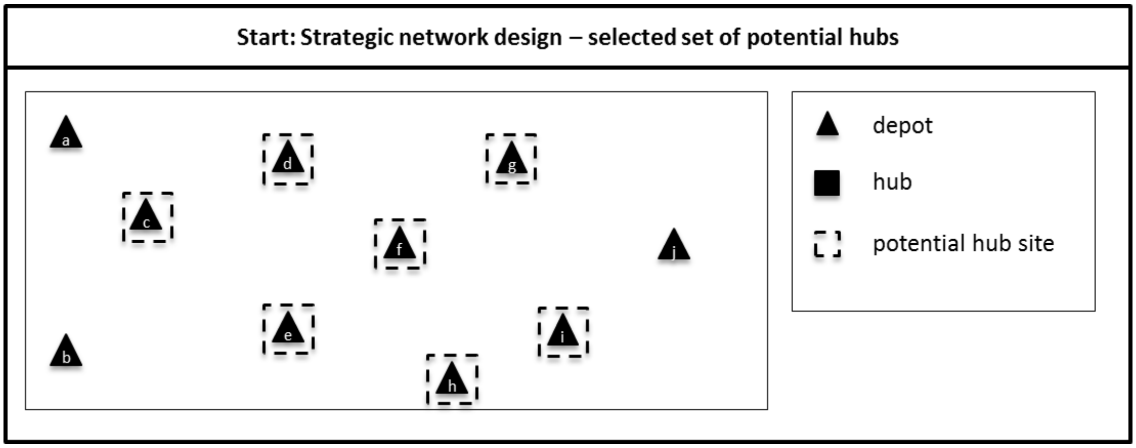
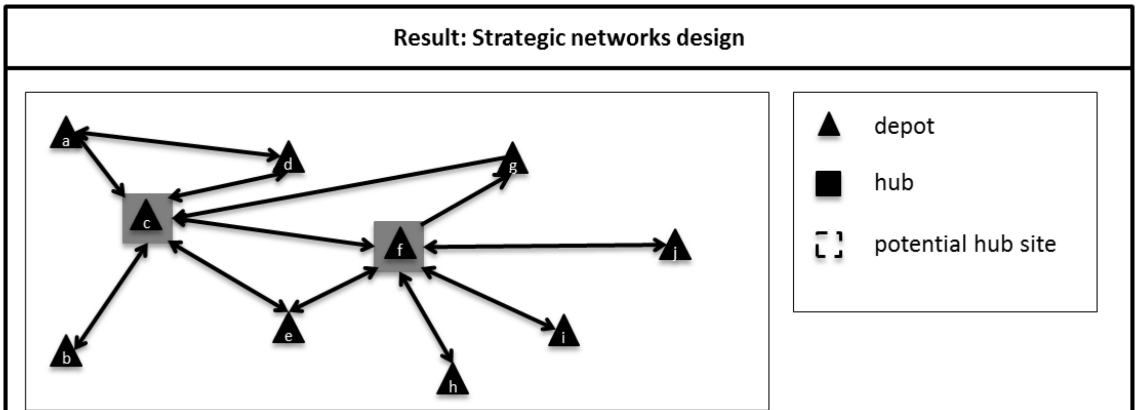


Figure B: for strategic network design, potential hub locations are identified first.



Paths from depot to depot

orig	dest	serv	loc 1 →	loc 2 →	loc3 →	loc4
a	b	exp	a →	hub c →	b	
a	c	exp	a →	hub c →	c	
a	d	exp	a →	d		
a	e	exp	a →	hub c →	E	
a	f	exp	a →	hub c →	hub f →	f
a	g	exp	a →	hub c →	hub f →	g
a	h	exp	a →	hub c →	hub f →	h
a	i	exp	a →	hub c →	hub f →	i
a	j	exp	a →	hub c →	hub f →	j
...	...	...	...	...	...	...
g	a	exp	g →	hub c →	a	
...	...	...	...	...	...	...

Sectors

from	to	flow
a	hub c	210
a	d	30
hub c	a	295
d	a	90
b	hub c	440
hub c	b	345
...	...	
hub c	hub f	735
hub f	hub c	560
...	...	
g	hub c	100
...	...	

Figure C: the outcome of strategic network design is the chosen hub configuration, and a draft of the network setup via the selection of paths and resulting sectors.

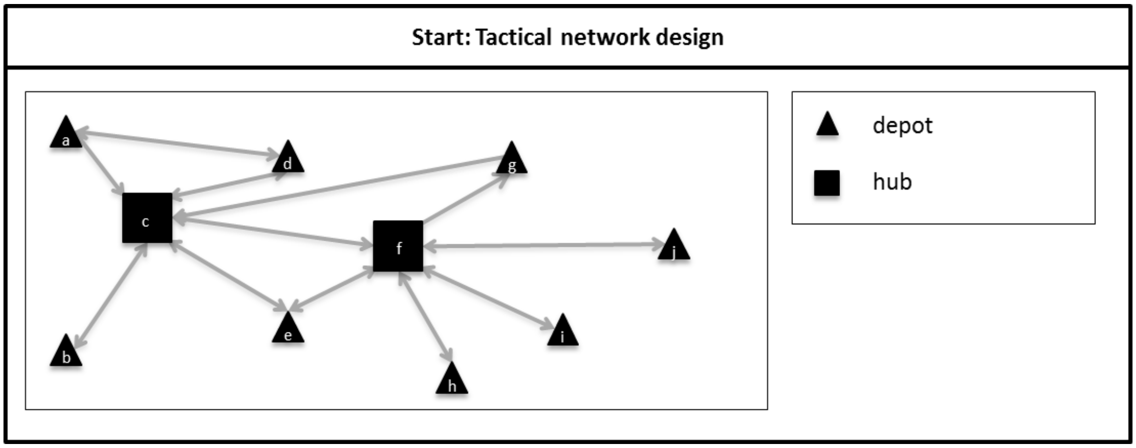
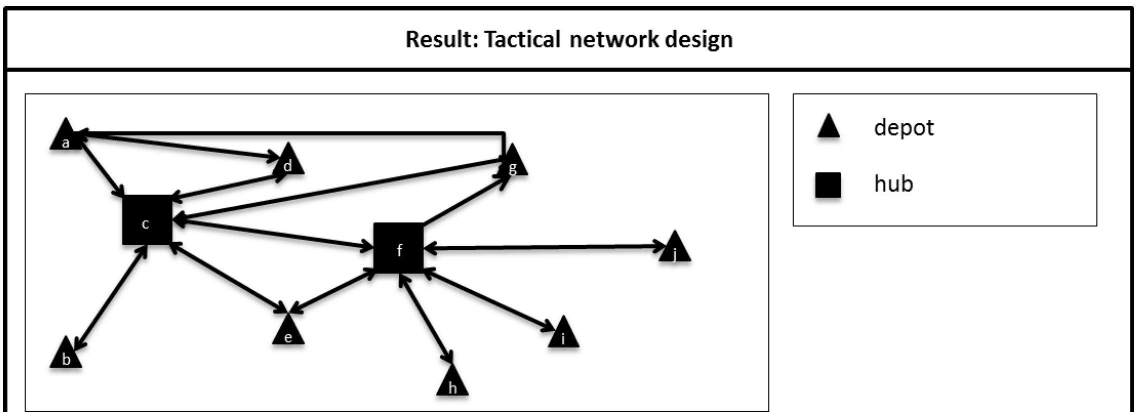


Figure D: the strategic network design is a starting point for tactical network design; hub locations are given at this stage, and a review of paths and sectors is done at a more detailed level by the inclusion of times.



Routes from depot to depot

orig	dest	serv	route sequence	from	to	dep. time	arr. time
a	b	exp	1	a	hub c	20:00	22:00
a	b	exp	2	hub c	b	02:00	06:00
a	h	exp	1	a	hub c	20:00	22:00
a	h	exp	2	hub c	hub f	23:00	04:00
a	h	exp	3	hub f	h	05:00	07:00
...	...	...		...	...	...	...
a	g	exp	1	a	g	20:00	04:00
...	...	...		...	...	...	...
h	g	exp	1	h	hub f	20:00	23:00
h	g	exp	2	hub f	g	04:00	07:00
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

Movements

from	to	dep. time	arr. time
a	hub c	20:00	22:00
a	d	21:00	02:00
hub c	a	05:00	07:00
d	a	20:00	01:00
b	hub c	20:00	02:00
hub c	b	02:00	06:00
...	...	...	...
hub c	hub f	23:00	04:00
hub f	hub c	23:00	04:00
...	...	...	...
hub f	g	04:00	07:00
...	...	...	...

Figure E: the outcome of tactical network design is the selection of routes and movements.

2.2 Refinement of the conceptual model based on earlier research

The problems that we consider in this dissertation are known as (special variants) of the network design problem. Recent overviews on network design in express networks are given by Alumur & Kara (2009), Wieberneit (2008). Overviews on network design in general are given by Revelle & Eiselt (2008), Melo et al. (2009), Campbell & O'Kelly (2012), and Farahani et al. (2013) .

In Section 2.2.1 we give a formulation of the general network design problem. Fundamental assumptions in network design solutions are discussed afterwards in Section 2.2.2. Some notes on the computational complexity of the problem are provided in Section 2.2.3. We end this section by a conclusion on focus areas for research on the network design problem in express networks (Section 2.2.4).

2.2.1 Formulation of the general network design problem

Various network design models have been formulated for different purposes and a unifying view on them has been presented in (Magnanti & Wong, 1984). The general network design formulation as presented in (Magnanti & Wong, 1984) starts by the definition of a network represented by a graph ($G = (N, A)$), that consists of a set of *nodes (N)* and a set of directed or undirected *arcs (A)*. The nodes in an express network are formed by the depots, airports, and hub locations and the arcs, which are directed in the situation of an express network, represent the (possible) sectors in a road network and the (possible) legs between locations in an air network.

Furthermore, (Magnanti & Wong, 1984), defines a set of commodities (K) for which a flow of packages, R_k, has to be shipped from its point of origin $O(k)$, to its point of destination, denoted by $D(k)$. If there is only one commodity that needs shipment, the network design is referred as a *single-commodity* network design in contrast to *multi-commodity* network designs that serve a variety of commodities. Typically, express networks have a strongly multi-commodity nature because each ods-pair needs to be accounted individually. Note that this clearly distinguishes the express business from any other type of business: where general location distribution systems focus on delivery of a product from a random production location to a customer, express businesses need to deliver specific packages from sender to receiver within tight time restrictions.

The network is designed via the definition of variables for the discrete arc decisions and the continuous flow decisions. The binary variable y_{ij} denotes if arc (i, j) is chosen as part of the network's design, where i and j denote nodes. The continuous variable f_{ij}^k denotes the flow of commodity k on arc (i, j), and restrictions on the flows that traverses an arc are denoted by the parameter K_{ij}. The network design formulation can then be written as (Magnanti & Wong, 1984):

$$min \quad \phi(f, y) \tag{2.1}$$

$$\text{s.t.} \quad \sum_{j\in\{N\}}\left(f_{ij}^k\right) - \sum_{i\in\{N\}}\left(f_{ij}^k\right) = \begin{cases} R_k & \forall i = O(k) \\ -R_k & \forall j = D(k) \\ 0 & \forall (i,j) | i \neq O(k), j \neq D(k) \end{cases} \quad \forall k \in \{K\} \tag{2.2}$$

$$\sum_{k\in\{K\}} f_{ij}^k \leq K_{ij} y_{ij} \quad \forall (i,j) \in \{A\} \tag{2.3}$$

$$(f, y) \in S \qquad (2.4)$$

$$f_{ij}^{k} \geq 0, y_{ij} \in \{0,1\} \qquad \forall (i,j) \in \{A\}, k \in \{K\} \qquad (2.5)$$

The objective function $\phi(f, y)$ represents the cost that is to be minimised in the network, like unit routing cost (related to f_{ij}^{k}) and fixed cost to set-up the network (related to y_{ij}). Constraints (2.2) denote the flow conservation constraints that provoke flows to originate or destine at locations other than its origin or destination. Constraints (2.3) regard the available arc capacity (which might be set to a non-restricting size for uncapacitated problems). The Constraints (2.4) denotes a class of additional restrictions (S) that can be added to the general network design formulation, for example to limit the number of arcs chosen.

Strategic network design

The generic formulation of Magnanti & Wong (1984) represents what we classified as tactical network design decisions. The strategic network design decisions can be formulated by the introduction of additional constraints that associate node design variables with the arcs that are directed into/out of that node. This problem was introduced in literature by O'Kelly (1986) and a first quadratic integer problem formulation was presented by O'Kelly (1987). In the *p-hub* location problem, the number of hub locations is restricted, i.e. p hub locations are to be chosen in the network. In his later work, O'Kelly (1992) introduced fixed cost to the hub location problem removing the necessity to select a predetermined number of hubs. This type of optimisation is also known as the *fixed charge* hub location problem, where the decision on the number of hub locations to use is part of the optimisation problem. The most common objective in strategic network design problems is cost minimisation, though other variants have been proposed as well. These concern minimisation of the largest transport time (Kara & Tansel, 2001), minimisation of the number of hubs ((Kara & Tansel, 2003), (Tan, et al., 2007), (Yaman, et al., 2007), (Alumur & Kara, 2008)), or maximisation of the total freight to be delivered to customers within a certain time bound (Yaman, et al., 2008). The strategic network design problem is also known as the *hub location problem*.

Service network design

A special class of network design formulations targets the planning of schedules and support decisions related to if and when fleets depart. This class is known as *dynamic service network design* (Crainic, 2000). To reflect scheduling operations, a time dimension must be introduced into the formulation, which can be achieved by the formulation of a time-spaced network in which each node in the network not only represents a facility but also a time at which the facility is visited. Such dynamic networks are presented in (Farvolden & Powell, 1994), (Pedersen, et al., 2009), (Root & Cohn, 2008). The resulting graph is similar to the network design formulation above, but is significantly larger as a result of these time dimensions. In practice, these networks are either huge in their size or the single time periods represent big time segments. Wieberneit (2008) concludes, based on a review of different service network design approaches, that one first should analyse if possible simplifications of the problem are possible, or even better, to avoid time space networks. And in case that time space networks cannot be omitted, the advice is to use a solution approach that uses reduction techniques.

In the next section we discuss some typical design variants that are used in the operation of express networks.

2.2.2 Fundamental assumptions in network design approaches

The strategic and tactical network design problem in general has been studied in literature for a while. Several fundamental assumptions on the nodes, arcs, cost and service definitions have been proposed. In this section we provide an overview of these assumptions. An illustrative overview of the classical and extended network design typologies is provided in Figure 7.

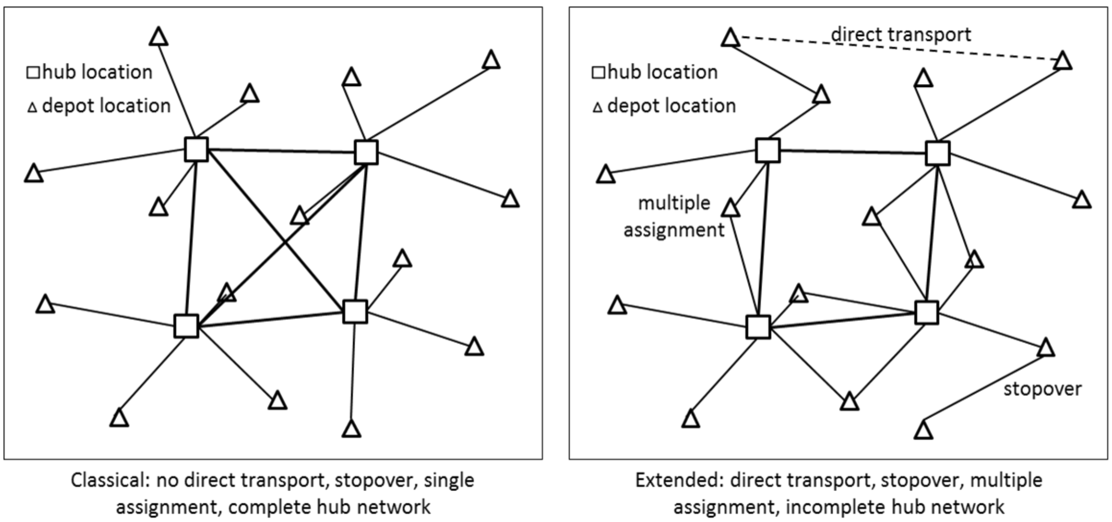


Figure 7: The classical hub location problem and the extended hub location problem with direct transport, stopover, multiple assignment and an incomplete hub network (Essen, et al., 2014).

Assumptions on the nodes

The assumptions on the nodes that we discuss below concern the hub location nodes. Note that assumptions on other nodes (e.g. depots and airports) are irrelevant for the network design scope covered in this dissertation.

Regarding the nodes in the network, the main assumption concerns restrictions on the amount of flow that can be handled by each type of node. If there is no limitation on the hub capacities, the hubs are said to be *uncapacitated* while the situation in which the total throughput at a hub is restricted is known as the *capacitated* variant. Additionally, *single hub* networks and *multiple hub* networks are distinguished as the number of hubs determines a large part of the complexity and thus the possible solution strategies.

Assumptions on the arcs

There are several assumptions found in literature that specify the arcs that can be used during network design. These can best be classified based on the function of the nodes, i.e. either being an origin/destination of a flow or a hub node at which flows are consolidated. We can then distinguish three types of nodes for which particular assumptions hold.

Firstly, we specify arcs between origin/destination nodes. Note that when flows of an ods-pair traverse directly from the particular origin node to the particular destination node, this is known as a *direct path/route*; an alternative way to use a direct arc is to combine the transport of flows of several origins to a hub, or vice versa, to serve a multiple of destinations via a certain hub node. In this alternative way to use a node to node arc, the arc is said to be used in a *stopover path/route*.

The second type of arcs concerns the arcs from origin/destination nodes to hub nodes. When the final network design is restricted to select only one arc from an origin/destination node to a hub node, it is known as a *single allocation* network. The alternative in which *multiple allocations* are allowed was introduced by Campbell (1994) and a variant on this concerns the *r-allocation* networks, in which at most r hub locations can serve an origin/destination node, (Yaman, 2011).

The last set of arcs relates to the arcs between the hub nodes. When it is assumed that all hubs are connected, the network is known as a *complete* network. If there is only partially connectivity between the hubs, the hub network design is referred as an *incomplete* hub network.

Additionally, arcs are sometimes restricted in the amount of flow that it can serve and these types of problems are referred as *capacitated* compared to the unrestricted variant which is referred as the *uncapacitated* problem. Unfortunately, the term *uncapacitated network design* is used for restricted arc problems as well as restricted node problems.

Assumptions related to the cost function

There are several levels of detail considered in the cost function that is used in network design approaches. The most accurate cost function specifies the real cost to operate a hub location and the transportation cost in the network.

Hub location cost concern fixed cost to operate a hub location at a given site, and variable cost relate to the flows of packages that are handled in the hub. In the tactical network design problem, the hub locations are fixed, so that fixed hub cost can be left out of scope in the design phase.

Transportation cost relates to kilometres driven/miles flown, vehicle or aircraft cost, and man-hour cost of the drivers or crew. There is an additional level of complexity in the calculation of transportation cost, as both vehicles/aircrafts and drivers/crew have to return to their origin base at regular times; additionally, legislation poses restrictions on drivers and crew working hours. The most accurate representation of transportation cost would result when tours or flights are created within the restrictions posed by legislation. In practice however, we generally see a simplified cost function.

The first network design formulation as discussed in (O'Kelly, 1987) used a unit cost to traverse flows over an arc, and applies an *economies of scale discount* (denoted by $\propto$, where $0 < \propto < 1$) to inter-hub arcs as larger flows are expected during inter-hub transport due to consolidation at the hubs. O'Kelly & Bryan (1998) stated that the inclusion of an exogenously determined discount applied to all inter-hub arcs regardless of the differences in the flows travelling across them, oversimplifies the problem. The authors claim that the cost has to be presented by a non-linear function such that marginal travel cost decreases as flows increase. The non-linear cost function is afterwards approximated by a piece-wise linear cost function. Other researchers noted that the use of a (discounted) linear cost function based on unit cost particularly underestimates cost on arcs that traverse low volumes of packages. Typically, these low package flows are in the express practice observed between origin/destination nodes and hubs and on inter-hub arcs over longer distances. One of the approaches proposed in literature (e.g. (Campbell, 1994)) regarding the cost reflection between origin/destination nodes and hubs is the use of flow thresholds and fixed cost on the transports between origin/destination nodes and hubs. Other researchers, like Podnar et al. (2002), questioned the use of discounting of only inter-hub arcs, and proposed to use a discounting to all arcs traversing flows larger than a certain threshold. A last approach found in literature is proposed by Zäpfel & Wasner (2002) and Wasner & Günter (2004), who suggest to use a cost calculation based on the individual transports, depending on the distances and the number of required vehicles.

Assumptions related to service definitions

The extent to which service definitions are covered during network design differs in the approaches found in literature. As Wieberneit (2008) suggested, simplifications or reduction techniques are desirable to work with networks of realistic sizes. There are two approaches that are known from literature. In the first approach, timing aspects are relaxed by making use of a *service coverage* restriction, which specifies the total time that is available to transport packages from an origin node to a destination node. This simplifies the network structure, but lacks detailed scheduling information. The other approach found in literature is the use of a fixed *hub window*, particularly used in single hub network design. In that situation, all package flows of an origin node need to arrive at the hub location before the start of the hub window, the sorting occurs during the hub window, and the delivery to the destination nodes occurs afterwards. The advantage of this method is that it reduces the complexity so that scheduling information can be retrieved. The drawback is that tight connections define the hub windows thereby diminishing the timing advantage that less tight connections may use; additionally, this method is mainly useful in single hub network design as another level of complexity arises when connectivity between hubs needs to be guaranteed as well.

In this section we provided a general overview of strategic and tactical network design problems for express service providers. Additionally, details on general assumptions taken are provided in this section of the dissertation.

2.2.3 A note on the computational complexity of the network design problem

The strategic network design problem that we consider in this dissertation is also known as the uncapacitated hub location problem and is proven to be NP-hard. Mirchandani & Francis (1990) proved this by showing that the uncapacitated facility location problem, which is a known NP-hard problem, is a special variant of the uncapacitated hub location problem. The facility location problem has two kinds of decision variables, the first being the variable that indicates whether a facility is opened and the second assigns each customer to an opened facility. Cost are incurred for each facility that is used as well as for the assignment of customers to facilities. For details on the facility location problem we refer to (Dashkin, 1995), (Eiselt & Marianov, 2013). Mirchandani & Francis (1990) have shown that this problem is a special variant of the uncapacitated hub location problem when customers are considered as depots, facilities are considered as hubs and cost for interhub transport is zero.

The tactical network design problems that we consider in this dissertation for road and air network design are also known as capacitated network design problems, and these type of problems are classified as strongly NP-hard. We proof this as follow. Consider the problem known as the network loading problem, a special variant of the capacitated network design problem, in which the variable flow cost is zero and facilities of fixed capacity are available to carry the flow. In our situation, this would mean that cost only relate to the movements operated in road networks or the flightlegs operated in air networks; any other cost like hub handling cost or fixed cost of aircrafts is equal to zero. These (load) facilities (i.e. the movements or flightlegs) need to be installed on the arcs in the network. A special variant of this is the two facility design problem (TFDP) in which only two (load) facilities are available: a low capacity and a high capacity variant. Mirchandani (1989) showed that the TFDP is strongly NP-hard. As the TFDP is a special variant of the general network loading problem and as such a special variant of the capacitated network design problem that we consider in this dissertation, the latter is proven to be strongly NP-hard as well.

As we have shown that the problems considered in this dissertation are (strongly) NP-hard, heuristics will be used in order to solve real-life problem situations. This will be discussed in more detail when defining the scientific model in the chapters that follow.

2.2.4 Conclusion: focus areas for research on the network design problem in express networks

In this section we reviewed existing literature on network design. Based on earlier research, we define three main focus areas that are important success factors for implementation of the outcomes of our research in the practice of express service providers:

- **Inclusion of express network typologies:** network design solutions should be able to deal with all specific network typologies operated in express networks, being: direct transport, stopover, multiple assignment, and an incomplete hub network;

- **Appropriate cost reflection:** the solution should reflect cost to operate express networks in an appropriate way;

- **Incorporation of scheduling aspects:** the solution method needs to support the service network design problem by providing timetables for fleets operated in express networks.

There are a few remarks that can be made regarding these focus areas in relation to the research objectives that are stated in this dissertation.

Firstly, note that the value of each success factor needs to be considered in relation to the period of decisions made. As discussed earlier, the creation of time schedules is way too extensive for the strategic nature of decisions on hub locations. As long as the cost to generate time schedules in a later phase is reflected well enough, strategic decisions can be made.

Secondly, it can be remarked that the cost reflection in tactical network design basically results from fleet scheduling operations so that the level of inclusion of express network typologies and the level of consideration of scheduling aspects are the key aspects of tactical network design.

Thirdly, we would like to remark that the service network design problem has been examined in air network design research in a number of studies. The similar problem in road network design has almost gone unremarked. For that reason, we decided to devote two research studies to this topic. In our first research we consider the tactical road network design problem with two simplifications: we consider all express network typologies except stopovers; only a single vehicle type is available for transportation. In the second research we release both simplifications. Additionally, we learned from our first research study that hub operations need to be incorporated to some level and studied this requirement in our second research on tactical road network design.

2.3 The last stage in conceptualisation of our research: the definition of research questions

In this dissertation we aim to develop OR models that support a cost efficient operation of an express line-haul network with guaranteed service level agreements, and focus on the applicability of our work in the daily practice of one of the big four express service providers. Our work is focussed on the design of road and air networks, and for both, extensions of existing methods are proposed that better suit the operation of express service providers in general. Particularly, we focus on the inclusion of express network typologies, a better reflection of cost, and the creation of a timetable for vehicles and aircrafts. We are now ready to state the research questions that belong to each research objective.

Strategic road network design:

Research objective I: Develop a method that supports the decision on hubs to be operated in an express road network such that total cost is minimised under tight service requirements.

Research question Ia: How can cost that result from the operation of vehicles in the network and handling at hub locations be reflected in a strategic road network design solution?

Research question Ib: How can we include all different express network typologies simultaneously in a strategic road network design solution that reflects operating cost well?

Tactical road network design:

Research objective II: Develop a method that designs the set of movements and supports (refined) routing decisions to achieve service commitment at minimum cost in an express road network.

Research question II: How can tactical road network design solutions be enriched such that all packages are routed through the express road network at minimal cost in terms of hub handling and operated fleet?

Tactical air network design:

Research objective III: Develop a method that supports routing decisions and designs flight schedules at minimum cost for multiple hubs in express air networks.

Research question III: How to design a multiple hub air network solution method that supports detailed routing and flight scheduling decisions and gives guidance to hub operations in terms of hub sort windows and total package flow to be handled?

Research question III: How can air network design solutions make efficient use of pre-sorted ULDs in order to reduce hub handling time and hub sorting capacity?

The remainder of this dissertation is divided into three main sections. In the first section, Part I, we present the scientific model and solution for strategic and tactical road network design. The second section, Part II, presents the scientific model and solution for tactical air network design. In the last section of this dissertation we present our conclusions and reflect on our work, and point to directions for further research.

Part I - Road network design

3 Heuristics for the uncapacitated hub location and network design problem with a mixed vehicle fleet and regional differentiation

This chapter is based on the following journal paper:

Heuristics for the Uncapacitated Hub Location and Network Design Problem with a Mixed Vehicle Fleet and Regional Differentiation

Van Essen J; Meuffels W; Aardal K; Fleuren H
Submitted to a scientific journal

3.1 Abstract

The uncapacitated hub location and network design problem for express service providers has been addressed in several papers. Our research is motivated by two observations in hub network design at these providers: firstly, the classical approximation of the cost function in hub network design does not reflect real cost when deploying the designed hub network; second, classical assumptions on network typologies do not suit the design of express networks. We therefore extended the uncapacitated hub location and network design problem with regional optimisation as observed in express networks (i.e. the inclusion of direct transport, multiple allocation, stopover, and an incomplete hub network) in combination with an improved cost function based on a mixed vehicle fleet. Two heuristics are presented to produce good-quality solutions to these two problems, namely Genetic Algorithms and Simulated Annealing. The heuristics are tested on benchmark data from the Civil Aeronautics Board (CAB) and on modified instances from an express service provider. This shows that the extended model reduces cost by an improved selection of hub locations in certain regions.

3.2 Introduction

Express service providers offer services to transport packages (i.e. parcels, flows, or pieces of freight) from senders to receivers, with predefined delivery times. Typically, point-to-point package flows are small, so express service providers use several sorting centres to distribute flows over vehicles operated between the sorting facilities. Usually local sorting centres, called *depots*, are used to sort the flows after collection, or to distribute the flows to vehicles taking care of the delivery to customers. These local collection and delivery operations are referred to as the *pickup and delivery process*. Long-distance transport between the depots occurs via larger sorting centres, called hubs, during the *network process*. The scope of this chapter concerns the network process.

The hub location and network design problem

The problem of locating hubs in a network is called the hub location and network design problem, which was introduced by O'Kelly (1986). The problem can be formulated in a graph, in which notes denote depots and hubs, and arcs denote the transportations in the network. Decisions are to be taken concerning the selection of hubs and arcs, to organise the transport of packages from *origin depot*, the depot that collects the packages at the sender, to *destination depot*, the depot that organises the delivery to the receiver. Classical models on the design of these types of networks consider the following four main assumptions, (Nickel, et al., 2001), (Gelareh & Nickel, 2011):

- The hub-level network is a complete graph.
- Using inter-hub arcs has a lower price per unit than using depot-hub arcs; that is, inter-hub arcs benefit a discount, $0 < \alpha < 1$.
- Direct arcs between depots are not allowed.
- The triangle inequality holds in the cost structure and cost is proportional to the distance.

Additionally, we observe that most classical models assume *single* assignment from depot to hub locations, instead of connecting depots to more than one hub (*multiple* assignment). Moreover, the classical problem formulation has no restrictions on the capacities of the hub locations, and is referred to as the *uncapacitated* hub location problem. Note that some researchers unfortunately also used the term uncapacitated to denote that the amount of flows that traverse via arcs is unlimited. We now present the observations that motivated our research objective.

Research objective: express network typologies and refined cost function

The express company that inspired our research offers services covering a country, and observes diffused flow patterns, where larger flows of packages originate in industrial areas and smaller flows of packages in more rural areas. Hence, network cost to tranship all packages in a country is strongly related to the available vehicles and their capacities related to the sizes of the package flows. At some depots, package flows are large enough to assign multiple vehicles to transport packages to hubs, and these vehicles may even be routed to different hubs. While at other depots, even single vehicles show underutilisation from depot to hub and either combined transport from/to depots might be used to reduce cost or smaller vehicles are operated. Similar reasoning holds for the transport between hubs: for some hub pairs enough packages can be consolidated to arrange direct transportation from hub to hub, but for other pairs of hubs, further consolidation at an intermediate hub location strongly reduces cost.

We noticed that as a result of typical packages flows in express networks, the network typologies operated by these providers differ in design compared to classical network designs by allowing:

- *direct transport* between depots, i.e. packages that are transported from origin depot to destination depot without sorting and/or consolidation at a hub;
- *multiple assignment* from a depot to hubs, i.e. depots that receive/deliver their packages from/to more than one hub;
- *stopover* transports, i.e. depot-depot transports that are not used for direct delivery of packages but for further consolidation from/to a hub; and
- an *incomplete hub network,* i.e. a network in which not all hubs are connected and in which some hubs further consolidate transports at intermediate hub locations.

These network typologies are illustrated in Figure 8. Also, as a result of the package flows and corresponding network typologies that are operated, the general assumption of a discounted cost function does not reflect operating cost in an express network, which results from the dispatching of a heterogeneous fleet of vehicles. Our research objective is therefore to solve the uncapacitated hub location and network design problem with the inclusion of express network typologies and a good reflection of operating cost.

Chapter outline

In Section 3.3, we elaborate on the extensions that we apply on the classical hub location and network design problem and make a reference to the existing literature. Thereafter, we present the problem formulation and solution methods in Sections 3.4 and 3.5 respectively. Afterwards, we apply our solution methods on various datasets in Section 3.6 and conclude in Section 3.7 with a summary of our research and directions for further investigations.

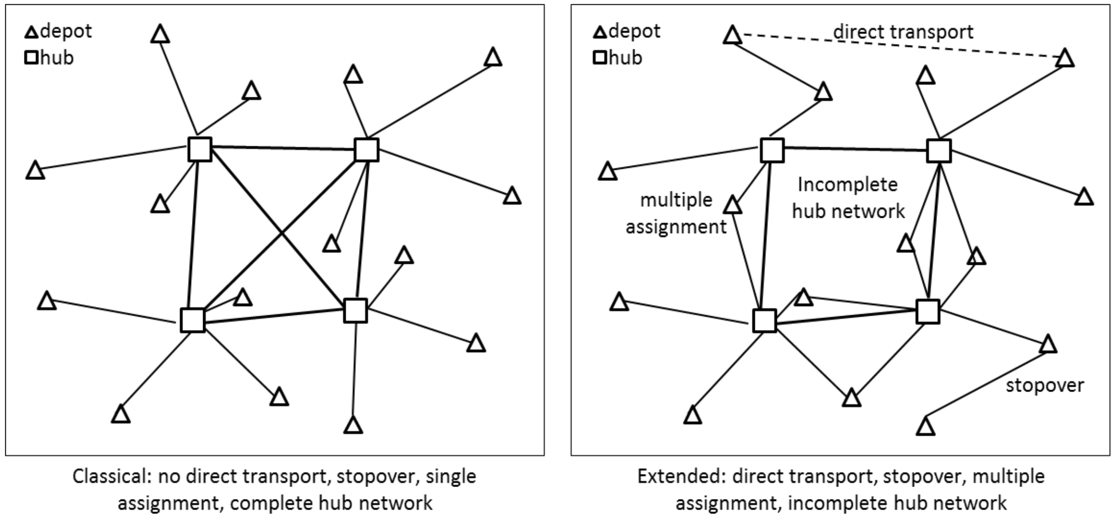


Figure 8: The classical hub location problem and the extended hub location problem with direct transport, stopover, multiple assignment and an incomplete hub network.

3.3 Literature review

For the origin of the hub location and network design problem we refer to (Campbell & O'Kelly, 2012) and recent overviews on the problem are provided in (Alumur & Kara, 2008), and (Farahani, et al., 2013). Below, we discuss related literature regarding the cost functions used in hub network designs and we review literature that considered some of the express network typologies.

The cost function

The refinement of the cost function releases the assumption of discounted economies of scales on inter-hub arcs. A few researchers have relaxed this assumption, like O'Kelly & Bryan (1998), who claimed that the inclusion of the discount factor oversimplified the real transportation cost, and suggested to use a piece-wise linear cost function. Other approaches to improve cost calculations on the transport from depots to hubs concern the proposals of fixed thresholds and fixed cost on these transports ((Campbell, 1994), (Podnar, et al., 2002)). More refined cost functions are used in the work of Zäpfel & Wasner (2002) and Wasner & Günter (2004). In these papers, transport cost result from the cost of individual transport, which depend on the distances and the number of required vehicles. The research of Zäpfel & Wasner (2002) and Wasner & Günter (2004) has a strong focus on locating depots and a small number of hubs, in which even business knowledge is applied to reduce the problem size. Meuffels et al. (2010) has used a similar cost function, but only uses this for the design of the network and not for the hub location decisions.

Extended network typologies

Firstly note that some researchers have included direct transports, like the work of (Aykin, 1994). Typically, restrictions are set on the allowed direct transports. (Aykin, 1994) for example specifies a set of depot pairs for which direct transport is allowed.

Multiple allocations from depots to hubs are more seen in hub location problems, but are also generally accompanied by restrictions on the depots that commit for multiple assignments. This extension of the classical problem was first introduced in literature by Campbell (1994), and several researchers have worked on methods to solve the problem. An overview on the solution approaches for the problem introduced by Campbell (1994) is presented in (Alumur & Kara, 2008). An example of restrictions that are used when designing multiple assignment problems to reduce network complexity and fixed cost, is the suggestion of Campbell (1994) to apply flow thresholds and fixed cost for connecting depots to hubs. Later on, Yaman (2011) designed a model with r-allocations from depots to hubs in which each depot can be served by a maximum of r hub locations.

Stopover transport is found in most literature regarding the hub location and particularly the network design problem of air transport (e.g (Kuby & Gray, 1993), (Armacost, et al., 2002)). In road transport, this is a less-seen approach, though some researchers have included these ways of transport, like Yaman et al. (2007) who applied this on the latest arrival hub location problem in a model that focusses on the throughput time in an express network (not on cost).

Lastly to discuss is the relaxed assumption of a complete hub network. Little research has been conducted on the hub location problem with incomplete hub networks. One of the few examples is given in (O'Kelly & Miller, 1994), who presents different hub typologies including typologies with full or partial connectivity between hubs. Campbell & Krishnamoorthy (2005) and Campbell et al. (2005) introduced the hub arc location problem, modelling incomplete hub networks in which the established inter-hub connections is to be decided and a discounting of unit flows is only allowed for a limited number of these connections. Yoon & Current (2008) relaxes the assumption of a complete hub network by formulation of the problem with fixed and variable arc cost, where the fixed cost represent the cost to provide a service and the variable cost are incurred with flows that use the transport connection.

Research contribution

Note that although some of the introduced network typologies for express networks have been considered in earlier research, we have not found any research that combined all these express network typologies in one design. In this chapter, we design the uncapacitated hub location and network design problem with extended network typologies in one design. The cost function that we use is in line with the approach used in (Zäpfel & Wasner, 2002) and (Wasner & Günter, 2004). By using this cost function, we do not need to take any further restrictions on the considered network typologies. In the next section we formulate the problem, and afterwards our solution method is presented.

3.4 Problem formulation

The uncapacitated hub location and network design problem can be formulated as follows: let L be a set of locations, consisting of a subset of depots D and a subset of (potential) hub locations H. Recall that customer flow collection and delivery occurs at depots, while hubs only consolidate and tranship depot flows, so that it is possible that a node location is both in the subsets of depots and hubs. Moreover, the flows to be transported between depot $d_1 \in D$ and depot $d_2 \in D$ are given and denoted as $w_{d_1 d_2}$.

To transport these flows, a fleet F of different vehicle types is available, and we assume a sufficient amount of each vehicle type f. Each vehicle has a corresponding load capacity q_f and cost per unit distance c_f^{var}. The extension of the cost function now results by multiplication of this variable cost by the number of vehicles that are deployed between a pair of locations, and the variable $y_{l_1 l_2 f} \in \mathbb{N}$ is introduced to denote this number of vehicles.

Regarding the basic network typologies, i.e. the typologies in the classical formulation, the following decisions are made:

- The decision which hubs to use, by introduction of the binary variable $\bar{g}_h \in \{0,1\}$.
- The assignment of depots to these hubs, via variable $g_{dh} \in \{0,1\}$; the single assignment restriction is forced by a constraint that takes the sum of this variable for all hubs and sets it equal to one.
- The decision on the path of the flows from depots via hubs, by using the variable $x_{d_1 d_2 h_1 h_2} \in [0,1]$; $d_1, d_2 \in D$ and $h_1, h_2 \in H$.

The extension of the classical formulation can then be achieved in the following way:

- The extension of direct connections between depots results from the introduction of the decision variable $x_{d_1 d_2}^D$, denoting the fraction of package flows that follows a direct route from depot d_1 to depot d_2.
- Multiple assignment from depots to hubs is arranged by releasing the binding restriction of the variable g_{dh}.
- In order to include stopover transports and an incomplete network with three-hub routes, the routing variable $x_{d_1 d_2 h_1 h_2}$ is extended, That is, we introduce a variable $\dot{x}_{d_1 d_2 d_3 d_4 h_1 h_2 h_3}$ ($\in [0,1]$; $\forall d_1, d_2, d_3, d_4 \in D$ and $h_1, h_2, h_3 \in H$), that denotes a route from depot d_1 making a stopover at depot d_2 on the way to hub h_1; it then visits hubs h_2 and h_3 on the way to destination d_4 via a stopover at depot d_3. Additionally, the variable $\dot{g}_{d_1 d_2}$ is used to denote that $d_1 \in D$ and depot $d_2 \in D$ are served in the same stopover route.

In Appendix A we present the basic and extended model; note that the basic model formulation can be derived from the extended model by setting $d_2 = d_1, d_3 = d_4$ and $h_2 = h_1$ or h_3 for the routing variable $\dot{x}_{d_1 d_2 d_3 d_4 h_1 h_2}$ and restricting the variable $x_{d_1 d_2}^D$ to attain a value of zero. Also, the forcing constraint for single assignment has to be added.

3.5 Solution methods

The computational complexity in the uncapacitated hub location and network design problem comes from the routing variable: for the smallest real real-world network that we present in Section 3.6.3, with 20 depots, 148 depot-depot pairs with a positive flow and 30 possible hub locations, the routing variable in the basic situation with two hub-routes results already in more than 100K possible routes that need to be evaluated. If we allow stopovers at one intermediate depot and three-hub routes, more than 1 billion routes are possible. Mirchandani & Francis (1990) have proven that the basic uncapacitated hub location problem is already NP-hard. The proof follows by showing that the facility location problem, which is known to be NP-hard, is a special case of the uncapacitated hub location problem in which only decisions are to be made for depot to hub transports and not for inter-hub transport. We therefore decided for a heuristic approaches to find good feasible solutions.

Furthermore, one of the observations we drew during earlier research on this topic, is the observation of the existence of a pool of hub configurations that perform very similar at cost level. That is, it is clear that if an express company operates too less hub locations, distances from depots to hubs are large, making underutilised transport from depot to hub expensive. On the other hand, too many hub locations reduce the sizes of flows between hubs and thus reduce utilisations on inter-hub transports. Though, between these configurations with too less or too many hubs, there exists a range of hub configurations that result in almost similar cost figures. This is especially seen in situations where hub configurations can be shifted clockwise, as illustrated in Figure 9. Note that the existence of a “bathtub curve” in location design problems is recognised in other research as well. For example, Simchi-Levi et al. (2003) concluded based on this concept that a supply chain can be prepared for disruptions by investing in more locations than the optimal number, and thus investing in redundancy, without significantly increasing supply chain cost. Based on this observation, we designed an approach in which we first derive a pool of solutions that perform similar on cost in a simplified network configuration. Afterwards, the best hub configuration for express networks is derived by evaluating the pool of solutions for detailed express network typologies.

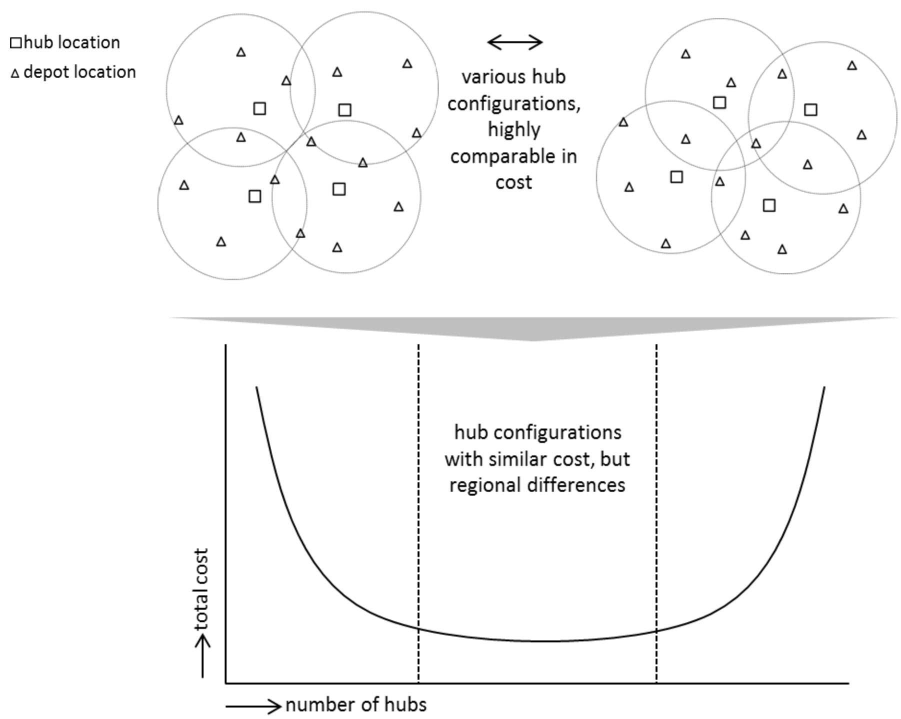


Figure 9: Observation of a “bathtub curve”: the hub location and network design of express service providers results in a range of solutions that are comparable in cost, and show regional differences only.

In summary, our solution method follows the following sequential approach:

i. Pre-processing phase: derive the fraction of flows that are routed directly.
ii. Global optimisation phase: determine a pool of solutions that provide the number and location of hubs, for which cost differences are small.
iii. Local optimisation phase: use the pool of global solutions to find the best solution based on local configurations with stopovers, multiple allocation and three hub routes.

Each phase is discussed in detail below.

3.5.1 Pre-processing phase

In the pre-processing phase we derive the fraction of flows that is routed directly, which occurs if at least a full vehicle of the largest size can be deployed from depot to depot. The reason to allow directs only when a full vehicle can be operated, is because direct connections are sensitive to flow fluctuations, as low utilisation is expensive. Note that a full direct vehicle is always cost efficient, as a hub route will cause additional handling cost and a detour in driving kilometres and thus cost. The remaining flows are to be routed via at least one hub location.

3.5.2 Global optimisation phase

In this step, the basic model formulation of the uncapacitated hub location and network design problem is solved with a heuristic approach. Particularly, we compare two solution methods, namely Genetic Algorithms (GA) and Simulated Annealing (SA).

Each method will derive the number and location of hubs, by providing a solution represented by a string of zeros and ones. In this, the string length represents the number of possible hub locations and each entry corresponds to a particular hub location; a value one then indicates that the hub is chosen in the solution while a value of zero indicates that the potential hub is not selected. Besides, cost of each solution is provided; this is the result of allocating each depot to its nearest established hub providing the flows from depots to hubs, between hubs, and from hubs to depots. The cost is then calculated by deriving the optimal number of vehicles for each particular flow between a pair of locations, and multiplying this with the unit transport cost and the distance travelled.

Both GA and SA are used to create a pool of solutions of size Z, which will be evaluated for regional potential in the next phase. Below we describe the implementation of each method in more detail.

3.5.2.1 Genetic Algorithms

GA can be described as a probabilistic search, which imitates the process of natural selection and evolution (see (Goldberg, 1989)). We implemented this as follows:

i. Creation of a(n initial) population – The first individual of the population is generated in a greedy way[3] and the remaining $Z - 1$ individuals are randomly generated[4].

[3] Firstly, the best 1-hub location is determined, by enumerating every possible hub location. This hub will be established and the best of the remaining hubs is added to the solution. If the cost of this 2-hub solution is lower than the cost of the

ii. Repeat M_1 times:
   a. Select two parents for reproduction by a binary tournament selection.
   b. A child solution is generated by first applying a single two-point crossover operator[5] to the selected parents. The crossover procedure is followed by a mutation procedure. Each bit in the offspring can be changed from zero to one, or vice versa, with probability P_1.
   c. For each generated offspring solution, the total cost is calculated. The weakest individual in the population, i.e., the individual with highest cost, is replaced by the child solution, if the child is fitter than this particular individual. The child is only added to the population, if an equal individual is not already present in the population. Therefore, each individual in the population is unique.

The pool of solutions equals the population after finalisation of M_1 evolutions.

3.5.2.2 Simulated Annealing

SA is a generic probabilistic meta-algorithm for locating a good approximation of the global optimum of the objective function in a large search space (see (Collins, 1988)). The implementation is as follows:

i. The initial solution is generated in a greedy way[3], similar as the implementation in the solution method of Genetic Algorithms.
ii. Set an initial temperature T and a reduction factor δ with $0 < \delta < 1$.
iii. Repeat M_2 times:
   a. Reset the value of T to its initial temperature.
   b. Repeat M_3 times:
   c. Select a neighbour solution S_0. A neighbour solution S_0 is a solution where one hub is added to and/ or removed from the current solution S.
   d. Let $\Delta = f(S') - f(S)$, where $f(S)$ denotes the objective value of S. If $\Delta \leq 0$, set $S \leftarrow S'$, else set $S \leftarrow S'$ with probability $e^{-\Delta/T}$.
   e. Set $T \leftarrow \delta T$.

The Z best solutions are stored, and during the whole process of SA it is validated if the current solution is better than one of the individuals present in the population. Again, a solution is only replaced if it is not yet present in the population.

1-hub solution, the 2-hub solution is accepted. This continues until the $n + 1$ hub solution has higher costs than the n-hub solution. The n-hub solution is the first individual in the population.

[4] Each individual is the solution with lowest costs out of Z' randomly generated solutions, where Z' is called the size of the subpopulation. The number of hubs to be opened in a randomly generated solution is equal to $0.25 * cardinality\ of\ set\ D$. The hubs to be opened in a generated solution are chosen one by one by selecting them uniformly from the hub locations that are not yet opened.

[5] Note that there can be numerous ways to apply crossover operations. We have chosen for a single two-point crossover operator as this suits the concept of combining regional best solutions.

3.5.3 Local optimisation phase

For the pool of solutions, local improvements are made based on the defined structures. This is done for each individual solution, and eventually, the single best solution is selected. Below, we discuss each local improvement in more detail.

3.5.3.1 Depot to hub allocation

In the global optimisation phase, each depot is allocated to its nearest hub. In this phase, we improve this assignment, firstly by applying a local search strategy to improve the single assignment and second by the allowance of multiple allocations from depots to hubs.

Improved single allocation

The single allocation strategy from depots to hub is improved by a local search algorithm that works as follows for an individual solution. For each depot, the cost of shifting it from its current hub to one of the other established hubs in the solution is calculated. Then, the depot with highest cost saving applies a shift to its most favourable hub location. This process is repeated until no shift moves with reduced cost can be found.

Multiple allocation

Afterwards, the multiple allocation from depots to hubs is evaluated. Based on the previous steps, the route from depot d_1 to depot d_2 is known. We allow multiple connections and hence a bypass of a hub, if there is enough flow to fill the largest vehicle from/to that hub and cost reduces. This is illustrated in the figure below. Note: the threshold for creation of multiple allocations, which can easily be relaxed, is set to prevent unexpected cost due to fluctuations in daily flows.

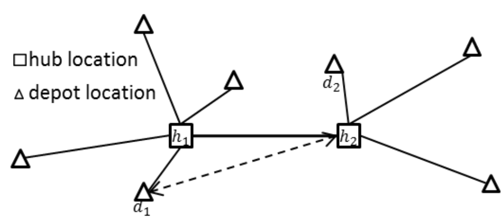


Figure 10: Multiple allocation from depot d_1 to hub h_2 is allowed, if this reduces cost and flows are large enough to create a full vehicle load.

As in the previous step, the depot that provides the highest cost reduction is reassigned first and this process is repeated until no further cost reductions are found.

3.5.3.2 Stopovers

Given the assignment of depots to hubs, stopovers can be determined. Stopovers are created for all depots in a region of a hub, i.e. $\{\forall d | g_{dh} = 1\}$. There are two practical assumptions regarding stopovers. Firstly, as stopovers are used because of low flows, it is assumed that either all or none of the depot-hub flow is transported via a stopover. Second, as the available time in express networks is limited, it is assumed that a stopover is only allowed to visit one intermediate depot location. The problem is solved for each hub location, with a MIP-formulation as presented in Appendix B.

3.5.3.3 Incomplete hub network

It can be more cost effective to allow more than two hubs on a route from depot to depot. Due to the restricted time limits in express networks, we assume that at most three hubs are used in a route. Based on the routes of the flows between depots, hub flows can be derived. We then formulate a mixed integer programming model to reduce the number of vehicles to transport flows between hubs; the corresponding formulation can be found in Appendix B.

3.6 Computational results: a case study

In this section, we compare the results of our two heuristics for the global optimisation and the impact of the local optimisations on these solutions. We implemented our heuristics in AIMMS 3.8 and ran them on an Intel Core2 CPU 6400 2.13 GHz with 2 GB RAM. The CPLEX 11.0 solver is used (a) to calculate the number of vehicles of each type needed to transport a particular flow, (b) to determine the stopovers, and (c) to determine the routes from hub to hub.

To study the quality of the solutions of the two heuristics, we first present three LP-relaxations in Subsection 3.6.1 Besides, our model is tested on the well-known dataset of the Civil Aeronautics Board (CAB) in Subsection 3.6.2 and additionally, we test the performance of the heuristics on three modified networks of a large international express service provider in Subsection 3.6.3. We proceed then by some sensitivity analysis on the individual extensions of the local optimisation in Subsection 3.6.4, and discuss the sensitivities of the model parameters in Subsection 3.6.5.

Note: in the next sections we use the term "basic" to denote that the model has run the pre-processing phase, the global optimisation phase and the local optimisation phase only for the improved depot to hub allocation; "extended" is used to refer to results of the model in which all optimisation phases have run.

3.6.1 LP-relaxations

In order to study the quality of the solutions of the heuristic selection of the hubs, it is important to have a good lower bound on the optimal solution. The initial LP-relaxation, referred to as LP-relaxation I, of the model formulation used in the global optimisation phase is weak; we therefore strengthen the constraint set by the addition of constraints that ensure that at least one vehicle has to be assigned to each connection with a positive flow of packages. Further tightening of the relaxation is obtained by estimating the minimum amount of vehicles required. LP-relaxation II refers to the relaxation with these additional constraints. Both relaxations are discussed in Appendix C.

Both LP-relaxations are tested on data of Network A, for which details can be found in Section 3.6.3. Results are determined for optimisation on the full data set of potential hubs, and in addition, we tested the LP-relaxations on a data set in which only the hubs selected by the heuristics for the basic model are included (HHS). The results are shown in Table 1 where the CPU time is given in seconds, when not stated otherwise. The IP Solution HHS denotes the optimal integer solution for the problem for the constrained hub set, and is used as reference data.

No feasible solutions for LP-relaxation II were found when the full set of hubs was used. When we only consider the hubs selected by the heuristics, we see, in the third column of Table 1 (Heuristic Hub Set (HHS) – Best value), that the lower bound of the LP-relaxation is strengthened by the additional constraints. The gap between the values of the LP-relaxation and the optimal solution, given in the fifth column, has decreased from 28% to 7%.

	Complete hub set		**Heuristic Hub Set (HHS)**		**IP Solution HHS**	
	Best value	CPU time (s)	Best value	CPU Time (s)	Best value	CPU Time (s)
LP-relax. I	65	3,549	72	1.09	100	290
LP-relax. II	-	>64 hours	93	2.25	100	87

Explanation of the headers:

"Complete hub set": LP-solution for all potential hubs; "Heuristic Hub Set (HHS): LP-solution for subset of hubs, subset (HHS) follows from solution pool GA/SA; "IP Solution HHS": IP-solution to HHS.

Best value": value of the lowest cost solution by model formulation LP-relaxation. I, LP –relaxation II; "CPU time": runtime of the model.

Table 1: Results of the LP-relaxation for Phase II of the optimisation.

The optimal integer solution found for the subset of hubs is the same as the value found by the heuristics after local improvement of the single assignment from depots to hubs. Therefore, it seems that the heuristics with this local improvement finds the optimal network design with single allocation. We refer to this formulation as the basic problem, which comes close to the classical problem formulation. Further on, the heuristic solutions are compared to LP-relaxation II where the subset of hubs is used, because this LP-relaxation gives the strongest lower bound, while the IP Solution cannot be found in reasonable time for larger problem sizes.

3.6.2 CAB-data

In this section, we compare the results of the presented solution methods to the well-known Civil Aeronautics Board (CAB) dataset, which is a dataset introduced by O'Kelly (1987) and regularly used to test research on the p-hub location problem. It is based on airline passenger flow between 25 US cities in 1970 and consists of 25 depots and 25 possible hub locations; subproblems of size $|L| = 10, 15, 20$ are derived based on the main set of nodes. Although the p-hub location problem differs from the problem considered in this research due to the restriction of opening exactly p hubs, the problem can be translated to the uncapacitated hub location problem considered in this work as follow.

In order to transform the p-hub location problem to an uncapacitated hub location problem, we first need to analyse outcomes of the p-hub location for various values of p. Best known solutions for the p-hub location problem are derived in a study by Skorin-Kapov et al. (1996), for a number of locations $p = 1, 2, 3$ and 4; the lowest cost hub configuration with either 1, 2, 3, or 4 hubs is referred to as the best known solution for comparison. Besides, as the problem presented by O'Kelly (1987) uses unit transport cost instead of vehicles types, we will assume in our modelling approach that a single vehicle type is available for which the capacity and cost per unit distance equal 1 while setting the distance equal to the unit cost.

We run the global and local phases of our model approach, however during local optimisation we only allow the improved single allocation to run. All other local strategies are irrelevant, as the unit-flow vehicle has no economies of scale and thus results in a high number of connections to obtain shortest paths. Table 2 presents the results.

	Best known solutions		Genetic Algorithm			Simulated Annealing		
L	**Best solution**	**Hub conf.**	**Best value**	**%dev**	**CPU time (s)**	**Best value**	**%dev**	**CPU time (s)**
10	931	4	931	0.00	3	931	0.00	6
15	1,307	4	1,307	0.00	8	1,307	0.00	17
20	1,210	4, 20	1,210	0.00	17	1,210	0.00	35
25	1,257	4, 8, 20	1,257	0.00	31	1,257	0.00	68

Explanation of the headers:

"L" : number of nodes in scope ($L \in \{10,15,20,25\}$);"Best solution": cost of the (known) optimal hub location; "Hub conf.": hub configuration that provides lowest cost.

Genetic algorithm/Simulated annealing: "Best value": value of the lowest cost solution based on either GA or SA; "%dev": percentage deviation from the optimal solution; "CPU time": runtime of the model.

Table 2: Results of the models applied to the well-known CAB-dataset.

From the results it can be concluded that both implemented heuristics find the optimal solutions on the CAB dataset. Note that Genetic Algorithms is slightly faster than Simulated Annealing regarding CPU time on these instances.

3.6.3 Real world data

In this section we present the results of our model for the three networks as described in Table 3. In this, it should be noted that the results are indexed, to assure confidentiality. Also vehicle capacities are relative to each other as presented in the table below. For example, when we have two vehicles with capacity 5,000 and 10,000 respectively, we denote the capacity of the largest vehicle by 100 and the capacity of the smallest vehicle by 50. The data of these networks is given in Table 3.

We compare the results of LP-relaxation II with the results of the Genetic Algorithm and the Simulated Annealing approach (see Table 4). As in the previous paragraph, we run the heuristics with the global optimisation and the local optimisation regarding the improved single allocation; this is referred as the basic model approach. Besides, we run the approach that includes all local features for the heuristics and refer to this as the extended model approach.

	Network A	Network B	Network C
Nr. depots	20	54	80
Nr. potential hubs	30	16	16
Nr. depot-depot pairs	148	1,874	5,974
Nr. vehicles types	3	1	2
Vehicle capacities	18, 50, 100	100	58, 100
Vehicle cost	67, 83, 100	100	67, 100

Table 3: Overview of the test cases - three different network configurations are available for testing.

Note: in order to guarantee confidentiality of the data, only indexed results are shown, in which the result of the LP-relaxation II is set to 100; in Network C however, no solution is found for the LP-relaxation, so that the basic model solution (which is similar for both heuristics) is considered as a base. Moreover, as the computation time for multiple hub routes increases, the solver is interrupted if the relative difference of the objective function (i.e. the gap between the bound of the LP-relaxation II and the best integer solution found) is within 2%.

Network		LP-relaxation II	GA-Basic	GA-extended	SA-basic	SA-extended
A	Best value	100	108	90	109	90
	CPU time (s)	2	26	79	57	127
	Hub conf.	A1, A2, A3, A4	A1, A2, A3, A4	A1, A3, A5	A1, A2, A3, A4	A1, A3, A5
B	Best value	100	105	90	105	91
	CPU time (s)	5,544	74	37,805	165	43,877
	Hub conf.	B1, B2, B3, B4, B5, B6	B1, B2, B3, B4, B5, B6	B1, B2, B4, B5, B6, B7, B8, B9	B1, B2, B3, B4, B5, B6	B1, B2, B4, B5, B6, B7, B8, B10
C	Best value	-	100	92	100	91
	CPU time (s)	>68 hours	151	14,748	378	26,540
	Hub conf.	-	C1, C2, C3, C4, C5, C6, C7	C1, C2, C3, C4, C5, C6, C7, C8, C9	C1, C2, C3, C4, C5, C6, C7	C1, C2, C3, C4, C5, C6, C7, C8, C10

Explanation of the headers:

"LP-Relaxation II": LP-solution to the problem achieved with LP formulation LP-Relaxation II; "GA-basic"/"SA-basic": solution found with GA or SA after pre-processing, global optimisation, and depot to hub allocation; "GA-extended"/"SA-extended": solution found with GA or SA after pre-processing, global optimisation, local optimisation.

Best value": value of the lowest cost solution based on either GA or SA; "CPU time (s)": runtime of the model in seconds; "Hub conf.": selected hub set.

Table 4: Results of the models applied to modified data of an express service provider.

Firstly note that the gap between the cost of LP-relaxation II and the basic model is within a range of 10% for both Networks A and B. This is a relatively small gap, which strengthens the belief that the heuristics perform well. For Network C, the computation time of LP-relaxation II was longer than 68 hours. During this time, no feasible solution was found, and therefore, no lower bound is obtained. GA and SA give the same solution for the basic model of all three networks, but the computation time of SA is longer than the computation time of GA.

For the extended model, the solutions of GA and SA differ, although still being close. In general, cost decrease significantly by the inclusion of the extended features. In Network A, we see that the final hub configuration has one hub less. Detailed analyses showed that stopovers had reduced the need for an additional hub. In Network B, the extended model has added small supportive hubs in two regions to consolidate regional flows. These supportive hubs are connected to a few main hubs of the network, a clear example of an incomplete hub network. The difference between GA and SA is a region in which two hubs make only little difference in total cost figures. In Network C, one of the main hubs of the basic solution located in an outer region of the basic solution becomes a supportive hub, connecting to two new main hubs that are closer to the country centre.

3.6.4 Sensitivities of the individual extensions

To explain the observed results on the real-world data, we tested each extension individually. Regarding the local search procedure, we see that the improved single allocation strategy contributes to a cost reduction between 1 and 2% in each network. Besides, it is observed that the final hub configuration is influenced by this strategy. On the other hand, we observed that multiple allocations have lower impact: although cost reduces slightly, hub configurations remain unchanged. Similarly, we see that direct transports have low impact on the final configuration chosen. Stopovers and the inclusion of three hubs in a route do have huge impact. In regions with a lot of small flows we observe that a supportive hub might be introduced, that connects to a few main hubs. On the other hand, if a few depots have low flows, these are combined in stopovers, which result in shifted hubs (in general more to the centre of a country). Both options are therefore very useful to decide on the location of hubs in certain regions.

3.6.5 Sensitivities of the parameter settings

In this section we discuss the sensitivities of the parameters for both heuristics. The parameters that are varied are:

Genetic Algorithm	Simulated Annealing
$Z = 30$	$T = 0.25f(S)$
$Z' = 5$	$\delta = 0.95$
$P_1 = 0.01$	$M_2 = 2.5 * cardinality\ of\ set\ H$
$M_1 = 200 * cardinality\ of\ set\ H$	$M_3 = 200$

Table 5: Used parameter settings of the heuristics.

Recall that Z is the size of the population and Z' the size of the subpopulation used to create an initial population. Besides, P_1 denotes the mutation probability and M_2 the number of iterations that the Genetic Algorithm will do. For Simulated Annealing, T is the initial temperature, which is based on the objective function value $f(S)$ of the initial solution S. More, δ denotes the reduction factor, M_2 the number of iterations and M_3 the number of subiterations.

Below, we discuss the performance of the heuristics when executing the GA heuristics 50 times with a different seed for each of the individually changed parameters. For each test, we determine the percentage of runs that obtain the lowest cost solution. This percentage is denoted as "%Best value" in the table below. Note: the performance is tested for Network A with the characteristics as described in Table 3.

The sensitivity analysis shows that the population size may result in shifts from the optimal solution; when this parameter is set too low, the population might converge to a local optimum. If the population is too large, the number of iterations M_1 needs to be increased to obtain convergence to the optimum value. The subpopulation size, required to generate the initial population, has low influence: only if this value is set to a very low value of one, some runs do not show the optimal value. The mutation probability has high impact on the performance of the heuristic. If this value is set too low, the evolution of the solution is not finished after performing the number of iterations set. On the other hand, if mutation is stimulated too much, convergence to a local optimum might occur more often. Lastly, we tested the number of iterations. This is a parameter that has a strong impact if it is set too low, because some time is needed to evolve hub configurations. However, at some point further iterations do not add to any further improvements.

Sensitivity of the parameter settings of the Genetic Algorithms implementation

Changed values	%Best value	CPU Time
Base (Table 5)	100	32
$Z = 20$	96	30
$Z = 40$	90	36
$Z' = 1$	98	31
$Z' = 10$	100	33
$P_1 = 0.001$	27	32
$P_1 = 0.1$	94	35
$M_1 = 100 * cardinality\ of\ set\ H$	76	20
$M_1 = 300 * cardinality\ of\ set\ H$	100	44

Table 6: Sensitivities of the parameters for the Genetic Algorithms implementation.

For Simulated Annealing, it is more difficult to find parameter settings that provide an optimal solution for each of the 50 runs. However, the chosen parameter settings perform good enough, as near-optimal solutions show similarities on global level. Regarding the temperature, it is seen that a too high temperature reduces the chance to leave a local solution, and hence does result in a lower performance. Similarly, if the temperature decreases too fast, it is very likely to end up in a local solution. On the other hand, if the temperature slows down too slowly, more iterations might be required to come to an optimal solution. More iterations obviously result in a higher chance to obtain the optimal solution, but this is strongly related to the required solution time. Hence we choose to set this value of M_2 equal to 2.5, as this seems to give a nice balance between solution time and optimality. The same reasoning applies to the number of subiterations.

Sensitivity of the parameter settings of the Simulated Annealing implementation

Changed values	%Best value	CPU Time
Base (Table 5)	94	64
$T = 0.1f(S)$	94	64
$T = 0.5\ f(S)$	84	63
$\delta = 0.85$	92	64
$\delta = 0.95$	20	62
$M_2 = 2 * cardinality\ of\ set\ H$	88	51
$M_2 = 3 * cardinality\ of\ set\ H$	96	75
$M_3 = 100 * cardinality\ of\ set\ H$	80	49
$M_3 = 300 * cardinality\ of\ set\ H$	96	78

Table 7: Sensitivities of the parameters for the Simulated Annealing implementation.

3.7 Conclusions and Recommendations

In this chapter, we relaxed the classical assumptions of the uncapacitated hub location and network design problem regarding directs, multiple allocations, stopovers and an incomplete hub network, combined with a cost function that improves the estimation of cost to deploy such a network. These extensions are particularly useful to distinct solutions that are similar on overall cost for the generic high level problem but outperform in specific regions. Two heuristics were implemented, Genetic Algorithms and Simulated Annealing, to evaluate the impact of these improvements. Especially the extensions of stopovers and an incomplete hub network have a strong regional impact in selection of the hub locations.

When comparing the two methods, we may conclude that both perform well and show approximately same solutions on global and local level. Computation times are almost similar as well, though GA is slightly faster than SA.

Lastly we would like to conclude with some topics for further research. We presented a sequential approach in which we first derive a global solution and afterwards apply local optimisation. Although this can be motivated by the observation that global solutions are similar on cost level and divide a country in regions where hubs hardly differentiate on global cost level, it might be interesting to see the impact of an approach that combines these sequential steps. The largest challenge here is to reduce computation times of the MIP-problems resulting from the inclusion of stopovers and three hub routes; these formulations include a knapsack problem which is an NP-hard problem. Further research might replace these formulations by a heuristic approach.

Another topic for further research concerns the inclusion of time restrictions, as express service providers have only a limited time for transportation. Although some restrictions are included in our model (i.e. the number of hubs in a route is limited to three and only one intermediate location is allowed in a stopover), explicit time restrictions might add a significant improvement for express businesses to end up with solutions that support the guaranteed service levels.

Lastly, note that the problem formulation can be enriched by the inclusion of fixed and variable hub cost.

3.8 Appendix A

Parameters:

$w_{d_1 d_2}$ non-negative package flows between depot $d_1 \in D$ and depot $d_2 \in D$,

w_d^{orig} denotes the total amount of flow originating at depot $d \in D$, i.e., $\sum_{d_2} w_{d d_2}$,

w_d^{dest} denotes the total amount of flow destined at depot $d \in D$, i.e., $\sum_{d_1} w_{d_1 d}$,

$\bar{d}_{l_1 l_2}$ distance between nodes $l_1 \in L$ and $l_2 \in L$,

q_f the capacity of vehicle $f \in F$,

c_f^{var} the cost per unit distance for vehicle $f \in F$.

Decision variables:

$\bar{g}_h$ a binary variable that equals one if hub $h \in H$ is established, and zero otherwise,

g_{dh} a binary variable that equals one if depot d $\in D$ is allocated to hub $h \in H$, and zero otherwise,

$x_{d_1 d_2}^D$ variable that denotes the fraction of package flows from depot d_1 to depot d_2 that is routed via a direct connection, where $d_1, d_2 \in D$,

$x_{d_1 d_2 h_1 h_2}$ variable that denotes if the flow follows route $d_1 \to h_1 \to h_2 \to d_2$, and zero otherwise, where $d_1, d_2 \in D$ and $h_1, h_2 \in H$,

$\dot{x}_{d_1 d_2 d_3 d_4 h_1 h_2 h_3}$ variable that equals one if the flow follows route $d_1 \to d_2 \to h_1 \to h_2 \to h_3 \to d_3 \to d_4$, where $d_1, d_2, d_3, d_4 \in D$ and $h_1, h_2, h_3 \in H$,

$\dot{g}_{d_1 d_2}$ binary variable that equals one if depot d_2 is visited on the route from depot d_1 to a hub, and zero otherwise,

$y_{l_1 l_2 f}$ number of vehicles used on arc $(l_1, l_2) \in L$ and type $f \in F$.

Mathematical formulation of the basic model:

$$\min \quad \sum_{l_1, l_2 \in \{L\}; f \in \{F\}} \left(c_f^{var} * \bar{d}_{l_1 l_2} * y_{l_1 l_2 f} \right) \tag{3.1}$$

$$\text{s.t.} \quad \sum_{d \in \{D\}} (g_{dh}) = 1 \qquad \forall d \in D \tag{3.2}$$

$$g_{dh} \leq \bar{g}_h \qquad \forall d \in D; h \in H \tag{3.3}$$

$$\sum_{h_1, \in \{H\}} \left(x_{d_1 d_2 h_1 h_2} \right) \leq \bar{g}_{h_2} \qquad \forall d_1, d_2 \in D; h_2 \in H \tag{3.4}$$

$$\sum_{h_2, \in \{H\}} \left(x_{d_1 d_2 h_1 h_2} \right) \leq \bar{g}_{h_1} \qquad \forall d_1, d_2 \in D;\ h_1 \in H \tag{3.5}$$

$$\sum_{d_2 \in \{D\}} \left(w_{d_1 d_2} * g_{d_1 h_1} \right) = \sum_{f \in \{F\}} \left(q_f * y_{d_1 h_1 f} \right) \qquad \forall d_1 \in D;\ h_1 \in H \tag{3.6}$$

$$\sum_{d_1 \in \{D\}} (w_{d_1 d_2} * g_{d_2 h_2}) = \sum_{f \in \{F\}} (q_f * y_{d_2 h_2 f}) \quad \forall d_2 \in D;\ h_2 \in H \quad (3.7)$$

$$\sum_{d_1, d_2 \in \{D\}} (w_{d_1 d_2} * x_{d_1 d_2 h_1 h_2}) \leq \sum_{f \in \{F\}} (q_f * y_{d_2 h_2 f}) \quad \forall h_1, h_2 \in H \quad (3.8)$$

$$g_{dh}, \bar{g}_h \in \{0,1\} \quad \forall d \in D; h \in H \quad (3.9)$$

$$x_{d_1 d_2 h_1 h_2} \geq 0 \quad \forall d_1, d_2 \in D; h_1, h_2 \in H \quad (3.10)$$

$$y_{l_1 l_2 f} \in \mathbb{N} \quad \forall l_1, l_2 \in L, f \in F \quad (3.11)$$

The objective function minimises the total line-haul cost to operate the network. Constraints (3.2) ensure the single assignment of depots to hubs. Depots can only be assigned to hubs that are used, which is forced by Constraints (3.3). Constraints (3.4) and (3.5) are the routing constraints which follow from depot to hub assignment (as all hubs are connected). When flows are routed between pairs of locations, enough vehicle capacity should be available; this is the purpose of constraints (3.6)-(3.8). The constraints (3.9)-(3.11) define the decision variables.

Mathematical formulation of the extended model:

$$\min \quad \sum_{l_1, l_2 \in \{L\}; f \in \{F\}} (c_f^{var} * \bar{d}_{l_1 l_2} * y_{l_1 l_2 f}) \quad (3.12)$$

$$\text{s.t.} \quad x^D_{d_1 d_4} + \sum_{d_2, d_3 \in \{D\}; h_1, h_2 h_3} (\dot{x}_{d_1 d_2 d_3 d_4 h_1 h_2 h_3}) = 1 \quad \forall d_1, d_4 \in D \quad (3.13)$$

$$\sum_{d_2, d_3 \in \{D\}; h_2, h_3 \in \{H\}} (\dot{x}_{d_1 d_2 d_3 d_4 h_1 h_2 h_3}) \leq \bar{g}_{h_1} \quad \forall d_1, d_4 \in D;\ h_1 \in H \quad (3.14)$$

$$\sum_{d_2, d_3 \in \{D\}; h_1, h_3 \in \{H\}} (\dot{x}_{d_1 d_2 d_3 d_4 h_1 h_2 h_3}) \leq \bar{g}_{h_2} \quad \forall d_1, d_4 \in D; h_2 \in H \quad (3.15)$$

$$\sum_{d_2, d_3 \in \{D\}; h_1, h_2 \in \{H\}} (\dot{x}_{d_1 d_2 d_3 d_4 h_1 h_2 h_3}) \leq \bar{g}_{h_3} \quad \forall d_1, d_4 \in D;\ h_3 \in H \quad (3.16)$$

$$\sum_{d_2 \in \{D\}} (\dot{g}_{d_1 d_2}) = 1 \quad \forall d_1 \in D \quad (3.17)$$

$$\dot{g}_{d_1d_2} \leq \dot{g}_{d_2d_2} \qquad \forall d_1, d_2 \in D \quad (3.18)$$

$$\sum\nolimits_{d_3,d_4\in\{D\};h_1,h_2,h_3\in\{H\}} \left(\dot{x}_{d_1d_2d_3d_4h_1h_2h_3}\right) \leq \dot{g}_{d_1d_2} \qquad \forall d_1, d_2 \in D \quad (3.19)$$

$$\sum\nolimits_{d_1,d_2\in\{D\};h_1,h_2,h_3\in\{H\}} \left(\dot{x}_{d_1d_2d_3d_4h_1h_2h_3}\right) \leq \dot{g}_{d_3d_4} \qquad \forall d_3, d_4 \in D \quad (3.20)$$

$$\sum\nolimits_{h_1,h_2h_3\in\{H\}} \left[\sum\nolimits_{d_3,d_4\in D} \left(w_i^{orig} * \dot{x}_{d_id_jd_3d_4h_1h_2h_3}\right) + \sum\nolimits_{d_1,d_2\in\{D\}} \left(w_j^{dest} * \dot{x}_{d_1d_2d_id_jh_1h_2h_3}\right)\right] + \left(w_{d_id_j} * x_{d_id_j}^D\right) \leq \sum\nolimits_{f\in\{F\}} \left(q_f * y_{d_id_jf}\right) \qquad \forall d_i, d_j \in D \quad (3.21)$$

$$\sum\nolimits_{d_1,d_2,d_3,d_4\in\{D\}} \left[\sum\nolimits_{h_3\in\{H\}} \left(w_{d_1d_4} * \dot{x}_{d_1d_2d_3d_4h_ih_jh_3}\right) + \sum\nolimits_{h_1\in\{H\}} \left(w_{d_1d_4} * \dot{x}_{d_1d_2d_3d_4h_1h_ih_j}\right)\right] \leq \sum\nolimits_{f\in\{F\}} \left(q_f * y_{h_ih_jf}\right) \qquad \forall h_i, h_j \in H \quad (3.22)$$

$$\sum\nolimits_{d_1,d_3,d_4\in\{D\};h_2,h_3\in\{H\}} \left(w_{d_1}^{orig} * \dot{x}_{d_1d_2d_3d_4h_1h_2h_3}\right) \leq \sum\nolimits_{f\in\{F\}} \left(q_f * y_{d_2h_1f}\right) \qquad \forall d_2 \in D; h_1 \in H \quad (3.23)$$

$$\sum\nolimits_{d_1,d_2,d_4\in\{D\};h_1,h_2\in\{H\}} \left(w_{d_4}^{dest} * \dot{x}_{d_1d_2d_3d_4h_1h_2h_3}\right) \leq \sum\nolimits_{f\in\{F\}} \left(q_f * y_{d_3h_3f}\right) \qquad \forall d_3 \in D; h_3 \in H \quad (3.24)$$

$$\bar{g}_h \in \{0,1\} \qquad \forall h \in H \quad (3.25)$$

$$\dot{g}_{d_1d_2} \in \{0,1\} \qquad \forall d_1, d_2 \in D \quad (3.26)$$

$$x_{d_1d_4}^D, \dot{x}_{d_1d_2d_3d_4h_1h_2h_3} \geq 0 \qquad \forall d_1, d_2, d_3, d_4 \in D; h_1, h_2, h_3 \in H \quad (3.27)$$

$$y_{l_1l_2f} \in \mathbb{N} \qquad \forall l_1, l_2 \in L, f \in F \quad (3.28)$$

The objective function minimises the total line-haul cost to operate the network. Constraints (3.13) ensure that the total flow of packages between two depots is transported via one or more routes. This can be a direct connection and/or a route via one or more depots and hubs. Constraints (3.14)-(3.16) enforces a hub to be established if a route visits this hub. Constraints

(3.17) allocate each depot to another depot (for the stopover requirement). Each depot is allocated to exactly one depot, since $\dot{g}_{d_1 d_2}$ is a binary variable. A depot can also be allocated to itself such that no other depot is visited on the way from this depot to its allocated hub. Constraints (3.18) ensures that a stopover can visit at most one depot on the route from depot to hub and from hub to depot respectively. Constraints (3.19) and (3.20) make sure that the right fraction of flow is assigned to a route. There can only be flow between depot d_1 and depot d_2 if depot d_1 is allocated to depot d_2. Constraints (3.21)-(3.24) determine the number and type of vehicles needed on a connection. Constraints (3.21) determine this for connections between depots, Constraints (3.22) for connections between hubs, Constraints (3.23) for connections from depot to hub, and Constraints (3.24) for connections from hub to depot. The variables are defined by Constraints (3.25) – (3.28)

3.9 Appendix B

MIP-model for stopovers

Parameters:

$D'_h \subseteq D$ subset of depots in the region of hub $h \in H$,

$\overline{w}_{dh}^{orig}$ denotes the total amount of flow originating at depot $d \in D$ and routed to hub $h \in H$,

$\overline{w}_{dh}^{dest}$ denotes the total amount of flow destined at depot $d \in D$ and routed from hub $h \in H$.

Decision variables:

$\tilde{x}_{d_1 d_2}$ a stopover route from depot $d_1 \in D$ to depot $d_2 \in D$.

Mathematical formulation – for a given hub $h \in H$:

$$\min \quad \sum_{d_1,d_2 \in \{D'_h\}; f \in F} \left(c^{var} * d_{d_1 d_2} * y_{d_1 d_2 f}\right) + \sum_{d \in \{D'_h\}; f \in F} \left(c_f^{var} * d_{dh} * y_{dhf} + c_f^{var} * d_{hd} * y_{hdf}\right) \tag{3.29}$$

$$\text{s.t.} \quad \sum_{d_2 \in \{D'_h\}} \tilde{x}_{d_1 d_2} = 1 \qquad \forall d_1 \in D'_h \tag{3.30}$$

$$\tilde{x}_{d_1 d_2} \leq \tilde{x}_{d_2 d_2} \qquad \forall d_1, d_2 \in D'_h \tag{3.31}$$

$$\tilde{x}_{d_1 d_2} * \overline{w}_{d_1 h}^{orig} \leq \sum_{f \in \{F\}} \left(q_f * y_{d_1 d_2 f}\right) \qquad \forall d_1, d_2 \in D'_h, d_1 \neq d_2 \tag{3.32}$$

$$\tilde{x}_{d_1 d_2} \overline{w}_{d_1 h}^{dest} \leq \sum_{f \in \{F\}} \left(q_f * y_{d_2 d_1 f}\right) \qquad \forall d_1, d_2 \in D'_h, d_1 \neq d_2 \tag{3.33}$$

$$\sum_{d_1 \in \{D_h'\}} \tilde{x}_{d_1 d_2} \overline{w}_{d_1 h}^{orig} \quad \leq \quad \sum_{f \in \{F\}} \left(q_f * y_{d_2 h f} \right) \qquad \forall d_2 \in D_h' \quad (3.34)$$

$$\sum_{d_1 \in \{D_h'\}} \tilde{x}_{d_1 d_2} \overline{w}_{d_1 h}^{dest} \quad \leq \quad \sum_{f \in \{F\}} \left(q_f * y_{h d_2 f} \right) \qquad \forall d_2 \in D_h' \quad (3.35)$$

$$\tilde{x}_{d_1 d_2} \in \{0,1\} \qquad \forall d_1, d_2 \in D_h' \quad (3.36)$$

$$y_{l_1 l_2 f} \in \mathbb{N} \qquad \forall l_1, l_2 \in \{h, D_h'\}; f \in F \quad (3.37)$$

The objective function of the MIP-problem consists of three terms. The first term determines the transportation cost for connections between depots, and the second and third term determine the transportation cost for connections between depots and hubs. Constraints (3.30) ensure that the total amount of flow is transported via one stopover. Constraints (3.31) ensure that there can only be one stop on the route from a depot to a hub. Constraints (3.32)-(3.35) determine the number and type of vehicles used on each connection. Constraints (3.32) and (3.33) determine this for connections between depots, and Constraints (3.34) and (3.35) determine this for connections between depot and hub. Constraints (3.36) state that the variable $\tilde{x}_{d_1 d_2}$ must be a binary variable and constraints (3.37) state the variable $y_{l_1 l_2 f}$ must be a non-negative integer variable.

<u>Incomplete hub network (three hub routes)</u>

Parameters:

$\overline{w}_{h_1 h_2}^{hubs}$ the flows to be transported between hubs $h_1 \in H$ and $h_2 \in H$.

Decision variables:

$\hat{x}_{h_1 h_2 h_3}$ the fraction of the volume from hub $h_1 \in H$ to hub $h_3 \in H$ that is transported via hub $h_2 \in H$.

Mathematical formulation:

$$\min \quad \sum_{h_1, h_2 \in \{H\}; f \in F} \left(c_f^{var} * d_{h_1 h_2} * y_{h_1 h_2 f} \right) \qquad (3.38)$$

$$\text{s.t.} \quad \sum_{h_2 \in \{H\}} \left(\hat{x}_{h_1 h_2 h_3} \right) \quad = \quad 1 \qquad \forall h_1, h_3 \in H \quad (3.39)$$

$$\sum_{h_3 \in \{H\}} \left(\overline{w}_{h_i h_3}^{hubs} * \hat{x}_{h_i h_j h_3} \right) + \sum_{h_1 \in \{H\}} \left(\overline{w}_{h_1 h_j}^{hubs} * \hat{x}_{h_1 h_i h_j} \right) \quad \leq \quad \sum_{t \in T} \left(q_f * y_{h_i h_j f} \right) \qquad \forall h_i, h_j \in H \quad (3.40)$$

$$\hat{x}_{h_1h_2h_3} \geq 0 \qquad \forall h_1, h_2, h_3 \in H \quad (3.41)$$

$$y_{l_1l_2f} \in \mathbb{N} \qquad \forall l_1, l_2 \in H; f \in F \quad (3.42)$$

The objective function minimises the transportation cost for connections between hubs. Constraints (3.39) ensure that the total amount of flow is transported from hub h_1 to hub h_3 via one or more routes. Constraints (3.40) determine the number and type of vehicles needed on the connections between hubs. Constraints (3.41) force variable $\hat{x}_{h_1h_2h_3}$ to be a non-negative continuous variable, and Constraints (3.42) state that variable $y_{l_1l_2f}$ must be a non-negative integer variable.

3.10 Appendix C

Note that the weakness in relaxation of the problem results from the variable $y_{l_1l_2f}$; due to the relaxation, vehicles can be used partially and cost is accounted partially as well. We introduced two sets of inequalities to tighten the outcome of the relaxed problem situation.

In order to tighten the LP-relaxation of the basic model the following constraints are added to the problem:

$$\sum_{f\in F}\left(y_{d_1h_1f}\right) \geq g_{d_1h_1} \qquad \forall d_1 \in D, \bar{w}_{d_1h}^{orig} > 0; h_1 \in H \quad (3.43)$$

$$\sum_{f\in f}\left(y_{d_2h_2f}\right) \geq g_{d_2h_2} \qquad \forall d_2 \in D, \bar{w}_{d_2h}^{dest} > 0; h_2 \in H \quad (3.44)$$

$$\sum_{f\in F}\left(y_{h_1h_2f}\right) \geq x_{d_1d_2h_1h_2} \qquad \forall d_1, d_2 \in D, w_{d_1d_2} > 0; h_1, h_2 \in H \quad (3.45)$$

Here $\bar{w}_{d_1h}^{orig} = \sum_{d_2\in\{D\}}\left(w_{d_1d_2}\right)$ and $\bar{w}_{d_2h}^{dest} = \sum_{d_1\in\{D\}}\left(w_{d_1d_2}\right)$. These sets of constraints ensure that at least one vehicle must be assigned to each connection with a non-negative flow of packages. Constraints (3.43) ensure this for connections from a depot to a hub, Constraints (3.44) for connections from a hub to a depot, and Constraints (3.45) for connections from a hub to a hub.

The LP-relaxation can be further tightened by changing the above constraints to the following constraints, where $\bar{q} = \max_f\left[q_f\right]$.

$$\sum_{f\in F}\left(y_{d_1h_1f}\right) \geq \left\lceil\frac{\bar{w}_{d_1h_1}^{orig}}{\bar{q}}\right\rceil g_{d_1h_1} \qquad \forall d_1 \in D; h_1 \in H \quad (3.46)$$

$$\sum_{f\in f}\left(y_{d_2h_2f}\right) \geq \left\lceil\frac{\bar{w}_{d_2h_2}^{dest}}{\bar{q}}\right\rceil g_{d_2h_2} \qquad \forall d_2 \in D;\ h_2 \in H \quad (3.47)$$

$$\sum_{f\in F}\left(y_{h_1h_2f}\right) \geq \left\lceil\frac{w_{d_1d_2}}{\bar{q}}\right\rceil x_{d_1d_2h_1h_2} \qquad \forall d_1, d_2 \in D; h_1, h_2 \in H \quad (3.48)$$

The total amount of flow that needs to be transported on a connection is divided by the capacity of the largest vehicle. This number denotes the minimum number of vehicles needed to transport this flow. Constraints (3.46) state this for connections from a depot to a hub, Constraints (3.47) for connections from a hub to depot, and Constraints (3.48) for connections from a hub to a hub. The LP-relaxation of the model extended with these constraint sets is denoted by LP-relaxation II.

4 Enriching the tactical network design of express service providers with fleet scheduling characteristics

This chapter is based on the following journal paper:

Enriching the Tactical Network Design of Express Carriers with Fleet Scheduling Characteristics

Meuffels, I; Fleuren, H; Cruijssen, F; Van Dam, E
Flexible Services Manufacturing Journal (2010), Vol. 22, Is. 1-2, pp. 3-35

4.1 Abstract

Express service carriers provide time-guaranteed deliveries of packages (i.e. documents, parcels, or pieces of freight) via a network consisting of depots and hubs. In this, depots take care of the collection and delivery of packages, and hubs have the function to consolidate packages in between the depots. The tactical network design problem assigns depots to hubs, determines arcs between hubs, and routes packages through the network. Afterwards, fleet scheduling creates a schedule for vehicles operated in the network. The strong relation between flow routing and fleet scheduling makes it difficult to optimise the network cost. Due to this complexity, fleet scheduling and network design are usually decoupled. We propose a new tactical network design model that is able to include fleet scheduling characteristics (like vehicle capacities, vehicle balancing, and drivers' legislations) in the network design. The model is tested on benchmark data based on instances from an express service provider, resulting in significant cost reductions.

4.2 Introduction

Express service providers offer time-guaranteed deliveries of packages. Direct transport from sender to receiver is the fastest way of transport but this is in general not cost efficient. Therefore, express service providers operate a network in which packages of many customers are consolidated. Packages of several senders are consolidated at depots, transported to other depots via the line-haul network and finally delivered to the consignees. We will now briefly describe how the express supply chain is organised. Then a description of network design is given followed by a discussion on fleet scheduling. At the end of this introduction, our research goals are stated.

4.2.1 Express supply chain

The first depot at which a package arrives after pickup is called the *origin depot* (or *origin)* of the package; the depot from where the package is delivered to the consignee is called the *destination depot* (or *destination*) of the package. The transport of packages between origin depot and destination depot is called *line-haul*. Origin and destination depot form an *od-pair*. For these od-pairs, several *services* are offered, defined in terms of promised delivery dates and times of the packages. Packages of an od-pair with the same service can always be transshipped together during line-haul transport. The number of packages of one service to be transshipped between two depots is called the *flow* of the *origin-destination service pair* (*od-service pair*); the total flow of packages to be transported between two depots is called the flow of the od-pair.

Cut-off times form the connection between the pickup and delivery process and the line-haul process and guarantee the on-time delivery of packages. That is, all packages of one service collected in the pickup process have to be processed and loaded into line-haul vehicles before the *collection cut-off time* of the corresponding service; the line-haul transport starts afterwards. The line-haul vehicles have to arrive at the destination depots before the *delivery cut-off time* of the corresponding service. The line-haul vehicles are unloaded after arrival at the destination depot and packages are processed such that the final delivery to consignees can start afterwards. Flows in the line-haul network are either directly transported between depots or consolidated at *hub* locations. A hub is a sorting centre serving depots and other hubs. Hubs in the express network are crucial in making fast and reliable connections. A *direct route* between depots can be established if there is enough flow to create a (nearly) full vehicle load between two depots. A direct route can also be used when none of the hub routes is able to meet the service requirements of the corresponding od-pair. A *hub route* is a route from depot to depot visiting hubs in between; note that hub routes result in detours of flow routing.

Express service providers can use road or air modes in their line-haul transport. Generally, road transport is preferred because of the lower cost involved. Air transport is used to establish services that cannot be offered by road transport. Considering cost, it is clear that fleet cost dominate in the design of air networks. In road networks, fleet cost is an important cost component though other cost components (like handling cost) are important as well, and the trade-off between these cost components determines the final network. Besides, the cost of a single aircraft is in general much higher than the cost of a vehicle, so that it is more costly to use an additional aircraft compared to the use of an additional vehicle. This difference is also illustrated in the design of both networks: while models on air network design focus on fleet routing, road network design focuses on flow routing. The resulting flight schemes are in general the same each day independently of flow size while road schemes slightly differ per day, i.e. the main schedule in a road network is fixed, though in case of large flows it is possible to gain some additional vehicle capacity (but against a higher, but still profitable, price). A second difference in the design of both networks is the capacity of the fleet: aircraft capacities are mostly higher than vehicle capacities, so that in air networks each route is performed by one aircraft, while in road networks multiple vehicles may be scheduled along the same route. Single aircraft routing in combination with a constant aircraft fleet often leads to the introduction of sort windows at hubs in the air network in order to guarantee services: that is, all aircrafts arrive before the start of the sorting process and leave after sorting has taken place. On the contrary, vehicles in road networks arrive and leave hubs at several moments in time. Finally, air networks often face additional restrictions on the possibilities of transport (like time slots at airports, runway constraints, aircraft landing constraints, etcetera) that are not found in road networks. In this chapter, we focus on road transport and our modelling approach is sophisticated to such a network. An overview of a typical express supply chain for road transport is given in Figure 11.

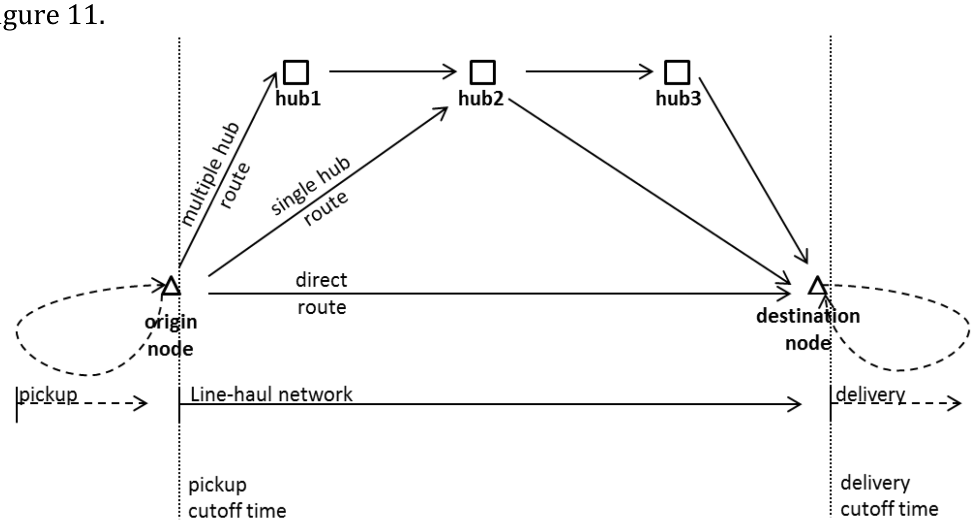


Figure 11: Overview of the express supply chain, with the pickup and delivery process and the network process. Flow can follow a direct route, or a (single/multiple) hub route.

4.2.2 Hub network design

Consolidation at hub locations was introduced in the literature by O'Kelly (1986). The construction of a line-haul network is better known as the hub network design problem. Generally, there are two decision levels in hub network design problems. The strategic hub network design problem of express service providers decides on the number and location of hubs in the line-haul network. The tactical hub network design problem concerns the assignment of depots to hubs, determines arcs (i.e. line-hauls) between hubs, and routes flows through the network.

In general, strategic and tactical network design discussed in the literature focuses on minimisation of the sum of unit transport cost. It is generally assumed that consolidated transport between hub locations benefits from economies of scale such that unit transport cost of inter-hub flows can be discounted. The main restrictions in both strategic and tactical network design are *flow conservation* and *service commitment*. Flow conservation requires that all flow has to be transported between depots; service commitment requires that flows are transported within predefined time limits. It is often assumed that the hub network is complete when a link between every hub pair is established, and that no direct routes are allowed, (Alumur & Kara, 2008). Besides, some literature assumes *capacitated* hub locations that can only deal with a limited amount of flow (e.g (Aykin, 1994), (Melkote & Dashkin, 2001)).

4.2.3 Fleet scheduling

After tactical network design, vehicle schedules need to be created such that the flow can be transported. An important aspect of fleet scheduling is the inclusion of *waiting times* (Kara & Tansel, 2001): a vehicle can only depart once the flow scheduled on that vehicle has arrived and been processed. In particular, waiting times are important in case of the last vehicle moving via a certain arc. Flows can only be consolidated when there is enough time available for consolidation. This is illustrated in Figure 12: the cut-off times imply that there are only 10 hours available to transport flows (a,b) and (c,d); as a result, consolidation of inter-hub flows is not possible. Note that cut-off times not only define the available time of transport, but also define the moment of transport.

Another important aspect of fleet scheduling is *vehicle balancing*: since express service providers operate on a daily basis, the number of incoming and outgoing vehicles should be balanced for every location. A third aspect in fleet scheduling that needs attention concerns *drivers' legislations*. Maximum driving times and prescribed breaks may not be violated. If the driving time between two locations exceeds the maximum driving time of one driver, a second driver is required resulting in additional cost. In the network design literature, the problem of fleet scheduling and balancing is referred to as the *fleet scheduling problem*.

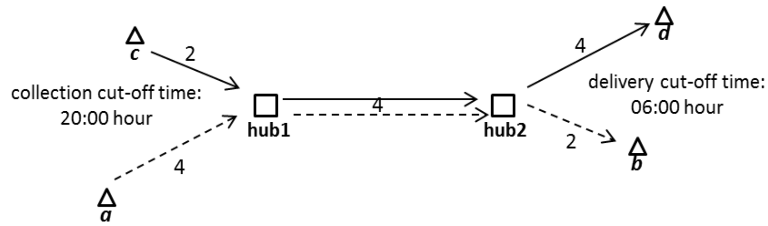


Figure 12: Due to the limited time in the network, it is not always possible to consolidate flows; the flows of $(\boldsymbol{a}, \boldsymbol{b})$ can not be consolidated with the flows of $(\boldsymbol{c}, \boldsymbol{d})$ at $\boldsymbol{hub1}$. Note: number above arcs denote driving times.

4.2.4 Research goals

This chapter concerns the tactical network design in road transport of express service providers. The research is inspired by practical considerations not yet dealt with in the literature.

4.2.4.1 Cost function: plainly linear

The first extension on the existing literature concerns the cost function, which in practice turns out to be more complex than generally seen in the literature. In the latter, the cost function results from unit transport cost and inter-hub transport is discounted. However, O'Kelly & Bryan (1998) claim that the inclusion of an exogenously determined discount applied to all inter-hub arcs regardless of the differences in the flows travelling across them, oversimplifies the problem. The authors claim that the cost has to be presented by a non-linear function such that marginal travel cost decreases as flows increase. The non-linear cost function is afterwards approximated by a piece-wise linear cost function.

We agree with O'Kelly (1986) that hub consolidation results in economies of scale compared to direct driving, but like O'Kelly & Bryan (1998) we disagree with the traditional discounting of inter-hub transport only. As in (O'Kelly & Bryan, 1998), we observe that unit transport cost decreases as flows increase, however, we will not apply a discounting of flows but determine vehicle movements explicitly both towards hubs and in-between hubs. In this approach, we assume that only one vehicle type is available, which hardly limits practical applications[6]. Besides, note that the model that is proposed can easily be extended to relax this assumption when needed (more routes need to be generated in that case).

The cost in our network design incorporates the plainly linear cost function, since we explicitly determine vehicle movements. This approach is applied to all arcs in the network, so it is not limited to inter-hub arcs only. The discounting of only inter-hub arcs was also questioned by

[6] There are several reasons that support this assumption. Firstly, because unit cost of medium size vehicles are about 85% of large size vehicles, while the capacity is halved. This means that medium size vehicles will only be used in case of very small flows. Besides, if medium size vehicles are scheduled at (approximately) full vehicle loads, daily fluctuations may require a second medium size vehicle, resulting in higher transport cost in the end. A second reason to apply only one vehicle type comes from subcontracting of transport, which occurs regularly at the express carrier in scope. For subcontracting, vehicle movements are combined in so-called *tours* that can be driven by the same vehicle. There are several classifications of these tours, and the better the tours, the lower the cost of subcontracting. We will not elaborate on the details of tour generation since it is behind the scope of this research. However, an example of a low-cost tour is a 1-day tour (total driving time about 9 hours) that starts and ends at the same location and can be driven by one driver satisfying drivers breaks prescribed by regulation. It is easy to understand that if we would use medium trucks towards hubs and large trucks between hubs, tour generation becomes more difficult since separate tours need to be created towards hubs or in-between hubs. The cost of subcontracting these tours is in general much higher than the cost savings achieved by using trucks of medium size and hence we assume that only one vehicle type is available.

Podnar et al. (2002), who proposed a discounting applied to all arcs traversing flows larger than a certain threshold. Since the dispatching of vehicles is subject to the size of the flow in our network design, we do not need to make any further assumptions on this.

An illustration of the two functions found in literature and the plainly linear cost function proposed in this chapter, can be found in Figure 13, where the figure on the left shows the total cost and the figure besides the average cost as a function of the total flow. In this figure, 'linear' refers to the cost function found in the traditional literature, 'non-linear' refers to the cost function proposed by O'Kelly & Bryan (1998), and 'plainly linear' reflects the vehicle dependent cost function.

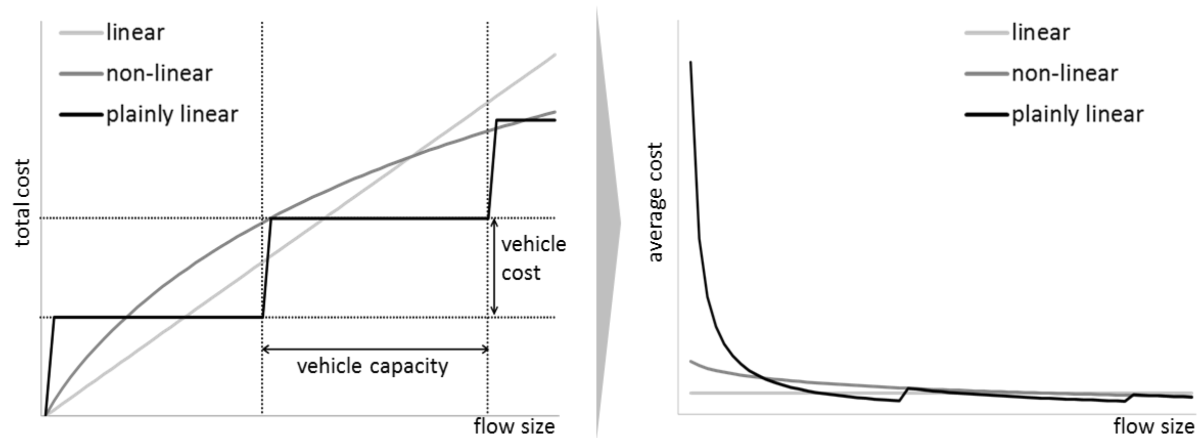


Figure 13: Total cost versus average cost: only for large flow sizes, linear, non-linear and plainly linear cost function show similar unit transportation cost.

4.2.4.2 Cost function: additional cost components

Besides, we improve the reflection of real-world cost made by express service providers by the inclusion of some other cost components. One of these is vehicle balancing cost. Crainic (2002) describes the need to move empty vehicles because of the imbalances that exist in trade flows that result in discrepancies between vehicle supply and demand in various zones or nodes in the network. Since balancing cost forms a substantial part of the total costs in an express network, we include this cost in our network design. A second cost component that we add to our design concerns the cost of a second driver. Drivers' legislations may not be violated, so additional cost is made when a second driver is required. The last cost component that we take into account concerns variable handling cost at hub locations. Note that we do not need to include fixed hub cost, since hub locations are given in the tactical network design

4.2.4.3 Note: road versus air networks

It should be noted that (some of) the cost aspects discussed above are captured in the literature on design of air networks. However, to the best of our knowledge, no literature on the design of road networks has included these cost components in their modelling.

Express carriers offering next day services face tight time constraints. The literature discusses the usage of a cover radius, (Kara & Tansel, 2003), which is a bound on transport time. However, the available time to transport flows depends on the service definition. The tactical network design model presented in this chapter uses cut-off times to derive the available time to transport flows. In this way multiple services can be included. However, during network design it is not checked whether flows can be combined in a truck. This is done in the heuristic that is run afterwards.

4.2.4.4 Assumptions on routing

Routes that are allowed in our model can be varied, so long as service requirements can be satisfied with respect to the cut-off times. We therefore do not have to assume a complete hub network, nor exclude direct routing. Besides, depots are not restricted to be connected to a *single* hub node, so we allow *multiple assignment* of depots to hubs. However, each depot is directly assigned to a hub node, that is, no *stopovers* at other depots are made.

4.2.4.5 Assumptions on hub nodes

Finally, we assume that hub locations can handle a limited amount of flow because hub locations are fixed and given in the tactical hub network design.

The remainder of this chapter is organised as follows. Section 4.3 gives an overview of the literature on hub network design. The modelling approach is presented in Section 4.4. Two network design models are presented, a traditional model and a new model. A fleet scheduling heuristic is used to derive the final network cost so that the two models can be compared. The models are tested on data instances of an express service provider. The results are presented in Section 4.5. Finally our conclusions and directions for further research are given in Section 4.6.

4.3 Literature

This section briefly discusses the literature on the hub network design problem. Recent overviews on hub network design in express networks are given by Alumur & Kara (2009). Overviews on hub network design in general are given by Revelle & Eiselt (2008) and Melo et al. (2009).

Hub consolidation was introduced in the literature by O'Kelly (1986). In this work, O'Kelly introduced the concept of economies of scale on inter-hub flows: the idea is that flows between hubs might enjoy a discounted transport rate arising from the greater volume on these arcs. This is modelled by discounting unit transport cost for inter-hub flows. The first strategic hub network design model is a quadratic model presented by O'Kelly (1987). Afterwards, several researchers studied strategic and tactical hub network design and several variants of the problem are proposed. The strategic hub network design selects the locations of hubs in the network such that the sum of unit transport cost is minimised ((O'Kelly, 1992), (Aykin, 1994), (Aykin, 1995), (O'Kelly, et al., 1996)), the largest transport time is minimised ((Kara & Tansel, 2001)), the number of hubs is minimised ((Kara & Tansel, 2003), (Tan, et al., 2007), (Yaman, et al., 2007), (Alumur & Kara, 2008)), or the total number of packages delivered to customers within a certain time bound is maximised ((Yaman, et al., 2008)). This chapter focuses on the tactical hub network design. The remainder of this section concerns the literature on the tactical hub network design problem.

Kuby & Gray (1993) consider the tactical network design in air transport examining trade-offs and savings involved with stopovers and feeders towards a single air hub location. The authors observed that in real-world practices direct flights towards an air hub occur only occasionally: most flights stop over at several cities along their routes, and often feeder routes with smaller planes transfer loads to larger planes at intermediate cities. Therefore, a mixed-integer program is developed to design the least-cost single-hub air network including stopovers and feeders. In this, it is assumed that the hub location is already determined. The authors conclude that substantial improvements in cost, miles flown, load factor and number of aircraft can be achieved by using stopovers and feeders in the hub network, and that it is unrealistic to assume a network with only direct flights.

The tactical hub network design in air transport is further examined by Barnhart & Schneur (1996). Pickup and delivery aircraft routes and schedules are derived towards a single hub node. Each aircraft route begins at the hub, visits a set of destination airports followed by an idle period, then visits a set of origin airports before returning to the hub. The idle time in between can be used for ferrying (i.e. repositioning of aircrafts). Earliest pickup and latest delivery times are used at the airports. Associated with the hub is a cut-off time, which is the latest time an aircraft may arrive at the hub. Three service levels are defined in these models: next-day service (24 hours), second day service (48 hours) and deferred service (3-5 days).

A system that determines aircraft routes, fleet assignments and package routings simultaneously has been described by Armacost et al. (2004). Like Barnhart & Schneur (1996), pickup and delivery routes towards a single air hub are derived including time windows for pickup and delivery. Armacost et al. (2002) and Armacost et al. (2004) use a composite variable formulation to solve a comparable model.

Multiple hub road networks are considered by Lin (2001). The author observes that vehicle balancing and drivers official work rules are important operating constraints in a cost-effective line-haul operating plan. The work afterwards considers the flow routing problem only, assuming that hub locations and fleet schedules satisfying the operational constraints are given. This problem can then be compared to a capacitated multi-commodity flow problem. Cost taken into consideration is unit transportation cost at the arcs and unit handling cost at hubs. To satisfy the service commitment constraint and capture for connectivity issues (see Figure 12) three hub windows are defined at which sorts can occur. Two algorithms, a Lagrangian Relaxation and implicit enumeration algorithm with ε-inequality are used to solve the flow routing problem.

Lin & Chen (2004) considers the integrated flow routing and fleet scheduling problem of an air-ground express carrier. Clusters of depots are created, in which each such cluster contains a hub location. Afterwards, secondary fleet routes are derived to transport packages between depots and hubs, and primary fleet routes are derived for transportation between hubs. Given primary and secondary fleet routes, flow routes are assigned to these fleet routes such that the service commitment constraint is satisfied. Connectivity issues are solved by assuming hub sorts to take place at given moments in time. Cost taken into consideration is fixed fleet cost, fleet transportation cost and location handling cost. Furthermore, it is assumed that there is insufficient demand under tight time restrictions to fill up vehicles or aircrafts, so that only one vehicle or aircraft can be dispatched on each fleet route.

Lin (2008) considers the integration of flow routing and fleet scheduling in a network which may contain stopovers and directs. Cost components taken into account are fixed fleet cost, fleet transportation cost, balancing cost, and location handling cost. Again, fleet routes are derived and each such route can be performed by one vehicle or aircraft. Compared to their work in Lin & Chen (2004), no clustering operation is performed although the model will still assign a depot location to one hub. However, the hub used for inbound operations can differ from the hub used for outbound operations, though all inbound (outbound) flow will use the same route to (from) the hub location. Hub sorts are presented to deal with connectivity issues in order to satisfy service commitment. A feasible fleet route plan is determined and afterwards flow routes are derived.

A comparison of the discussed literature on the important design aspects can be found in Table 8. In this table, “NDtrad” and “NDnew” refers to the models that are discussed in the next section.

4.4 Modelling

The modelling presented in this section solves the network design and fleet scheduling problem in two steps. Firstly, a tactical network design model is run to derive flow routes. The tactical network design models that are used are discussed in Section 4.4.1. Two models are proposed, the first model is a traditional model that discounts economies of scale on inter-hub flow routing, and will be used for benchmarking. The second model is new and includes fleet scheduling characteristics in network design. In order to compare the results, a fleet scheduling heuristic is solved to determine the network cost (Section 4.4.2). The fleet scheduling heuristic uses the flow routes found by one of the network design models. However, the heuristic can also be applied on existing routes of an express service provider. Afterwards, a balancing model is run to derive the repositioning of vehicles. Final output of the model is fleet schedules and network cost. Figure 14 gives an overview of the modelling approach.

Table 8: Literature comparison on general characteristics, design configuration, and cost components.

	General			Design configuration								Cost components					
	strategic or tactic	mode of transport	hub network	direct routes	non-hub assignment	stopovers (at non-hubs)	max nr. of hub touches	hub capacity	hub sort window	consolidation effect	service commitment	fixed hub cost	variable hub cost	labour cost	fixed transp. cost	unit transp. cost	balancing cost
NDtrad	tactical	road	multiple	no	multiple	no	2	uncapacitated	no	discount inter-hub flows	yes	no	no	yes	yes	yes	no
NDnew	tactical	road	multiple	yes	multiple	no	unlimited	capacitated	no	based on fleet size	yes	no	yes	yes	yes	yes	yes
(Alumur & Kara, 2008)	strategic	road	multiple	no	single	no	3	uncapacitated	no	discount inter-hub flows	yes, by a time bound	yes	yes	no	yes	yes	no
(Armacost, et al., 2004) [(2002)]	tactical	air	multiple	no	single	yes	1	capacitated	yes	based on fleet size	yes	no	yes ([no])	yes	yes	yes	yes
(Aykin, 1995) [(1994)]	strategic	road	multiple	yes	multiple	no	2	uncapacitated [(capacitated)]	no	discount from/to and inter-hub flows	no	no ([yes)]	no	no	no	yes	no
(Barnhart & Schneur, 1996)	tactical	air	single	no	single	yes	1	uncapacitated	yes	based on fleet size	yes	no	no	yes	yes	yes	yes
(Kara & Tansel, 2003)[(2001)]	strategic	road [air]	multiple	no	single	no	2	uncapacitated	no	discount inter-hub time	yes, by a time bound	no*	no*	no*	no*	no*	no*
(Kuby & Gray, 1993)	tactical	air	single	no	single	yes	1	uncapacitated	no	based on fleet size	yes	no	no	yes	yes	yes	no
(Lin, 2001)	tactical	road	multiple	no	multiple	no	2	capacitated	yes, multiple	based on fleet size	yes	no	yes	no	yes	yes	no
(Lin C-C, 2008)[(2004)]	tactical	road/ air	multiple	yes [(no)]	single	yes	unlimited	uncapacitated	yes, multiple	based on fleet size	yes	no	yes	no	yes	yes	yes [(no)]
(Melkote & Dashkin, 2001)	strategic	road	multiple	no	single	no	2	capacitated	no	none	no	yes	no	no	yes	yes	no
(O'Kelly, 1987)[(1996)]	strategic	road	multiple	no	single	no	2	uncapacitated	no	discount inter-hub flows	no	no	no	no	no	yes	no
(O'Kelly, 1992)	strategic	road	multiple	no	single	no	2	uncapacitated	no	discount inter-hub flows	no	yes	no	no	no	yes	no
(O'Kelly & Bryan, 1998)	strategic	road	multiple	no	multiple	no	2	uncapacitated	no	(dependent) discount inter-hub flows	no	no	no	no	yes	yes	no
(O'Kelly, et al., 1996)	strategic	road	multiple	no	multiple	no	2	uncapacitated	no	discount inter-hub flows	no	no	no	no	no	yes	no
(Podnar, et al., 2002)	strategic	road	multiple	yes	multiple	yes	unlimited	uncapacitated	no	discount flows above threshold	no	no	no	no	no	yes	no
(Tan, et al., 2007)	strategic	road	multiple	no	single	no	2	uncapacitated	no	discount inter-hub time	yes, by a time bound	no*	no*	no*	no*	no*	no*
(Yaman, et al., 2008) [(2007)]	strategic	road	multiple	no	single	no [(yes)]	2	uncapacitated	no	discount inter-hub time	yes, by a time bound	no*	no*	no*	no*	no*	no*

*model does not minimise on cost

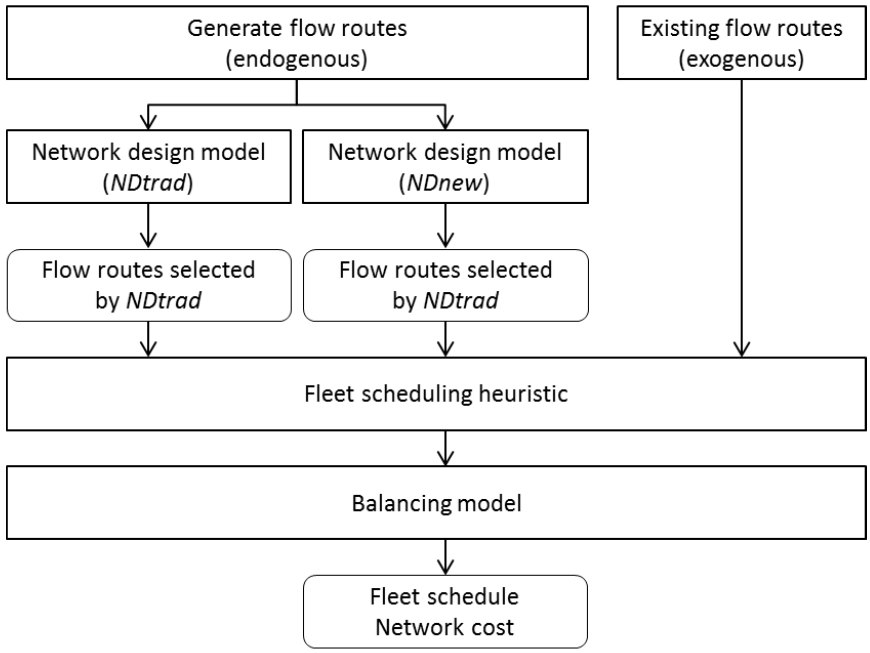


Figure 14: Overview of the modelling approach, resulting in flow routes, a fleet schedule, and the corresponding network cost.

4.4.1 Network design model

The network design starts with a set of locations L containing hub locations $H \subset L$ and depot nodes $N \subset L$. Without loss of generality it is assumed that each location is either a hub location or a depot, i.e. $H \cap N = \emptyset$. Depot i offers services s to customers guaranteeing a delivery time of packages received at the depot before the collection cut -off time c_{is}^{o} ; in order to satisfy the service, the package has to be delivered at the destination depot j before the delivery cut-off time c_{js}^{d}. It is assumed that services between depots are only offered to the customers if their service requirements can be met. That is, the available time between collection cut-off time and delivery cut-off time of the od-service has to be larger than the driving time between these locations. The total flow of packages of service s from depot i to depot j is denoted by f_{ijs}.

Recall that we assumed that there is only one vehicle type available to transport flows. The capacity of this vehicle is equal to v units of flow. Note that vehicle capacity and flow need to be expressed in the same unit (e.g. weight, volume, packages, etcetera). Vehicles move via the arcs A of the network; the start location of an arc is denoted by $s_a \in L$ and the end location is denoted by $e_a \in L$. The distance of arc $a \in A$ is given by d_a.

Drivers' legislations should be taken into account when determining the drivers' cost, since maximum driving times and prescribed breaks may not be violated. When the driving time between two locations exceeds the maximum driving time of one driver, a second driver is required and this cost has to be incorporated. We follow the European Regulations (EUR-lex, 2006) that prescribe a maximum driving time of 9 hours and an uninterrupted break of no less than 45 minutes, after a driving period of 4.5 hour. The total cost of a vehicle moving via arc a is denoted by C_a and includes vehicle transport cost and (second) drivers' cost. All required cost information is available.

For each pair of depots (i, j) with a positive flow $\sum_s f_{ijs} > 0$, routes[7] are generated. A route $r \in R$ is created via the arcs a of the network; the parameter u_{ra} equals 1 if route r uses arc a and 0 otherwise. Since route r starts at depot location i and ends at a depot location j it can only be used to satisfy services of the corresponding pair of depots. Besides, the route can only be used for service s of the depot-pair if it can leave depot i after the collection cut-off time (c_{is}^{o}) and arrives at depot j before the delivery cut-off time (c_{js}^{d}) taking transport time and hub sorting time into account. This results in a parameter p_{ijsr} that equals 1 if a route can be used to serve service s of od-pair (i, j) and 0 if it cannot. See Figure 15.

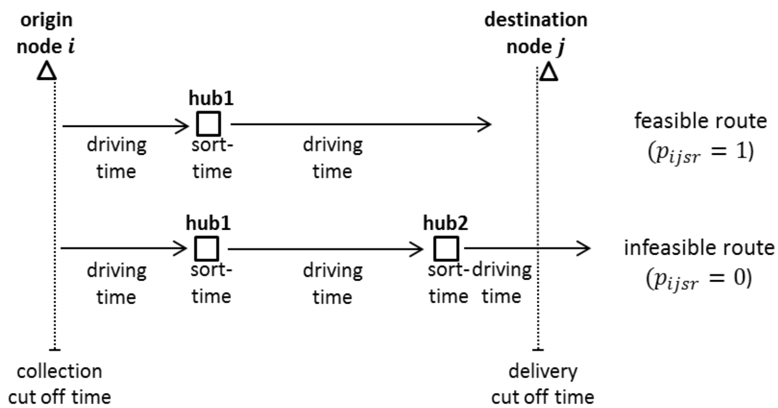


Figure 15: Route generation: routes that start after the collection cut-off time and arrive at the destination depot before the delivery cut-off time are feasible; all other routes are infeasible and not considered in the network design models *NDtrad* and *NDnew*.

A *direct route* $i \rightarrow j$ is a route that only uses a depot-depot arc (i.e. $s_a, e_a \in N$). It is not allowed to pass depots other than the origin and destination depot of the route, i.e. a route $i \rightarrow j \rightarrow k$ with $i, j, k \in N$ is not allowed. A *single hub route* $i \rightarrow h_1 \rightarrow j$ is a route that uses a single depot-hub arc (i.e. $s_a \in N$ and $e_a \in H$) and a single hub-depot arc (i.e. $s_a \in H$ and $e_a \in N$). A *multiple hub route* $i \rightarrow h_1 \rightarrow h_2 \rightarrow j$ can pass more than one hub and uses one depot-hub arc, one or more hub-hub arcs (i.e. $s_a, e_a \in H$) and one hub-depot arc. Routes are categorised by their number of hub touches n $(n = 0,1,2,3, \ldots)$; a route with n hub touches is referred to as type H_n route. Note that H_0 refers to a direct route, H_1 refers to a single hub route and $H_n, n > 1$ denotes multiple hub routes. The modeller can indicate which routes should be taken into account in the model. In general, when we say that we include H_n-routes, all possible routes with n hub touches are generated for each service of an od-pair. However, note that routes that cannot meet the requirements of an od-service are not included in the set of routes that are fed into one of the models.

It is assumed that each service of an od-pair can be satisfied by at least one of the routes generated. If none of the hub routes is able to meet the service requirements, the flow has to be routed directly from origin depot to destination depot. This flow that has to be routed directly

[7] Note: what we refer here as a route, is actually a path as introduced in Chapter 1. The timing information becomes available after fleet scheduling.

because of tight time constraints is denoted by f_{ijs}^{D} and can be determined in a preprocessing phase of one of the network design models presented in the next sections. If some services of an od-pair (i,j) have to be routed directly while others can be routed via a hub route, it is possible to allow these services to use this direct route as well. In this case, either all flow of these services can be routed via this direct route or only part of the flow can use this route. This is discussed in more detail in the sections below. The remaining flow for which a route has to be determined by one of the network design models is denoted by f_{ijs}^{R} and equals $f_{ijs} - f_{ijs}^{D}$. An overview of the parameters is given below.

L	set of locations, index l
$N \subset L$	set of depots, index i, j
$H \subset L$	set of hub locations, index h
S	set of services, index s
c_{is}^{o}	origin cut-off of depot i service s
c_{js}^{d}	destination cut-off of depot j service s
f_{ijs}	total flow from depot i to depot j service s
f_{ijs}^{D}	flow from depot i to depot j of service s that has to be routed via a direct route
f_{ijs}^{R}	flow from depot i to depot j of service s that is routed via hubs
A	set of arcs, index a
s_a	start location of arc a
e_a	end location of arc a
d_a	distance of arc a
v	capacity of a vehicle
C_a	cost of one vehicle v moving via arc a
R	set or routes, index r
u_{ra}	1 if route r uses are a and 0 otherwise
p_{ijsr}	1 if od-pair (i,j) can use route r for service s and 0 otherwise

Sections 4.4.4.1. and 4.4.1.2 present the network design models that are used to determine the flow routes.

4.4.1.1 Network design model: traditional model

This section discusses a traditional model of the tactical network design problem of an express service provider. We call this model *traditional* since it reflects aspects that generally are incorporated in network designs seen in the literature (see Table 8). However, two aspects differ from traditional models, namely service commitment and the depot-hub assignment. We chose to satisfy the service commitment constraint here, since service commitment has highest priority in the express business, and network design that does not satisfy this restriction is of no value to the express company. Moreover, the new network design also has to satisfy this requirement, so that a comparison can only be made if this requirement is also incorporated in the traditional model. Due to this requirement and the short time to satisfy services, it is not always possible to connect a depot to only one hub, and hence we allow multiple assignments from depots to hubs. Unit transport cost of a vehicle moving via arc a follow from dividing the cost of one vehicle moving via that arc by the capacity of the vehicle, i.e. $\frac{1}{v}C_a$. To incorporate economies of scales on inter-hub flow routing, a factor α_a is included such that $\alpha_a \leq 1$ for inter-hub arcs and $\alpha_a = 1$ for non-hub arcs.

As described above, it is possible that some flow has to be routed directly because of tight time constraints. It is possible that some flow of an od-pair has to be routed directly while other services of the od-pair can be satisfied by a hub route. In that case, we assume that all flow of the od-pair is routed directly. The parameters f_{ijs}^D and f_{ijs}^R are updated accordingly. Note that either f_{ijs}^D or f_{ijs}^R is equal to 0.

The network design model chooses one route for each service s of od-pair (i,j) with a positive flow f_{ijs}^R. The variable x_{ijsr} equals 1 if od-service (i,j,s) uses route r. The flow conservation constraint can now be modelled as

$$\sum_{r\in\{R\}} p_{ijsr}x_{ijsr} \quad = \quad 1 \qquad \forall i,j \in N, s \in S, f_{ijs}^R > 0$$

The total costs of the network design can be formulated as

$$\sum_{i,j\in\{N\};s\in\{S\};r\in\{R\}} \sum_{a\in\{A\}} \left(\frac{1}{v}\alpha_a u_{ra} C_a\right) f_{ijs}^R x_{ijsr}.$$

An overview of the model and additional parameters and variables is given below. This network design model is referred to as “NDtrad”.

Parameters

α_a discount factor on arc a for economies of scale

Variables

x_{ijsr} 1 if od-pair (i,j) uses route r for service s, 0 otherwise

NDtrad-model

$$\min \quad \sum_{i,j\in\{N\};s\in\{S\};r\in\{R\}} \sum_{a\in\{A\}} \left(\frac{1}{v}\alpha_a u_{ra} C_a\right) f_{ijs}^R x_{ijsr}. \tag{4.1}$$

$$\text{s.t.} \quad \sum_{r\in\{R\}} p_{ijsr}x_{ijsr} \quad = \quad 1 \qquad \forall i,j \in N, s \in S, f_{ijs}^R > 0 \tag{4.2}$$

$$x_{ijsr} \in \{0,1\} \qquad \forall i,j \in N, s \in S, r \in R \tag{4.3}$$

Note that the model enumerates on all feasible routes, selecting the least cost route for each od-pair and service type combination. The traditional model is formulated in this way so that it can be compared to the model formulation of the new network design model presented in the next section.

4.4.1.2 Network design model: new model

Instead of incorporating a scaling factor for economies of scales, an upper bound on economies of scales can be obtained by determining the minimum number of vehicles required to transport the flows. This network design model selects a route for each service of an od-pair; the routes that can be selected need to satisfy the service requirements of the corresponding service of the od-pair. Since each chosen route is feasible, the model results in a minimum number of vehicles to transport the flows. If time constraints are tight, more vehicles are needed to transport the flows. In case of loose time constraints, the number of vehicles determined by the network design model is sufficient to transport the flows. The model therefore results in an upper bound on achievable economies of scale.

If some flow of an od-pair must be routed directly because of tight time constraints the remaining capacity on the used vehicles are available for transporting flow of the od-pair that could be routed via a hub route (i.e. f^R_{ijsr}). However, this flow only uses this direct route if there is enough time for consolidation. The parameters f^R_{ijsr} and f^D_{ijsr} are updated accordingly. Note that f^R_{ijsr} and f^D_{ijsr} can be larger than 0 at the same time.

The network design model again chooses one route for each service s of od-pair (i,j) with $f^R_{ijsr} > 0$. As in Section 4.4.1.1 the variable x_{ijsr} is used to denote that od-pair (i,j), service s uses route r. The flow conservation constraint is again modelled as

$$\sum_{r\in\{R\}} p_{ijsr} x_{ijsr} \quad = \quad 1 \qquad \forall i,j \in N, s \in S, f^R_{ijs} > 0.$$

The total vehicle capacity on each arc has to be sufficient to transport the flow using that arc. By y^R_a we denote the number of vehicles needed to transport the flows f^R_{ijs} via arc a. This results in the constraint

$$\sum_{i,j\in\{N\};s\in\{S\};r\in\{R\}} u_{ra} f^R_{ijs} x_{ijsr} \quad \leq \quad v y^R_a \qquad \forall a \in A.$$

The parameter $\bar{y}^R_a$ denotes the number of vehicles required to transport direct flows, and equals $\lceil f^D_{ijs}/v \rceil$. The required number of repositioning vehicles moving via arc a is denoted by y^B_a. The vehicle balancing constraint now becomes

$$\sum_{a\in\{A|S_a=l\}} (\bar{y}^D_a + y^R_a + y^B_a) \quad = \quad \sum_{a\in\{A|S_e=l\}} (\bar{y}^D_a + y^R_a + y^B_a) \qquad \forall l \in L.$$

In practice, the amount of flow that can pass through a hub is limited to the capacity of the hub. We assume that hub h can handle at most Q_h units of flow. Note that it is never optimal to handle flows more than once in a hub, so that the restriction of capacitated hub locations is non-restrictive if $Q_h \geq \sum_{ijs} f^R_{ijs}$. Since routes are generated, it is known if hub h is passed by a route r; this is denoted by the parameter q_{rh} that equals 1 if route r uses hub h and is equal to 0 otherwise. The hub capacity constraint is modelled as

$$\sum_{i,j\in\{N\};s\in\{S\};r\in\{R\}} f_{ijs}^{R} q_{rh} x_{ijsr} \quad \leq \quad Q_h \qquad \forall h \in H.$$

The total costs are the sum of the variable hub cost and the cost of vehicles moving via the arcs of the network. The hubs that are passed by a route are known so that the variable cost of one unit of flow using route r can be derived. This cost is denoted by C_r^H. Some express service providers subcontract vehicle movements (a discussion of subcontracting can be found in (Krajewska & Kopfer, 2009)). As a result, repositioning vehicle movements are sometimes bought at a lower rate when subcontractors can use the movement for other purposes. Now, the total costs of the network follow as (with repositioning vehicles discounted by a factor $\gamma \leq 1$)

$$\sum_{r\in\{R\}} C_r^H f_{ijs}^R x_{ijsr} + \sum_a C_a(\bar{y}_a^D + y_a^R + \gamma y_a^B).$$

An overview of the model and additional parameters and variables is given below. This network design model is referred to as “NDnew”.

Parameters

γ	discount factor of repositioning vehicles moving via arc a
Q_h	maximum amount of flow which can pass through hub h
q_{rh}	1 if route r uses hub h and 0 otherwise
C_r^H	variable hub cost of using route r
$\bar{y}_a^D$	number of vehicles moving via arc a to transport flows f_{ijs}^D

Variables

x_{ijsr}	1 if od-pair (i,j) uses route r for service s, 0 otherwise
y_a^R	number of vehicles moving via arc a to transport flows f_{ijs}^R
y_a^B	number of repositioning vehicles moving via arc a

NDnew-model

$$\min \quad \sum_{r\in\{R\}} C_r^H f_{ijs}^R x_{ijsr} + \sum_{a\in\{A\}} C_a(\bar{y}_a^D + y_a^R + \gamma y_a^B) \qquad (4.4)$$

$$\text{s.t.} \quad \sum_{r\in\{R\}} p_{ijsr} x_{ijsr} = 1 \qquad \forall i,j \in N, s \in S, f_{ijs}^R > 0 \qquad (4.5)$$

$$\sum_{i,j\in\{N\};s\in\{S\};r\in\{R\}} u_{ra} f_{ijs}^R x_{ijsr} \leq v y_a^R \qquad \forall a \in A \qquad (4.6)$$

$$\sum_{a\in\{A|S_a=l\}} (\bar{y}_a^D + y_a^R + y_a^B) = \sum_{a\in\{A|S_e=l\}} (\bar{y}_a^D + y_a^R + y_a^B) \qquad \forall l \in L \qquad (4.7)$$

$$\sum_{i,j\in\{N\};s\in\{S\};r\in\{R\}} f_{ijs}^R q_{rh} x_{ijsr} \leq Q_h \qquad \forall h \in H \qquad (4.8)$$

$$x_{ijsr} \in \{0,1\} \qquad \forall i,j \in N, s \in S, r \in R \qquad (4.9)$$

$$y_a^R, y_a^B \in \mathbb{N}_0 \qquad \forall a \in A \qquad (4.10)$$

4.4.2 Fleet scheduling heuristic

The network design models of Section 4.4.1 determine a route for each od-service (i,j,s). The fleet scheduling heuristic presented in this section determines the real number of vehicles required to transport the flow and derives fleet schedules. Post-processing determines repositioning cost once fleet schedules are created. This heuristic is used to test relative performance of *NDtrad* and *NDnew*.

The fleet scheduling heuristic uses the following rules for vehicle departures via arc a:

- a vehicle can depart if its departure is critical for the service requirements of one of the od-services for which flow is loaded on the vehicle;
- a vehicle can depart if all flow to be transported via arc a is available;
- a vehicle can depart if it has a full vehicle load.

The heuristic uses an *event list E* of possible departures. All flow is assumed to be available at the origin depot at the collection cut-off time. These collection cut-off times c_{is}^{o} are the first possible departure times that are added to the event list. The second group of possible departure times that are added are the so-called *critical departure times*. All flow has to be available at the destination depot before the delivery cut-off time of its corresponding service. Since the flow route x_{ijsr} of od-service (i, j, s) is known, the latest departure time at each arc in the route can be determined via backwards computing, by starting at the delivery cut-off time taking into account transport time and sorting time. The latest departure time of od-service (i, j, s) at location s_a of arc a is called the critical departure time, denoted by t_{ijsa}^{crit}. The last group of events, the availability time of flow at hub locations, results from the arrival of a vehicle: flow that arrives at a hub location needs further transport and this transport is possible after sorting. The time at which arrived flow can leave the hub location is called the *availability time*; the availability time of od-service (i, j, s) to be further transported via arc a is denoted by t_{ijsa}^{avail}. Every time a vehicle is scheduled to depart, flow can be transported. If there is more flow available than the capacity of the vehicle, flow with the earliest critical departure time at the corresponding departure location has highest priority to use this vehicle and is transported to the next location.

The heuristic starts with the first event time e in the event list. Then it checks: (1) do there exist arcs with flow having reached a critical departure time? If there is some flow, vehicles are scheduled to depart and the flow is transported. If there is none, the next question is: (2) do there exist arcs for which all flow has arrived? If there exists such an arc, vehicles are scheduled and flows are transported to the next location. Finally, it is checked: (3) do there exist arcs at which a full vehicle can be loaded? If this is the case, a vehicle departs and the flow arrives at the next location. Afterwards, e is removed from the event list and the next event in the event list is considered. The heuristic terminates when all flow has arrived at its destination depot.

Note that vehicle departures are caused because of flow arrivals in step (2) and (3): in step (2), the last flow to be transported via an arc has arrived, and in step (3), flow arrives resulting in a full vehicle load. However, in step (1), a departure does not need to be instigated by the arrival of flow. It might be that some flow is waiting for other flows to arrive, but at some moment (the critical time) it can no longer wait. Then, a vehicle is scheduled to transport this flow. However, this vehicle could already leave at the moment the last flow, which is transported by this vehicle, arrived. This time is referred to as time e^*. Note that e^* can be the availability time of the flow that causes the critical departure, or the availability time of other flow that arrived at this arc (after the arrival of the flow causing the critical departure). The vehicle is scheduled at time e^*, which can be earlier than the critical event time (and therefore also earlier than the current event time, i.e. $e^* < e$). If this vehicle arrives at a hub location, the flow needs further transport. Recall that this flow becomes available for further transport at time t_{ijsa}^{avail}. Now notice that it is possible that the flow becomes available before the current event time e, because the vehicle might have been scheduled before this time. Since this could impact vehicles already scheduled between t_{ijsa}^{avail} and e, these vehicle departures need to be reconsidered. Therefore, the heuristic turns back in time so that the event list restarts at $e = e^*$. All vehicle departures scheduled after e^* are cancelled and the flow is pushed backwards accordingly. This step is referred to as a *reset* of the event list. An example of a reset is given in Figure 18 in the Appendix and an overview of the heuristic is given in Figure 16.

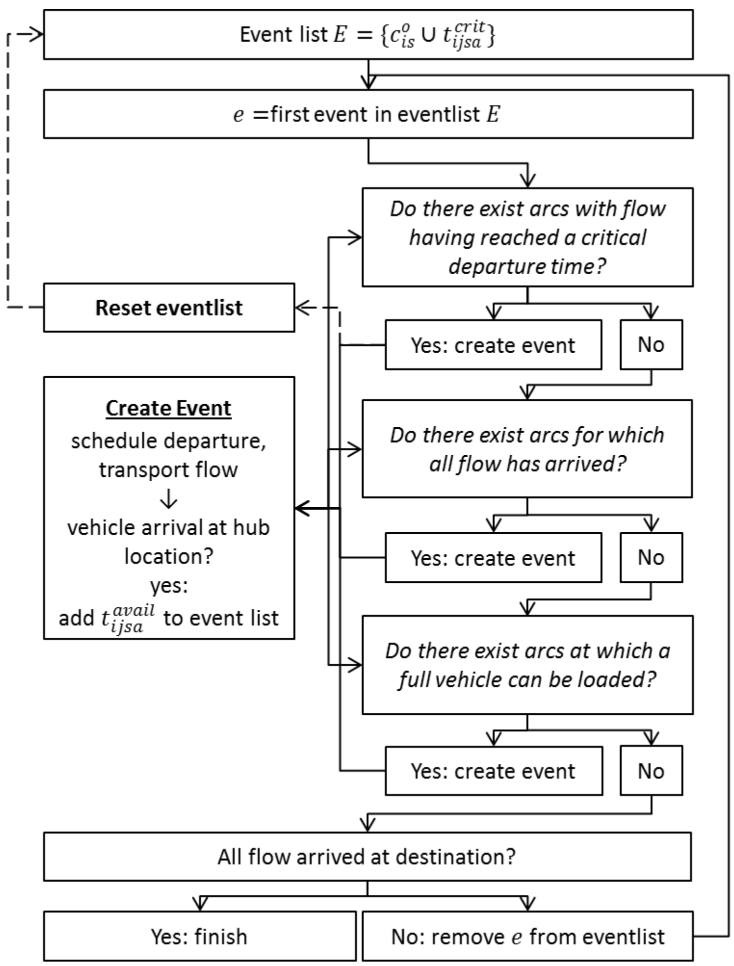


Figure 16: Fleet scheduling heuristic based on three main decisions (critical departure, all flow arrived, and full vehicle load).

Vehicle balancing cost is not determined in the network design model *NDtrad*; the network design model *NDnew* determines vehicle balancing cost but due to tight time constraints, the real number of vehicles required to transport the flows can be higher. Therefore, vehicle balancing cost needs to be determined.

The fleet scheduling heuristic results in a number of vehicles moving via each arc; this is denoted by $\bar{y}_a$. The required number of balancing vehicles moving via each arc (i.e. y_a^B) needs to be derived. The balancing constraint becomes

$$\sum_{a\in\{A|S_a=l\}}(\bar{y}_a + y_a^B) \quad = \quad \sum_{a\in\{A|S_e=l\}}(\bar{y}_a + y_a^B) \qquad \forall l \in L.$$

Then, the repositioning cost needs to be minimised so that the resulting model becomes as following

Parameters
γ discount factor of repositioning vehicles moving via arc a
$\bar{y}_a$ number of transportation vehicles moving via arc a

Variables
y_a^B number of repositioning vehicles moving via arc a

Balancing model

$$\min \quad \sum_a \gamma C_a y_a^B \tag{4.11}$$

$$\text{s.t.} \quad \sum_{a\in\{A|S_a=l\}}(\bar{y}_a + y_a^B) \quad = \quad \sum_{a\in\{A|S_e=l\}}(\bar{y}_a + y_a^B) \qquad \forall l \in L \tag{4.12}$$

$$y_a^B \in \mathbb{N}_0 \qquad \forall a \in A \tag{4.13}$$

4.4.3 Remarks

Cut-off times are used to determine the available time to transport flows in the network design models. However, the moment of transport is not taken into account (i.e. the possibility to combine flows in time is not checked during network design). Note that including time moments in modelling flow routes would dramatically increase the number of routes possible, since no assumptions are made on departure moments at hub locations. Therefore, the fleet scheduling heuristic is run to estimate the possibility to combine flows in time; if flows cannot be combined, additional vehicles are required. The resulting cost after fleet scheduling are therefore in general higher than the cost found after network design, but the difference in cost depends on the routes given to the heuristic. As a result, a suboptimal solution of the network design model could give lower cost after fleet scheduling than the optimal solution of the network design model.

4.5 Computational study

The research was inspired by practical considerations of an express carrier. This section presents the results of the models applied to modified instance data of the express service provider.

Data instances were created for two geographies (Geography A and B) which are based on actual countries. Data instances define the number and location of depots and hubs, the cut-off times and the services offered between depots. An overview of the characteristics of the geographies can be found in Table 9. Geography A has 31 depots and Geography B has 37 depots; in both geographies, four hubs are available. Note that the largest distance and the average distance between od-pairs are larger in Geography B than in Geography A. There is a positive flow between each pair of depots in Geography B while there are only 750 od-pairs with a positive flow in Geography A. In the latter, there is no flow between 180 od-pairs. However, the total flow in Geography A is larger than in Geography B.

In both geographies, two services were defined: services s_1 and s_2. In both geographies, 80% of the total flow is of service s_1 and the remaining 20% has service s_2. Packages with an s_1-service are available at the origin depot before 20:00h and have to arrive at the destination depot before 07:00h in both geographies. In Geography A, s_2-packages are available at the origin depot before 21:00h and have to arrive at the destination depot before 06:00h. In Geography B, s_2-packages are available at the origin depot before 20:00h and have to arrive at the destination depot before 07:00h two days later. Note that s_2 is a faster service in Geography A, but a slower service in Geography B.

For both geographies, three cases are constructed varying in the demand for each service. For every geography, the total demand is the same in each case, however the geographical spread differs. The first case, *Case1*, describes the situation in which there is equal demand for each service (i.e. f_{ijs_1} is the same for each (i, j) and f_{ijs_2} is the same for each (i, j)). *Case2* considers the situation with moderate fluctuations in demand for each service. Finally, *Case3* describes the situation with strong differences in demand for each service. The latter can be interpreted as a situation where a group of depots represents net senders generating large flows to be transported to net receivers, while there is only small demand vice versa. In both geographies, 15 depots are indicated as net senders and the remaining depots are net receivers. Note that the equality in demand as assumed in *Case1* does not often occur in reality. Fluctuations as created in the second and third set of cases, occurs regularly, where the intensity of these fluctuations is subject to the country in scope. For example, fluctuations as in *Case2* are found in countries as Germany or France. On the other hand, intensive fluctuations as in *Case3* occur for example in Turkey, where the economic activity in the West is much stronger than the economic activity in the East.

Balancing movements are discounted by 10% of a transport movement cost (i.e. $\gamma = 0.90$) and variable hub cost is €0.05 per kg (i.e. $C_r^H = 0.05$). Hub capacities are assumed to be non-restrictive and 60 minutes of sorting time is needed at each hub location.

Section 4.5.1 compares the results found by the traditional and the new network design model. Sensitivities of the new network design model are discussed in Section 4.5.2. Section 4.5.3 concludes with some remarks on computation time, routes generated, and vehicle types used.

4.5.1 Comparison of the results

This section compares the results found by using *NDtrad*-routes or *NDnew*-routes. The *NDtrad* model is used for benchmarking; we therefore chose $\propto$ in each case such that the *NDtrad* model gave lowest cost (considered values of $\propto$ are $\propto = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$) to ensure a fair comparison. Of course, in practice $\propto$ is input. The resulting values of $\propto$ are 0.8 in all cases of

Geography *A*, $\propto = 0.00$ in *Cases B1* and *B2*; *Case B3* shows lowest cost when $\propto = 0.20$. The results of *NDnew* are found by including direct routes, 1-hub routes, 2-hub routes, and 3-hub routes (i.e. we generate all feasible routes visiting respectively at most zero, one, two, or three hub locations). 4-hub routes are excluded since none of these routes could meet the service requirements. Note that only 1-hub and 2-hub routes are included in the traditional model, since we assumed that only two hub touches are allowed. Since we do not exclude any hub-hub arc, this assumption is non-restrictive, because the triangle inequality indicates that a 2-hub route is always cheaper than a 3-hub route.

Table 10 shows for each instance the results found by using *NDtrad*- or *NDnew*-routings and the percentage difference between them. The first column in the table displays the total cost; afterwards the total number of vehicle movements and the corresponding distance driven are presented. The last two columns of the table show the average number of hub touches per kg of flow, and the average hub throughput respectively. A cost breakdown in the three cost components (i.e. balancing cost, variable hub handling cost, and transport cost) is displayed in Figure 19, on the left, and the division of flow over the kind of routes can be found in Table 11 and Figure 19.

Comparison of the cost shows that in all cases cost can be reduced by using the routes proposed by the new network design. On average, total cost can be reduced with 5.0% in Geography *A* and with 1.1% in Geography *B*. Recall that our new cost function includes three cost components: transport cost, variable hub cost, and balancing cost. The main cost savings are achieved by reducing variable hub cost and balancing cost: on average, 17.5% of the variable hub handling cost can be saved and 18.1% of the balancing cost can be saved. Transport cost decreases in half of the cases, while it increases in cases *A1*, *B1*, and *B2*.

We see a decrease in number of vehicle movements in all cases except for *CaseB1*. On average, we see that the total number of vehicle movements is reduced by 2.1%. The resulting total distance driven is on average reduced by 2.1%.

The changes in cost and vehicle movements are caused by changed flow routings. In general, the new routings show less hub touches. On average, the traditional model results in 18.4% of direct routes, 55.5% of 1-hub routes, and 26.1% 2-hub routes; the new model results in 25.1% direct routes, 63.0%1-hub routes, 10.8%2-hub routes, and 1.2%3-hub routes (see Table 11). The resulting average number of hub touches and the average hub throughput are reduced with 17.5% when using the new network design model. Less hub routing immediately implies less hub handling cost. Balancing cost is reduced due to the inclusion of direct routes.

The results of this section are referred to as the 'base' results of the *NDnew* model.

4.5.2 Sensitivities NDnew-routings

This section shows the sensitivities of the results of *NDnew*-routings to kind of routes, hub capacities, variable hub cost, transport cost, and balancing cost. The results are compared based on cost and route usage. The "Cost overview"-figures that are shown, present cost divided in transport cost, variable hub cost, and balancing cost. The "Routing overview"-figures show the percentage of flow that is routed per kind of route (i.e. direct route, 1-hub route, 2-hub routes, or 3-hub routes).

4.5.2.1 NDnew-routings – Sensitivity to kind of routes

Here we show the results found by *NDnew*-routing, when varying the kind of routes. The kind of routes that can be included are direct routes, 1-hub routes, 2-hub routes, and 3-hub routes. Note that the 'base' results assume that all kind of routes may be used. The results are averages over the cases and are displayed in Figure 20.

When we only allow direct routes, the cost is €712,653 which is about 2.6 times[8] as high as the cases in which hub routes are allowed. However, note that including hub routes reduces transport cost but on the other hand leads to increasing handling cost and balancing cost. Including 2-hub routes reduces the cost with 4.7% when direct routes are allowed and with 6.2% when direct routes are not allowed. Apparently, inter-hub flow routing is profitable. However, the inclusion of 3-hub routes has only small impact: on average, 0.4% of the cost can be saved. When we compare the results of allowing direct routes to the results in which direct routes are not allowed, it can be seen that the cost on average are 1.1% higher if direct routes are forbidden. Besides, transport cost remains almost at the same level but both variable hub handling cost and vehicle balancing cost are higher.

Concluding, hub routing leads to a large cost saving. However, in the used geographies, the cost effect of including 3-hub routes is only small when 1-hub and 2-hub routes are included. When allowing direct routings together with hub routings, cost can be further reduced due to lower handling cost and balancing cost. This implies that it is favourable to use more direct routes than in the traditional model.

4.5.2.2 NDnew-routings – Sensitivity to hub capacities

This section shows the results of limiting hub capacities. Recall that the 'base' results assume non-restrictive hub capacities. The results are averages over the cases and are displayed in Figure 21.

Since maximum hub capacities were unknown, these capacities were first derived as follows. In each case, the maximum hub capacity was derived as the maximum hub throughput found for one of the hubs in the 'base' results. We refer to this as Q_{base}. Afterwards, fractions θ of these hub capacities were taken as bound on the maximum hub capacity, i.e. $Q_h = \theta Q_{base}$ for all hubs h. This was done for fractions $\theta = 0.20, 0.40, 0.60, 0.80$ and 1.00 of the maximum hub capacity.

It is easily understandable that cost decreases when θ increases. This is caused by decreasing transport cost. However, balancing cost and variable hub cost increase. In case only 20% of the hub capacity is available, cost is 37% higher than in the 'base' results. However, the cost level stabilizes as soon as 60% of the hub capacity is available. Apparently, the flow can be spread more evenly over the hub locations so that cost savings from consolidation can still be achieved. This can also be seen from the routings: only an additional 5.1% of the flow is routed directly in case $\theta = 0.60$ when comparing to the 'base' results.

Concluding, it can be said that there is a strong relation between hub routing and hub capacities. In particular, when hub capacities are restrictive, more direct routing is used, resulting

[8] If we compare to a situation in which only medium size vehicles are used in a network with only directs, this factor is slightly lower, namely 2.2.

in higher cost. However, cost stabilizes as soon as 60% of the hub capacity is available. Note that fixed hub cost is left out of consideration, although this is likely to depend on capacity.

4.5.2.3 NDnew-routings – Sensitivity to variable hub cost

The sensitivity of the results was tested against varying variable hub cost C_r^H. The results are averages over the cases and are displayed in Figure 22.

From the results of varying hub cost it can be concluded that there is a strong relation between variable hub cost and flow routing: when variable hub cost increases, hub routing is less preferred. As a result, required hub capacities are strongly impacted by variable hub cost: there is an overcapacity of 29.8% when variable hub cost increases to €0.25 (compared to €0.05).

4.5.2.4 NDnew-routings – Sensitivity to transport cost

Finally, the influence of the varying transport cost C_a was investigated. Again, the results are averages over the cases and are displayed in Figure 23.

The model was run for increasing transport cost C_a. These cost were increased to $1.5C_a$, $2.0C_a$, $2.5C_a$, $3.0C_a$, and $3.5C_a$. Compared to the 'base' results (i.e. $1.0C_a$), we see that all cost components increased. It was expected that increasing transport cost would result in more hub routing, since consolidation of flows reduces the number of vehicle movements. But the results show that only a small percentage of the flow uses more hub routing: when transport cost are 3.5 times as high, direct routing decreases only with 2.8% and 1-hub routing decreases with 1.7%. That means that direct routing and 1-hub routing is still attractive even when transport cost is high. Finally note that hub capacities need to increase: when we multiply C_a with 3.5, we see an increased hub capacity of 8.3% compared to Q_{base}.

Two effects need to be taken into consideration to explain the impact of transport cost on hub routing: firstly, more hub routing results in higher total transport cost due to increasing variable hub cost and the detour of flow; second, more hub routing results in lower total transport cost as a result of more consolidation. Note that varying transport cost has both a positive and negative effect on the total transport cost. The results show that there is only small impact of increasing transport cost on hub routing. That indicates that the cost savings of more consolidation are only small compared to the increasing cost due to the detour of flow routing and the increasing variable hub cost. As a result, direct and 1-hub routing is still profitable even when transportation cost strongly increases.

4.5.2.5 NDnew-routings – Sensitivity to discounting of balancing cost

Finally, the sensitivity of the results was tested against varying discounting of balancing cost γ. In the 'base' results, it is assumed that balancing cost is discounted with 10% ($\gamma = 0.90$). The results are again averages over the cases and are displayed in Figure 23.

The results are shown for decreasing discounting of balancing cost. Considered values of γ are 0.00, 0.25, 0.50, 0.75, and 1.00. From the cost it can be seen that cost increases when γ increases, but the largest differences are caused by increasing balancing cost. However, there is a small increase in transport cost when balancing cost increases. Besides, observe that more flow is routed directly. This can be explained as follows. Suppose some flow can be routed via a hub route, consolidated with other flows so that it does not generate additional transport cost (it only generates variable hub cost). However, due to imbalances of flow, a repositioning vehicle is required to drive between origin depot and destination depot. On the other hand, the flow can be routed directly, but in that case a vehicle has to be scheduled resulting in additional transport cost. No repositioning vehicle is required in that case. It is obvious that the first option, routing via hub locations, is preferred when repositioning is (strongly) discounted; the last option is preferred otherwise.

Concluding, it can be said that there is only a small dependency between direct routing and balancing cost.

4.5.3 Remarks

The cases in this section are run on an Intel(R) Core 2 CPU, 2.00 GHZ, and 2.00 GB of RAM. The traditional network design model is run in a few seconds, an overview of run times of the new network design is given in Figure 17 and shows that 70% of the cases were run within 15 minutes of computation time. 25% of the cases take longer than one hour of computation time, but this is mainly the result of restricting the capacity of flows that can pass through a hub. For all these cases, the best solution found within an hour is within one percent of the optimal solution. The fleet scheduling heuristic is on average run in eleven minutes.

The number of routes that are generated a priori, depends on the type of routes that are included. The maximum number of H_n-routes that are included is equal to $S * N * (N-1) * \prod_{h=1}^{n}(H-n+1)$. However, no routes are generated if there is no flow between an od-pair, and also routes that do not meet the service restrictions are excluded. The maximum number of routes to be generated in the new network design of Geography A equals 76,260 but the actual number of routes fed into the model equals 6,615. In Geography B, the maximum number of routes equals 109,924 and the actual number of routes equals 50,14. For the traditional network design, these numbers (maximum versus actual) are 29,760 and 4,772 in Geography A and 42,624 and 22,030 in Geography B.

In the introduction we discussed the use of medium size vehicles towards hubs and large size vehicles between hubs. Some tests are done based on this assumption, to validate the comparison between the traditional and new network design model. All tests that are done, show that the cost savings achieved by using medium trucks in the new network design are at least as high as the cost savings achieved by the traditional network design. This could also be expected, since the new network design uses less consolidation compared to the traditional network design.

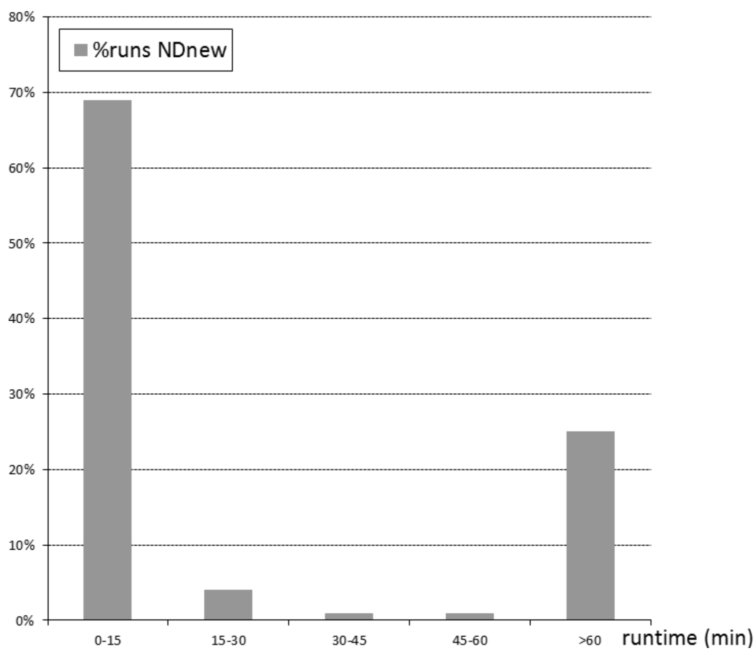


Figure 17: Resulting runtime *NDnew*.

This section showed the results of applying the network design models on modified data of an express service provider. The next section states our conclusions and recommendations for further research.

4.6 Conclusions and directions for further research

This chapter proposed a new tactical network design model for express service providers.

The model was tested on modified instance data of an express service provider. Test cases were created for two geographies, and for each such geography three test cases were generated varying in the geographical spread of demand. In each test case, cost savings could be achieved if routes proposed by the new network design were used instead of the traditional routes. The first geography showed an average cost saving of 5.0% and the second geography showed an average cost saving of 1.1%. The main cost savings were achieved by reduced variable hub cost and reduced balancing cost.

Furthermore, the sensitivity analyses showed that the cost was 2.6 times as low when consolidation is used to transport flows compared to only direct driving. These savings can still be achieved even when only 60% of the hub capacity is available. Of all cost components, variable hub cost influences hub routing the most: increasing variable hub cost leads to a strong decreasing hub routing. Higher balancing cost leads only to a small increase in direct routing; higher transport cost results in a small increase in (multiple) hub routing.

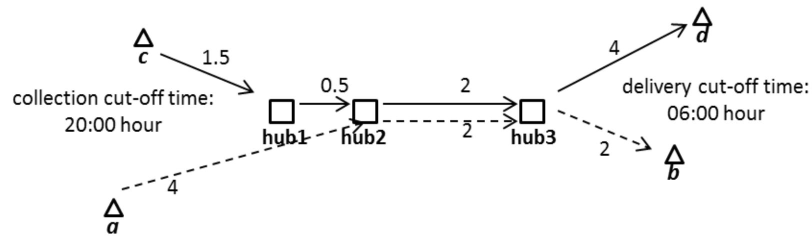
This chapter showed cost reductions by including fleet scheduling characteristics in the tactical network design of express service carriers. The models were tested on modified instance data of two geographies. The results of the geographies differ; it should be further investigated how characteristics of geography affect the routings. Furthermore, it would be interesting to consider the network design of multiple countries at once (e.g. the network design of Europe). We suggest a column generation approach to reduce computation time when scaling up to data instances of this size.

Final fleet schedules were derived after flow routing. Further cost reductions are expected if fleet schedules and flow routings are determined simultaneously. This research focussed on the tactical network design of express service providers. More research needs to be done to show the impact of fleet scheduling on the strategic network design of express service providers.

4.7 Appendix

Critical times:

(a, b) at a $= 20{:}00$,
hub2 $= 00{:}00$,
hub3 = 03: 00
(c, d) at $c = 22{:}00$,
hub1 $= 00{:}30$,
hub2 $= 02{:}00$,
hub3 = 05: 00



Note: numbers above arcs denote driving times,
hub processing time is 1 hour.

Assume: flows: (a, b) and (c, d),
no full vehicle load at arrival at one of the hubs,
not all flow has arrived at arrival at one of the hubs.

Vehicle movements created, without a reset of the eventlist:

start		end	timing (hours)			transport of
depot a	→	hub2	20:00	→	00:00	(a, b)
depot c	→	hub1	20:00	→	21:30	(c, d)
hub2	→	hub3	00:00	→	02:00	(a, b)
hub1	→	hub2	00:30	→	01:00	(c, d)
hub2	→	hub3	02:00	→	04:00	(c, d)
hub3	→	depot b	03:00	→	06:00	(a, b)
hub3	→	depot d	05:00	→	06:00	(c, d)

Flow available:

location	time	flow
depot a	20:00	(a, b)
depot c	20:00	(c, d)
hub2	00:00	(a, b)
hub1	22:30	(c, d)
hub3	03:00	(a, b)
hub2	02:00	(c, d)
hub3	05:00	(c, d)

Vehicle movements created, with a reset of the eventlist:

start		end	timing (hours)			transport of
depot a	→	hub2	20:00	→	20:00	(a, b)
depot c	→	hub1	20:00	→	21:30	(c, d)
~~hub2~~	~~→~~	~~hub3~~	~~00:00~~	~~→~~	~~02:00~~	~~(a, b)~~
hub1	→	hub2	22:30	→	23:00	(c, d)
hub2	→	hub3	00:00	→	02:00	$(a, b), (c, d)$
hub3	→	depot b	03:00	→	06:00	(a, b)
hub3	→	depot d	05:00	→	06:00	(c, d)

Flow available:

location	time	flow
depot a	20:00	(a, b)
depot c	20:00	(c, d)
hub2	00:00	(a, b)
hub1	22:30	(c, d)
~~hub3~~	~~03:00~~	~~(a, b)~~
hub2	02:00	(c, d)
hub3	05:00	$(a, b), (c, d)$

Figure 18: Example of a reset of the eventlist; a vehicle movement from hub2 to hub3 can be saved by the reset.

	Geography A	Geography B
nr. of depots	31	37
nr of hubs	4	4
nr. of od-pairs	750	1,332
largest distance	909	1,428
average distance	372	626
total flow	750,000	350,000
%$\boldsymbol{s_1}$flow	80	80
%$\boldsymbol{s_2}$flow	20	20
$\boldsymbol{c^o_{is_1}}$	day1-20:00	day1-20:00
$\boldsymbol{c^d_{is_1}}$	day2-07:00	day2-07:00
$\boldsymbol{c^o_{is_2}}$	day1-21:00	day1-20:00
$\boldsymbol{c^d_{is_2}}$	day2-06:00	day3-07:00

Table 9: overview characteristics of Geographies *A* and *B*.

	Total cost (€)	Total nr. of movements	Total distance (kms)	Average nr. hub touches per kg
CaseA1 - NDtrad	224,636	359	155,511	1.03
CaseA1 - NDnew	219,307	350	155,425	0.89
CaseA1 - difference	**-2.40%**	**-2.50%**	**-0.10%**	**-13.80%**
CaseA2 - NDtrad	228,519	368	157,661	1.07
CaseA2 - NDnew	216,042	353	153,637	0.86
CaseA2 - difference	**-5.50%**	**-4.10%**	**-2.60%**	**-19.90%**
CaseA3 - NDtrad	236,317	392	165,476	1.04
CaseA3 - NDnew	219,351	363	156,540	0.86
CaseA3 - difference	**-7.20%**	**-7.40%**	**-5.40%**	**-16.90%**
CaseB1 - NDtrad	307,510	582	292,951	1.11
CaseB1 - NDnew	306,751	596	293,830	0.99
CaseB1 - difference	**-0.20%**	**2.40%**	**0.30%**	**-10.80%**
CaseB2 - NDtrad	314,175	606	296,546	1.28
CaseB2 - NDnew	308,065	599	298,396	0.82
CaseB2 - difference	**-1.90%**	**-1.20%**	**0.60%**	**-36.30%**
CaseB3 - NDtrad	310,906	597	299,770	0.93
CaseB3 - NDnew	307,796	596	297,650	0.87
CaseB3 - difference	**-1.00%**	**-0.20%**	**-0.70%**	**-7.00%**
CaseGeography A - difference	**-5.00%**	**-4.70%**	**-2.70%**	**-16.90%**
CaseGeography B - difference	**-1.10%**	**0.40%**	**0.10%**	**-18.00%**
CaseOverall - difference	**-3.00%**	**-2.10%**	**-1.30%**	**-17.50%**

Table 10: comparison of the results for the different cases that are run, regarding cost, movements, distance and average hub touches.

	H0 (%)	H1 (%)	H2(%)	H3 (%)
NDtrad – average	18.4	55.5	26.1	0.0
NDnew – average	25.1	63.0	10.8	1.2

Table 11: comparison of the results for the different cases that are run, for categorised routes based on hub touches.

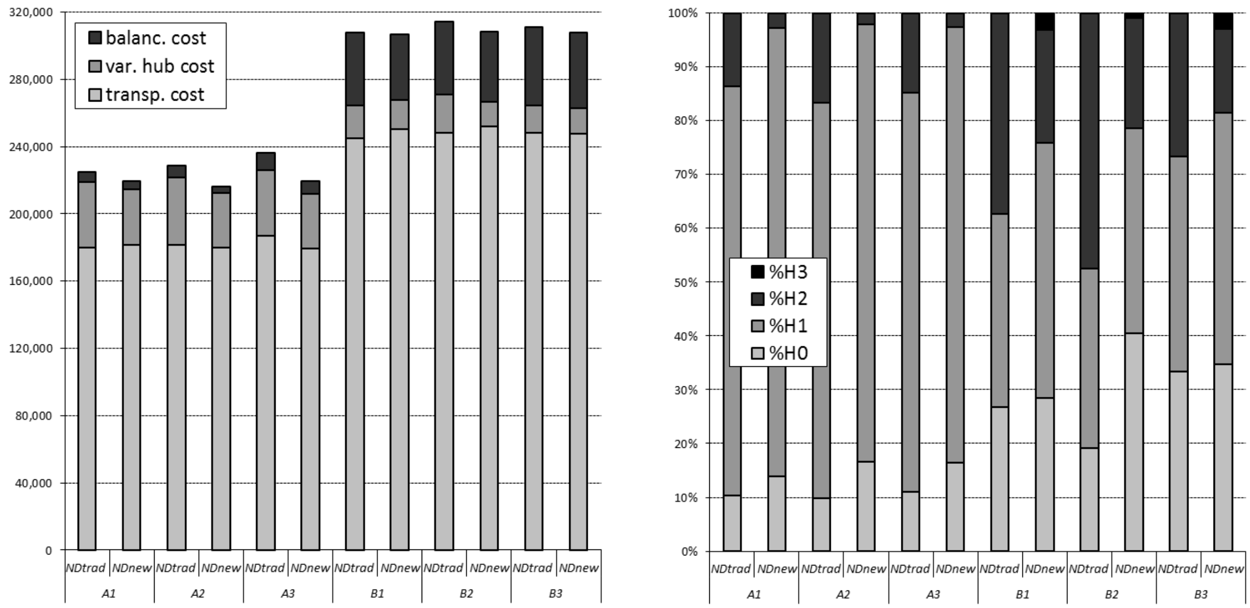


Figure 19: Comparison of the results for *NDtrad* and *NDnew* based on total cost and route categorisation for the different cases run per Geography.

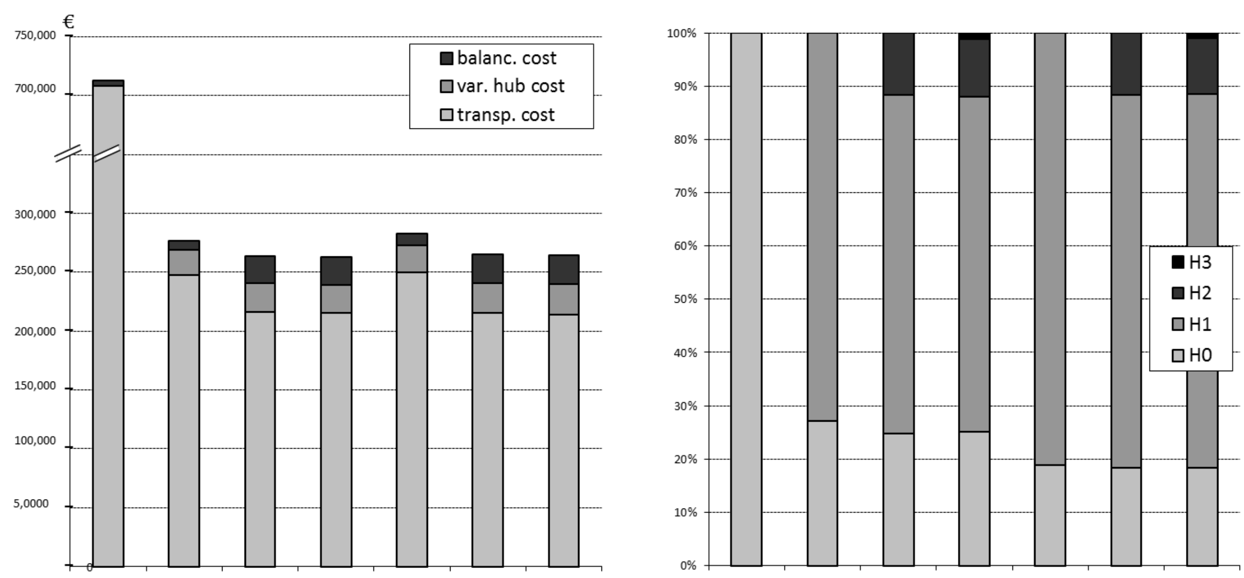


Figure 20: Sensitivity kind of routes - comparison of the results for *NDtrad* and *NDnew* based on total cost and route categorisation.

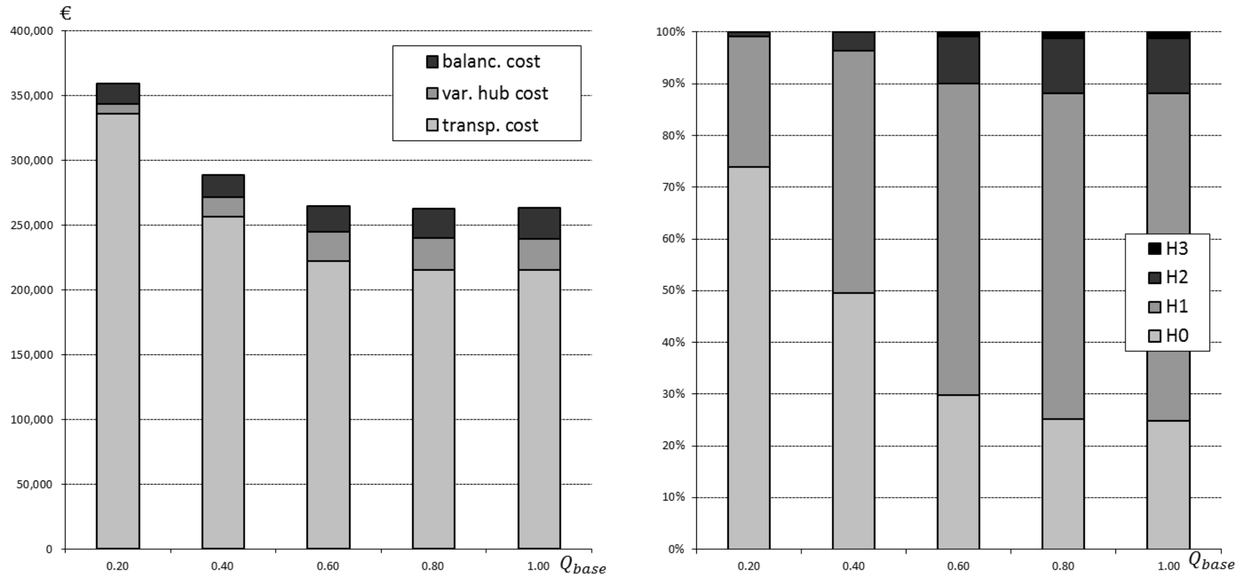


Figure 21: Sensitivity hub capacities - comparison of the results for *NDtrad* and *NDnew* based on total cost and route categorisation.

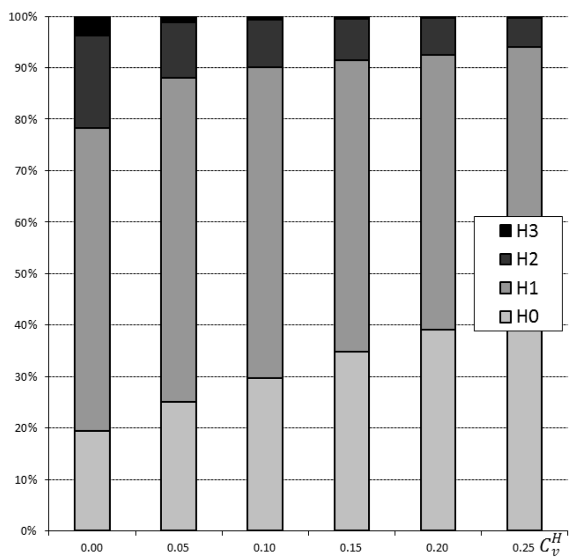


Figure 22: Sensitivity variable hub cost - comparison of the results for *NDtrad* and *NDnew* based on route categorisation.

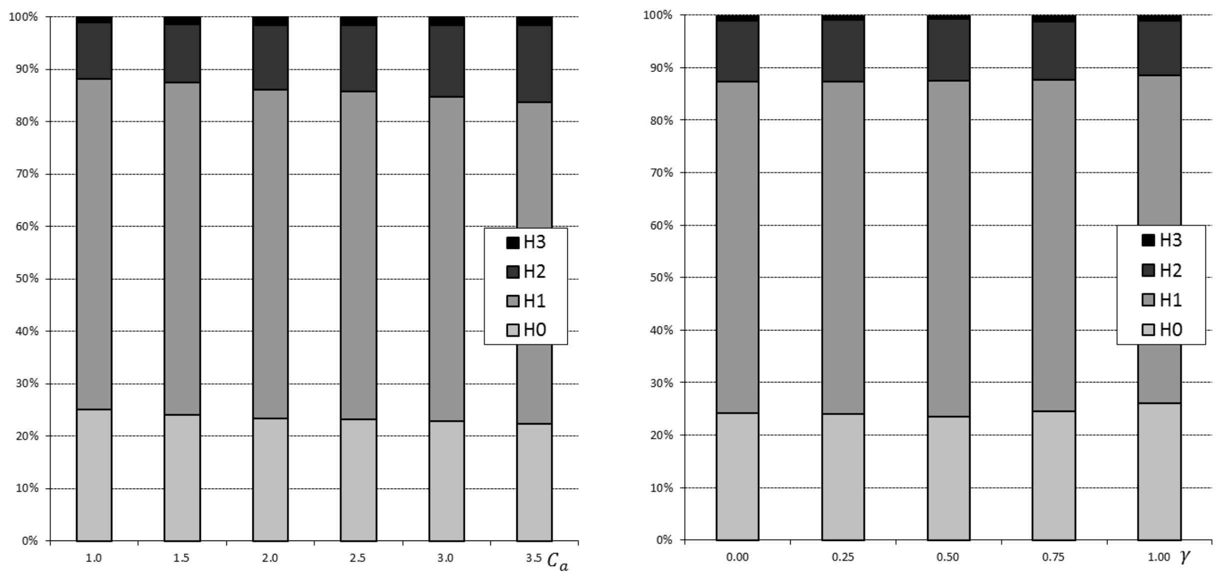


Figure 23: Sensitivity transport and balancing cost - comparison of the results for *NDtrad* and *NDnew* based on route categorisation.

5 Scheduling movements in the network of an express service provider

This chapter is based on the following journal paper:

Scheduling Movements in the Network of an Express Service Provider

Louwerse, I; Mijnarends, J; Meuffels, I; Huisman, D; Fleuren, H
Flexible Services Manufacturing Journal (2013), DOI 10.1007/s10696-013-9171-x

5.1 Abstract

Express service providers manage shipments from senders to receivers under strict service level agreements. Such shipments are usually not sufficient to justify a single transportation, so it is preferred to maximise consolidation of these shipments to reduce cost. The consolidation is organised via depots and hubs: depots are local sorting centres that take care of the collection and delivery of the packages at the customers, and hubs are used to consolidate the transportation between the depots. A single transportation between two locations, carried out by a certain vehicle at a specific time, is defined as a movement. In this chapter, we address the problem of scheduling all movements in an express network at minimum cost. Our approach allows imposing restrictions on the number of arriving/departing movements at the hubs so that sufficient handling capacity is ensured. As the movement scheduling problem is complex, it is divided into two parts: one part concerns the movements between depots and hubs; the other part considers the movements between the hubs. We use a column generation approach and a local search algorithm to solve these two sub-problems, respectively. Computational experiments show that by using this approach the total transportation costs are decreased.

5.2 Introduction and research motivation

An express service provider offers time-guaranteed delivery of packages (i.e. documents, packages, or pieces of freight) to its customers. This concerns collection of the packages at the customers, transportation via the network of the express service provider, and delivery of the packages at the receivers. Local sorting centres called depots take care of the collection and delivery of the packages to the senders and receivers. It is not cost efficient to transport all packages directly from their origin depot to their destination depot, as usually the number of depots is large and the amount of packages that needs to be transported between each pair of depots is relatively small. Therefore, hubs are used to consolidate the transport between depots. At the hubs, packages are unloaded from the arriving vehicles, sorted, and loaded on the departing vehicles.

Express service providers offer multiple service types to their customers. The service types differ in the date and time at which packages need to be at their destination depot. We define a flow as all packages between a pair of depots with the same service type. The origin cut-off time denotes the latest time a flow should be available at its origin depot, while the destination cut-off time denotes the latest time a flow has to arrive at its destination depot. For each flow, a route describing the sequence of locations visited when transporting the flow from its origin depot to its destination depot, has to be determined. The cut-off times determine which routes are possible for each flow.

Several vehicle types are used to transport the packages along the routes. The vehicle types differ in speed, capacity and cost per kilometre. The preferred vehicle type depends on the number of packages that needs to be transported and the time available to transport the packages. In an express network, a number of vehicles move between the different types of locations according to a given schedule. One such transportation between two locations done by a certain vehicle type on a given time is called a movement. The challenge of scheduling all the movements in an express network at minimum cost is referred to as the movement scheduling problem.

The movement scheduling problem considered in this chapter originates from one of the largest express service providers in the world, as it was expected that its current scheduling process could be improved. The current process is not supported by any modelling tools and movement schedules are improved locally, based on practical experiences. This research aims to develop a model that solves the movement scheduling problem.

For practical reasons, it might be desired to restrict the changes in the number of arrivals and departures at the hub locations. To ensure that the new movement scheme can be implemented on short term, the current hub capacities should be sufficient to handle the new movement scheme. In this research, we present additional constraints to take this into account. However, our goal is to find the optimal movement scheme, which can be obtained by relaxing these restrictions.

The solution approach that is presented in this chapter divides the movement scheduling problem into two parts, such that for each part a well-known mathematical formulation can be developed. The first part concerns scheduling the movements between depots and hubs; we present a composite variable formulation for this problem and the column generation algorithm used to solve it. The second part concerns the problem of scheduling the movements between hubs, which is formulated as a network loading problem and solved with a local search algorithm.

The remainder of this chapter is organised as follows. In Section 5.3 we discuss literature related to our problem, and we introduce the overall solution approach. The mathematical formulations are given in Section 5.4. Section 5.5 describes the solution approach in more detail. In Section 5.6 the results of several experiments based on data of an express service provider are presented. The conclusions and some recommendations for further research are mentioned in Section 5.7.

5.3 Literature review and solution framework

The movement scheduling problem has similarities to the network design problem, which concerns the decision how to move the packages from their origins to their destinations and how to route the vehicles. The movement scheduling problem can be seen as an extension of the service network design problem on three levels. First of all, the time component has to be integrated in the model. To accomplish this, the planning horizon is discretised and the physical network is replicated for each period. Such dynamic networks are presented in (Farvolden & Powell, 1994) and (Pedersen, et al., 2009). The second feature that has to be included is the general empty balancing of vehicles, also called repositioning. We refer to (Dejax & Crainic, 1987) for an overview of models for empty vehicle repositioning. The third feature concerns the design variables, which are the variables concerning the allocation of capacity to the arcs. In most network design models, these are binary variables. However, in our model several vehicles can be assigned to an arc, resulting in integer variables. Crainic (2000) describes a model with integer variables modelling discrete choice design decisions.

To our knowledge, there is no literature presenting problem structures corresponding to our problem. However, several papers deal with problems that have similarities with the problem described in this chapter. Such problems are presented in (Armacost, et al., 2002), (Armacost, et al.,

2004) and (Barnhart & Schneur, 1996). Two simplifications are made in these papers. Firstly, the presented models concern a one-hub structure. Second, sort start and end times at the hub are introduced as the latest time packages can arrive at and leave from the hub, respectively. As a result of these sort start and end times two separate problems can be solved: one to determine the depot-hub routes and one to determine the hub-depot routes. Usually, the networks of the express service provider that motivate our research concern more than one hub without a distinct hub sort as vehicles arrive and depart within the same time window, so that the presented models are insufficient.

Other problems related to our problem are the multi-depot vehicle routing problem (MDVRP) and the pickup and delivery problem with time windows and transhipment (PDPTWT). Baltz et al. (2007) describe the MDVRP as the problem of assigning customers to depots and finding optimal service tours, given a set of depots and a set of customers who must be served from these depots. A polynomial time approximation scheme (PTAS) exists for the MDVRP, but because it cannot efficiently solve larger real-world instances, heuristics are usually used (see for instance (Cordeau, et al., 1997) for a tabu search heuristic). The aim of the pickup and delivery problem with time windows (PDPTW) is to construct optimal vehicle routes to serve transportation requests between pickup and delivery locations. Mitrović-Minić & Laporte (2006) investigates the usefulness of transhipment points for the PDPTW, allowing transportation requests to be served by two vehicles, and present a heuristic to solve the problem.

The research presented in this chapter concerns a multi-hub express network in which we determine the routings of the flows, the vehicles to be used and the schedules of the resulting movements. As we want to construct routes that go via several hubs, it is not realistic to use hub windows. Besides, we want to take into account a heterogeneous vehicle fleet, different service types, exact timing of movements and packages, and routes via several depots and hubs. Because of these additional aspects, formulating our problem in one model results in a comprehensive formulation which will be too difficult to solve. Based on the structure of the problem, we divide it into several sub-problems. For each sub-problem a well-known mathematical formulation can be developed.

5.3.1 Sub-problems

Recall that for each flow a route has to be determined and that the cut-off times determine the routes that are possible. Two types of routes can be distinguished: direct routes and non-direct routes. In case the cut-off times are very tight and any possible route via at least one hub is infeasible, a direct route between the origin and destination depot of the flow is used. Large flows between depots can also be a reason to use a direct route. In any other case, the packages can be consolidated and routed via one or several hubs.

Non-direct routes visit one or more intermediate locations, of which at least one is a hub. Non-direct routes are divided in routes from depots to hubs (DH-routes), routes between two hubs (HH-routes), and routes from hubs to depots (HD-routes). An important difference between the DH/HD-routes and the HH-routes is that all packages in a flow need to use the same DH- and HD-route, but they can use different HH-routes. For each flow exactly one DH- and one HD-route have to be chosen. When determining the HH-routes, also the exact moment in time the packages move between the hubs is important. Because of these differences, the problem of finding the non-direct routes is divided into two parts: the first part concerns movements from depots to hubs and vice versa (DH/HD-problem) and the second part concerns the movements between the hubs (HH-problem). As flows are routed via sequences of DH-, HH- and HD-routes, the problems are connected to each other. An example of a network with hubs, depots and routes is shown in Figure 24.

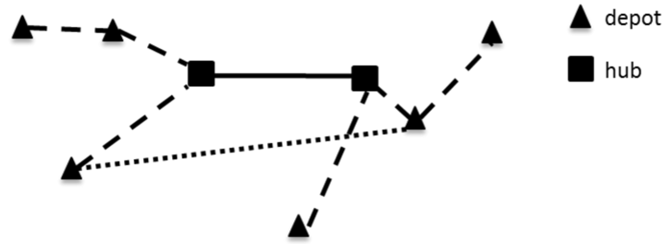


Figure 24: Example of a network. HH-routes are represented by solid lines, DH/HD-routes by dashed lines and direct routes by dotted lines.

A DH-route starts at a depot, possibly visits other depots to load packages, and ends at a hub. During this process, packages can only be loaded and not unloaded, as sorting of packages occurs at the hubs. Similarly, HD-routes start at a hub and visit one or more depots where packages are unloaded. In this process, packages can only be unloaded. If a DH-route or a HD-route visits several depots, some processing time has to be taken into account at these intermediate depots to load or unload packages. Once a DH-route ends and the packages arrive at the hub, the packages are unloaded, sorted, and loaded on a vehicle that moves either to another hub or to the destination depot. A composite variable approach as described in (Armacost, et al., 2002) is used to formulate the DH/HD-problem as a set partitioning problem. This formulation has the flexibility to handle practical constraints that cannot be easily incorporated in traditional formulations. We solve DH/HD-problem with a column generation algorithm. Barnhart et al. (1998) show that column generation can be successfully applied to set partitioning problems with a large number of variables.

HH-routes are used to route the packages between the hubs. A HH-route visits at least two hubs. The first and last hubs on the HH-route are the hub at which the DH-route ends and the hub at which the HD-route starts, respectively. At each hub along a HH-route it is possible to both unload and load packages. The HH-problem is formulated as a network loading problem. Network loading problems (NLPs) model the design of capacitated networks with no variable flow cost, and with facilities of fixed capacity available to carry flow. For an overview of models and algorithms of network design problems we refer to (Magnanti & Wong, 1984) and (Magnanti & Mirchandani, 1993). Despite the importance of the NLP in a variety of settings (freight transportation, telecommunications industry), the available research on it is limited. The primary focus in the literature has been on the uncapacitated variant. The algorithms and heuristics developed for the uncapacitated problem, as for instance presented in (Holmberg & Hellstrand, 1998), do usually not work very well for the capacitated problem. The gap between the integer solution and the linear programming solution is small for the uncapacitated variant, but not for the capacitated variant as shown by Leung et al. (1990). We will use a local search algorithm to solve the NLP.

To transport the packages via a non-direct route, sequences of DH-, HH- and HD-routes have to be constructed. Two routes can be connected if the end location of the first route and the start location of the second route are the same, and the end time of the first route is before the start time of the second route.

5.3.2 Overall algorithm

Figure 25 shows the overall algorithm used to solve the movement scheduling problem. Firstly, the direct routes are determined. A direct route is used if there are enough packages between two depots to fill a vehicle of the largest type, or if a direct route is the only possibility to let packages arrive at their destination depot in time. In the former case a direct route is optimal, for the latter it is the only possibility to meet the service requirements. Therefore, the direct routes are determined beforehand and not reconsidered later in the solution process. In the next phases of the overall algorithm, consecutively the DH/HD-problem and the HH-problem are solved, and the repositioning cost is determined. In Section 5.5, the column generation used to solve the DH-problem and the local search algorithm used to solve the HD-problem are discussed. The repositioning problem is described in Section 5.3.3.

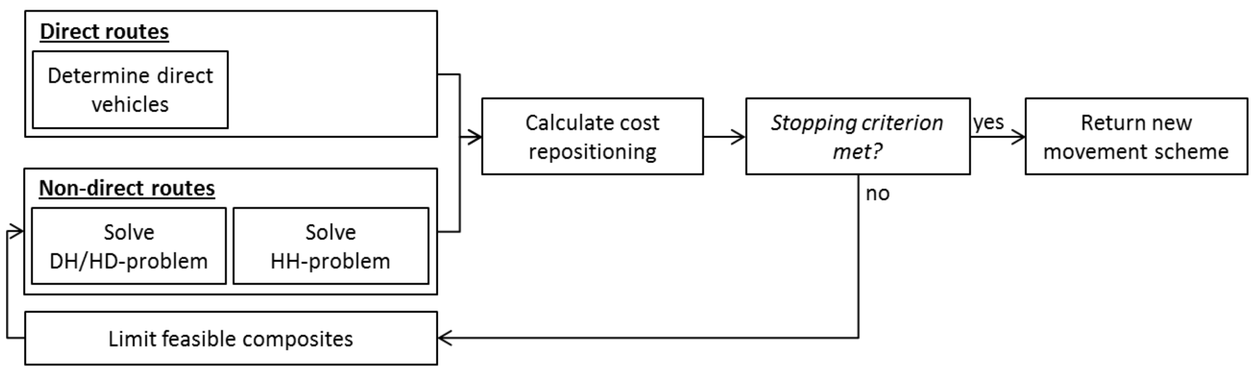


Figure 25: Overall algorithm.

The sub-problems are solved sequentially which means that the solution of one sub-problem depends on the solutions of one or several other sub-problems. Therefore, there has to be some interaction between the sub-problems and an iterative procedure is defined. A stopping criterion is defined and as long as the stopping criterion is not met, a new movement scheme is created. We try to improve the overall solution by adding additional constraints to the DH/HD-problem such that some of the composites used in the current solution are not feasible in the next solution. We decide which constraints to add to the DH/HD-problem based on the solution of the HH-problem. It is relatively expensive to transport the flows that are on a hub movement of a small vehicle type or a hub movement with a low utilisation as the distances between the hubs are relatively large. Therefore, we try to decrease the total transportation cost by restricting these flows to be transported via only one hub.

5.3.3 Repositioning problem

Because the movement schedule is carried out on a daily basis, the number of vehicles that start at each location at the beginning of the planning horizon and the number of vehicles that end at each location at the end of the planning horizon have to be equal. If this is not the case, some vehicles need to be repositioned which results in additional cost. Once the DH/HD-problem and the HH-problem are solved, the repositioning cost has to be determined.

The objective function of the repositioning problem is to minimise the total repositioning cost. The constraints of the model specify that for each location and each vehicle type the number of vehicles repositioned from the location minus the number of vehicles repositioned to the location has to be equal to the surplus of vehicles at this location at the end of the day. This means that the formulation of the repositioning problem represents a flow problem. By solving the corresponding linear programming relaxation, we obtain the optimal solution.

5.4 Mathematical formulations

This section describes the assumptions of the movement scheduling problem and presents the mathematical formulations of the DH/HD-problem and the HH-problem.

5.4.1 Assumptions

First of all, it is assumed that all locations are given and that the function of a location, i.e. whether a location is a depot or a hub, cannot be changed. Second, services are given and as express service providers offer guaranteed services to their customers we assume that these services are feasible with respect to the cut-off times. The third assumption is that the number of vehicles available of each type is infinite. If there would be a limited number of available vehicles of each type, additional restrictions can be added to the model and our solution approach would still be applicable. Fourth, it is assumed that the costs of the different vehicle types are such that it is always cheaper to use one vehicle of a larger type than two vehicles of a smaller type. This assumption does not affect the basic aspects of our solution approach. In case this assumption does not hold, only minor changes are required for calculating the minimal transport cost in the shortest path problem of the column generation algorithm and for determining the initial solution of the local search algorithm. Moreover, the capacity of the hubs can be finite, defined by additional restrictions that impose a maximum number of movements that can arrive at and depart from a hub.

For all packages that are not routed directly, a non-direct route, consisting of a DH-route, a HH-route and a HD-route, needs to be selected. Because of modelling reasons, it is assumed that all packages in a flow have to be transported via the same DH- and HD-route. However, in general this assumption is not restrictive as the total flows at the depots are usually low so that only a few vehicles are used. For the flows between the hubs, the assumptions are less strict as the total flows between the hubs are larger and therefore flows between hubs are often consolidated. Hence, it is assumed that the packages in a flow can use different HH-routes.

To limit the possible number of DH- and HD-routes, there are some restrictions on the depots and hubs that can be on the same route. The maximum number of depots on a DH/HD-route is two. Hence, each DH- and HD-route consists of one or two depots and one hub. Besides, only depots and hubs that are located near each other can be on the same route. These restrictions can be validated as in general the timing in express networks is tight, so that it is unlikely that the DH- and HD-routes can visit more than two depots or that they can go via locations that are not located near each other. These restrictions can easily be relaxed without altering the solution approach. However, this results in a significant increase in the size of the problem and the computation time. For the HH-routes, we do not restrict the number of hubs the packages can visit in a route.

5.4.2 DH/HD-problem

The DH/HD-problem is to assign each flow to a DH- and a HD-route such that all packages arrive at their destination depot in time, feasible combinations of DH- and HD routes are chosen, and the transportation cost is minimised. The problem can be formulated using the composite variable approach as described in (Armacost, et al., 2002). The composites represent a combination of vehicles on a route with sufficient capacity to transport a feasible combination of flows. Compared to general network design problems in which both flow variables and design variables have to be modelled explicitly, the use of composite variables results in a dramatic decrease in problem size as the flow variables are removed as explicit decision variables and embedded in the design variables. Using composite variables, the problem can be modelled as a set partitioning model. Armacost et al. (2002) show that such a model formulation is stronger than the conventional network design model formulation by showing that the integer programming formulations are equivalent and that the composite variable formulation has a stronger linear programming relaxation.

An overview of the notation of the model is given below.

Sets

$L = D \cup H$ set of locations, index i
$H \subset L$ set of hubs, index h
$D \subset L$ set of depots, index j, k
V set of vehicle types, index v
S set of services, index s
P set of DH-composites, index p
Q set of HD-composites, index q

Parameters

r_h maximum number of DH- and HD-movements at hub h
c_p cost of DH-composite p
d_q cost of HD-composite q
m_{phv} number of vehicles v arriving at hub h on DH-composite p
n_{qhv} number of vehicles v departing from hub h on HD-composite q
v_{pjis} 1 if DH-composite p transports flows s from depot j to location i; 0 otherwise
w_{qijs} 1 if HD-composite q transports flows s from location i to depot j; 0 otherwise
e_{pq} 1 if DH-composite p can be used in combination with HD-composite q; 0 otherwise

Variables

x_p 1 if DH-composite p is used; 0 otherwise

y_q 1 if HD-composite q is used; 0 otherwise

Parameters e_{pq} are used to ensure that feasible combinations of DH- and HD-composites are chosen. DH-composite p can be used before HD-composite q if there is enough time available between the end time of composite p and the start time of composite q to transport the packages between the end hub of composite p and the start hub of composite q, and to process the packages at the hub(s). The DH/HD-problem can be formulated as follows.

Mathematical formulation:

min $\sum_{p\in\{P\}}(c_p * x_p) + \sum_{q\in\{Q\}}(d_q * y_q)$ (5.1)

s.t.

$\sum_{p\in\{P\}}(v_{pjis} * x_p) = 1 \quad \forall i \in L, j \in D, s \in S$ (5.2)

$\sum_{q\in\{Q\}}(w_{qijs} * y_q) = 1 \quad \forall i \in L, j \in D, s \in S$ (5.3)

$\sum_{v\in\{V\}}(\sum_{p\in\{P\}}(m_{phv} * x_p) + \sum_{q\in\{Q\}}(n_{qhv} * u_p)) \leq r_h \quad \forall h \in H$ (5.4)

$\sum_{p\in\{P\};q\in\{Q\}}(v_{pjks} * w_{qkjs} * e_{pq} * x_p * y_q) = 1 \quad \forall j,k \in D, s \in S$ (5.5)

$x_p, y_q \in \{0,1\} \quad \forall p \in P, q \in Q$ (5.6)

The objective function (5.1) minimises the cost of transporting the packages between the depots and hubs. Constraints (5.2) and (5.3) specify that one DH-composite and one HD-composite have to be selected for each flow. A maximum on the number of DH- and HD-movements at each hub is described by Constraints (5.4). Constraints (5.5) ensure that for each flow that has to be transported between two depots a feasible combination of a DH- and a HD-composite is selected. Quadratic Constraints (5.5) can be linearised by introducing binary variables z_{pq}, defined for each possible combination of a DH- and a HD-composite. Using these variables, Constraints (5.5) can be replaced by the following set of constraints (5.7) – (5.9):

$\sum_{p\in\{P\};q\in\{Q\}}(v_{pjks} * w_{qkjs} * e_{pq} * z_{pq}) = 1 \quad \forall j,k \in D, s \in S$ (5.7)

$z_{pq} \leq x_p \quad \forall p \in P, q \in Q$ (5.8)

$z_{pq} \leq y_q \quad \forall p \in P, q \in Q$ (5.9)

5.4.3 HH-problem

The hub problem is to transport the flows between the hubs, at minimum cost and given the DH- and HD-route of the flow. We model the hub problem as a network loading problem. The time dimension is explicitly taken into account by using a time-expanded directed graph. Each node in the graph represents a hub in a given time period. Directed arcs are added to link nodes. An arc connecting two nodes representing the same hub represents freight waiting at the hub to be loaded onto an outbound vehicle. An arc connecting two nodes representing different hubs represents a transportation service between two hubs. The set of arcs is divided in several subsets, one for each vehicle type.

Sets

N	set of nodes, index i, j
$N^h \subset N$	set of nodes representing hub h
A	set of arcs, index (i, j)
F	set of flows, index f
S	set of services, index s
P	set of DH-composites, index p
Q	set of HD-composites, index q

Parameters

n_h	maximum number of vehicles at hub h
k^f	total number of packages of flow f
c_{ij}	cost of moving one vehicle over arc (i, j)
q_{ij}	capacity of one vehicle on arc (i, j)

Variables

x_{ij}^f	number of packages of flow f on arc (i, j)
y_{ij}	number of vehicles on arc (i, j)

The origin and destination hub of flow f are represented by $O(f)$ and $D(f)$, respectively. Below, the arc-based formulation of the hub problem is presented.

$$\min \quad \sum_{(i,j)\in\{A\}} \left(c_{ij} * y_{ij}\right) \tag{5.10}$$

$$\text{s.t.} \quad \sum_{j\in\{N\}} \left(x_{ij}^f\right) - \sum_{j\in\{N\}} \left(x_{ji}^f\right) = \begin{cases} k^f & if\ i = O(f) \\ -k^f & if\ i = D(f) \\ 0 & otherwise \end{cases} \quad \forall f \in F \tag{5.11}$$

$$\sum_{f\in\{F\}} \left(x_{ij}^f\right) \leq q_{ij} y_{ij} \quad \forall (i,j) \in A \tag{5.12}$$

$$\sum_{i\in\{N^h\}}\left(\sum_{j\in\{N\backslash N^h\}}(y_{ij}) + \sum_{j\in\{N\backslash N^h\}}(y_{ji})\right) \leq n_h \qquad \forall h \in H \quad (5.13)$$

$$x_{ij}^f \in \mathbb{R}_+ \qquad \forall f \in F, (i,j) \in A \quad (5.14)$$

$$y_{ij} \in \mathbb{Z}_+ \qquad \forall (i,j) \in A \quad (5.15)$$

The objective function (5.10) minimises the total transportation cost. Constraints (5.11) are the flow conservation constraints, and Constraints (5.12) specify that the flow on an arc cannot exceed the capacity of the vehicles on the arc. A maximum number of vehicles that can arrive at and depart from hub h is described by Constraints (5.13). The integrality of the variables x_{ij}^f is ensured by the integrality of the vehicle capacities.

5.5 Solution methods

In this section we discuss the column generation algorithm used to solve the DH/HD-problem and the local search algorithm used to solve the HH-problem. We refer to (Lübbecke & Desrosiers, 2005) and (Aarts & Lenstra, 2003) for a general discussion of topics related to column generation and local search.

5.5.1 Column generation algorithm

The algorithm starts with creating a set of initial columns for the DH/HD-problem. We use the following heuristic to do this. Each flow goes either via the hub closest to the origin depot and the hub closest to the destination depot, via one of these hubs or via the fastest hub (i.e. the hub that results in the earliest arrival time at the destination depot).

We apply column generation to the linear programming relaxation of the DH/HD-problem (5.1-5.6), without Constraints (5.5). Let π_{jis}^*, μ_{ijs}^* and σ_h^* be the optimal values of the dual variables associated with Constraints (5.2), (5.3), and (5.4) respectively. The reduced cost for non-basic composites p and d are:

$$\theta_p = c_p - \sum_{i\in\{L\};j\in\{D\};s\in\{s\}}\left(v_{pjis} * \pi_{jis}^*\right) - \sum_{v\in\{V\};h\in\{H\}}\left(m_{phv} * \sigma_h^*\right) \qquad \forall p \in P \quad (5.16)$$

$$\lambda_q = d_q - \sum_{i\in\{L\};j\in\{D\};s\in\{s\}}\left(w_{qijs} * \mu_{ijs}^*\right) - \sum_{v\in\{V\};h\in\{H\}}\left(n_{qhv} * \sigma_h^*\right) \qquad \forall q \in Q \quad (5.17)$$

The pricing problems of the DH- and HD-problems can be solved separately. The problems are very similar and we only present the pricing problem of the HD-composites in this section. Each flow is represented by nodes in the graph. These nodes are called flow nodes. Besides, a source node and several sink nodes are added. The source node does not represent a physical location, but is added to ensure a start location for each shortest path. The sink nodes represent the hubs. The source is connected to all flow nodes and the flow nodes are connected to two sinks (representing the two nearest hubs). Besides, each flow node is connected to each other flow node that can be in the same composite. This means that each flow node is only connected to nodes representing the same origin depot and a different destination location and/or service type, or nodes representing one of its two nearest depots.

Figure 26 shows a representation of such a directed acyclic graph for the DH-problem for a network with two depot locations (j and k), one service type (s), and one hub (h). The graph can be seen as consisting of two similar sub graphs: an upper graph and a lower graph. In the upper graph, there is a node from the source to each flow node. Besides, each flow node is connected to the other flow node representing the same origin depot and to the hub. In the lower graph, each flow node is connected to the other flow node representing the same origin depot and to the hub. Besides, the flow nodes in the upper graph are connected to the flow nodes in the lower graph representing the other depot.

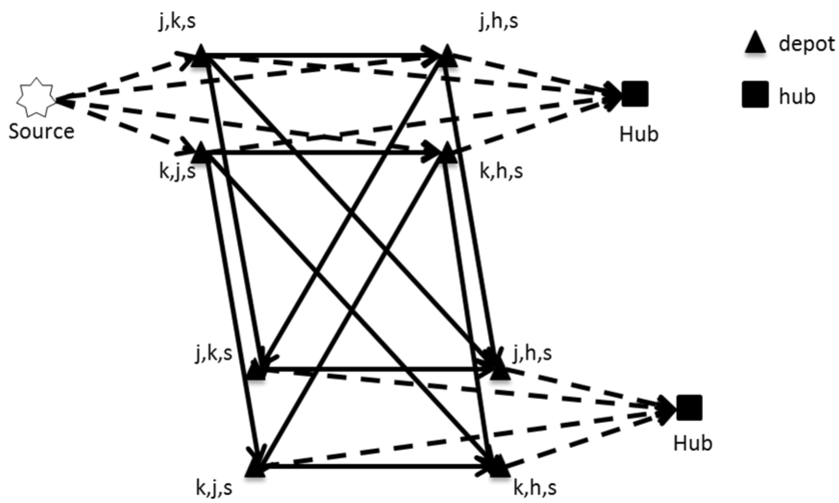


Figure 26: Directed acyclic graph.

Paths in the graph go from the source via one or several flow nodes to a hub node. Each path represents a feasible composite. An approximation of the reduced cost of a composite is obtained as the path cost, if the arc weights are given as follows:

- An arc from the source to a node n representing flow from depot j to location i and service s is labelled with cost $-\pi^*_{jis}$ plus the minimum cost of transporting flow f_{jis}.
- An arc from node n to node n' representing flow from depot j' to location i' and service s' is labelled with cost $-\pi^*_{j'i's'}$ plus the minimum cost of transporting flow $f_{j'i's'}$.
- An arc from node n to a sink node m representing hub h is labelled with cost $-\sigma^*_h$.

Equation (5.17) shows that the reduced cost of a DH-composite consists of the cost of transporting the total flow on the composite, minus the dual variables of the flows on the composite and the dual variable of the hub at which the composite ends. As we do not know beforehand which flows will be on the composite, we cannot take the exact cost of the composite into account in the shortest path problem. To take the cost to some extent into account, we add the minimum cost of transporting the flows to the arc weights. This minimum cost is defined as the cost of a vehicle of the largest type divided by the capacity of this vehicle, and multiplied by the distance between the origin depot of the flow and the nearest hub of this depot.

Including an arc in a composite means including the node at which the arc terminates in the composite. If a path only visits flow nodes in the upper sub graph, the composite contains one depot. If a path visits flow nodes in both the upper and the lower sub graphs, the composite contains two depots. A shortest path problem is solved to find paths with negative reduced cost.

To solve the shortest path problem, we use the single-source shortest path algorithm for directed acyclic graphs as described in (Cormen, et al., 2001). The algorithm produces a shortest path from the source node to every other node in the graph. Note that we are only interested in the shortest paths from the source to the hubs and that the algorithm produces several columns with negative reduced cost. In each iteration, several columns representing a shortest paths between the source and a hub can be added if they have negative reduced cost.

5.5.2 Local search algorithm

Given a solution of the DH/HD-problem, the HH-problem is solved. The origin and destination hub, as well as the arrival time at the origin hub and the departure time from the destination hub of each flow are given by the solution of the DH/HD-problem. We solve the HH-problem with a local search algorithm.

In the initial solution, all packages go directly from their origin hub h^o to their destination hub h^d. The time package p arrives at its origin hub is denoted as α_p, and the latest time it has to leave its origin hub to arrive at its destination hub on time is denoted as β_p. A vehicle is scheduled if there are enough packages available at the origin hub to fill a vehicle of the largest type, or at the time some packages have to leave the origin hub to arrive at the destination hub in time. If there are enough packages to fill a vehicle of the largest type, the packages with the earliest departure times from the destination hub are assigned to this vehicle first. Packages that do not necessarily have to leave the origin hub are scheduled on the movement as long as there is capacity left. The vehicle type chosen is the smallest vehicle type such that the capacity of the vehicle exceeds the number of packages that needs to be scheduled.

The steps used to determine the movements of the initial solution between each pair of origin and destination hub are summarised below.

1. Denote the set P as all packages p between h^o and h^d.
2. Set $t_0 = min_{p\in\{P\}}(\alpha_p)$ and $t_1 = min_{p\in\{P\}}(\beta_p)$.
3. Set the current time $t = t_0$ and the total flow at h^o waiting to be transported $f = 0$.
4. While $t \leq t_1$do:

a. Add all packages p for which $t = \alpha_p$ to f.
b. Check whether f is large enough to fill one or more vehicles of the largest type. Schedule the movement(s) and subtract the packages assigned to the vehicle(s) from f.
c. If there are packages p in f for which $t = \beta_p$, schedule a movement and subtract the packages from f.
d. Set $t = t + 1$.

Solutions in the neighbourhood of the current solution are obtained by performing small modifications to the current solution. We define small modifications as rerouting packages that go from hub h^o to hub h^d in the current solution to going from hub h^o via hub h^i to hub h^d in the new solution. Once packages are rerouted, some movements between the hub pairs h^o- h^d, h^o - h^i and h^i - h^d have to be rescheduled. In the neighbourhood search, we do not consider packages that are scheduled on movements consisting of the largest vehicle type with no capacity left. The reason is that it is always most cost efficient to use the shortest route and a vehicle of the largest type if there are enough packages.

The movements that are candidates to be rescheduled are the movements of the smaller vehicle types, and the movements with a low utilisation (defined as the ratio of the flow on the movement and the capacity of the movement). To start each neighbourhood search, we order the movements within each group of vehicle types by increasing utilisation, starting with the group of movements with the smallest vehicle type. Denote this movement set as M.We start with the first movement $m \in M$ and reschedule this movement via hubs h^i for which the driving time between hubs h^o and h^i, the processing time at hub h^i and the driving time between hubs h^i and h^d does not exceed the available time between the arrival of the packages at hub h^o and the departure from hub h^d. It would be computationally too intensive to consider each movement $m \in M$ and each intermediate hub h^i. Therefore, we start with the first movement $m \in M$ and consider for this movement all possible intermediate hubs h^i. As long as no movement scheme with lower cost is found, the algorithm continues with the next movement; again considering all possible intermediate hubs. We denote with z the current solution, with z^* the best solution found so far, and with $c(z)$ the cost of solution z. The steps of the local search algorithm are summarised below.

1. Set $z^* = z$.
2. While $c(z) \leq c(z^*)$ do:
   a. Get the first movement $m \in M$.
   b. For each possible intermediate hub $h^i \in \{H \backslash h^o, h^d\}$ do:
      i. Reroute the packages scheduled on m via h^i.
      ii. Calculate the cost of the new movement scheme z.
      iii. Set $z^* = z$ if $c(z) < c(z^*)$.
   c. Remove m from M.

We have found a (local) optimal solution if, at the end of the algorithm, the set M is empty and $c(z) \leq c(z^*)$.

5.6 Computational results

We use three networks to test the solution approach presented in Section 5.3. In this section we first describe the test networks and the parameter settings. Next, numerical results are presented, which show that our solution approach can be used to decrease total transportation cost. We used CPLEX 12.1 on a 3.0 GHz processor with 3 GB of memory to solve the LP and MIP models in the column generation algorithm, and the repositioning model.

5.6.1 Networks and parameter settings

The solution approach is applied to three data instances. The first data instance originates from the Global Optimisation Game (GO-Game). This game is designed to show the challenges faced by an express service provider and is based on a simplified network. For more information on this game, we refer to (Meuffels, et al., 2010). The other two test networks are modified data instances of an express service provider.

Table 12 shows the main characteristics of the networks. The processing times, and the number of services and vehicle types is the same for all networks. The number of locations, packages and flows differ. While the GO Network is mainly introduced for illustrative purposes, Network 1 and Network 2 are representable networks of express service providers. Figure 27 in the Appendix shows a graphical representation of the GO Network, which is available online (http://www.tnt-ortec-game.nl).

Two service types are defined: premium and normal. Table 13 shows that the number of packages with premium service is small compared to the number of packages with normal service. The origin cut-off times of the packages of normal service are earlier than the cut-off times of the packages of premium service. The opposite holds for the destination cut-off times.

Characteristics	GO network	Network 1	Network 2
Locations	10	44	60
Hubs	2	4	10
Processing time hubs (in minutes)	60	60	60
Processing time depots (in minutes)	15	15	15
Services	2	2	2
Vehicle types	2	2	2
Number of packages	100,000	800,000	1,100,000
Number of flows	180	1,088	1,245
Maximum distance	537	909	1,215
Average distance	254	369	539

Table 12: Characteristics of the networks.

Characteristics	GO network	Network 1	Network 2
Normal			
%packages	85	87.5	90
Origin cut-off time	Day 1, 20:00	Day 1, 18:00	Day 1, 18:00
Destination cut-off time	Day 2, 07:00	Day 2, 11:00	Day 2, 12:00
Premium			
%packages	15	12.5	10
Origin cut-off time	Day 1, 20:00	Day 1, 20:00	Day 1, 20:00
Destination cut-off time	Day 2, 06:00	Day 2, 08:00	Day 2, 10:00

Table 13: Characteristics of the service types.

In the column generation algorithm we specify a maximum number of columns that can be generated per iteration. Besides, we limit the number of columns in the restricted master problem by deleting in each iteration the columns of which the reduced cost are positive and above a certain threshold. These columns are kept in a separate pool. At the beginning of each iteration we check whether some columns in this pool have negative reduced cost and should be added again to the restricted master problem. We run experiments with different values for the threshold for removing columns from the network and for the maximum number of columns added per iteration. Based on these experiments, we set the maximum number of columns added in each iteration to 25 for all three networks. For the GO Network the threshold for removing columns is set to 250, for Network 1 and Network 2 this threshold is set to 500. For the GO Network and Network 1 we obtain the best results using these settings. For Network 2 we are not able to run experiments with higher values for these thresholds due to memory issues. We expect that increasing the values of the thresholds results in better solutions.

5.6.2 Numerical results

As our goal is to obtain the optimal movement scheme, we relax the constraints that limit the number of movements at each hub. The cost of the original solution and the cost of the best solutions found by our solution method are shown in Tables 14, 15 and 16.

	Total	Direct	DH/HD	HH	Repositioning
Original solution	87,879	5,916	67,884	5,208	8,871
Initial solution	84,893	34,473	33,663	13,671	3,086
Iteration 1	84,893	34,473	33,663	13,671	3,086
Iteration 2	85,463	34,473	35,226	12,369	3,395
Iteration 3	85,194	34,473	37,041	11,067	2,613
Iteration 4	85,289	34,473	38,133	9,765	2,918
Iteration 5	86,258	34,473	38,625	9,765	3,395

Table 14: Transportation cost, GO network.

	Total	Direct	DH/HD	HH	Repositioning
Original solution	46,895	689	25,406	15,468	5,333
Initial solution	40,866	569	23,053	13,703	3,541
Iteration 1	40,866	569	23,053	13,703	3,541
Iteration 2	42,032	569	24,649	12,894	3,921

Table 15: Transportation cost, Network 1.

	Total	Direct	DH/HD	HH	Repositioning
Original solution	164,863	9,993	70,501	60,901	23,468
Initial solution	138,061	6,735	49,063	55,745	26,518
Iteration 1	134,190	6,735	46,708	55,745	25,002
Iteration 2	141,480	6,735	57,853	51,891	25,001

Table 16: Transportation cost, Network 2.

The original solution represents the cost of the movement schemes similar to the current movement schemes of the express service provider. In the current movement schemes, most packages are transported via the hub closest to the origin depot and the hub closest to the destination depot, and routes via several depots are not possible.

The initial solution is the movement scheme without applying the column generation algorithm to find better DH- and HD-composites. The initial set of DH- and HD-composites is determined as described in Section 5.5.1.

The cost of the best movement schemes found with our solution method are 3.4, 12.9 and 18.6 % lower than the cost of the original solutions for respectively the GO Network, Network 1 and Network 2. Applying the column generation algorithm to the initial solutions of the GO Network and Network 1 does not result in a decrease in cost. When the column generation algorithm terminates because no additional columns with negative reduced cost can be identified, the gap between the lower bound and the solution of the DH/HD-problem was 0 and 3.2% for respectively the GO Network and Network 1. This indicates that the best solution found by our solution method is close to the optimal solution of the DH/HD-problem. Applying the column generation algorithm for Network 2, the DH/HD cost decreases by 4.8% and the total cost by 2.8%. The gap between the lower bound and the solution of the DH/HD-problem was about 50%. As the gap is large, we expect that better solutions can be obtained with higher threshold values.

The last columns of the tables show the repositioning cost. Comparing the repositioning cost of the original solution and the solutions computed with our solution approach, it is shown that the repositioning cost of the original solution is much higher for the GO Network and Network 1. Between the different solutions of our method the cost varies less. This indicates that the impact on the repositioning cost of changes in the DH/HD- and HH-routes is small.

For each network, the time needed to compute the initial solution is only a few milliseconds. The total running time of the algorithm for the GO Network is two seconds. For Network 1, the running times of the first and second iteration are 188 and 74 seconds, and for Network 2 these are equal to 273 and 471 seconds. Hence, the best solution of the GO Network and Network 1 are found within one second, and the best solution for Network 2 is found in less than five minutes.

In the overall algorithm, we first solve the DH/HD-problem and subsequently the HH-problem and the repositioning problem. This means that we find a good solution for the DH/HD-problem but that this solution influences the HH-cost and the repositioning cost, and that the total transportation cost can be relatively high. We developed an iterative procedure with the goal to decrease the total transportation cost by rescheduling the movements with low utilisation in the HH-problem. The experiments show that this procedure does result in a decrease in HH cost, but that this decrease is offset by an increase in DH/HD-cost. For the three networks, the movement scheme with the lowest total transportation cost is either the movement scheme of the initial solution or the movement scheme found in the first iteration.

For Network 1 and Network 2, ten additional cases were constructed by varying the flows of the networks. The number of packages of a flow is randomly generated from a normal distribution with mean equal to the number of packages of this flow in the original network and standard deviation equal to a quarter of the mean. Also for the simulated cases of Network 2 we are only able to apply the column generation algorithm with low threshold values due to memory issues. Hence, we do not show the results of these experiments. For Network 1 the results of the simulated cases, shown in Table 17, are similar to the results shown in Table 15. The minimum, average and maximum gap between the lower bound and the solution of the DH/ HD-problem in the first iteration are equal to 2, 2.6 and 3.3 %, respectively. Only the first iteration is presented as also for the simulated case we are not able to improve the movement scheme by using the iterative procedure. For a specification of the cost for each separate case, we refer to Table 18 in the Appendix.

	Total	**Direct**	**DH/HD**	**HH**	**Repositioning**
Initial solution	41,203	676	23,080	13,647	3,800
Iteration 2	40,978	676	22,949	13,647	3,706

Table 17: Average transportation cost, simulated cases of Network 1.

5.7 Conclusions and recommendations

This chapter proposes a solution approach to schedule movements in the network of an express service provider. Hereby, we simultaneously take into account a heterogeneous vehicle fleet, different service types, exact timing of movements and packages, and routes via several depots and hubs. The solution approach consists of a column generation algorithm, a local search algorithm and an iterative procedure.

The solution approach is tested on three data instances, of which one is a simplified network and the other two are modified instances of networks of an express service provider. We show that using our solution approach the total transportation cost is decreased by 3.4, 12.9 and 18.6 %, respectively.

The results indicate that our solution approach is well suitable to apply to small and middle-sized networks. Due to memory issues we are not able to apply the solution approach with the optimal parameter settings to the large network. However, the results look promising and we expect that by using another implementation of the solution approach good results can be obtained for large networks as well.

For our test instances, the best movement scheme is always found in the initial solution or in the first iteration. This means that the computation time needed to find the best solution is limited. In case longer computation times are allowed as well, the iterative procedure can be used to decrease the transportation cost in subsequent iterations of the algorithm. Considering different rules for the iterative procedure is a subject for further research. Besides, another design of the algorithm could be investigated. For instance, what the effect on total transport cost would be if the order in which the sub-problems in the overall algorithm are solved is changed.

Another topic for further research is the pricing problem of the column generation algorithm. Firstly, we use an approximation algorithm to solve the pricing problem. To prove optimality, the pricing problem has to be solved exactly. Second, in the shortest path problem we use an approximation of the transportation cost of the composites as it is not possible to calculate the exact cost beforehand without knowing which flows are on the composite. A topic for further research is whether it is possible to reformulate the pricing problem and take the exact transportation cost into account.

5.8 Appendix

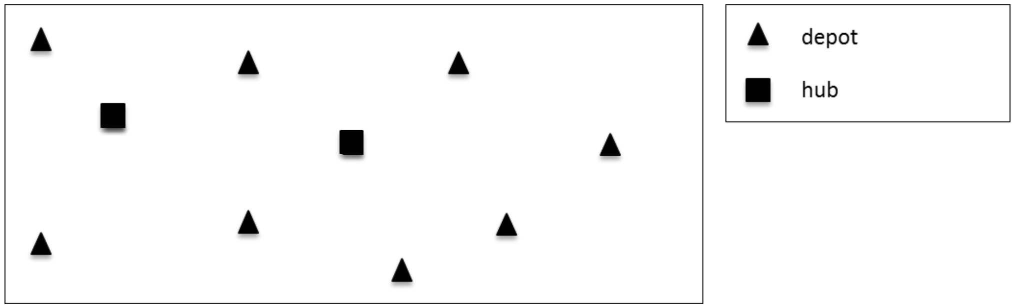


Figure 27: Graphical representation of the GO network.

	Total	Direct	DH/HD	HH	Repositioning
Case 1					
Initial solution	40,292	1,069	22,529	13,114	3,579
Iteration 1	40,292	1,069	22,529	13,114	3,579
Case 2					
Initial solution	41,251	760	23,413	13,548	3,530
Iteration 1	41,251	760	23,413	13,548	3,530
Case 3					
Initial solution	41,682	760	23,198	14,256	3,468
Iteration 1	41,242	760	22,936	14,256	3,291
Case 4					
Initial solution	41,594	569	23,128	13,670	4,231
Iteration 1	41,136	569	22,866	13,670	4,031
Case 5					
Initial solution	40,448	569	23,128	13,388	3,363
Iteration 1	40,448	569	23,128	13,388	3,363
Case 6					
Initial solution	40,769	569	22,776	13,548	3,876
Iteration 1	40,262	569	22,513	13,548	3,632
Case 7					
Initial solution	41,230	569	23,135	13,755	3,771
Iteration 1	41,230	569	23,135	13,755	3,771
Case 8					
Initial solution	42,071	760	22,866	14,258	4,187
Iteration 1	41,668	760	22,603	14,258	4,047
Case 9					
Initial solution	41,517	569	23,164	13,548	4,236
Iteration 1	41,517	569	23,164	13,548	4,236
Case 10					
Initial solution	41,171	569	23,462	13,381	3,758
Iteration 1	40,733	569	23,200	13,381	3,583

Table 18: Transportation cost, simulated cases of Network 1.

Part II - Air network design

6 Reduced hub handling in multiple hub air network design for express service providers by the introduction of pre-sorted unit loading devices

This chapter is based on the following journal paper:

Reduced Hub Handling in Multiple Hub Air Network Design for Express Providers by the Introduction of Pre-Sorted Unit Loading Devices

Meuffels I; Van Dijk T; Fleuren H
Submitted to a scientific journal

6.1 Abstract

This research contributes to the design of (multiple) hub air networks and is inspired by an express service provider who observed an increasing pressure on hub sorting capabilities in their current single hub air network. To resolve the pressure on the hub in the network, a multiple hub network configuration was considered, and to reduce handling at consecutive hub locations in such a network, the idea of pre-sorted unit loading devices (ULDs) arose. This research presents a method to design multiple hub air networks and decides on routing of the flows, the creation of pre-sorted ULDs and the corresponding sorts of the flows, the sort windows at the hub locations, and the number, types and routings of the aircrafts that provide the service. To the best of our knowledge, none of earlier research has considered this problem so far. In this chapter, we present a three-stage solution method, with a pre-processing phase to decide on the routings of the flows and the hub sort windows; a flight-optimisation phase that decides on the number, types, and routings of the aircrafts; and finally, a ULD-optimisation phase that optimises the ULDs to be used. In each phase, mixed integer programming models are presented, and the three phases can be iterated to come to a final best solution. The method is illustrated by a case study that resembles real world data of the express service provider and shows that a two hub network can be operated with hubs that require half of their single-hub-network capacity.

6.2 Introduction

Express service providers offer transportation services that guarantee a predefined delivery date and time to their customers. In their door-to-door service, packages (i.e. documents, parcels, or freight) are collected at the customer and sorted at a local sorting centre, called a *depot*. At these depots, packages towards other local sorting centres are consolidated and transported via one or more global sorting centres, which are called *hubs*. Hubs sort toward the local sorting centre that serves the receiving customer at the promised delivery date and time.

The process of collection and distribution of packages at the customers handled by the depots is known as the *collection and distribution* (or *pickup and delivery*) process in express networks, while the transportation process between local sorting centres is known as the *network* (or *line-haul*) process. To guarantee on-time deliveries of packages, the available time for each process and their connection is specified by the definition of a *cut-off* time. That is, the *origin cut-off* time specifies the time at which the collection process has to be finished, while the *destination cut-off* time specifies the time at which the distribution can start. The time in between these cut-off times denotes the available time to transport packages between the depots. Typically, cut-off times coincide with business working days, i.e. collections may occur till business closure so that origin cut-off times lie after five in the evening while distribution occurs in the morning, so destination cut-off times are specified early morning. This means that the *network* process has to occur overnight for next day services. Besides, the mode of transport for each process is strongly related to the available time. Collection and distribution is a process that typically is executed by road; that is, in general vans are used to take care of these types of transport, although recently the use of motorcycles and even (electrical) bicycles becomes more popular. In the network process, the most common ways of transport are road and air, though some providers use trains or shiploads to transport their less urgent shipments.

The construction of an efficient and reliable express supply chain offers many challenges on strategic, tactical and operational level. The research conducted in this chapter is motivated by challenges observed in the air network of one of the four largest express service providers in Europe. In recent years, this express service provider observed that the characteristics of the flows in their air network have changed as the number of packages to be transhipped grew, but the average weight per package dropped down. This has set pressure on hub sorting capabilities. This has raised questions about the design of the network and a multiple hub network was considered as an alternative to resolve pressure on hub handling.

Operating a multiple hub air network involves a number of design decisions, like "how many hubs do we need", "where do we need them", "what capacities do we need", "how should we operate our aircrafts from/to and between these hubs"? The possible hub configurations are in practice restricted, as the number of airports in Europe that allow night operations is limited; moreover, it is preferred that the current hub is part of the network. As a result, the hub location decisions can be covered by scenario analyses if it is possible to design the network for operating air movements.

The network should provide an all-point service between the airports and a common network design principle for this is hub-and-spoke (Yan, et al., 2013). However, the available time in the network is very limited and a back-of-the-envelope calculation shows that on-time delivery comes in danger when packages have to be sorted more than once. On the other hand, flows that originate/destine at an airport are too low to operate flights to multiple hubs in a cost efficient way. This has led to the idea of pre-sorted unit loading devices (ULDs), being containers used to load packages in single units that can be loaded at once in an aircraft. When pre-sorting is used, packages can visit two hubs but are only sorted at one hub, i.e. either packages are pre-sorted from the origin airport to the second hub, or the first hub sorts the packages by destination airport. An illustration of the four possible routes of ULDs for a pair of airports and two hubs is provided in Figure 28; note that the first option denotes that the ULDs are sorted at each hub, while the other three options denote variants with pre-sorting.

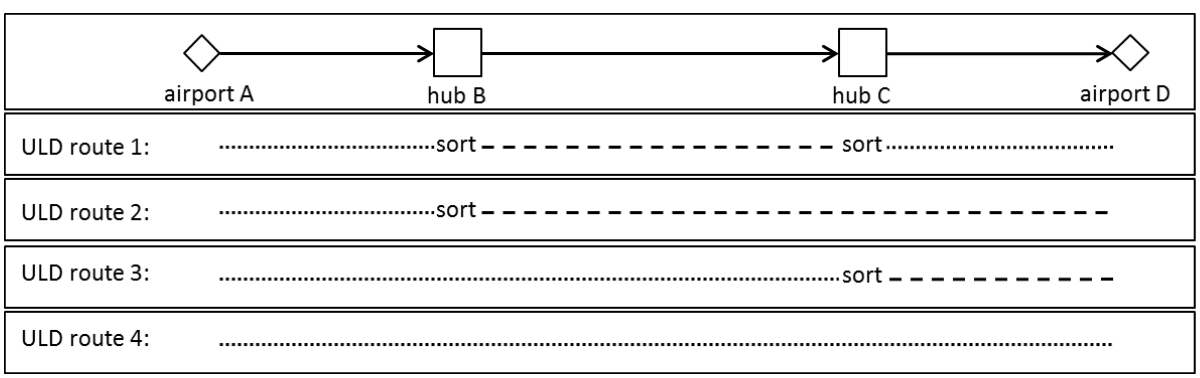


Figure 28: Illustration of four possible ULD routes and the corresponding hub sorts at which each ULD is broken down and/or build.

The research presented in this chapter supports tactical scenario analyses on multiple hub air networks respecting hub capacities, while optimising the number and schedules of the airplanes to be operated as well as the number and types of (pre-sorted) ULDs. Strategic decisions regarding the number and location of hubs are input for this research. In the next section we present a literature review on the design of air networks followed by our research contribution in Section 6.4. Afterwards, we present our modelling approach in Section 6.5. Some computational results are given in Section 6.6 and our conclusions and recommendations are presented in Section 6.7.

6.3 Literature overview

Several studies have proposed different solution approaches to design an air network in which packages are sorted and redistributed among flights at only one hub. Kuby & Gray (1993) are the first to explore the trade-offs and savings involved with stopovers and feeders in the design of the air network of Federal Express Co., which operates a network where most flights to and from the hub make one or more stopovers. Kuby and Gray (1993) design a mixed-integer program to design the minimum cost single hub air network assuming that the hub location is known, and show significant cost reduction compared to the pure hub-and-spoke network. Barnhart & Schneur (1996) use a column generation approach to design a single hub network as well, in which each airport is served by a single stop of an aircraft. Kim et al. (1999) design a multimodal network, but without inter-hub transport. In this, a main air hub is used for service satisfaction while regional hubs can be used for particular flows of packages. Their model includes a number of operational constraints including hub capacity restrictions and fleet balance, and a column generation technique with row generation optimisation and heuristics is used for solving real-world problem sizes. The computational complexity in solving conventional network design formulations results from the two levels of decision variables, being those of the aircraft routing and those for the package flow decisions. Hence, Armacost et al. (2002) present a solution method reducing a decision level by means of composite variables. In order to do so, each airport is a priori connected to an air hub and hub sorting capacity constraints are removed. The model has been used to design the air network of UPS and has resulted in millions of savings for this company (Armacost, et al., 2004). Barnhart & Shen (2004) extend the composite variable approach to cope with premium one day shipments and deferred second-day shipments. However, due to practical complexity, tactical planning (of plans) for next-day and second-day are designed sequentially, where the fleet positions resulting from next-day planning are used for second day planning.

In contrast to the single-sort network designs presented above, Ngamchai (2007) discusses air network design with a two-stage operation, in which some of the packages are sorted twice at two distinct hubs and the required sorting capacity at each hub is considered a design variable. Two solution methods are used: a column generation method is implemented to optimise the problem by means of linear programming (LP) relaxation, in which the resulting model is then embedded into a branch-and-bound approach to generate an integer solution. However, for solving realistic problem sizes, the model is intractable and a Genetic Algorithm approach is used to solve larger instances; the latter is separated into two parts, of which the first step is a grouping representation to assign airports to hubs and the latter an aircraft route representation for aircraft route cycles. The grouping representation uses hub territories that result from the hub windows and corresponding feasible assignments of airports to the hubs. Airports lying in more than one hub's territory are randomly assigned to a hub location and service is ensured by a main hub, having a connection to each individual airport. Each package will be sorted twice at two distinct hubs if its origin airport and destination airport are located on different hub's territory; still, some packages are sorted only once when their origins and destinations lie in the same hub's territory. Afterwards, aircraft route cycles for a given fleet are designed given predefined hub windows and sorting capacities.

Finally, it is worthwhile to mention that none of the existing literature on single or multiple hub air network design has dealt with the use of (pre-sorted) unit loading devices.

6.4 Research contribution

In this chapter, we present a method that enables an air express service provider to design the tactical network. Compared to existing literature, our main contribution is the design of a multiple hub air network, which can be operated without the necessity of a main hub for connectivity reasons. It is expected that cost reduce, as in the situation of the provider in consideration, a single aircraft in general can tranship all flows of an airport, so that two aircraft visits are expensive in definition. But, service time is limited, so pre-sorted unit loading devices are used to pass hub sorts in order to keep up with guaranteed deliveries.

Designing the tactical network concerns decisions regarding the number and types of aircrafts that are needed to tranship all the flows, as well as decisions on how to operate these aircrafts. In addition, we support decisions on the number and type of ULDs that are carried by each aircraft. For each such ULD, it is decided where it has to be built or broken down, where the latter implies that a sort has to occur. The advantage of a sort is that packages can be redistributed, but the disadvantage is clearly the negative impact on time.

To assure that there is enough time to sort packages and to connect the aircrafts at the hub locations, we design a method to decide on the hub windows to set. To the best of our knowledge, none of the literature so far has described a method to decide on these hub windows. In this research, we do.

Lastly, our research supports a number of routing decisions. Firstly, it supports decisions on which depots are sending and receiving flows from particular airports. Secondly, removing the necessity of a main hub for connectivity requires decisions on the hub visits in a route from airport to airport. And thirdly, aircraft routing decisions are to be made. These decisions are all part of the approach presented in this chapter.

In the next section we present our model approach for the tactical network design problem that includes all mentioned decisions.

6.5 Model approach

In this section we present our approach to design multiple hub air networks without assuming that each airport is connected to a main hub and introduce pre-sorted ULDs to make the service instead. There are eight main decisions to take, being the decision…

a. … which airport serves a depot;
b. … on the flow's hub visits and corresponding sorts;
c. … on hub windows;
d. … on the flow routes;
e. … on the number of aircrafts per type to use;
f. … on the route and time schedule of each selected aircraft;
g. … on the number and type of ULDs assigned to each aircraft;
h. … on the sort location of each ULD.

In Section 6.5.1 we presented a literature overview of the network design problem underlying the flight scheduling problem. Earlier research has considered solution methods regarding the decisions (d), (e), and (f) and concluded that the resulting problem becomes impractical for real-world problem sizes. In our approach, we have to decide on five other topics as well, so we knew we had a major challenge to solve.

The approach we finally decided to take is the result of a number of discussions with and analyses of the network of the express service provider. Based on these discussions, two observations are taken into consideration. Our first observation is that aircraft (routing) cost are leading in the network design. Especially the fixed fleet cost is significant compared to any other cost in the network. ULD cost and hub handling cost have lower impact on the final network design. Second, we observed that ULD loading is a weaker constraint in the sense that it is strongly related to fluctuating daily package sizes and crew loading capabilities, which are hard to express in theoretical numbers. Small ULD-overloads might in practice be feasible and should be evaluated in more detail, as it might drive cost in an unnecessary way.

Combining these observations, we decided to take a sequential approach in which each step is designed to include just enough details about the overall problem to make the sequential approach good enough while keeping problem sizes manageable. The first step is a pre-processing phase (I) in which we design the decisions on the depot-airport assignment, flow's hub visits and corresponding sorts, and the hub windows. Afterwards, the main network design phase (II) is run, in which the remaining decisions (d) – (h) are made. The main phase may provide improvements in the decisions of the pre-processing phase, and might iterate to include these improvements.

The main network design phase also consists of two consecutive steps, referred to as the Flight Optimisation phase (II-I) and the ULD Optimisation phase (II-II). In the flight optimisation phase, the decisions are covered regarding fleet sizes and schedules as well as flow routes ((d) – (f)). The ULD Optimisation phase that is run afterwards is in fact an evaluation step in which ULD configurations and load assignments are validated ((g), (h)). The user can decide whether the evaluation step has provided a feasible solution or not, and in the latter case information will be passed to the first optimisation step and the process is repeated until user acceptance.

A summary overview of the approach that we take to solve the tactical network design problem of an express service provider is given in Figure 29. Note that each optimisation phase that we consider contributes to existing literature. Earlier research has considered problems related to the decisions taken in the flight optimisation phase, however, the problem considered in this research extends the existing literature in this field by the addition of hub sorts. Similar, the decisions considered in the pre-processing and ULD optimisation phase are not discussed in literature at all.

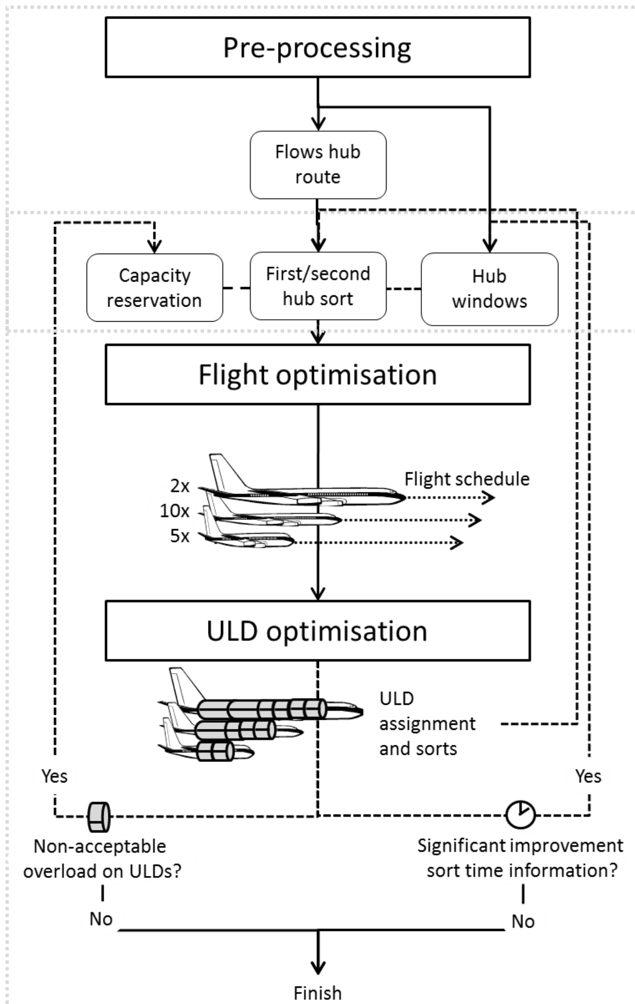


Figure 29: Overview of the model approach, consisting of a pre-processing step and the derivation of a flight schedule by flight optimisation and ULD optimisation.

We will now proceed discussing each particular phase in more detail, where we start with the main network design formulation in Section 6.5.1, and afterwards present the pre-processing phase in Section 6.5.2 as it follows part of the main model formulations.

6.5.1 Main network design: flight optimisation and ULD optimisation

In this section, we discuss the main network design problem formulations, which include a flight optimisation step (see Section 6.5.1.1) and a ULD optimisation step (see Section 6.5.1.2).

6.5.1.1 Flight optimisation

Flight optimisation concerns the optimisation of the fleet size and corresponding flight schedule, according to a number of restrictions. These restrictions are captured in a mixed integer programming formulation, which we will now discuss in detail.

Consider a network of A airports and H hubs ($A \cup H \in L$, the set of locations) in which package flows have to be transported. For this, a fleet of heterogeneous aircrafts F is available with associated fixed cost, c_f^{fixed}, and variable cost, $c_{l_1 l_2 f}^{var}$, of operation between a pair of locations $l_1, l_2 \in L$; in particular, the variable cost (c_r^{var}) of a specific route $r \in R$ of a specific aircraft is the sum of the individual transports between each pair of airports visited in such a route. The indicator ι_{fr} is in this used to denote that fleet f belongs to route r. The main decisions regard the number of aircrafts to operate γ_f and the routes flown y_r by each aircraft. If there exists an imbalance in the number of incoming and outgoing aircrafts at a location, repositioning cost are involved, and the particular repositioning of an aircraft between a pair of locations is given by the variable $y_{l_1 l_2 f}^B$.

There are a number of restrictions regarding the fleet that can be operated. These include a minimum number of aircrafts to be used $\bar{\gamma}_f^{min}$, and a maximum number of aircrafts available, $\bar{\gamma}_f^{max}$, and in some situations a particular aircraft has to be operated between a pair of locations, $\bar{\gamma}_{l_1 l_2 f}$; the parameter $\kappa_{l_1 l_2 r}$ denotes that a route $r \in R$ is operated between locations $l_1, l_2 \in L$. Possible routes are generated for each aircraft type, and each such route is only valid if it respects operational constraints like maximum flying ranges, airport opening/closing times, ramp sizes, etcetera. The routes are generated in a pre-processing step of the optimisation, and are classified as inbound routes (R^I), which are routes that start at an airport, possibly visit intermediate airports on the way, and end at a hub; inter-hub routes (R^H), starting and ending at a hub; and outbound routes (R^O), that start at a hub, possibly visit some intermediate airports, while ending at an airport. An illustration of the different type of routes is provided in Figure 30. As stated earlier, the binary parameter ι_{fr} provides information regarding the fleet type that is used to operate route r, while parameters β_{lr}^{start} and β_{lr}^{end} denote the start and end location l of a route r respectively. The binary parameter ϑ_{ar} is used to denote that a particular route r visits airport a.

Furthermore, the number of aircraft visits at an airport might be restricted: $\bar{\vartheta}_a^{Imax}$ denotes the maximum number of aircrafts that may visit airport a for pickups of flows (i.e., inbound routes) while $\bar{\vartheta}_a^{Omax}$ denotes the maximum number of aircrafts that may visit this airport for delivery of flows (i.e. outbound routes).

Moreover, services are to be satisfied, which means that all packages should be transported from their origin airport to their destination airport. The size of these package flows is denoted as $w_{a_1 a_2 k}$ where $a_1 \in A$ is the origin of the package flow and $a_2 \in A$ the destined airport of the package flow. Lastly, $k \in K$ denotes flow types, i.e. document, parcel, freight, which are distinguished because of the differing associated hub sorts; that is, packages are sorted automatically, while document and freight sorts require (different) manual sort capacities. At this stage of the optimisation, the hubs that are visited on the way from airport a_1 to airport a_2 and flow type k are given. Suppose these hubs are h_1 and h_2 in sequential order, we distinguish the flows inbound to a hub (from a_1 to h_1), between hubs (from h_1 to h_2), and outbound from a hub (from h_2 to a_2) by the following parameters in respective order: $\overline{w}^{orig}_{a_1 h_1 k}$, $\overline{w}^{hubs}_{h_1 h_2 k}$, and $\overline{w}^{dest}_{a_2 h_2 k}$.

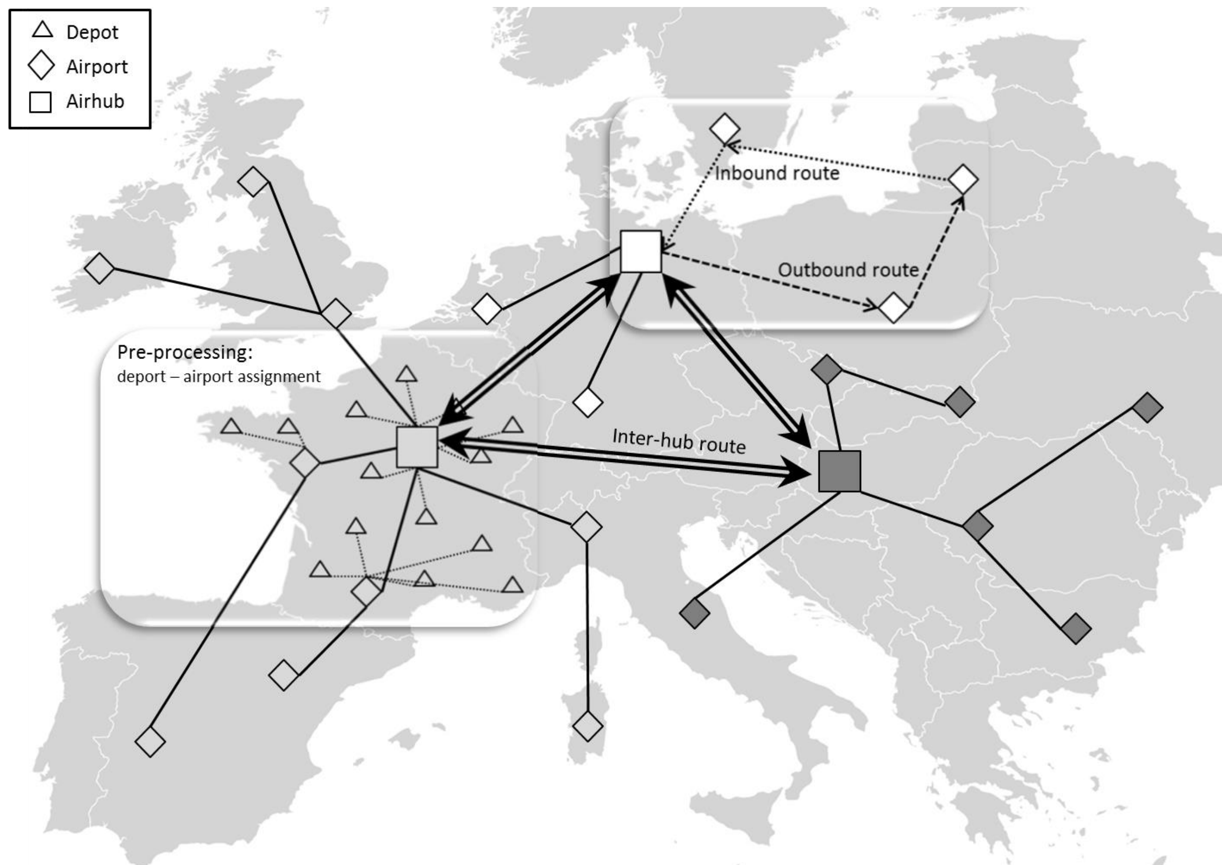


Figure 30: Illustration of the network design with three different types of aircraft routes: inbound routes, outbound routes and inter-hub routes; additional illustration of the depot to airport assignment.

To guarantee service satisfaction, the operated routes should provide enough capacity to tranship the flows. We follow the approach of Armacost et al. (2002) and define extreme routes to incorporate the flow routing decision in the aircraft route decision, by providing information regarding the maximum flow that can be served by a particular route. We denote the maximum flow on an extreme inbound route $\bar{\bar{r}} \in \bar{\bar{R}}^I$ towards hub h_1 by $\bar{\bar{w}}^{orig}_{a_1 h_1 k \bar{\bar{r}}}$, similar the maximum flow of an extreme inter-hub route $\bar{\bar{r}} \in \bar{\bar{R}}^H$ is denoted as $\bar{\bar{w}}^{hubs}_{h_1 h_2 k \bar{\bar{r}}}$ between hubs h_1 and h_2, and the maximum flow of an extreme outbound route $\bar{\bar{r}} \in \bar{\bar{R}}^O$ from hub h_2 is provided by $\bar{\bar{w}}^{dest}_{a_2 h_2 k \bar{\bar{r}}}$. The sum of the maximum flows of the (extreme) routes that are used, $\bar{\bar{y}}_{\bar{\bar{r}}}$, should be larger than or equal to the flow that has to be fulfilled to satisfy services. Lastly, note that the connection between routes and extreme routes is established by the binary parameter $\lambda_{r\bar{\bar{r}}}$. In this, $\lambda_{r\bar{\bar{r}}}$ equal to one means that all flows of at least one stop in the route are completely transhipped in this aircraft, while a fractional value denotes that part of the package flows are covered in this aircraft, while another aircraft covers the remaining part. For more details regarding the extreme route formulation, we refer to the prior work of Armacost et al. (2002).

Besides, we have to assure that hub capacities are not exceeded, that is, the runway capacity and the flow sort capacity may not be exceeded. For the runway capacity constraint, we use the information of the hub window at each hub location. The start of the hub window defines the time at which all inbound aircrafts have to be arrived at the hub to connect to inter-hub transport, while the end of the hub window defines the time at which packages are sorted and ready for transport towards the destination airports. That means that all inbound flights have to arrive before the start of the hub window, and all outbound flights leave after the end of the hub window. We divide the time before the start of the hub window and after the end of the hub window in T periods and model its capacity in an implicit way. This is done by looking at cumulative capacities of the runway and earliest arrivals of inbound routes and latest departures of aircraft outbound routes. This is illustrated in Figure 31 for the runway capacity before the start of the hub window and thus for the inbound flights. We denote the cumulative capacity of the runway before the start of the hub window as $\bar{\mu}^{I}_{ht}$ and after the end of the hub window as $\bar{\mu}^{O}_{ht}$; the binary parameter μ^{I}_{hrt} is equal to one if the earliest arrival of the aircraft flying route r and arriving at hub h is after or at period t and zero otherwise, and similarly, the binary parameter μ^{O}_{hrt} indicates that the latest departure of an outbound aircraft route r is at or after period t at hub h.

Note that aircrafts flying between hubs are not considered in this restriction. The main reason is that the number of aircrafts that are expected to be used between hubs is only very small, so that no issues are foreseen at this stage. Though, a similar approach could be taken for modelling of the inter-hub flights as well, but at this time this would only add unnecessary complexity to the model.

The sort capacity restriction follows a similar approach as the runway capacity restriction, using an implicit way to model these capacity restrictions. That is, we have a cumulative sort capacity at the hub location, and a cumulative required capacity based on the arrivals of the flows. However, the latter is a little bit more complex than the situation of aircraft arrivals, due to the extreme route formulation. The extreme routes make a maximum reservation of capacities, so if one extreme route can provide a capacity of 70% of the total flow of an airport and another route can serve 80% of the flow of the airport, the total flow capacity is 150% which is clearly an overestimation of the real flows. However, this only occurs if more than one aircraft picks up flows at an airport, which is expected to occur at a few airports only. Hence, we expect that the implicit approach is good enough to consider restrictions on hub capacities (especially as overestimated capacity reservation is a conservative way of modelling). Moreover, this should be seen in the light of the capacity figures, which are estimated and hard to express in theoretical numbers. There are two remaining aspects to be mentioned. Firstly, note that we will distinguish flows that have to be sorted before the start of the (inter)hub window, as these flows need further transport to another hub. The flows that have to be sorted are predetermined in the pre-processing phase, and are denoted as $\bar{\bar{\omega}}^{1st}_{lhk\bar{\bar{r}}t}$, and the hub sort capacity of what is called the first sort is denoted as $\bar{\omega}^{1st}_{hkt}$. Similar, flows that arrive by an inter-hub aircraft route or flows that only pass one hub because both origin airport and destination airport are within the same hub region, are to be sorted in the second sort ($\bar{\bar{\omega}}^{2nd}_{lhk\bar{\bar{r}}t}$) and the corresponding hub capacity is $\bar{\omega}^{2nd}_{hkt}$. The second aspect concerns the impact of the runway capacity on the hub sorts. Arrivals of flows are assumed to occur at the earliest arrival of the aircraft, but due to the runway capacity these flow arrivals might be delayed a little bit. E.g. in the illustration of Figure 31, the flows of the second aircraft are assumed to arrive in the third time period, instead of the third time period. For this, we only expect small delays (each time period describes shifts of a few minutes), so that the impact of the implicated modelling of the runway might be neglected in light of the estimation of the hub sorts.

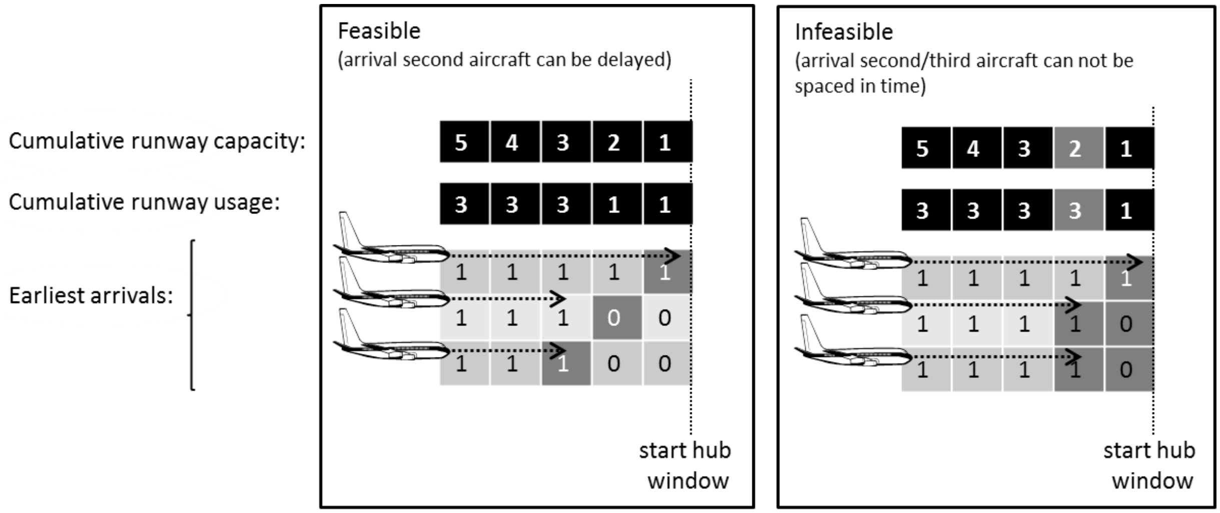


Figure 31: Illustration of the implicit way to model the runway capacities by considering cumulative runway arrivals and cumulative runway capacities at the hub location.

We now present an overview of the parameters and variables of the model as well as the mixed integer programming formulation of the flight optimisation problem:

Parameters

Node data

L set of locations, union of airports and hubs

A set of airports

H set of hub locations

Hub capacities

$\bar{\mu}^{I}_{ht}$ available runway capacity for inbound aircrafts at hub $h \in H$ at time $t \in T$

$\bar{\mu}^{O}_{ht}$ available runway capacity for outbound aircrafts at hub $h \in H$ at time $t \in T$

$\bar{\omega}^{1st}_{hkt}$ maximum flow sort capacity during the first sort at hub $h \in H$ for flow type $k \in K$ at time $t \in T$

$\bar{\omega}^{2nd}_{hkt}$ maximum flow sort capacity during the second sort at hub $h \in H$ for flow type $k \in K$ at time $t \in T$

Location visit restrictions

$\bar{\vartheta}^{Imax}_{l}$ maximum number of visits by inbound routes at location $l \in L$

$\bar{\vartheta}^{Omax}_{l}$ maximum number of visits by outbound routes at location $l \in L$

Fleet type data

F fleet type

Time data

T set of time periods

Cost data

c^{fixed}_{f} fixed cost for the use of one aircraft of fleet type $f \in F$

$c^{var}_{l_1 l_2 f}$ variable cost for transport from location $l_1 \in L$ to location $l_2 \in L$ for fleet type $f \in F$

c_r^{var} cost of a route $r \in R$; note that the fleet type $f \in F$ is known per route

Restricted fleet type data

$\bar{\gamma}_f^{min}$ minimum number of fleet type $f \in F$ that has to be operated

$\bar{\gamma}_f^{max}$ maximum number of fleet type $f \in F$ that might be operated

$\bar{\gamma}_{l_1 l_2 f}$ number of fleet type $f \in F$ that has to be operated between locations (l_1, l_2)

Route data

Subsets of (extreme) routes are denoted with a superscript I for inbound routes, H for inter-hub routes and O for outbound routes; the subscript h is used to denote subsets for a particular hub

R route, i.e. a sequence of visited locations by a particular aircraft

$\bar{\bar{R}}$ extreme routes

$\lambda_{r\bar{\bar{r}}}$ 1 if extreme route $\bar{\bar{r}} \in \bar{\bar{R}}$ belongs to route $r \in R$

β_{lr}^{start} 1 if route $r \in R$ starts at location $l \in L$, 0 otherwise

β_{lr}^{end} 1 if route $r \in R$ ends at location $l \in L$, 0 otherwise

ϑ_{ar} 1 if route $r \in R$ visits airport $a \in A$

ι_{fr} 1 if route $r \in R$ is operated by fleet type $f \in F$, 0 otherwise

$\kappa_{l_1 l_2 r}$ 1 if route $r \in R$ uses fixed arc (l_1, l_2) and 0 otherwise

μ_{hrt}^{I} cumulative use of the runway at hub $h \in H$ by inbound route $r \in \{R^I\}$ at time $t \in T$

μ_{hrt}^{O} cumulative use of the runway at hub $h \in H$ by outbound route $r \in \{R^O\}$ at time $t \in T$

Flow data

K flow types that desire different handling procedures at the hubs

$w_{l_1 l_2 k}$ total flow of flow type $k \in K$ that needs to be transported from airport location $l_1 \in L$ to location $l_2 \in L$

$\bar{w}_{lhk}^{orig}$ all flow originating at location $l \in L$ and routed to hub $h \in H$ with flow type $k \in K$

$\bar{w}_{lhk}^{dest}$ all flow destined at node $l \in L$ and routed to hub $h \in H$ with flow type $k \in K$

$\bar{w}_{h_1 h_2 k}^{hubs}$ the flows to be transported between hubs $h_1 \in H$ and $h_2 \in H$ with flow type $k \in K$

$\bar{\bar{w}}_{lhk\bar{\bar{r}}}^{orig}$ the maximum flow originating at location $l \in L$ with flow type $k \in K$ and routed to hub $h \in H$ that can be transhipped by a single execution of extreme route $\bar{\bar{r}} \in \bar{\bar{R}}$

$\bar{\bar{w}}_{lhk\bar{\bar{r}}}^{dest}$ the maximum flow destined at location $l \in L$ with flow type $k \in K$ and routed to hub $h \in H$ that can be transhipped by a single execution of extreme route $\bar{\bar{r}} \in \bar{\bar{R}}$

$\bar{\bar{w}}_{h_1 h_2 k\bar{\bar{r}}}^{hub}$ the maximum flow that can be transhipped between $h_1 \in H$ and $h_2 \in H$ of flow type $k \in K$ that can be transhipped by a single execution of extreme route $\bar{\bar{r}} \in \bar{\bar{R}}$

$\bar{\bar{\omega}}_{lhk\bar{\bar{r}}t}^{1st}$ the maximum flow originating at location $l \in L$ with flow type $k \in K$, routed to hub $h \in H$, that can be transhipped by a single execution of extreme route $\bar{\bar{r}} \in \bar{\bar{R}}$ and arrives at the hub at time $t \in T$ and has to be sorted during the first hub sort

$\bar{\bar{\omega}}_{lhk\bar{\bar{r}}t}^{2nd}$ the maximum flow originating at location $l \in L$ with flow type $k \in K$, routed to hub $h \in H$, that can be transhipped by a single execution of

	extreme route $\bar{\bar{r}} \in \bar{\bar{R}}$ and arrives at the hub at time $t \in T$ and has to be sorted	during the second hub sort

Variables

γ_f	number of aircraft of fleet type $f \in F$ that is operated
y_r	number of times that route $r \in R$ is operated
$y^B_{l_1 l_2 f}$	number of aircraft of fleet type $f \in F$ deployed at leg (l_1, l_2) for empty repositioning
$\bar{\bar{y}}_{\bar{\bar{r}}}$	real variable to denote the use of extreme route $\bar{\bar{r}} \in \bar{\bar{R}}$

Mixed integer programming formulation

$$min \quad \sum_{f \in \{F\}} \left(c_f^{fixed} * \gamma_f \right) + \sum_{r \in \{R\}; f \in \{F\}} \left(c_r^{var} * \iota_{fr} * y_r \right) + \sum_{l_1 l_2 \in \{A,H\}; f \in \{F\}} \left(c^{var}_{l_1 l_2 f} * y^B_{l_1 l_2 f} \right) \tag{6.1}$$

$$s.t. \quad \sum_{\bar{\bar{r}} \in \{\bar{\bar{R}}^I\}} \left(\bar{\bar{w}}^{orig}_{a_1 hk \bar{\bar{r}}} * \bar{\bar{y}}_{\bar{\bar{r}}} \right) \geq \bar{w}^{orig}_{a_1 hk} \qquad \forall a_1 \in A;\ k \in K;\ h \in H \tag{6.2}$$

$$\sum_{\bar{\bar{r}} \in \{\bar{\bar{R}}^H\}} \left(\bar{\bar{w}}^{hubs}_{h_1 h_2 k \bar{\bar{r}}} * \bar{\bar{y}}_{\bar{\bar{r}}} \right) \geq \bar{w}^{hubs}_{h_1 h_2 k} \qquad \forall h_1, h_2 \in H;\ k \in K \tag{6.3}$$

$$\sum_{\bar{\bar{r}} \in \{\bar{\bar{R}}^O\}} \left(\bar{\bar{w}}^{dest}_{a_2 hk \bar{\bar{r}}} * \bar{\bar{y}}_{\bar{\bar{r}}} \right) \geq \bar{w}^{dest}_{a_2 hk} \qquad \forall a_2 \in A;\ k \in K;\ h \in H \tag{6.4}$$

$$\sum_{\bar{\bar{r}} \in \{\bar{\bar{R}}\}} \left(\lambda_{r\bar{\bar{r}}} * \bar{\bar{y}}_{\bar{\bar{r}}} \right) = y_r \qquad \forall r \in R \tag{6.5}$$

$$\sum_{r \in \{R^I\}} \left(\iota_{fr} * y_r \right) \leq \gamma_f \qquad \forall f \in F \tag{6.6}$$

$$\sum_{r \in \{R^H\}} \left(\iota_{fr} * y_r \right) \leq \gamma_f \qquad \forall f \in F \tag{6.7}$$

$$\sum_{r \in \{R^O\}} \left(\iota_{fr} * y_r \right) \leq \gamma_f \qquad \forall f \in F \tag{6.8}$$

$$\sum_{l_1, l_2 \in \{A,H\}} \left(y^B_{l_1 l_2 f} \right) \leq \gamma_f \qquad \forall f \in F \tag{6.9}$$

$$\gamma_f \geq \bar{\gamma}_f^{min} \qquad \forall f \in F \tag{6.10}$$

$$\gamma_f \leq \bar{\gamma}_f^{max} \qquad \forall f \in F \quad (6.11)$$

$$\sum_{r\in\{R\}}(\kappa_{l_1 l_2 r} * \iota_{fr} * y_r) \geq \bar{\gamma}_{l_1 l_2 f} \qquad \forall l_1, l_2 \in \text{L}; f \in F \quad (6.12)$$

$$\sum_{r\in\{R\}}\left(\beta_{lr}^{start} * \iota_{fr} * y_r\right) + \sum_{l_2\in\{L\}}\left(\beta_{lr}^{start} * y_{ll_2 f}^{B}\right) = \sum_{r\in\{R\}}\left(\beta_{lr}^{end} * \iota_{fr} * y_r\right) + \sum_{l_1\in\{L\}}(\beta_{lr}^{end} * y_{l_1 l f}^{B}) \qquad \forall l \in \text{L}; f \in F \quad (6.13)$$

$$\sum_{r\in\{R^I\}}(\vartheta_{ar} * y_r) \leq \bar{\vartheta}_a^{Imax} \qquad \forall a \in A \quad (6.14)$$

$$\sum_{r\in\{R^O\}}(\vartheta_{ar} * y_r) \leq \bar{\vartheta}_a^{Omax} \qquad \forall a \in A \quad (6.15)$$

$$\sum_{r\in\{R^I\}}(\mu_{hrt}^{I} * y_r) \leq \bar{\mu}_{ht}^{I} \qquad \forall h \in H; t \in T \quad (6.16)$$

$$\sum_{r\in\{R^O\}}(\mu_{hrt}^{O} * y_r) \leq \bar{\mu}_{ht}^{O} \qquad \forall h \in H; t \in T \quad (6.17)$$

$$\sum_{\bar{\bar{r}}\in\{\bar{\bar{R}}^I\};l\in\{L\}}\left(\bar{\bar{\omega}}_{lhk\bar{\bar{r}}t}^{1st} * \bar{\bar{y}}_{\bar{\bar{r}}}\right) \leq \bar{\omega}_{hkt}^{1st} \qquad \forall h \in H; k \in K; t \in T \quad (6.18)$$

$$\sum_{\bar{\bar{r}}\in\left\{\bar{\bar{R}}^I,\left(\bar{\bar{R}}^H | \sum_{r\in\{R^H\}}(\beta_{hr}^{end} * \lambda_{r\bar{\bar{r}}})=1\right)\right\};l\in\{L\}}\left(\bar{\bar{\omega}}_{lhk\bar{\bar{r}}t}^{2nd} * \bar{\bar{y}}_{\bar{\bar{r}}}\right) \leq \bar{\omega}_{hkt}^{2nd} \qquad \forall h \in H; k \in K; t \in T \quad (6.19)$$

$$\gamma_f \in \mathbb{N}_+ \qquad \forall f \in F \quad (6.20)$$

$$y_r \in \mathbb{N}_+ \qquad \forall r \in R \quad (6.21)$$

$$y_{l_1 l_2 f}^{B} \in \mathbb{N}_+ \qquad \forall l_1, l_2 \in \text{L}; f \in F \quad (6.22)$$

$$\bar{\bar{y}}_{\bar{\bar{r}}} \in \mathbb{R}_+ \qquad \forall \bar{\bar{r}} \in \bar{\bar{R}} \quad (6.23)$$

As discussed, the objective is to minimise cost of the network that is operated. Constraints (6.2) – (6.4) provide the flow conservation constraints that assure that all flows are transported via inbound, inter-hub transport if necessary, and outbound transport. Constraints (6.5) relate the extreme routes to the aircraft routes assuring that the number of aircrafts used is integer. The Constraints (6.6) – (6.9) determine the number of aircrafts used per type and Constraints (6.10) and (6.11) make sure that this number is within the minimum and maximum bounds. The legs that have to be operated are given in Constraints (6.12) and Constraints (6.13) make sure that the fleet is balanced at each location. In the constraints that follow, the number of aircrafts that visit an airport during inbound/outbound flights is limited, see Constraints (6.14) and (6.15) respectively. The next constraints, Constraints (6.16) and (6.17), concern the runway capacity restrictions, and finally, Constraints (6.18) and (6.19), assure that enough sorting capacity is available. Lastly, variable restrictions are provided in Constraints (6.20) – (6.23).

In the next section, we discuss the ULD optimisation step in detail, providing information on the number and type of unit loading devices transported by each aircraft and the hubs at which these are unpacked and sorted.

6.5.1.2 ULD optimisation

In this section we present the ULD optimisation modelling. In order to do so, let us first introduce some terminology and assumptions with respect to the ULD routing possibilities.

Flows of packages are transhipped via an inbound route, inter-hub route (if necessary), and an outbound route. During inbound/outbound flights, stops at intermediate airports might be made. If flows are transhipped in ULDs, these ULDs are first loaded at the origin airports, and afterwards it is common to load or break down such ULDs only at hubs, to prevent any loss of time and additional handling at intermediate airport stops. Note that this means that flows of different airports are not transhipped in the same ULD during inbound/outbound flights and some capacity is lost because the package flows of airports do not match the exact capacity of ULDs.

For each airport-airport pair, we then define four possible ULD routes to tranship the flows in which we assume that flows are routed via a maximum of two hubs: (1) flows are sorted at each hub, (2) flows are only sorted at the first hub, (3) flows are only sorted at the second hub, and (4) flows are sorted at none of the hubs. Figure 28 in the introduction illustrates these four possible routes that can be used. Flows that arrive in ULDs that are to be sorted are said to arrive in mixed ULDs while we denote flows that arrive in ULDs that are bypassing the sort said to arrive in pre-sorted ULDs.

ULD routes can only be created along the flight routes of the aircrafts. Hence, we define for each aircraft the flight-legs being the particular legs flown by an aircraft. For example, if an aircraft flies from airport A to hub B to hub C to airport D, we define the flight-legs (A, B), (B, C) and (C, D). Note that each aircraft has unique flight-legs, that is, if another aircraft flies the same route, the similar legs, e.g. (A,B), have unique flight-leg codes[9]. We denote the total set of flight-legs as $\check{F}$ and the set of ULD types is denoted by U with corresponding ULD flow capacity $\check{u}_{uk}$.

Each flight-leg $\check{f}$ has a corresponding flow capacity $\check{q}_{\check{f}k}$, which denotes the capacity of the flow of packages of flow-type k that can be transhipped by the aircraft, where the total flow between two locations is denoted as $w_{l_1 l_2 k}$ (and $w_{l_1 l_2 k}^{NR}$ for non-restricted flows). Besides, each aircraft has a corresponding aircraft capacity regarding the number of ULDs that can be carried by

[9] Note: to prevent redundancy in the optimisation, two flight-legs operated in the same timeframe are replaced by a super-flight-leg", having the same capacity as the sum of the individual flight-legs.

an aircraft ($\ddot{\check{y}}_{u\check{f}}$), which is justified as each ULD requires the same amount of space (i.e. it is not possible to replace a ULD by two smaller ULDs). However, some aircrafts may have a main deck and a lower deck allowing different types of ULDs. To distinguish these situations we define compartments in an aircraft and for each such compartment $v \in V$ we define the number of ULD spaces as $\bar{\sigma}_{v\check{f}}$ for flight-leg $\check{f}$ and the required space to place a single ULD $u \in U$ as σ_{vu}. The belly of an aircraft is small and can therefore only contain packages and documents, but no pieces of freight. Hence, we model the belly as a specific compartment with its own type of ULDs[10] ($u \in U \backslash U^{NR}$), and we classify these ULDs as "restricted", as only part of the flow types can be covered by such ULD types.

The building blocks of the routes are segments, which denote all the flight-legs between sorts. Let us illustrate the concept of flight-legs and segments along the four ULD-routes of Figure 28 in which each route uses flight-legs (A, B), (B, C) and (C, D). The first ULD-route has a segment corresponding to each individual flight-leg, as a sort occurs at hubs B and C. Contrarily, the second and third ULD-route contain only two segments, where ULD-route 2 uses segments (A, B) and (B, D), and ULD-route 3 uses segments (A, C) and (C, D). The last ULD-route contains only one segment (A, D), as all hub sorts are bypassed. The total set of segments is denoted as the set $\check{B}$ and the parameter $\check{b}_{\check{b}\check{r}}$ is used to denote that a segment $\check{b} \in \check{B}$ is contained in ULD-route $\check{r} \in \check{R}$. Besides the parameter $\ddot{\check{b}}_{\check{b}\check{f}}$ is used to indicate that a flight-leg $\check{f} \in \check{F}$ belongs to a segment $\check{b} \in \check{B}$ and $\check{\delta}_{\check{f}\check{r}}$ is used to denote that flight-leg $\check{f} \in \check{F}$ belongs to ULD-route $\check{r} \in \check{R}$. Additionally, the parameter $\bar{\check{y}}_{u\check{b}}$ is used to denote the ULD capacity per ULD type of a segment. Note: if a route visits one hub only, virtual flight-legs are created at that particular hub node. E.g. if the flows of the above figure are routed via hub B only, corresponding flight-legs are (A, B), (B, B) being the virtual flight-leg, and (B, D).

The model will select a ULD-route for each flow of packages instigating the hub sorts that have to occur. We use the parameters $\beta^{1st}_{h\check{r}}$ and $\beta^{2nd}_{h\check{r}}$ to denote that a ULD-route results in a sort at hub $h \in H$, where the first sort parameter denotes that ULD-route $\check{r} \in \check{R}$ has flows to be sorted during the first sort and the second parameter denotes the same for the second sort. Besides, the parameters $\breve{\omega}^{1st}_{l_1 l_2 hk\check{r}t}$ and $\breve{\omega}^{2nd}_{l_1 l_2 hk\check{r}t}$ are used to denote the flow size to be sorted at each hub sort. Although the end of the hub window is presumed to be fixed, the start of the hub window can be optimised via the decision variable $\check{t}^{start}_{ht}$. The resulting cumulative flow sort capacity is denoted as $\bar{\pi}^{1st}_{hkt_1t_2}$ and $\bar{\pi}^{2nd}_{hkt}$, where the capacity of the first hub sort depends on the chosen start of the hub window and the second sort capacity is known by assumption of the fixed end of the hub window.

The main goal of the model is to validate if the flight optimisation model has resulted in a feasible solution when ULDs are used, and of course to create a ULD loading plan. The objective of the model is to minimise the exceeded ULD-capacity, which is preferably equal to zero. These excesses are measured per sector for all flow-types ($\varphi^{all}_{\check{b}k}$) and additionally for the non-restricted flow-types ($\varphi^{NR}_{\check{b}k}$) and weightings can be applied to each excess ($\bar{\varphi}^{all}_{k}$, $\bar{\varphi}^{NR}_{k}$). The weightings are

[10] Note: in reality, the belly is a space were ULDs do not fit. However, for modelling purposes, we deal with the belly as being a normal compartment, with a specific type of ULDs, representing the available belly space and restrict these ULDs to contain parcels and documents only.

determined by counting the number of times a segment is used in a ULD-route as this provides an indication of the popularity of a segment; however, the user is free to change the weightings when preferred. To obtain the exceeding ULD capacity, a ULD route has to be selected for each flow. The decision regarding the ULD-route that is selected, is denoted as $\check{x}_{a_1 a_2 k \check{r}}$. Besides ULDs have to be assigned, so for each segment the number of ULDs of each type is derived by $\check{y}_{u\check{b}}$ (and $\check{y}_{u\check{b}}^{virt}$ for a virtual segment). During the optimisation, it might be possible to improve the hub windows that are set earlier. In order to do so, the variables $\check{z}_{\check{f}t}$ and $\check{\pi}_{\check{r}t}^{2nd}$ are introduced, where the first denotes that a flight-leg departs in period $t \in T$ and the second denotes that a ULD-route arrives at a second sort at time $t \in T$.

We now present an overview of the parameters and variables of the model as well as the mixed integer programming formulation of the flight optimisation problem:

Parameters

Arc data

$\check{F}$	set of flight-legs
$\check{B}$	set of segments, consisting of one or more flight-legs

Fleet type data

$\check{q}_{\check{f}k}$	flow capacity of flight-leg $\check{f} \in \check{F}$ for flow-type $k \in K$

ULD data

U	set of ULD types
U^{NR}	subset of ULD types that can be used by any flow type, i.e. non-restricted ULDs, $U^{NR} \in U$
$\check{u}_{uk}$	ULD flow capacity of ULD-type $u \in U$ for flow-type $k \in K$
$\bar{\varphi}_k^{all}$	weighing factor maximum exceeding ULD-capacity of flow type $k \in K$
$\bar{\varphi}_k^{NR}$	weighing factor maximum exceeding ULD-capacity of non-restricting flow type $k \in K$

Restricted ULD-data

$\check{\check{y}}_{u\check{b}}$	maximum number of ULDs of type $u \in U$ for segment $\check{b} \in \check{B}$
$\dddot{y}_{u\check{f}}$	maximum number of ULDs of type $u \in U$ for flight sector $\check{f} \in \check{F}$

Compartments

V	set of compartments
σ_{vu}	compartment space required in compartment $v \in V$ to store one ULD of ULD-type $u \in U$
$\bar{\sigma}_{v\check{f}}$	compartment space available in compartment $v \in V$ of flight-leg $\check{f} \in \check{F}$

Route data

$\check{R}$	set of ULD routes
$\beta_{h\check{r}}^{1st}$	1 if ULD-route $\check{r} \in \check{R}$ visits hub $h \in H$ during the first sort
$\beta_{h\check{r}}^{2nd}$	1 if ULD-route $\check{r} \in \check{R}$ visits hub $h \in H$ during the second sort
$\check{b}_{\check{b}\check{r}}$	1 if ULD-route $\check{r} \in \check{R}$ uses segment $\check{b} \in \check{B}$, 0 otherwise
$\dddot{b}_{\check{b}\check{f}}$	1 if flight-leg $\check{f} \in \check{F}$ belongs to segment $\check{b} \in \check{B}$
$\check{\delta}_{\check{f}\check{r}}$	1 if flight-leg $\check{f} \in \check{F}$ belongs to ULD-route $\check{r} \in \check{R}$

Flow data

$w_{l_1 l_2 k}$	flow from location $l_1 \in L$ to $l_2 \in L$ with flow type $k \in K$

$w_{l_1 l_2 k}^{NR}$ flow from location $l_1 \in L$ to $l_2 \in L$ with flow type $k \in K$ that can be placed in any type of ULD

Composite flow and routing data

$\breve{\omega}_{l_1 l_2 hk\check{r}t}^{1st}$ the maximum flow from location $l_1 \in L$ to $l_2 \in L$ of flow type $k \in K$, routed to hub $h \in H$, that can be transhipped by a single execution of ULD-route $\check{r} \in \breve{R}$ and arrives at the hub at time $t \in T$ and has to be sorted during the first hub sort

$\breve{\omega}_{l_1 l_2 hk\check{r}t}^{2nd}$ the maximum flow from location $l_1 \in L$ to $l_2 \in L$ of flow type $k \in K$, routed to hub $h \in H$, that can be transhipped by a single execution of ULD-route $\check{r} \in \breve{R}$ and arrives at the hub at time $t \in T$ and has to be sorted during the second hub sort

Hub capacity

$\bar{\pi}_{hkt_1 t_2}^{1st}$ maximum flow sort capacity during the first sort at hub $h \in H$ for flow type $k \in K$ at time $t_1 \in T$when the hub window starts at time $t_2 \in T$

$\bar{\pi}_{hkt}^{2nd}$ maximum flow sort capacity during the second sort at hub $h \in H$ for flow type $k \in K$ at time $t \in T$

Other data

M “big” number

Variables

ULD exceeding decisions

$\varphi_{\check{b}k}^{all}$ exceeded ULD-capacity at segment $\check{b} \in \breve{B}$ and flow type $k \in K$

$\varphi_{\check{b}k}^{NR}$ exceeded ULD-capacity at sector $\check{b} \in \breve{B}$ for non-restricted products and flow type $k \in K$

Routing decisions

$\check{x}_{l_1 l_2 k\check{r}}$ fractional value to denote if ULD-route $\check{r} \in \breve{R}$ is used to tranship flows of flow type $k \in K$ from location $l_1 \in L$ to location $l_2 \in L$

ULD loading plan

$\check{y}_{u\check{b}}$ number of ULDs of type $u \in U$ for sector $\check{b} \in \breve{B}$

$\check{y}_{u\check{b}}^{virt}$ number of ULDs of type $u \in U$ for virtual sector $\check{b} \in \breve{B}$

Hub window

$\check{\tau}_{ht}^{start}$ 1 if the start of the hub window for inter-hub transport at hub $h \in H$ occurs at time $t \in T$

Time decisions

$\check{z}_{\check{f}t}$ 1 if flight-leg $\check{f} \in \breve{F}$ departs in period $t \in T$

$\breve{\pi}_{\check{r}t}^{2nd}$ part of the flows using ULD-route $\check{r} \in \breve{R}$ that arrives at the second hub sort at time $t \in T$

Mixed integer programming formulation

$$\min \quad \sum_{k\in\{K\};\check{b}\in\{\check{b}\}}\left((\bar{\varphi}_k^{all} * \varphi_{\check{b}k}^{all}) + (\bar{\varphi}_k^{NR} * \varphi_{\check{b}k}^{NR})\right) \tag{6.24}$$

$$s.t. \quad \sum_{\check{r}\in\{\check{R}\}}\left(\check{x}_{a_1a_2\check{r}}\right) \geq 1 \qquad \forall a_1, a_2 \in A \tag{6.25}$$

$$\sum_{a_1,a_2\in\{A\};\check{r}\in\{\check{R}\}}\left(w_{a_1a_2k} * \check{\delta}_{\check{f}\check{r}} * \check{x}_{a_1a_2k\check{r}}\right) \leq \check{q}_{\check{f}k} \qquad \forall \check{f} \in \check{F}; k \in K \tag{6.26}$$

$$\sum_{a_1,a_2\in\{A\};\check{r}\in\{\check{R}\}}\left(w_{a_1a_2k} * \check{b}_{\check{b}\check{r}} * \check{x}_{a_1a_2k\check{r}}\right) \leq \varphi_{\check{b}k}^{all} + \sum_{u\in\{U\}}\left(\check{u}_{uk} * (\check{y}_{u\check{b}} + \check{y}_{u\check{b}}^{virt})\right) \qquad \forall \check{b} \in \check{B}; k \in K \tag{6.27}$$

$$\sum_{a_1,a_2\in\{A\};\check{r}\in\{\check{R}\}}\left(w_{a_1a_2k}^{NR} * \check{b}_{\check{b}\check{r}} * \check{x}_{a_1a_2k\check{r}}\right) \leq \varphi_{\check{b}k}^{NR} + \sum_{u\in\{U^{NR}\}}\left(\check{u}_{uk} * (\check{y}_{u\check{b}} + \check{y}_{u\check{b}}^{virt})\right) \qquad \forall \check{b} \in \check{B}; k \in K \tag{6.28}$$

$$\varphi_{\check{b}k}^{all} \leq M * \left(\sum_{u\in\{U\}}\left(\check{u}_{uk} * (\check{y}_{u\check{b}} + \check{y}_{u\check{b}}^{virt})\right)\right) \qquad \forall \check{b} \in \check{B}; k \in K \tag{6.29}$$

$$\varphi_{\check{b}k}^{NR} \leq M * \left(\sum_{u\in\{U^{NR}\}}\left(\check{u}_{uk} * (\check{y}_{u\check{b}} + \check{y}_{u\check{b}}^{virt})\right)\right) \qquad \forall \check{b} \in \check{B}; k \in K \tag{6.30}$$

$$\check{y}_{u\check{b}} + \check{y}_{u\check{b}}^{virt} \leq \check{\check{y}}_{u\check{b}} * \sum_{a_1,a_2\in\{A\};k\in K;\check{r}\in\{\check{R}\}}\left(\check{b}_{\check{b}\check{r}} * \check{x}_{a_1a_2k\check{r}}\right) \qquad \forall \check{b} \in \check{B}; u \in U \tag{6.31}$$

$$\sum_{\check{b}\in\{\check{b}\}}\left(\ddot{\check{b}}_{\check{b}\check{f}} * (\check{y}_{u\check{b}} + \check{y}_{u\check{b}}^{virt})\right) \leq \ddot{\check{y}}_{u\check{f}} \qquad \forall \check{f} \in \check{F}; u \in U \tag{6.32}$$

$$\sum_{\check{b}\in\{\check{b}\};u\in\{U\}}\left(\ddot{\check{b}}_{\check{b}\check{f}} * \sigma_{vu} * (\check{y}_{u\check{b}} + \check{y}_{u\check{b}}^{virt})\right) \leq \bar{\sigma}_{v\check{f}} \qquad \forall \check{f} \in \check{F}; v \in V \tag{6.33}$$

$$\sum_{t\in\{T\}}\left(\check{\tau}_{ht}^{start}\right) = 1 \qquad \forall h \in H \tag{6.34}$$

$$\check{z}_{\check{f}t_2} \leq \sum_{t_1\in\{T|t_1\leq t_2\}}\left(\check{\tau}_{ht_1}^{start}\right) \qquad \forall \check{f} \in \check{F}; h \in H; t_2 \in T \tag{6.35}$$

$$\sum_{t\in\{T\}}\left(\check{z}_{\check{f}t}\right) = 1 \qquad \forall \check{f} \in \check{F} \tag{6.36}$$

$$\sum_{a_1,a_2\in\{A\};k\in\{K\};t\in T}\left(\check{x}_{a_1a_2k\check{r}} * \breve{\omega}^{2nd}_{a_1a_2hk\check{r}t}\right) \quad = \quad \sum_{t\in T}\left(\breve{\pi}^{2nd}_{\check{r}t}\right) \qquad \forall \check{r}\in\breve{R}; h\in H \quad (6.37)$$

$$\breve{\pi}^{2nd}_{\check{r}t_2} \quad \leq \quad \check{z}_{\check{f}t_1} \qquad \forall \check{r}\in\breve{R}, \check{f}\in\check{F}, t_1, t_2 \in \{T | t_1 + flighttime \leq t_2\} \quad (6.38)$$

$$\sum_{a_1,a_2\in\{A\};\check{r}\in\{\breve{R}\};t_2\in\{T|t_1\leq t_2\}}\left(\beta^{1st}_{h\check{r}} * \breve{\omega}^{1st}_{a_1a_2hk\check{r}t_2} * \check{x}_{a_1a_2k\check{r}}\right) \quad \leq \quad \sum_{t\in\{T\}}\left(\bar{\pi}^{1st}_{hkt_1t} * \check{\tau}^{start}_{ht}\right) \qquad \forall h\in H; k\in K; t_1\in T \quad (6.39)$$

$$\sum_{a_1,a_2\in\{A\};\check{r}\in\{\breve{R}\};t_2\in\{T|t\leq t_2\}}\left(\beta^{2nd}_{h\check{r}} * \breve{\omega}^{2nd}_{a_1a_2hk\check{r}t_2} * \breve{\pi}^{2nd}_{\check{r}t_2}\right) \quad \leq \quad \bar{\pi}^{2nd}_{hkt} \qquad \forall h\in H; k\in K; t\in T \quad (6.40)$$

$$\varphi^{all}_{\check{b}k}, \varphi^{NR}_{\check{b}k} \in \mathbb{R}_+ \qquad \forall \check{b}\in\check{B}; k\in K \quad (6.41)$$

$$\check{x}_{a_1a_2k\check{r}} \in [0,1] \qquad \forall a_1, a_2 \in A; k\in K; \check{r}\in\{\breve{R}\} \quad (6.42)$$

$$\check{y}_{u\check{b}}, \check{y}^{virt}_{u\check{b}} \in \mathbb{N}_+ \qquad \forall u\in U; \check{b}\in\check{B} \quad (6.43)$$

$$\check{\tau}^{start}_{ht} \in \{0,1\} \qquad \forall h\in H; t\in T \quad (6.44)$$

$$\check{z}_{\check{f}t} \in \{0,1\} \qquad \forall \check{f}\in\check{F}; t\in T \quad (6.45)$$

$$\breve{\pi}^{2nd}_{\check{r}t} \in [0,1] \qquad \forall \check{r}\in\breve{R}; t\in T \quad (6.46)$$

The objective (6.24) minimises the weighted exceeded ULD capacities subject to constraint set (6.25) – (6.26). Constraints (6.25) denote the flow conservation constraints, i.e. all flows have to be assigned to a ULD-route. Constraints (6.26) validate the total aircraft flow capacity per flow type, while Constraints (6.27) and (6.28) verify the flow capacity for the selected ULDs of the aircraft. Exceeded ULD capacities are only possible if a ULD is assigned to a segment, and this is declared by Constraints (6.29) and (6.30). Besides, ULDs can only be assigned to segments that are used in a chosen ULD-route, which follows from Constraints (6.31), and the number of ULDs assigned to a flight-leg is restricted by Constraints (6.32). Constraints (6.33) consider the available ULD capacity per compartment and flight-leg. Constraints (6.34) – (6.36) are used to select the start of the hub window and assure that inter-hub flights depart after the first sort has finished. As a result, flows arrive at the hub based on the aircraft arrivals/departures, which is provided in Constraints (6.37) and (6.38). The hub sort capacities are validated by Constraints (6.39) and (6.40), where the first set of constraints is used for the first hub sort and the second set of constraints is used for the second hub sort. The Constraints (6.41) – (6.46) is used to restrict the variables.

After ULD-optimisation, the user might decide if flight optimisation has to be rerun. A reason to repeat the optimisation steps is a possible exceed of the ULD capacity. In that case, the user might reserve some additional capacity by including additional dummy flows (separate parameter) to the optimisation. Besides, if the user has included optimisation of the hub window during ULD optimisation, a rerun might be applied with the hub window provided by the ULD optimiser.

In the next section we will discuss the pre-processing phase that determines the depot assignment to the airports, the airport assignment providing the flow's hub routes and corresponding sorts, and the initial hub windows.

6.5.2 Pre-processing phase

Prior to the creation of the flight schedule, a pre-processing step is used to create the desired inputs of the flight scheduling phase. The first pre-processing step is the derivation of the depot-airport assignment. Second, we derive the flow's hub routes and corresponding sorts, and last we decide on the initial hub windows. Each step will be discussed in more detail below.

Depot airport assignment

The depot to airport assignment is based on the selection of one of the following rules: (a) depot to nearest – the depot is assigned to the airport nearest to the depot (b) depot to best service airport – the depot is assigned to the airport that provides the latest departure at the origin depot and earliest arrival at the destination depot (where the depot can be assigned to different airports for inbound/outbound flows), or (c) depot to current – the depot is assigned to the airport at which it is currently assigned during the operation. Once these business rules are used, the user is free to make any changes in the assignment of depots to airports, before finalising this step.

Flow's hub routes, hub sorts and derivation of an initial hub window

The flow's hub routes and hub windows are derived once the depot to airport assignment is finalised. In order to do so, a simplified version of the flight optimisation is solved. This time, we will not follow the composite formulation but separate the decision on flow routes, aircraft flight routes, and last the decision on the hub window.

However, to remain with a tractable model, the set of aircrafts is reduced to include only the most common used aircraft type; more, this aircraft has an unlimited capacity. For each such aircraft, only one intermediate airport may be visited on the way from/to the hub. The flows of each origin airport are assumed to be flown to one particular origin hub, and similar each destination airport is assigned to a single destination hub.

Based on the locations that are visited on the way to the hub, the earliest arrival at the hub location can be derived, which is used to decide on the start of the inter-hub window. Besides an estimation of the sort time for each flight is made, and is added to the arrival of the particular flights. By backwards computing, the latest departure times of the outbound flights at the hubs are calculated as well, providing a restriction on the end of the hub window. Further it is assumed that an inter-hub flight has to be operated between each pair of hubs, so that the individual hub windows are related, i.e. the departure of the inter-hub flight has to occur after the start of the hub window at the first hub and the end of the hub window at the second hub has to be chosen after the arrival of all inter-hub flights including loading and unloading.

Final outputs of the pre-processing step which will be passed on to the fleet scheduling phase concern the hub windows that are chosen, but also the hubs that are visited from origin airport to destination airport. Let us now look at the outcomes of the modelling approach that we followed, by looking at the results of dataset used to illustrate the presented model approach.

6.6 Computational results

The model developed in this research is inspired by a real world express carrier. The model was tested on data of this carrier, but for confidentiality reasons, these results cannot be published. However, to illustrate the results of the model approach, a data set has been created that includes the characteristics of an express network without resembling the real world data.

Characteristics of the dataset

The dataset that is used contains 40 airports of which two airports will operate as a hub. The two hubs are chosen by evaluation of all pairs of airports and selection of the pair that minimises the total distance when each airport is connected to its nearest hub. This results in a hub in Paris (CDG) and a hub in Prague (PRG). It is assumed that all packages are available at the airports at 7 p.m. and have to be available at the destination airport at 7 a.m. in local time, so that collection and distribution can occur during business opening hours.

There are two types of aircrafts available to serve the transport of the package flows between the airports and the hubs, and three types of ULDs are available. The aircrafts have a main-deck and a belly, where two ULD types fit on the main-deck (Main-1 and Main-2) and a single type of ULD is available for the belly (Belly). The details of the aircrafts and the corresponding number of positions reserved for each type of ULD are provided in Table 19.

Aircraft-type	ULD-type	Nr. of ULDs	AC Capacity – Weight (kg)	AC Capacity – Volume (m3)
Aircraft A	Main-1	9	14,200	142
	Main-2	10		
	Belly	4		
Aircraft B	Main-1	15	21,600	216
	Main-2	15		
	Belly	6		

Table19: Details of the aircrafts, i.e. ULD-types, number of ULD-positions and capacity information in weight and volume.

The results

The model as described in Section 6.5 is run to design a flight schedule for the two hub configuration with Prague (PRG) and Paris (CDG). The optimisation is run until no significant improvement for the sort window is obtained and all ULD-overloads are solved, which took three iterations to conclude. The results are obtained on an Intel® Core™2 Quad CPU, with 16 GB RAM using the AIMMS 3.13 64-bit platform; the solver that is used is CPLEX 12.3 and in total it took 38 minutes to come to the final solution that is shown in Figure 32.

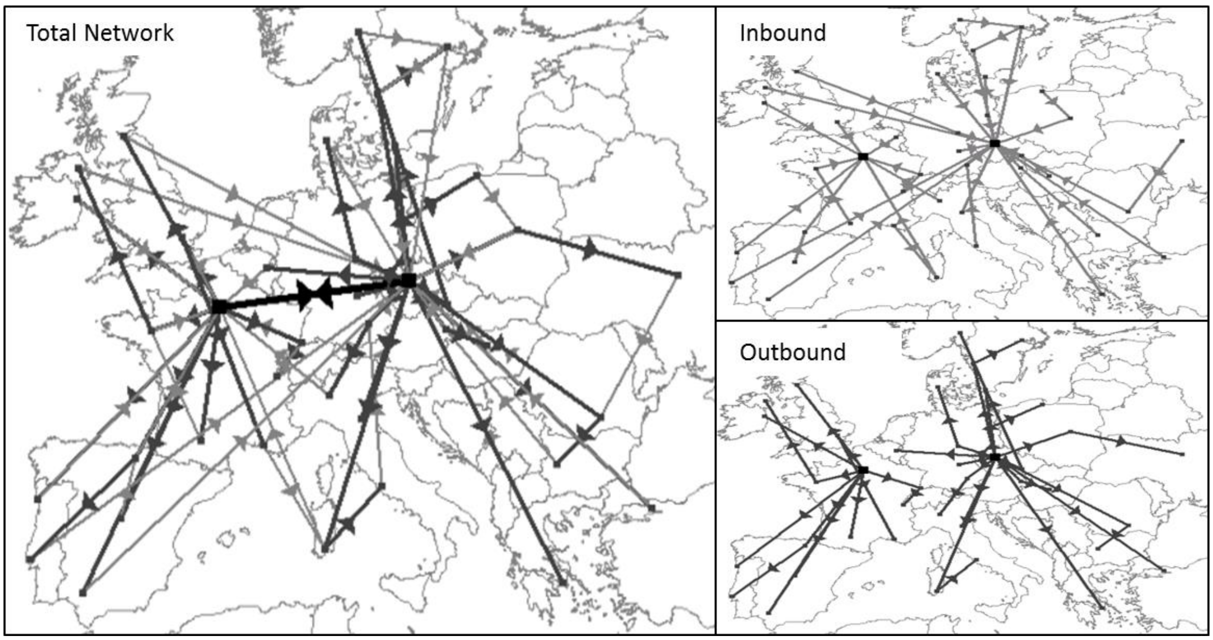


Figure 32: Two hub network structure with inbound (light grey), outbound (dark grey), and inter-hub (black) flights.

Table 20 provides some detailed KPIs regarding the selected hub window, the percentage of pieces that is pre-sorted per hub location, the number of aircrafts per type that is selected in each iteration, and the resulting ULD-overloads and cost of the network (as an index of the final outcome). During the first iteration, the hub window is shortened with 15 minutes in both PRG and CDG. As a result, the sort time for inbound flights is enlarged in the second iteration, so less pre-sorting is required and as a result the ULD-overload decreases in this iteration. However, apparently additional capacity is required to provide a schedule without ULD-overloads, so a smaller aircraft is replaced by a larger aircraft in the last iteration, solving the capacity issue.

The impact of the flight schedule on the throughput of flows at the two hub locations is shown in Figure 33. In Prague, 17 aircrafts arrive as inbound flights and additionally 5 aircrafts arrive as a result of inter-hub transport; in Paris, 9 aircrafts arrive as inbound flights where 7 aircrafts arrive as a result of the inter-hub flows from Prague. The resulting arrival patterns of the flow, for both pre-sorted flows and mixed flows is provided in Figure 33. Note that Prague has a late start hub window and as a result a significant number of pieces are sorted already at this hub and arrive in pre-sorted ULDs at Paris. On the other hand, Paris has an earlier hub window and has quite a number of pre-sorted ULDs on its inbound flights that need sorting at arrival at Prague.

Moreover, it is clear that a significant amount of pieces is pre-sorted, and the proposed schedule would not be feasible in case pre-sorting was not allowed. It is further worthwhile to mention that 79% of the total number of pieces visit Prague, but only 52% of the total number of pieces require a sort at Prague; similar, 64% of the pieces are handled by Paris, but only 32% of the total number of pieces is sorted at Paris. This means that 16% of the pieces need no sort at all as a result of pre-sorted ULDs! Moreover, the desired hub sorting capacities are lower compared to a situation in which only one of these hubs would be operated, so that the network can be operated by smaller sized hubs.

KPI		**Iteration 1**	**Iteration 2**	**Iteration 3**
Hub window	PRG	22:35 - 00:16	22:50 – 00:16	22:50 – 00:16
	CDG	21:25 - 01:26	21:40 – 01:26	21:40 – 01:26
%Pieces pre-sorted	PRG	40	37	34
	CDG	42	41	50
Number of aircrafts	Aircraft A	17	17	16
	Aircraft B	9	9	10
ULD-overload	(largest overload in m3)	Yes (2.46)	Yes (0.60)	No (0.00)
Cost	(by index)	99.64	99.97	100

Table 20: Aircraft details, i.e. ULD-types, number of ULD-positions, capacity information in weight and volume.

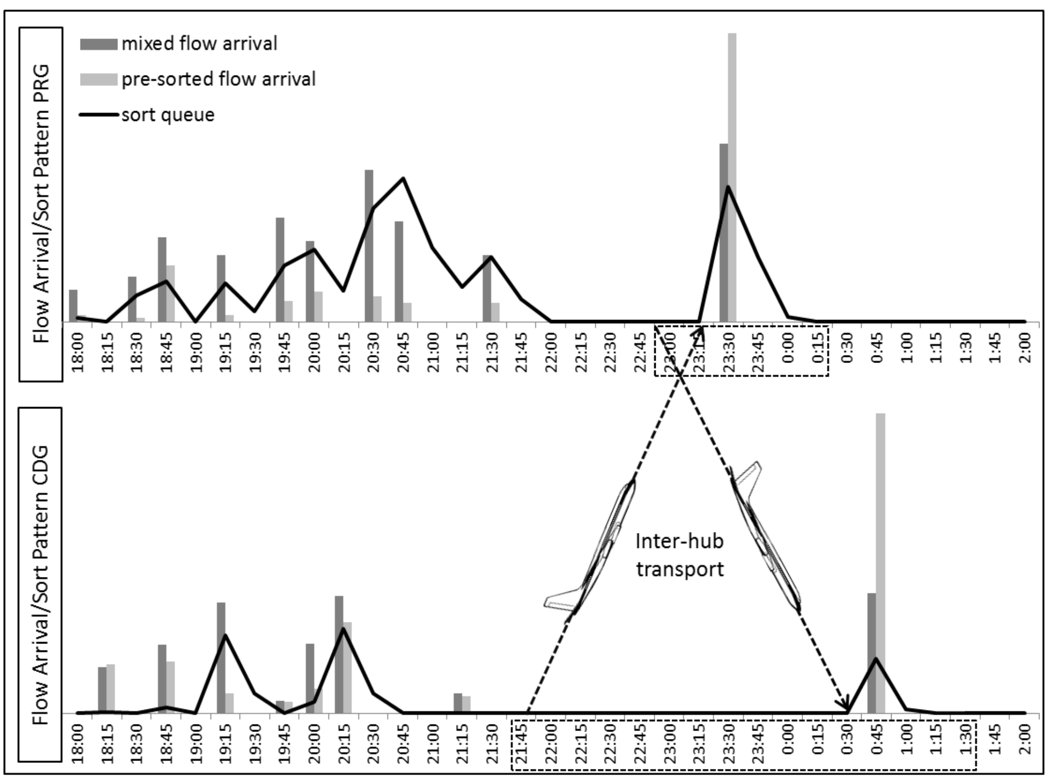


Figure 33: Flow arrival patter with mixed flow (i.e. sort is required) and pre-sorted flows. The queue represents the mixed flows that are waiting for a sort at the end of the time period.

6.7 Conclusions and recommendations

In this research, we presented a three-stage methodology to solve the multiple hub air network design problem of an express service provider with decisions on hub sorts.

The presented research contributes to the existing literature in several ways: it is the first to present a multiple hub air network without requiring a main hub for connectivity reasons. Furthermore, we are the first that incorporate the concept of (pre-sorted) unit loading devices. Additionally, a methodology is presented to derive hub windows. And lastly, we support a number of routing decisions. So far literature has considered the routing decisions for airport to airport flows and the routing of the aircrafts; we extended these routing decisions by depot to airport routing, extensions on the airport to air hub routing decisions, and decisions on the ULD-routings.

We solve the multiple hub air network design problem by the use of an iterative approach in which each step is designed to include just enough details about the overall problem to make the sequential approach good enough while keeping problem sizes manageable. The three steps are iterated until no further improvements are found, and additionally the user can intercept the iteration when he or she decides that the intermediate solution is acceptable. The sub-problems presented in each step of the iterations are solved by a mixed integer programming formulation and can be interrupted via one of the stopping criteria. We illustrated the methodology on a data set that resembles the real world characteristics of an express service provider, and the results showed that a good solution, without ULD overloads, can be obtained in three iterations that are solved within 38 minutes. Furthermore, significant sorting capacity reductions are found (about 50% of the singe-hub-network capacity) for the considered hub locations, showing the advantage of the introduction of pre-sorted ULDs.

The research could be improved by studying the possibilities to further integrate the sub-problems or extensions on the sub-problems. Particularly for the pre-processing sub-problem, further research might improve the solution, as the computational example showed that the final outcome could be improved by shifting the hub window in one of the iterations; also other test cases showed improvements in the hub sort window. Furthermore, flight optimisation and ULD-optimisation could be extended by additional routing strategies, like the exchange of ULDs at airports and three-hub routings among others.

Conclusion and Discussion

7 Main findings and conclusions

In this Chapter we present our main findings and conclusions that can be drawn from the research presented in this dissertation. According to the systematic view of problem-solving as presented by Mitroff, this chapter can be read as the path of validation of the scientific model against the original research objective and the path of feedback for future refinement of the conceptual model based on the found solutions.

7.1 Research objective I: strategic road network design

"Develop a method that supports the decision on hubs to be operated in an express road network such that total cost is minimised under tight service requirements."

The first research topic addressed the strategic road network design, which is concerned with decisions on the number and location of hubs. Typically, express networks have a few areas with significant flows to be transported but also many areas with small package flows. As a result, express networks operate network configurations that fit to these different types of flows. Large depot to depot flows might be organised via direct transports (a. direct transports). Depots with significant flows are served by multiple hubs (b. multiple assignment) in contrast to the depots served by a single hub, and sometimes several depot-hub flows are already combined on the way from/to the hub (c. stopovers). Similar, resulting hub to hub transports might either be driven directly or further consolidation might be applied (d. incomplete hub network).

Some of these extended network configurations have been considered in earlier research, but in general assumptions are made to deal with these types of configurations. For example, Aykin (1994) considered direct routings but restricts the potential direct transports and uses a discounting on inter-hub transport to encourage this type of transport. Multiple allocations between depots and hubs are found in literature, but either thresholds are introduced to restrict the number of depot-hub arcs (Campbell, 1994) or only a fixed amount of connections are allowed (Yaman, 2011). The concept of stopovers is uncommon in road network design, and as far as we know, only considered in the work of Yaman et al. (2007) who typically does not optimise on cost but on service. Similar to the design of a network with multiple allocations, the typology with an incomplete hub network is forced by the introduction of limitations on the arcs to use, either by allowing a limited discount on the inter-hub arcs (Campbell, et al., 2005) or by the introduction of fixed arc cost (Yoon & Current, 2008). Note that the typical configurations in practice are the result of the flows to be transported and the tightness of service level agreements. In our research we were aiming for a methodology that does consider these extended network typologies simultaneously.

Another refinement that we have made concerns the cost function. The cost estimations in literature are in general based on a single uniform vehicle type, where express service providers operate a multiple of vehicle types. Moreover, it is common that the cost estimation in literature follows the concept of economies of scale as introduced by O'Kelly (1987), though some researchers have debated this cost estimation like O'Kelly & Bryan (1998), Campbell (1994), Podnar et al. (2002), Zäpfel & Wanser (2002) and Wasner & Günter (2004). The level of consolidation in express networks is highly related to the tightness of the service level agreements, so the assumed economies of scale do not give an accurate representation of express line-haul cost. Further drawback of this cost function is the inability to distinguish among vehicle types that are operated simultaneously between pairs of locations.

Based on the above observations we stated our two research questions that belong to our research objective to support strategic decisions for hubs to be operated in an express road network. The first research question is stated as “How can cost that result from the operation of vehicles in the network and handling at hub locations be reflected in a strategic road network design solution?” and the second research question is “How can we include all different express network typologies simultaneously in a strategic road network design solution that reflects operating cost well?”.

As soon as our research questions were clear, we started thinking of a scientific methodology, so decisions on the techniques had to be made to support detailed decisions on tailored network designs while assuring that cost to operate such a network were reflected in the best possible way. In fact, the solution to both was found in the introduction of a step-wise cost function that follows the deployment of vehicles. By estimating cost in this way, it is inherent to the model to decide on the transports between depots, depots and hubs and in-between hubs, without making any assumptions beforehand.

Given that the basic problem is NP-hard, we knew that the problems to be solved are too large and complex for exact solution methods, so we decided to choose for a heuristic approach in which our detailed cost function could easily be embedded. We started with the design of a genetic algorithm: this technique combines two (parent) solutions and generates a new offspring based on these parent solutions. We felt this to be a natural way to find the best hub configuration, as hubs mainly serve consolidation in a region, and the combination of the best regional hub configurations would result in lowest operational cost. The technique seemed to work out well on scientific reference data and real world data, but to confirm this we thought about a second heuristic for comparison purposes. From the outcomes of the genetic algorithm but also based on our earlier experience, we learned that many times different hub configurations perform almost similar in cost. We were wondering whether the optimal solution indeed was found, or if we had ended at a locally optimal solution leaving potential for improvement. So to validate the outcome of the genetic algorithm, we aimed at a solution method that could leave local optima, and for that reason we decided to choose for simulated annealing. Based on the implementation of the outcomes, we concluded that both methodologies performed well and showed approximately the same network solutions.

The way we had incorporated the extended network configurations in our heuristics was the result of the following line of reasoning. Without any of the extensions on the network design, we already encountered that many hub configurations perform almost similar on cost. For that reason we expected that the classical network configuration without direct transports and stopovers, with a single allocation and a complete hub network would give a correct indication of the regions where hubs had to be located. Furthermore, we expected that the particular express network typologies are of a secondary effect in the selection of the final best hub configuration.

That is, solutions that are at a global scale similar in cost might significantly deviate when further detailing in network typologies is applied. So based on that idea, we designed a three step methodology: first derive the package flows for which hubs reduce cost due to consolidation, then determine the main hub configurations, and finally apply regional optimisations on a pool of similar global solutions to select the final best outcome.

Real world data instances were used to analyse the outcomes of the heuristic implementations. The results were verified against a relaxation of the exact problem with a promising outcome. Additionally, the outcomes of the basic network typology, before regional optimisation, were compared to the extended network typologies in which regional optimisation had been applied. Hub configurations indeed deviate when the extended network configurations were used with significant line-haul cost reductions as a result. Particularly, the extensions of stopover transports and an incomplete hub network cause these differences in selected hubs. All solutions could be achieved within reasonable runtimes, i.e. all domestic-scale networks were solved within 8 hours of runtime on a normal desktop computer.

To summarise, we extended literature on the strategic network design problem by enriching this problem with network typologies that allow directs, multiple allocations, stopovers and an incomplete network simultaneously in the network design. Additionally, we presented a cost function that is directly related to the deployment of vehicles and hence releases the classical assumption of economies of scales providing a better reflection of actual operating cost. All in all this led to network designs and hub configurations that significantly differ in structure and cost in certain regions compared to the designs in which these specific express network typologies were not included. Additionally, with our design we were able to solve the problem within manageable runtimes.

7.2 Research objective II: tactical road network design

"Develop a method that designs the set of movements and supports (refined) routing decisions to achieve service commitment at minimum cost in an express road network."

Once strategic decisions are clear, tactical decisions on the routings of flows and asset usage need to be made. In daily practice, we observed that the translation from the strategic to the tactical level is difficult: hardly any decision support tools are available that assist in network design changes, and most of the time such transitions are evaluated manually, which is a very time-consuming process. So our goal was to provide tools that enable the planning process at the tactical level.

From a research perspective we considered studies on this tactical network design problem which is better known as the service network design problem. Although the service network design problem can be formulated as a classical network design problem by the introduction of a time-space graph, it is known that this time-space graph grows rapidly in real world problem instances, so that solution methods might not be able to solve the problem. Unfortunately, there is hardly any research available that tackles this problem for real-world instances in the design of express road networks. Only in air network design, detailed scheduling operations were considered at the tactical level, like (Barnhart & Schneur, 1996), (Armacost, et al., 2002), (Armacost, et al., 2004), (Lin

& Chen, 2004), (Lin, 2008). However, the applicability of the methods used in these air studies is low for road network design because tactical air network design in literature only considers single hub networks and the use of a hub window. Both assumptions are too restrictive for the problem considered in road network design. On the other hand, road network design formulations only considered strategic network design problems, and the only timing aspect that was included in these studies, if any, is a cover radius which restricts the total driving time from an origin to a destination, without accounting for waiting time that may occur as a result of consolidation. Additionally, as we already noted in the previous section, the cost function considered in these studies were lacking a level of detail. So we identified a gap at the tactical road network design level, which we aimed to solve, to be able to make the transition from a strategic outcome to a tactical and operational plan.

The research question that we stated for this topic is "How can tactical network design solutions be enriched such that all packages are routed through the express road network at minimal cost in terms of handling and operated fleet?". As tactical network design for road line-haul transport had almost gone unremarked in literature, we decided to address this topic first in simplified form by considering all express network typologies except stopovers and considering only a single vehicle type. We discuss the outcomes of this basic tactical road network design in Chapter 4. Our second research on this topic releases both assumptions and considers the full tactical road network design problem. Additionally, we concluded from the basic research study that more attention should be given to the inclusion of hub operations and studied possibilities to improve the conceptual model for this during our second research. Outcomes of the full tactical road network design are provided in Chapter 5. Note that the tactical road network design problem is strongly NP-hard so that we have chosen for heuristic approaches in both solution directions for real world instances.

7.2.1 Basic tactical road network design

In our first research on the tactical road network design, we address the problem of routing and scheduling for express networks in which directs between depots might occur, depots are allowed to be connected to multiple hubs while hubs are not necessarily connected to all other hubs. To reduce the problem complexity we do not consider stopovers and take only one vehicle type into account. As the problem under consideration is still highly complex, we take a three-phase solution approach: in the first phase, we solve a more classical network design formulation, with an extended cost function to estimate the expected number of (balanced) movements; in the second phase, we evaluate the solution by developing a method to schedule movements; in the last phase we calculate fleet balancing cost to reflect the creation of driver tours. The movement scheduling heuristic that we developed for the second phase follows three basic rules: (a) schedule a movement if a full load can be created, (b) schedule a movement because all flows for the consecutive location are available, or (c) schedule a movement because of service reasons.

As the movement scheduling heuristic can be used to validate several sensitivities in tactical network design, we studied various network design concepts in more detail. First and foremost, we studied the sensitivity in the use of a more detailed cost function instead of a cost function that follows from assumptions on economies of scale. We concluded that the result of fleet schedules when detailed cost estimates were used during tactical planning were significantly better on cost than the traditional approach. Apparently, assumptions on economies of scale do not suit to the operational economies of scale in express networks, which is in line with our expectations, as we observed that consolidation in express networks is highly restricted by tight service level agreements. Afterwards, we analysed some routing sensitivities, the impact of direct driving and repositioning cost, and the dependencies between variable hub handling, unit transportation cost and hub capacities. The results are the following.

Firstly, we analysed the impact of possible variants in the allowed routings: direct transport, and transports via one, two or three hubs. The existence of economies of scale is proven as we clearly identified cost savings when two-hub routes were allowed in the network, compared to the solution where only one hub touch is allowed. Limited additional economies of scale were obtained by further consolidation between hubs, i.e. by allowing routes with three hub touches. However, it should be noted, that these limited cost reductions might be caused by the typical domestic network-sizes considered in the sensitivity analysis. We expect that similar analysis on large-scale domestic networks or continental networks would show higher economies of scale effects when three hub touches are allowed.

Secondly, we analysed the impact of direct driving and the sensitivity on the network design when including estimated repositioning cost. Both have limited impact on the design: direct driving concerns only a small number of od-pairs that have enough flows to qualify for direct transport. Balancing on the other hand is a factor that is inherent to the imbalances in flows which is an external factor that cannot be reduced in the network design phase. Note that there is a relation between direct driving and balancing cost: when the empty return transport is accounted at the same cost level as a loaded transport, this empty movement can be filled with even the lowest amount of packages that needs to be transported between the departure/arrival locations of the empty movement. In this way, handling cost can be reduced without an increase in line-haul cost as this cost needs to be made anyway.

Lastly, the most interesting sensitivity concerns variable hub handling cost, unit transportation cost and related to these the hub capacities. The routing is strongly dependent on variable hub handling cost: less hub touches are observed when the handling cost increases. The impact of unit transportation cost is less intuitive, as it includes two contradictory effects: higher unit transportation would desire more hub consolidation to reduce underutilised transports, but on the other hand, hub routing results in a detour in distance and combined with the additional variable handling cost, underutilised transport may be more efficient. Furthermore, we varied hub capacities to analyse the impact on the network design. We were surprised by the fact that reducing hub handling capacities has actually low impact on the total cost as long as a certain minimum capacity level is guaranteed. When analysing the results, we noted that there is a strong exchanging effect between hub handling and transportation cost: when hub capacities are unlimited, high consolidation is observed providing relatively high handling cost and relatively low transportation cost; when hub capacities are reduced, handling cost decreases while transportation cost increases with similar amounts. Apparently, there exists a range of tactical solutions that can be operated at similar cost levels.

In conclusion: we have developed a heuristic method that fills the existing gap at the tactical network design level, by scheduling movements in the network. This methodology has been used to show the importance of including a more detailed cost function during tactical network design. Variable hub handling has the largest impact on the design of the network, particularly regarding the level of consolidation. Similar fleet schedules can be created that interchange on handling and transportation cost. Secondary cost reductions are observed from direct driving, balancing and incomplete networks, so these might be considered as a secondary objective in the design of networks. As scheduling is only touched in literature in the design of air networks, we believe to be the first that deals with scheduling issues with respect to the design of express road networks in a multiple hub setting.

7.2.2 Full tactical road network design

Our next research objective, extends our prior work by considering some of the areas for improvement that we encountered or defined as scope extensions earlier. A first improvement that we would like to mention concerns the scheduling of heterogeneous vehicle types. Secondly, our prior work did not support network typologies with stopovers on the way from depot to hubs and from hubs to depots. As network typologies with stopover transports are used in express networks, we would like to extend our prior work on this topic. Lastly, our method should be able to restrict the peak in movement arrivals at hubs, as this was seen as a constraining factor by hub operations managers when (re)designing the network.

In our first research we defined a three-phase approach in which the scheduling part is an evaluation step that follows from the routing decisions made at tactical level. During this research, we were aiming for a scientific model that could consider integral routing and scheduling options. But as we still were aiming to solve real-world instances, we knew that we had to define a problem reduction technique to be able to find good solutions in realistic runtimes. As we learned from our first research on this topic that more emphasis should be given to hub operations, we also believed that it was a natural thought to decompose the design problem at the hubs. In that way, we could divide the problem in two manageable parts: the transports between depots and hubs and the transports in-between hubs. By iteratively solving one of the two parts while providing feedback to the other, we aimed to end up with a good solution.

We used a column generation approach for the depot to hub connections and a local search algorithm for inter-hub transport. We applied the methodology on real world data instances and showed that cost could be reduced by implementation of our method due to the use of multiple vehicle types and better consolidation resulting from stopovers. However, there were also two main drawbacks. The methodology is still too complex to solve larger size domestic or continental networks; additionally, we observed that the iterative methodology was not always powerful enough to reduce the cost of the total network during the iterations. We found that an improvement at the depot-hub part caused higher cost in the hub-hub part, or the other way around. Additionally, we concluded that further discussion on the inclusion of hub operations was required. A constraint on the total number of arrivals/departures at the hubs did not completely reflect the problem at hand. For example, the simultaneous arrival of vehicles does not necessarily cause problems, if there is enough time for sorting activities before the loading of the departing vehicles has to occur.

Based on these observations, we concluded that the conceptual model should be refined for this connectivity problem, which relates to the arrivals/departures of movements but also to the available door and sorting capacity at the hub under investigation.

In summary, we extended our earlier work by considering hub capacities via a limitation on the number of vehicle arrivals and departures at a hub location; however, further research on this topic is desired. Another contribution concerns the consideration of a heterogeneous fleet size and the consideration of the full scope of express network typologies in tactical network design with resulting cost reductions. To the best of our knowledge, none of these topics have been covered in literature regarding the design of road networks.

7.3 Research objective III: tactical air network design

"Develop a method that supports routing decisions and designs flight schedules at minimum cost for multiple hubs in express air networks."

Hub operations were also a discussion topic in the operation of the European air network. Particularly, the package flows that were traversing via European air shifted to lighter but more packages, so increasing pressure on hub handling was observed. To deal with the increasing pressure at the hubs, the idea arose to consider a multiple hub air network. However, a quick calculation showed that this would be infeasible when services had to be guaranteed as some packages might have to be sorted twice taking twice as much sorting time. This resulted in the idea of pre-sorted ULDS, being loading units that can be loaded at once in an aircraft. Pre-sorting ULDs could reduce handling in a single hub network as well as it might support multiple hub network designs to become time feasible.

In literature, we have seen several studies that considered air network design. Examples of these are (Kuby & Gray, 1993), (Barnhart & Schneur, 1996), (Kim, et al., 1999), (Armacost, et al., 2002), (Armacost, et al., 2004) and (Barnhart & Shen, 2004). All these studies have in common that they consider single hub networks; the concept of ULDs is not discussed in any of these. Additionally, we encountered that the operation of a multiple hub air network brings design challenges that had not or in limited form been addressed in literature so far. Designing a detailed air schedule for multiple hubs means that we have to make complex decisions with respect to the hub windows that provide connectivity to the different flights from and to airports but also connectivity to inter-hub transport needs to be guaranteed. Additionally, where in single hub networks all package flows are routed via that particular hub, we had to find a way to decide on which hub serves particular package flows. All in all, we had a wide range of challenges that were not yet covered in literature.

The only research that we have found on multiple hub air networks concerns the work of Ngamchai (2007), who discusses air network design that considers two-stage operations in which some of the packages are sorted at two distinct hubs. In the research of Ngamchai (2007), the sorting capacity at a hub is considered a design variable that needs to be determined. Furthermore, the set-up of the network design differs in the following aspects compared to the multiple hub configuration that we were considering: Ngamchai (2007) assumes a main hub to guarantee service, and the decision of hub windows to operate is not discussed in this research; additionally, the assignment of airports to hubs follows mainly by the definition of the hub window.

Based on this review of the problem situation and existing research we have stated two research questions: “How to design a multiple hub air network solution method that supports detailed routing and flight scheduling decisions and gives guidance to hub operations in terms of hub sort windows and total package flow to be handled?”, and, “How can air network design solutions make efficient use of pre-sorted ULDs in order to reduce hub handling time and hub sorting capacity?”.

We discussed with operational air and air-hub people how an acceptable outcome should look like. From these discussions we deduced that parameters on hub handling capacities are in practice hard to express in numbers and similar, that ULD loading capacities are fuzzy because they have a certain acceptance margin. This argumentation brought us to the decision to decompose the problem: firstly, create a basic flight schedule with a rough estimation of handling capacity and afterwards refine the results by building ULD assignments as an evaluation step thereafter. If the evaluation shows significant ULD-overloads or pressure on hub handling, the user can provide feedback to the first optimisation step and a new schedule will be generated.

Additionally, we had to make some enrichment in the basic scheduling of the flights, for the reason that we would like to evaluate multiple hub configurations as well. Based on our experience in road network design, we knew that decisions on routing and scheduling in a network with multiple hubs are tough. However, we used our operational knowledge on the expected set-up of such a multiple hub network: due to the package flows, it was expected that a single aircraft would be able to serve all flow at an airport. This immediately asks for the introduction of hub windows, as this is the only way to provide connectivity between all aircrafts in a network where point-to-point services have to be guaranteed. These two assumptions were used in a pre-processing phase, which defines the hub service areas as well as the operating hours of each such hub location.

We illustrated our approach on a data set that resembles real world characteristics of a European air network that requires overnight air service. The case study proved that the iterative method worked out very well: within three iterations, a feasible solution without any overloads on ULDs was obtained. Moreover, the required hub handling capacities at the considered hubs were far lower than observed in a single hub network. Lastly, we mention that the solving time is manageable: the case showed that the three iterations were run within 38 minutes of computation time on a normal desktop computer.

In summary, we introduced a methodology to design multiple hub air networks with restrictions on hub handling capacities and the introduction of (pre-sorted) unit loading devices. For single hub networks our contribution concerns the addition of ULD decisions: we make decisions on the number and type of ULDs in the network, and provide information on which packages are assigned to each ULD, and we decide if a sort has to occur at the hub location. These decisions are also made in multiple hub networks. Additionally, we created a flight schedule that provides connectivity at each hub location in a multiple hub setting. In order to do so, we introduced a hub window and the corresponding start and end time is outcome of the modelling.

8 Directions for further research

This chapter provides directions for further research, which can be used as input during the conceptualisation phase in future research. We first derive directions for future research based on a discussion of similarities and differences in road and air network design (see Section 8.1). Afterwards, we provide some directions for further research based on research limitations in Section 8.2. Both sections should be seen as general views on the feedback path of the systematic view of problem-solving of Mitroff. We close this chapter by a note on the path of implementation in Section 8.3.

8.1 Road or air network design: similar or different?

Network design problems in road and air transport are very similar in the decisions to be taken: in both situations, decisions have to be made on routing, consolidation, and asset usage to serve the demanded package transport from one location to another. However, when studying literature, we observed that chosen approaches to design the transport plan in road or air networks differ significantly. In practice, these differences are also observed in real-life planning. So what is the reason that these types of networks are designed in a different way? In this chapter we aim to motivate the differences that are observed both in scientific and practical design of road and air networks.

Network sizes: domestic, continental, and intercontinental networks

Firstly, we distinguish the three different sizes of networks that are operated by express service providers: domestic-scale networks (e.g. a network of the Netherlands, Germany, Spain, etcetera), continental-scale networks (e.g. the network in Europe) and intercontinental-scale networks. These network scales are typically distinguished due to demand for transport, type of transport and distances covered.

A first differentiator is the demand for transport in each such network: an A.T. Kearney report (A.T. Kearney, Inc, 2012) describes that about 91% of the shipments in the European CEP-market were related to domestic transports in the years 2009 to 2011. Clearly flows in domestic networks are higher than flows in continental and intercontinental networks. Unfortunately, no figures are available that relate the domestic, continental and intercontinental flow sizes in the CEP-market, but based on our experience, intercontinental transport is relatively small compared to domestic and continental package flows. Note that in general more flows in a network decreases unit transportation cost due to better utilisation.

A second differentiator in network sizes comes from the relative distance and corresponding service possibilities. Clearly, domestic networks in Europe are short distance networks and time-certain services can be offered by road transport. Air transport is in general too expensive to compete with road services and even service possibilities may be worse due to the limited number of airports as well as the unavailability of timeslots to organise air transports. So only in the larger domestic networks, air transports are used to offer time-certain services. Continental networks like a European network, are longer-distance networks in which road and air services may compete on service offering and cost. A fast and more expensive day-certain service can be offered by air, while a more economical but slower day-certain service can be offered by road. Clearly, intercontinental networks are large distance networks so that road transport is an uncommon mean of transport as it cannot provide attractive services. Intercontinental networks are therefore mainly served via air.

Ways to operate: owned, subcontracted, and outsourced operations

There are also different ways in which each type of network is organised. We distinguish three different ways to operate a network: owned, subcontracted or outsourced.

We use the term *owned transport* to denote that assets (i.e. the vehicles/aircrafts) are owned by the express service providers and drivers/crew are on payroll at the express service provider. In this situation, the express service provider defines the transports in terms of departure and arrival location and departure and arrival times. A *subcontracted transport* denotes a network in which the assets are owned by a third party and drivers are contracted via this third party as well. Though, the express service provider still defines the transports in terms of departures and arrival locations and times (i.e. the movements/flights). These specified movements/flights are dedicated for the transport of flows of the express service provider, but the means of transport may be used for transports of other parties at times that no movements/flights of the express service provider are specified. Lastly, we refer to *outsourced transport* to denote the situation that the assets are owned by a third party as well as drivers contracts, and additionally, the third party is in charge of the specification of departure and arrival locations and times. Outsourced transports are in general non-dedicated transports, i.e. only part of the space of a transport mean is hired by the express service provider (e.g. a block space agreement with an air freighter is an example of outsourced transport). Note that a subcontracted or outsourced transport is in general more expensive than an owned transport on a cost per kilometre rate. However, the risk to end with an idle mean of transport is taken by the third party for both subcontracted and outsourced transports. Third parties are willing to take that risk when expected revenues outweigh the cost, so attractive prices are only offered if there is enough opportunity to reallocate resources to other parties.

The basic conditions to choose for owned, subcontracted or outsourced operations is a combination of the total demand for transport in the network, and the availability of options for subcontracting and outsourcing. Typically, road transport is a market with a high number of players, and as flows in domestic and continental-scale networks for express service providers are significant enough, dedicated transports can be negotiated with subcontractors. Moreover, as a result of the subcontracting possibilities, there is in general flexibility to absorb fluctuations in flows, as additional (or cancelled) transports can be arranged with third parties. The level of (sub)contracting then results from the balance between taking the risk of idle or underutilised transports that are owned versus the higher price that is paid for a subcontracted transport. In contrast to the road market, only few players operate in the air transport market. As discussed, air transport competes with road transport on the continental scale in Europe, and the available outsourcing possibilities are too limited to offer the desired service levels. So the only option to provide this service is to operate either a subcontracted or owned network, if flows at least are significant enough to make this service attractive, which seems to be the case for the big four

European express companies. The options to attract additional resources in the air transport market are very low, and at the day of operation this is in general impossible because of the lacking supply but also because of the limited availability of timeslots at airports. As a result, third parties would bear the full risk of idle assets, which makes the prices for subcontracted transports unattractive. For that reason, it is common that continental air networks are operated in a fully owned manner. Note that such a network has low flexibility to react to demand fluctuations as the pool of assets is fixed. The same reasoning holds for subcontractor prices for intercontinental networks, so the options to consider here are either an owned or outsourced network. For the reason that the competing nature of road transport is missing at intercontinental scale, more flexibility in definition of services is possible. Additionally, flows are in general too low to justify the cost of owned transports, with some exceptions for key customer accounts (e.g. think of customers like Apple, Samsung, DELL, HP, and etcetera). As a result, intercontinental transports are in general outsourced, with few owned flights, and the best service option is offered based on the possibilities in the market.

Planning complexity

The planning of an owned or subcontracted network is similar in the sense that in both situations schedules are made by the service provider. However, if the possibility of subcontracting exists, there is much more flexibility at the operational planning day, and as a result, the tactical road network design planning results in a basic schedule that often is accompanied by agreements on adjustments that can be made in the operational plan by planners at a shorter term horizon, like a week or a day. When the network is completely owned and no subcontracting arrangements are in place as is the case in continental air network design, the tactical plan is far more detailed and has to be robust for package flow fluctuations. This tactical plan is a final schedule that in general is operated on each operational day, with only minor adjustments. The planning of an outsourced network completely differs from owned or subcontracted networks, for the reason that the express service provider is not in charge of scheduling the assets. As a result, the possible service that can be offered is in fact the result of the best possible offers that can be obtained by agreements with third parties. Typically, outsourced transports are in general used to offer day-certain services.

Furthermore note that the planning complexity is related to the size of the network, particularly regarding the number of hubs in each network. Typically, when travel times are relatively short due to the speed of transport used, single hub networks are operated. This holds for the operation of small domestic road networks (e.g. the Netherlands) but also for continental scale air networks. A network with multiple hubs is used for larger size domestics and in the situation of intercontinental transport. Clearly, the more hub locations are operated, the more complex the planning operations. First because decisions have to be made on the service area of each hub, i.e. there are multiple options to transport packages from origin to destination, and scheduling becomes complex, as not only depot-hub and hub-depot transports need to be synchronised, but also the transport between hubs. These synchronisation issues become much easier in single hub networks in which hub windows are introduced.

A last remark concerns the cost drivers that play a role in the planning complexity. Typically, air network design is driven by transportation cost, mainly due to cost of acquiring an aircraft but also because of the cost to operate flights (e.g. insurance cost, take-off and landing cost, maintenance cost, etcetera). Transport cost dominates in the design of such networks. For road networks, none of hub handling cost and transport cost is dominating; network design methodologies on road network design should include both cost components to arrive at the overall lowest operational cost level.

Road and air network design: can it be further aligned?

So far, we discussed the differences in road and air network design, but to what extent can approaches be aligned?

Is it possible to design road networks with the approaches used in air network design? And would it result in good solutions for the operation of a road network? Yes and no. Yes: the level of detail in which the planning is created in air network design is higher than the level of detail in road network design, and decisions on assets and flow routings are made in a more integral way. However, hub handling cost need to be added as this is an important cost driver in road network design; we expect that this is a relatively easy addition in the type of models used in air network design. The more complex extensions regard the use of multiple hubs and the assumption of hub windows which not necessarily provide best solutions in road network design. At this moment, we expect that we at least need technological advances in mathematics and computation times before design methodologies similar to air network design could be applied in road network design. But the advantages of the integral decisions on assets and flow routings would make such approaches more attractive to be considered when the mathematics and computational condition is fulfilled.

Similar questions for air network design: could we use approaches similar to these taken in road network design for air network design? For example, suppose that the conditions in international transport changes (e.g., demand for package flow transports increase, or more subcontracting options become available), which makes the option to consider owned or subcontracted transport relevant to consider. It can be expected that the current air network design approaches would be unable to solve the intercontinental air network design problem as these networks would typically be larger in size due to the use of multiple hubs. Unfortunately, we expect that road network design solutions would highly underestimate the cost in air networks, because the fixed cost of the aircraft is hard to incorporate in current approaches in road network design. Hence, we definitely expect that further research is needed when air networks become more complex.

Concluding remarks

In this section we discussed the different types of networks that are operated by express service providers and the corresponding planning characteristics, of which an overview is shown in Figure 34. We described differences in road and air networks and the corresponding way in which these networks are operated. This results in a statement on planning complexity, which relates to the level of detail in which tactical planning has to cope with scheduling operations and the complexity that results from the size of the network in terms of hubs. Research presented in this dissertation focussed on the design of domestic and continental networks, in which scheduling operations are part of the tactical network design level, and for that reason intercontinental transport is not further considered in this dissertation.

Furthermore, we expect that when technological advances in computational power and mathematical techniques become available, road networks might be designed in a similar way as the current approaches in air network design. However, when air networks become more complex, for example when introducing an owned intercontinental air network, further research is necessary, as approaches taken in road network design would definitely result in unnecessary high cost.

Network size*	Domestic	Continental		Intercontinental
Transport demand	High	Medium		Low
Mean of transport	Road transport	Road transport	Air transport	Air transport
Way of operation	Owned Subcontracted	Owned Subcontracted	Owned	Outsourced (Owned)
Type of schedule	Basic	Basic	Detailed	Optional
Nr. of hubs	Single or multiple**	Multiple	Single	Multiple
Cost drivers	Handling cost, transport cost	Handling cost, transport cost	Fixed aircraft cost, transport cost	Flow size

*network size reference for Europe; **depending on size of the network.

Figure 34: The different types and sizes of express networks and corresponding planning characteristics.

8.2 Directions for further research that follow from research limitations

This section discusses the following five directions for further research that resulted from current research limitations: (a) data availability and benchmarking, (b) data instances, (c) robustness, (d) implementation and improvement measurement, and (e) research scope. We will discuss these in detail below.

Data availability and benchmarking

The availability and quality of data is a process that often has slowed down the progress of research in this dissertation. Generally, this process involves discussions with experts on the desired data, gathering of the data, and a cleansing process on source data, to improve its quality and to make it available for use. Although data gathering was a time-consuming process, it also provided valuable research insights and potential for quick wins in practice.

Unfortunately, results of the methods applied to real data instances could not be shared with the research community for confidentiality reasons. And as the express business is a niche in the transport industry, no benchmarking data sets were available for testing and comparison with other methodologies. We therefore had to generate illustrative sets of data for the purpose of publication but with more or less the same characteristics as the real datasets. As these data sets are anonymised and hence available for sharing, we invite other researchers to use our data sets for future referencing. These data sets can be obtained on request.

Data instances

Our research was tested on real world data sets of TNT Express, but further research is needed to verify whether the conclusions from conducted studies would hold in a wider perspective. The significance of the methodologies presented could be further strengthened by further application on additional datasets of TNT Express (e.g. other geographies) but particularly its overall relevance would be confirmed when applied to data sets of other express companies.

Robustness

As data gathering and cleansing was already an extensive process, we decided to focus our research on a few static data sets that reflect average daily operations. However data quality is improving as businesses more and more invest in data quality, which creates the opportunity to make use of improved data sets and even real-time data. We think that further research on the design of express networks that uses the variation in data eases the implementation process of network design changes in everyday practice. The most important reason for the use of dynamic data concerns the capability to build a robust network design, that is more flexible to for example volume fluctuations (weekday versus weekend day, summer versus winter, holidays, etcetera) and network disturbances (like traffic jams, cloudy weather, etcetera).

Implementation and improvement measurement

One of the challenges in practice for network design is the implementation of the results and the measurement of performance indicators on cost and service. The difficulty stems from the fact that network design solutions in general redesign the current network and that implementation needs to be done via a gradual way of continuous improvements, which may take several years. This path of implementations has to be designed and is not always easy to create. The gradual way to redesign the network also makes it hard to measure performance improvements, as these come in a gradual way as well, and at times even cost increases might be accepted to achieve savings in a later stage.

Research scope

In this dissertation, we studied the line-haul network design problem for both road and air networks. Line-haul transport is however only part of the total supply chain of an express company. Extensions of the scope for pickup and delivery operations and a more detailed modelling of the processes at depots and hubs would be relevant and contributes to more integral decision making at supply chain level. One of the examples which we touched already in Research Objectives II is to consider the movement scheme in relation to hub operations: it could be very interesting to optimise the workforce at hubs and the design of the network at once. Another example would be the investigation of cut-off times: cut-off times now separate the PUD process and line-haul process but examining these processes simultaneously might show further saving opportunities. The latter becomes even more interesting in the light of the increasing trend in deliveries to consumers in the B2C segment, as deliveries to consumers shifts to early evening hours instead of business hours.

8.3 A note on the path of implementation for TNT Express

The overall objective of this research was the development of OR models that support a cost efficient operation of an express line-haul network with guaranteed service level agreements. Our research was part of a global optimisation program at TNT Express, raising the unique opportunity to study topics that can immediately contribute to practice. Discussions on research topics and conditions highly contributed to the relevance and applicability of research in practice, and we are very grateful having had that opportunity. Our research was often used as a first pilot, to illustrate strengths and weaknesses of methodologies, before being extended and incorporated in the tool suite used at TNT Express.

The research on strategic road network design as presented in Chapter 3 showed the strength of heuristics in strategic network design, but also pointed out the complexity of network design in the situation of incomplete hub networks. As these type of networks are typically used at express service providers when considering continental road network design, further research was needed. Combined with the insights from the enriched tactical network design methodology as presented in Chapter 4, a strategic network design model similar to the tactical network design presented in Chapter 4 is developed and now part of the standard tool suite at TNT Express. The tactical network design problem that enables to generate a basic fleet schedule as presented in Chapter 4 was further extended to cope with multi-day scheduling, and is incorporated in the tool suite of TNT Express as well. This methodology can be used independently, but also in a combination with strategic network design. The research as it is conducted in Chapter 5 on the tactical network design with hub operations was not yet convenient as it is not capable to design large-scale networks that TNT Express operates; we are currently investigating alternative methodologies to incorporate hub operations in strategic and tactical network design.

Lastly, the research on tactical air network design as presented in Chapter 6, has been used to study the alternatives of multiple hub air network design. This methodology is mainly used for what-if scenario analyses, and a similar single hub network design model is in use during everyday practice.

Final word of gratitude

All in all, I am proud on the wide range of research topics that I could work on, and trust that TNT Express has gained valuable insights during this research period by pointing directions for improvement and cost avoidance strategies.

I would like to express my word of gratitude to TNT Express and ORTEC, who both made it possible for me to conduct research as presented in this dissertation. Thank you for this great opportunity!

As a last remark, I would like to use the opportunity to thank my Ph.D. Committee members: prof. dr. Wout Dullaert, dr. ir. Cindy Kuijpers, dr. ir. ing. René Peeters, and prof. dr. T. van Woensel. I am grateful for the valuable feedback that I received during the predefense of this dissertation and grasped the opportunity to improve the quality of this dissertation.

Appendix
Business challenges at TNT Express

Notes to the reader:

Appendices A and B serve as background material to provide a broader understanding of challenges observed at express service providers. For that reason, the scope of these Appendices is the end-to-end supply chain of an express service provider. Furthermore it is remarked that these chapters are written from the perspective of TNT Express.

Lastly, please note that TNT Express partners with ORTEC when research projects are initiated that require special expertise in the field of Operations Research. Due to my appointment at ORTEC as an expert in the field of network design for both road and air express networks, I was strongly involved in all OR-related work discussed in the Appendices when it concerns the line-haul network design.

A. The design of express networks in a nutshell – Playing the Global Optimisation-Game (GO-Game)

This chapter is based on the following journal paper:

The Design of Express Networks in a Nutshell – Playing the Global Optimisation Game (GO-Game)

Meuffels, I; Fleuren, H; Poppelaars, J; Hoornenborg, H; De Rooij, F
OR News (2010), Vol. 39, Is. 2, pp. 6-8

A.1 Abstract[11]

TNT Express is one of the four largest express carriers serving several customers in more than 200 countries. They operate a large network containing many depots and hubs in order to serve all customer needs. Several thousands of vehicles and dozens of aircrafts are used on a daily basis, operating via strict schedules. The organisation of such a complex network is a challenging task and TNT Express recognises the benefits of operations research in fulfilling this task. This has led to a strong cooperation with universities and consultancy firms specialised in the field of operations research. The partnership with ORTEC, one of the largest providers of advanced planning and optimisation software solutions and consulting services, is an obvious result of their mission.

In order to explain the strength of Operations Research modelling and build understanding of the complexity in the design of an express network at management level at TNT Express, the idea arose to develop a game reflecting the challenges in such a design. A first prototype of this game was developed in close cooperation between prof. dr. ir. Hein Fleuren (Tilburg University) and ORTEC in AIMMS. AIMMS is an optimisation modelling system with integrated GUI builder. Later on, ORTEC further developed the game as a Java Applet so that it can be presented to a broader audience via the web (www.tnt-ortec-game.nl). A screenshot of the game is given in Figure 35. In the remainder, we will elaborate on the game.

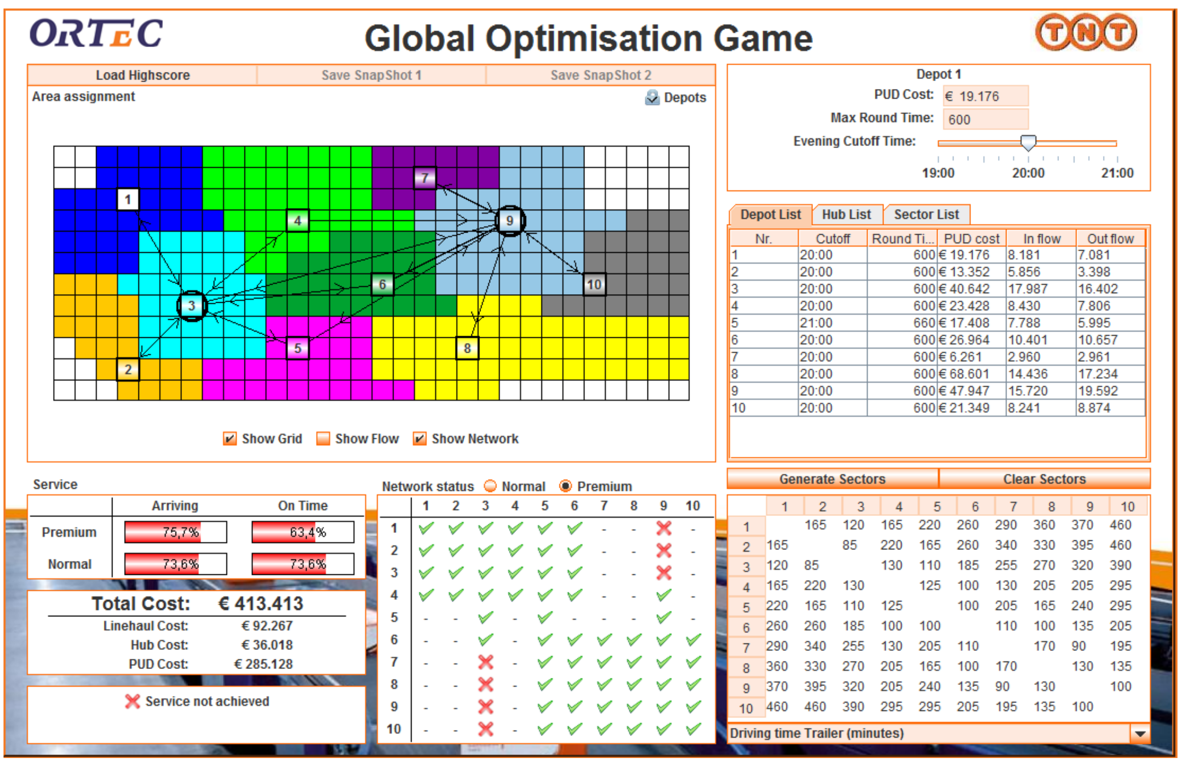


Figure 35: Screenshot of the GO-game with a two-hub network (hubs at depot locations three and nine).

[11] Some of the terminology introduced in this chapter might slightly differ from the general use of terminology in this dissertation; we decided not to align terminology in this chapter to maintain the connection with the online game.

A.2 The Global Optimisation-Game

In the express business shipments originate from the desire to move packages from a sender to a receiver. Express carriers take care of these shipments and in particular define services with guaranteed delivery date and time. In this, packages are picked up at the customer, transported via a line-haul network and finally delivered to the receiving customer. The transport is organised via a network consisting of depots and hubs. Depots consolidate the packages from senders and take care of the delivery of packages to the receiving customers. Hubs are large sorting facilities used to consolidate the transport of packages between the depots.

The GO-Game invites you to help the Express Company of GO-land, which is in desperate need for improvements of its operations. The objective is to construct a network in which all packages are transported at minimum cost.

Infrastructure

GO-land is displayed in the left corner of the game (see Figure 35). Currently, the Express Company uses ten depot locations that are located at the numbered squares. Each coloured square indicates a postal area that needs to be served and the colours correspond to the serving depot. The white areas are outside the border of GO-land. Besides, the Express Company would like to invest in some depots so that they can be used as hub locations. Although the company is not sure which depot locations to use as a hub and how many hubs are required, they know that a hub costs 20,000 Euros. Furthermore, it is known that the usage of multiple hubs results in efficiencies that payoff 30,000 Euros. The size of the hub location can differ between a capacity of 50,000 and 60,000 packages. The largest hub has an average processing time of 120 minutes per package while the smallest hub processes a package within 60 minutes. However, the unit cost of a consignment is lower in large hubs as a result of economies of scale (1.00 Euro) compared to small hubs (1.40 Euro).

GO services

The packages that need to be transported are called consignments. The Express Company of GO-land offers two different services for the delivery of these consignments, a *normal* service and a faster *premium* service. The demand for each service is constant each day and contains 100,000 consignments in total of which 85,000 require a normal delivery service and 15,000 a premium service. The origin postal area and delivery postal area of these consignments are known. To guarantee the on time delivery of *normal* packages, these packages have to be available at the delivery depot before 7:00 hour; consignments with *premium* services have to be available one hour earlier at 6:00 hour. The pickup of the packages needs to be finished for both services at a depot-specific time between 19:00 and 21:00 hour (referred to as the *cut-off* time).

Pickup and Delivery (PUD)

Once the packages to be delivered arrive at the depot, a sorting process starts. At 10:00 hour packages will be ready for delivery to the customers. Since the Express Company operates each day, the delivery process will be combined with the pickup of packages and this *pickup and delivery process* may last until the cut-off time of the corresponding depot has been reached. Figure 36 illustrates the timeframe corresponding to the pickup and delivery process.

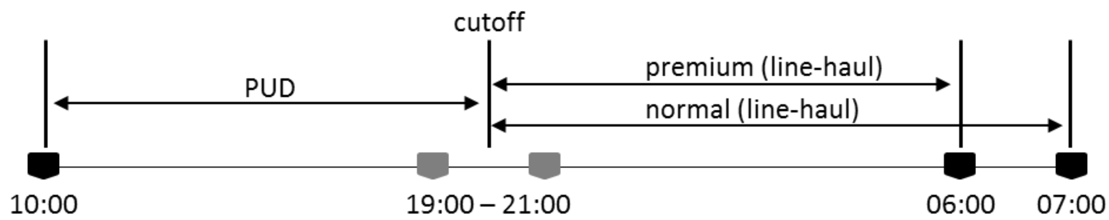


Figure 36: Time frame corresponding to the pickup and delivery process.

The pickup and delivery of packages is done by vans and the total number of vans required to serve an area depends on available time and the capacity of the vans. Each van has a maximum capacity of 60 consignments and a maximum speed of 70 kilometres per hour. The cost of a van is 0.5 Euro per kilometre. The distance from depot to postal area is equal to the Euclidean distance and each consignment is located at one kilometre further per consignment. The pickup or delivery of a single consignment takes about six minutes.

Sectors

The consignments are transported between depots via the line-haul network. This can be a depot-depot transport, depot-hub transport, hub-hub transport, or hub-depot transport. The transport can be arranged by the definition of sectors. Each sector has a start location, end location, start time, end time and vehicle type and the base cost to create a sector is 1,500 Euros. Variable cost of a sector depends on the vehicle type and the number of vehicles needed. There are two vehicles available: a truck with a capacity of 1,000 consignments, average speed of 70 kilometres per hour and cost of 3.00 Euros per kilometre; a van with a capacity of 100 consignments, average speed of 90 kilometres per hour and cost of 1.50 Euro per kilometres.

Routes

Consignments may be transported either via one or multiple sectors. A feasible route is a route that is able to transport packages from the service depot of the pickup area to the service depot of the delivery area that meets the service restrictions (i.e. does arrive before 6:00 hour for premium consignments and before 7:00 hour for normal consignments). Note that two consecutive sectors can only be used if the arrival location of the first sector equals the start location of the second sector (this location needs to be a hub location) and the time between arrival of the first sector and departure of the second sector needs to be at least the processing time at the corresponding hub location. In case of multiple feasible routes, the route that arrives the earliest at the end depot will be chosen, and in case of equal arrival times, the route with the least hub touches is selected.

A.3 Design Choices

The ultimate goal of the game is to design a network in which all consignments are transported at minimum cost. This can be achieved by a multiple of design choices, which will be discussed below.

Infrastructure

Although the company would like to keep the number of depot location fixed, it has some doubts about the exact location of these depots. You are invited to help them in finding the right locations of their depots and defining the service areas of each depot. Besides, the company believes that cost reduction can be attained by upgrading some depot locations to hub locations, so that further consolidation can find place. The number and location of these hubs needs to be investigated, as well as their size.

Services
The arrival time at the delivery depots is defined by the service offering. However, the cut-off times for pickup can be defined for each depot (between 19:00 and 21:00 hour). Note that an early cut-off time enlarges the time available for line-haul transport, but limits the time available for the pickup and delivery of packages. The reverse occurs in case of a late cut-off time. Can you help the company to decide on the best cut-off time?

Sectors
The sectors to transport the consignments need to be defined. There are a lot of decisions to be made here since the start and end locations need to be chosen, as well as the timings and the type of vehicle used.

A.4 Some concluding remarks

The GO-Game is a simplification of the challenges faced by express carriers. It was designed to show complex interactions between distinct parts of the supply chain. Many people within the company of TNT have been challenged to play the game, to experience its complexity and to see how their actions influence the supply chain as a whole. The insights gained in this way make people aware of the effects of their own actions in all day practices and shows the importance of good communication between the individual parts of the express network. The GO-Game is now a central part of TNT's strategic GO-Program, which aims to bring TNT a competitive advantage by using OR in its operations. The GO-Program also includes the GO-Academy, a two-year modular course for an annual cohort of about 50 people from TNT's operations around the world.

Besides of creating awareness within TNT's supply chain, the game was also developed to invite the OR community to help TNT in their mission of improved network design. Although the game simplifies reality, it nevertheless contains varies challenging operations research topics that are shown to be difficult to solve. For example, the location of depots can be seen as a *p-hub location problem* while the decision of hub locations can be seen as a *capacitated hub location problem*. The routing of the consignments is included in the *tactical network design problem* with *multiple assignments* to hub locations. Besides, *fleet scheduling decisions* need to be made in the creation of sectors. However, all these famous problems consider only part of the whole and become even more challenging as a result of their interactions. You are invited to play the game via www.tnt-ortec-game.nl, ENJOY!

B. Supply chain-wide optimisation at TNT Express

This chapter is based on the following journal paper:

Supply Chain-Wide Optimization at TNT Express

Fleuren H; Goossens C; Hendriks M; Lombard M.-C.; Meuffels I; Poppelaars J
Interfaces (2013), Vol. 43, Is. 1, pp. 5-20

B.1 Abstract

The application of operations research (OR) at TNT Express during the past seven years has significantly improved decision-making quality and resulted in cost savings of 207 million euros. The Global Optimisation-Program (GO-Program) initiative has led to the development of a suite of optimisation solutions to assist the operating units of TNT Express to improve their package delivery in road and air networks. To create and deploy these solutions, we established communities of practice (CoPs), at which internal and external subject matter experts meet three times annually at an internal conference. We also created a unique two-year learning environment, the GO-Academy, where employees of TNT Express are taught the principles, use, and deployment of optimisation techniques. As a result of these combined initiatives, OR is now an effective part of the core values at TNT Express.

B.2 Introduction

TNT Express N.V., one of the world's leading business-to-business express delivery companies, operates the largest express road and air network in Europe and air and road transportation networks in China, South America, Asia-Pacific, and the Middle East.

In this line of business, express delivery companies move packages (i.e., parcels, documents, or pieces of freight) from a sender to a receiver under various and guaranteed service level agreements, which specify delivery dates and times. Each service offering consists of collecting packages at a customer site, transporting them via a road and (or) air network, and delivering them to a recipient.

Each week, TNT Express delivers 4.7 million packages to recipients in over 200 countries, using a network of more than 2,600 facilities, a fleet of about 30,000 road vehicles and 50 aircraft, and a workforce of 77,000. Because of the highly volatile and competitive nature of the express delivery market, the company must ensure that its network is robust, agile, and able to effectively absorb demand fluctuations. Express delivery companies are focussed on achieving both cost efficiency and high levels of customer service, two goals that are often contradictory. The challenge is to design a supply chain that can effectively meet both criteria and manage this critical balance between them. Given that point-to-point flows are typically too low to justify a single transport over large distances, we prefer to maximize consolidation to reduce costs. Conversely, our ability to transport packages efficiently is restricted by the types of services we offer, which may vary considerably. For example, we offer time-definite express services with a guaranteed next-day delivery before 9 am, 10 am, 12 noon, or the end of the day, but we also offer day-definite services for less-urgent shipments that do not require next-day delivery. As a result, consolidating packages from similar collection origins can be difficult because the delivery time frames may vary considerably. Figure 37 illustrates our air and road supply chains.

After collection, packages are transported to depots, which are local sorting centres that manage the collection and delivery of customer packages. Hubs are large sorting facilities used to consolidate the transport of packages between the depots. TNT Express refers to the collection and delivery of packages at the depots as the pickup and delivery (PUD) process, while the transport between the depots is the network process. Cut-off times (i.e., due times) separate the PUD process from the network process: at the pickup cut-off time, the packages must be available at the origin depot for the network process; at the delivery cut-off time, the packages must be available at the destination depot for the delivery process. The PUD process encompasses the processing time at the depots, while the network process includes processing times at the hubs. Although we separate the PUD and network processes, implementing each process presents a number of challenges. For example, the PUD process includes assigning customer pickups to a particular depot, deciding on the number of vehicles required, and determining when the vehicles will visit the customer for collection or delivery. In the network process, a number of important decisions are made; these include sorting centres to be visited, sequence and time of departure, and the schedules needed for the vehicles connecting these locations. At the supply chain level, decisions about cut-off times (e.g., allocating more time to one process in favour of another) and the number and location of hubs and depots required must be made.

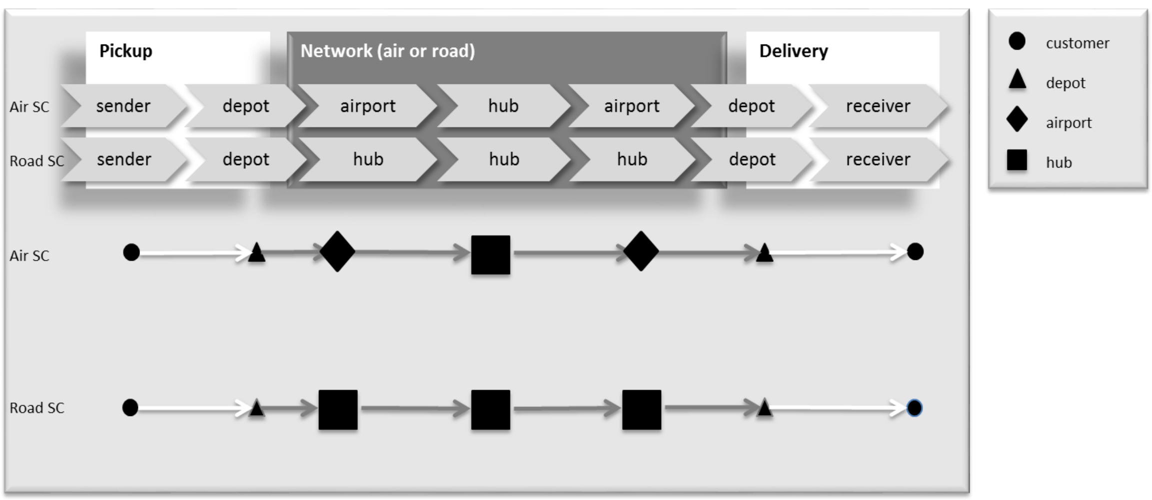


Figure 37: The TNT Express supply chain consists of road and air operations. The pickup and delivery (PUD) process concerns the collection and delivery of packages at the customers; the network process addresses the transportation of packages between depots.

B.2.1 Operations Research at TNT Express

In 2005, TNT Express embarked on its first operations research (OR) project. The initial strategy was to expand business activities rather than just focus on cost reductions and asset utilisation. Senior management understood that focussing on growth would not guarantee a competitive advantage in the long run. Triggered by a story on optimisation by Tilburg University professor Hein Fleuren, Marco Hendriks, Director of Strategic Operations & Infrastructure at TNT Express, sensed that quantitative methods should become the key enabler to increase the company's competitiveness. This awareness led to TNT Express' first OR project, which was aimed at optimising Italy's domestic road network. The results were promising: by rescheduling vehicles and reassigning packages, asset utilisation increased and transportation costs decreased by 6.4 percent. This initial success paved the way for the GO-Program and the close working relationship between TNT Express, Tilburg University, and ORTEC, an OR consulting and optimisation software provider that partners with TNT Express on optimisation activities.

The idea of incorporating OR into our decision making grew steadily. We were convinced that optimisation activities should be at the core of our business and that we should not organize these activities within a separate, large, centralised OR department. To achieve these goals, we established communities of practice (CoPs) and the GO-Academy. A CoP is a community of TNT Express business experts worldwide, which a central GO-team organises. CoP meetings, at which participants share best practices and optimisation knowledge and discuss these ideas with suppliers and academia members, are held three times each year. These meetings typically last two to three days and have 15 to 20 attendees. The GO-Academy is a two-year program designed to teach management and staff the principles of optimisation, which will enable them to recognise optimisation opportunities and develop a common global optimisation language. As the added value of OR gained visibility, the TNT Express board adopted OR to develop strategies to respond to the consequences of the economic crisis in 2008. Senior management understood that by applying OR across the business, it would be able to manage unit costs more effectively, while continuing to fulfill all service obligations. Unit costs had become important because they were rising steadily as a result of decreasing demand, increasing fuel prices, and more stringent environmental regulations.

B.2.2 The GO-Program

The goal of the GO-Program is to improve decision making throughout the TNT Express organisation and in each part of its supply chain. To address the challenges in networks, the PUD process, and the entire supply chain, we set up separate subprograms for each. These subprograms led to the creation of a portfolio of models, methodologies, and tools designed to solve the optimisation challenges we encountered. Given that each operating unit is at a different level of operational maturity, the solutions differ in optimisation complexity. At the lower end of the maturity scale, dashboards and guards help the operating units to analyse their actual performance and identify new optimisation opportunities. For the more mature units, we deploy advanced optimisation solutions.

These solutions are important in identifying and realising each GO-Subprogram cost saving; these savings were 207 million euros over the period 2008–2011: 132 million euros from the supply chain subprogram, 48 million euros from the networks subprogram, and 27 million euros from the PUD subprogram. The GO-Program also enabled us to reduce CO_2 emissions by 283 million kilograms—the CO_2 equivalent of 1,000 trucks traveling around the world seven times.

Although we expect more savings in the future, this chapter focusses on the solutions that have been in use for some time and therefore contributed the most to the reported savings. These solutions, which we describe in the sections below, are: TRANS in the networks subprogram, SHORTREC in the PUD subprogram, and DELTA supply chain in the supply chain subprogram.

We also describe how our use of OR has evolved to become a key component in decision making via the GO-Program. We illustrate this by describing the OR methods we applied and the benefits accrued and challenges encountered to date. We then highlight the GO-Academy—a game changer for successfully applying OR within our company. We conclude by describing our findings on applying OR at TNT Express.

B.3 Subprogram 1: TNT Express Routing and Network Scheduling (TRANS)

Network optimisation is concerned with optimising routes for the transportation of packages and vehicle tours. Within TRANS, our optimisation software, the network infrastructure (i.e., depots and hubs), the flows of packages to be transported, and the cut-off times are considered to be fixed. A transportation route defines the sequence of hubs, from the depot of origin through to the destination depot, including scheduled times of arrival and departure at the hubs, that a package will visit. A path is the simplification of a route, denoting only the sequence of hubs and excluding time information. Thus, multiple routes can operate along the same path. A tour describes the sequence of locations visited by a vehicle (and driver), including the times at which each location is visited. In general, tours start and end at the same location for the convenience of the driver. A movement connects two successive locations of a tour, with no intermediary stops. Characteristics of a movement are its departure and arrival times and the corresponding vehicle type. An empty movement is a repositioning of a vehicle that is not carrying packages. Figure 38 illustrates these definitions.

Figure 38: The path and route illustrate the movements of packages from their origin (e.g., Turin) to their final destination (e.g., Florence) and the vehicle movements. The path is the subset of a route that excludes drop-off times. A movement connects two successive locations and is a subset of a tour.

Because of the size of the networks that TNT Express operates, using a combined problem to determine the routes and tours would be too complex. Path generation is exponential relative to the number of locations; Italy's domestic network has about 100 depots and 10 hubs, resulting in over 35 billion possible paths for packages. Therefore, we separate the problem into several subproblems, each supported by a specific module in TRANS:

- The **service capability analyser** determines the fastest feasible routes based on the prespecified movements in the network. The resulting fastest possible service offerings are visualised on a map. Service implications of modifications to the movement scheme (i.e., the total set of movements operated in the network) are recalculated within a few seconds to support what-if analyses.

- The **routing module** is an extension of the service capability analyser. It generates a set of routes (not only the fastest), and assigns the packages to the movements of these routes. If packages cannot be assigned to movements, this is usually because of insufficient capacity, an issue that must be resolved. Conversely, if movements are underutilised, opportunities for improvements may exist. The routing module visualises the overutilisation and (or) underutilisation of movements; however, it does not automatically resolve these issues.
- The **movement heuristic module** constructs a new movement scheme based on the packages and their corresponding paths and assigns them to the resulting movements. The movement heuristic module can be used in combination with the optimal paths module to generate a new movement scheme, and can also be used to evaluate existing movement schemes.
- The **optimal paths module** determines the optimal paths for each package, given the current infrastructure of depots and hubs; it also considers service restrictions. In combination with the movement heuristic, this module supports the redesign of a network's arcs in its entirety.
- The **tour generation module** generates tours based on a movement scheme, including any empty movements. In general, empty movements are minimised to ensure that the tours generated are efficient.

All the above modules were designed and tested in close collaboration with the CoPs, who provided the business knowledge and requirements. Each module can be used as a standalone module or in combination with other TRANS modules; the results can be analysed via a range of graphical visualisations and key performance indicators (KPIs).

B.3.1 Operations Research techniques used in the TRANS modules

The problem solved by the service capability analyser and routing modules can be formulated as a multicommodity flow problem (Ahuja, et al., 1993) in a time-space network, because of the required connectivity of movements at all locations. Root & Cohn (2008) describe the setup of this time-space network; a node (f_i,t_i) corresponds to a facility (i.e., a depot or hub) f_i at a point in time t_i, and an arc between (f_1,t_1) and (f_2,t_2) represents the flow of packages from facility f_1 to facility f_2, departing at facility f_1, at time t_1, and arriving at facility f_2 at time t_2. Each arc represents a movement. To solve the problem within a reasonable run time, irrespective of network size, we use a heuristic approach. For route generation, we apply a branch-and-bound algorithm to generate a user-defined number of routes that will meet the service requirements (note that the service capability analyser needs only one route per origin-destination pair). The bounding rules are a combination of the number of hub touches (as low as possible), the arrival time at the destination depot (as early as possible), and the departure time at the origin depot (as late as possible). The branches result from the movement scheme; the occurrence of these branches is restricted to hub locations only, because the transfer of packages from one movement to another is permitted only at hub locations.

In addition to generating the routes, the route module assigns the packages to the routes that are generated. The assignment of packages to routes is based on route preferences and available capacity. In particular, the preferred route is taken for each set of packages (where the preference follows the same rules as the bounding rules), and the packages are allocated to route movements only if available capacity exists; if no capacity remains in any of the route movements, the second most-preferred route is selected; this assignment process is repeated until all packages are assigned to movements or until all routes have been evaluated.

The optimal paths module solves what the literature refers to as the tactical service network design problem. Crainic (2000) provides an overview of solution methods to solve this problem. At TNT Express, we implemented a mixed-integer programming formulation as proposed in (Meuffels, et al., 2010), where the number of paths to be generated is restricted. The movement heuristic module generates movements based on the paths of the packages using the heuristic that Meuffels et al. (2010) suggest. In the heuristic, any of the following three rules are used to schedule a movement: (1) when all packages are available at the departure location, (2) when a full vehicle movement can be created, or (3) because of the time restrictions. To generate tours based on a movement schedule, the tour generation module uses a set partitioning approach, as described in (Van Krieken, 2006).

B.3.2 Implementation of TRANS

Prior to implementing the TRANS solution, TNT Express analysts were using spreadsheets to conduct network optimisation analyses. However, because of the large size and complexity of the networks, they could analyse only small parts of the networks, which inevitably led to suboptimal solutions. Given that our analysts had been using spreadsheets for many years, we were reluctant to completely change their ways of working. Instead, we decided to develop a solution that was close to their spreadsheet environment, extended with several decision support modules to enable the analysts to work faster and more effectively on improving the performance of the networks.

The requirements and business logic for TRANS were developed and discussed in the CoPs and built by ORTEC and Tilburg University. The first modules developed were the service capability analyser and the routing module, which our analysts used to identify, visualise, and apply step-by-step improvements to our networks. This step-by-step approach led to strong user acceptance for the tool. The GO-Academy and the central GO-team were important enablers in creating awareness of and implementing this new way of working to improve network performance in our business units worldwide. TRANS has about two dozen users (members of the central network analysis team of a business unit or country) worldwide. As our analysts became more familiar with using quantitative models, we gradually increased the level of optimisation complexity deployed. Working with TRANS has become a best practice at TNT Express.

B.3.3 Benefits of TRANS

By using KPIs of the whole network, we gained many insights that we could not gain by using spreadsheets. By carefully considering factors such as empty kilometres and movement utilisation, the analysts were able to search for the most cost-effective means of transporting the packages.

The service capability analyser was frequently used to compare our commercial service offering and the capabilities of a network. Figure 39 shows an example of service improvements for the Barcelona, Spain depot. Our analysts discovered that some service offering deadlines were too tight, resulting in low levels of customer service. By analysing these types of scenarios in advance, we were able to avoid selling unachievable services to customers.

With the routing module, analysts can quickly evaluate changes to the movement schedule and easily apply any adjustments (e.g., alter a movement's time frame, adjust the vehicle type of a movement, change a movement's start or end location, or remove a movement from the schedule). Because it can generate more efficient tours, the tour generation module substantially reduced subcontracting costs and CO_2 emissions. We recently introduced the optimal paths and movement heuristic modules, which we are currently piloting in Italy; the preliminary results are encouraging.

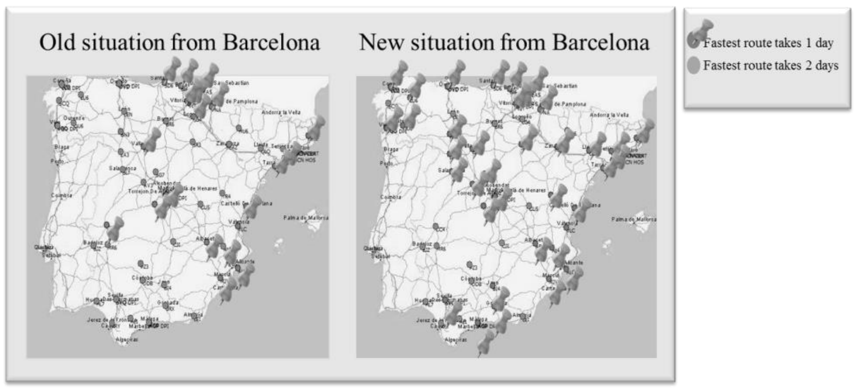


Figure 39: The service capability analyser can be used to improve service capabilities, as this figure illustrates for the depot in Barcelona, Spain. On the left, the figure shows the old situation (i.e., prior to using this analyser) with the locations for which a one-day service from Barcelona is available; on the right, the figure shows the new situation (i.e., after using the analyser) in which more locations and regions have a one-day delivery service available from Barcelona.

Thus far, we have used TRANS to analyse approximately 15 road networks. From 2008–2011, we accrued cost savings of 48 million euros, reduced kilometres driven by 69 million, reduced CO_2 emissions by 44 million kilograms. Some countries contributing to these savings used only the service capability analyser and routing modules—not the full suite of modules described above. Based on the achievements to date, we anticipate additional savings in the near future, particularly as the use of the more advanced optimisation modules becomes more widespread.

In addition to these quantitative benefits, we achieved a number of qualitative improvements, particularly in the area of service provision. This type of solution reduces the analysts' workload and provides them with opportunities to improve their analysis quality and their way of working. An example is the creation of annual schedules. Prior to using TRANS, analysts created a single annual schedule that included exceptions during peak periods. With TRANS, analysts can now create multiple schedules, including ones that can cope with the volume differences between workdays and weekends. Finally, TRANS changed the mindset of both analysts and managers: each group now focuses on searching for additional opportunities to optimise the network.

B.4 Subprogram 2: tactical planning in pickup and delivery (SHORTREC)

At TNT Express, PUD, which is planned at the depot level, impacts the first and last mile in the supply chain. Given that PUD accounts for more than 30 percent of operational costs, it is an important focus area of the GO-Program. At TNT Express, a round corresponds to a single vehicle starting at the depot, visiting customers in a certain sequence for collection or delivery of packages, and returning to the depot. Customer PUD rounds are determined during tactical round planning. Effectively organising the PUD process is challenging because millions of packages must be picked up and delivered each week.

The optimisation problem in the PUD process is to minimise the total pickup and delivery costs, while meeting all service level requirements. This implies minimizing the number of rounds (fixed cost) and the kilometres and hours driven for each round (variable cost). In PUD optimisation, the depot locations and their cut-off times are fixed. Constraints that must be considered are: vehicle capacity, service levels, driver regulations, and some softer constraints to ensure repetitiveness in the rounds and workload balancing. From an operational point of view, it is important to ensure that the daily pickup and delivery rounds remain consistent to prevent (1) disruptions to the sorting and loading processes at the depots, (2) increased workload, and (3) potential errors. The creation of consistent rounds increases customer satisfaction because the same driver visits the customer each time and can establish a positive working relationship with that customer. However, generating and executing similar rounds for each day of the week is difficult because of volume fluctuations. To deal with these challenges and support our analysts in PUD optimisation, we implemented a modified version of ORTEC's advanced vehicle routing and optimisation software, SHORTREC.

B.4.1 Operations Research techniques in SHORTREC

In logistics, the problem of generating rounds at minimum cost is known as the vehicle-routing problem (VRP), which Dantzig & Ramser (1959) studied first. Golden et al. (2008) provides a more recent overview. Given that the VRP problem is NP-hard, only small instances can be solved to optimality. Because of our problem sizes (e.g., the Rome depot handles 90,000 stops per week), combined with the additional nonstandard constraints to enhance productivity at the sorting facilities, we selected a heuristic optimisation approach in SHORTREC.

To ensure that the rounds are sufficiently robust to handle package flow fluctuations and to ensure a balanced workload across rounds, we introduced the concept of μ-zones*.* A μ-zone is a geographical area comprising a set of customer visits and a total average working time (i.e., drive time plus stoppage time) within a specific time bucket (e.g., an hour). Our preference is to establish visually attractive (i.e., nonoverlapping and convex) μ-zones. In the PUD process, a round traverses a series of neighboring μ-zones; each customer location in the μ-zone is visited, while driver regulations are satisfied. Volume fluctuations lead to changes in working time within a μ-zone (and, as a result, the working time of the round) and the utilisation of the vehicles. By reassignment of μ-zones to rounds, the change in working time of the round can be absorbed with minimal change to the overall round structure, preventing large changes in the depot sorting process. In cases of large fluctuations in packages, new rounds are added or removed. The μ-zones are created using the *k*-

means clustering algorithm (Kärkkäinen, 2006) and can be clustered into separate, geographic territories, each one served by a single PUD round (see Figure 40).

To ensure that a PUD round is both feasible and cost-effective, the round is evaluated in detail and optimised using the local search improvement algorithms in SHORTREC. The round might be improved by changing the sequence of customer visits within or between the μ-zones. If required, μ-zones can be exchanged between rounds. For an extensive description of the improvement algorithms used in SHORTREC, refer to (Kant, et al., 2008).

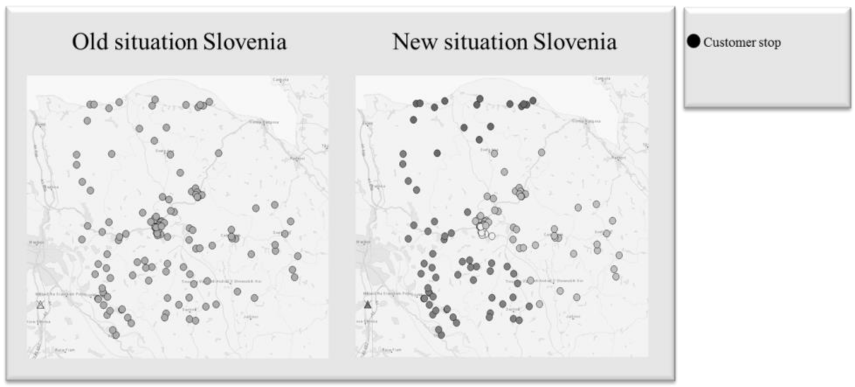


Figure 40: The figure shows μ-zone construction in SHORTREC for a part of Slovenia. On the left side, the old situation (i.e., prior to using SHORTREC) shows the stops to be clustered; on the right side, the new situation (i.e., after using SHORTREC) shows the resulting clusters—each with a different shading.

Because of the size of the problem instances, calculating distances and driving times was another challenge. The normal procedure is to calculate the distance and driving times, store the results in memory, and construct the rounds. However, this process requires a huge amount of memory (recall that the Rome depot has 90,000 stops per week). By using the concept of highway-node routing (Schultes, 2008), a method that exploits the layered structure of digital road networks, we were able to reduce computation times and memory consumption, allowing for on-demand calculation of distance and driving times. As a result, depot managers and analysts can optimise large instances on a normal desktop computer using SHORTREC.

B.4.2 Implementation of SHORTREC

Given that TNT Express has more than 2,000 depots, it was infeasible to implement SHORTREC in all locations at short notice. Therefore, we set a goal of optimising all the rounds in the depots of our main business at least once a year. Local staff and the central GO-team implemented the optimisation projects. To build trust and overcome resistance, the first step in the standard SHORTEC implementation procedure was to model the existing round structure of a depot. Supported by the graphical capabilities of SHORTREC, a member of the central GO-team demonstrated to the depot manager that the SHORTREC results using the current rounds were comparable to the actual costs, kilometres driven, and round structure. Schedule improvements were then generated using a standard set of optimisation scenarios. This standardised approach was developed in conjunction with the CoP members, carefully documented, and tested in various countries. The set of scenarios describes different ways of working in PUD and contains various scenarios, including evaluating combined PUD rounds, combining the pickups and deliveries of different types of packages (with regard to service and volume), or analysing the cut-off times at the depot. Each depot must apply this standard set of scenarios, allowing each to reach a high level of optimisation. This standardised implementation approach enabled us to rapidly deploy SHORTREC and disseminate PUD optimisation knowledge within the organisation.

B.4.3 Benefits of SHORTREC

SHORTREC has been deployed successfully in many European countries, including the UK, Germany, The Netherlands, Belgium, Italy, France, Spain, Austria, Denmark, Norway, Sweden, Portugal, and Greece, and has been used in various optimisation projects worldwide. This deployment is ongoing because our ultimate goal is to optimise each depot at least once a year.

During 2008–2011, 6 percent of the depots in Europe (260,000 rounds) have been optimised, resulting in 25 million euros in cost savings and an estimated accumulated CO_2 reduction of 11 million kilograms. With SHORTREC, we are now able to weigh the additional cost of creating visually attractive PUD rounds against the advantages that can be achieved in the sorting process at the depots. Moreover, because of the improved quality of the rounds, customer service has improved. Furthermore, daily improvements, in terms of our ability to cope with volume fluctuations, have been achieved as a result of the generated μ-zones. By using OR techniques, we are able to easily adapt decisions on the structure of rounds to absorb volume fluctuations by simply exchanging μ-zones between rounds.

B.5 Subprogram 3: supply chain optimisation (DELTA Supply Chain)

Because of the worldwide financial crisis at the end of 2008, TNT Express faced a strong decline in volumes, which continued until mid-2009. This drop in volume and associated revenue brought about an abrupt decline in air network performance, a problem that required an immediate solution. Because the air network forms a crucial part of our global service offering, we were impelled to start an end-to-end supply chain optimisation project that would reduce aircraft use, preserve future growth capabilities, but not worsen service. Based on the achievements of the GO-Program up to that time, we realised that we needed a tool that could support us in making strategic decisions and bring fact-based decision making to the board room. We decided to build the DELTA supply chain model, which would include every relevant detail of our supply chain and focus on reducing aircraft use as its first priority. Using the results of this model gave us the insight that we could decommission 12 of 59 airports and open 1 new airport, thus significantly reducing air transportation costs with minimal impact on customer service. More importantly, the results enabled us to survive the financial crisis. Stimulated by this success, the DELTA supply chain model became an important instrument in the development of our board's Vision 2015 strategy.

The DELTA model enables us to optimise our complete supply chain for a fixed depot and hub infrastructure, under varying volumes and ways of working (e.g., cut-off times, road and air transport). To the best of our knowledge, this is the only model in the express delivery industry that covers a complete air and road supply chain. We decided not to build one integrated model, but to design specific submodules to separately optimise the key components of the supply chain. By choosing to model the supply chain in this way, we could more easily understand and rely on the model's capabilities, which led to increased support for the decision-making process.

B.5.1 Operations Research techniques in the DELTA SC model

A typical DELTA model run begins with a volume-demand scenario. Cut-off times are imposed to secure the times required for both the PUD and network processes. The model aims to use road transport rather than air transport because the former generally results in lower costs and CO_2 emissions. The road network model determines the shortest paths for the packages by using the locations that may be visited as input. To incorporate network timing effects, we use hub time windows for the sorting activities, setting the latest arrival and earliest departure times to and from the hubs. Based on these time windows, we are able to determine if the service requirements for the specified packages can be met via road transportation. For packages that can be shipped by road, the number of required movements is calculated based on the routings of the packages. Any empty movements are determined using a classical transportation model that calculates the number of empty repositioning movements to estimate the cost of repositioning.

For the packages that are unable to meet the service requirements via the road network, we construct an air network using a separate model to create a minimum-cost air schedule between the airports in the network. The model starts by assigning depots to the airports in the air network based on one of two criteria: (1) best service (i.e., latest departure from and earliest arrival at the depots, based on the predefined earliest departures or latest arrivals at the airports), or (2) lowest cost (i.e., shortest distance between the airport and depot). Based on these assignments, the model determines the packages to be transported from the airport to the air hub and vice versa. Next, a mixed-integer programming problem is solved to determine the minimum-cost air schedule. The model ensures that sufficient aircraft capacity is available to carry all the packages, and it balances the number of incoming and outgoing aircraft per aircraft type at each location. The aircraft that can be used are restricted by a minimum and maximum number per aircraft type and the aircraft operating characteristics, such as maximum flying range, effective speed, cargo capacity, and landing restrictions. For airports, the model includes the following: the earliest permitted arrival or departure times, airport closing times, and the consideration that some airports do not permit multiple stops by TNT Express airplanes. Finally, for the air hub, the sorting time window is included, setting the latest arrival and earliest departure time for the aircraft, and the runway capacity at the air hub.

This model is based on the work of Armacost et al. (2002), with some additions to capture the specifics of the TNT Express operation. One main difference is the restriction of the number of stops at an airport. At some airports, we strongly prefer that all packages arrive and depart via one aircraft because this simplifies the handling process. A second difference is the inclusion of more detailed modeling of the runway capacity at the air hub. Instead of restricting the total number of arrivals and departures at the air hub, we would like to position them across time. Our approach is similar to the work of Barnhart & Schneur (1996). We also include the functionality to use a minimum or maximum number of aircraft per type, which may be used in the European Air Network, to cope with restrictions on fleet availability. For some situations, we even would have functionality to fix specific aircraft operations between two airports in the network. This request originated because the European Air Network is sometimes used in a combined setting for domestic network operations (in which TNT Express-owned aircraft are used).

With the road and air network complete, a binary integer programming model estimates the impact of the network movement arrival and departure times on the PUD cost. This model determines the optimal wave structure of every depot. We define a wave as a set of rounds that start and end at the same time at the depot. In the case of multiple arrivals at a depot, starting part of the rounds before all packages have arrived might be beneficial because this ensures that the rounds have a longer working day. This is particularly useful when packages destined for nearby customers might arrive at the depot at different times. If so, multiple rounds will have to be assigned to the same location area, resulting in larger average distances between stops. Similar advantages and disadvantages apply to collections whereby multiple departures occur in the network. The binary integer model determines the optimal number of waves required to balance both the length of rounds in a wave and the extra kilometres to be driven. In a final step, the model calculates the total cost and service KPIs of the complete supply chain to support management decision making.

B.5.2 DELTA SC model implementation challenges

Because of the board's tight schedule to rationalise air operations in 2008, substantial work was required within a very short time frame. Some team members were requested to completely clear their agendas of other activities for a number of weeks. A cross-functional strategic operations team, including modelling and optimisation specialists from both ORTEC and Tilburg University, was set up. The team monitored the progress of the project and became the platform for discussion and agreement on the many issues encountered during the project. The development of the DELTA supply chain model was a difficult task because the model covers all Europe and consists of about 650 depots, 90 hubs, and 150,000 origin-destination combinations. The challenge was to provide the right level of detail to support the board in its decision making, but not so much detail so as to render the results useless and distract the board from crucial insights and decision factors. Fortunately, work done in the previous years in each GO-Subprogram provided us with the experience we needed to agree on the relevant details of the model. However, because many team members were not yet familiar with strategic modelling, some members wanted to incorporate much more detail than was needed. As a result, a great deal of discussion and salesmanship ensued so that the team members would agree on the appropriate (strategic) level of detail that would be acceptable to each stakeholder.

Data gathering was another major challenge. Our experience led us to believe that the data gathering and verification exercise for this type of analysis could take months, even for one country. We were tasked with acquiring data for all Europe within only eight weeks.

Convincing people to execute the model was not an issue, because its major users were people in the strategy department of GO and consultants of ORTEC; however, building the decision makers' trust in using the model results was a challenge. They often did not immediately accept the initial results, mainly because certain details had been omitted from the calculations; however, more importantly, the new insights gained from the DELTA supply chain model violated their prior beliefs about operating the TNT Express supply chain. To show that the model's results were relevant and consistent, we formulated a number of scenarios, evaluated them, and explained them in detail. For example, to determine which airports to close, we generated and optimised more than 20 scenarios. Although the model showed consistent results for these scenarios, the air network analysts recalculated the results with very similar outcomes. As trust in the DELTA supply chain model grew, the TNT Express board became more confident in the results and decided to move forward based on the insights gained. It implemented the model results and used them to build its Vision 2015 strategy**.**

B.5.3 Benefits of the strategic analyses

The DELTA supply chain model has become a vital instrument for generating and analysing air network optimisation scenarios for various airport compositions. In 2008 and 2009, we opened 1 new airport and closed 12 of our 59 European airports; the impact on service levels was minor—less than 0.5 percent of the volume arrived more than one hour later (see Figure 41). Furthermore, we also eliminated six aircraft, three of which were expensive A300 aircraft. Naturally, this incurred some additional costs because of the longer distances driven between airports and depots. However, total net accumulated savings were 132 million euros and the CO_2 emissions reduction was 228 million kilograms. Achieving these reductions within such a short period improved our agility and ability to create value, even with volatile demand.

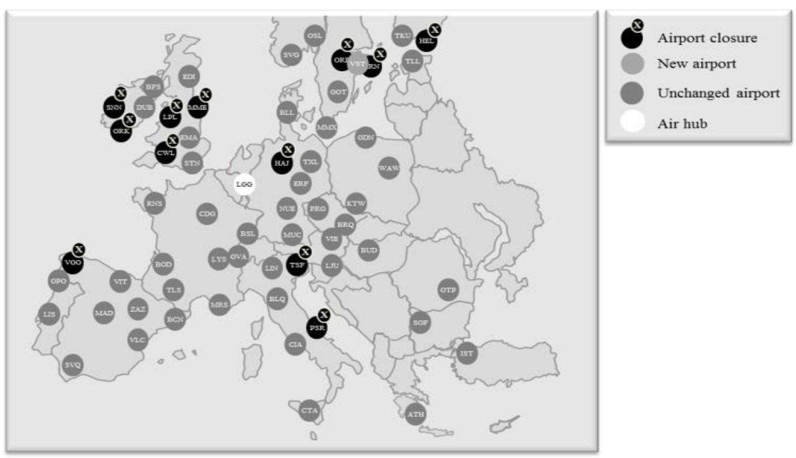


Figure 41: The figure shows the final result after management decisions; management decided to close 12 airports and begin TNT Express operations at 1 new airport.

To develop the TNT Express Vision 2015 strategy, a number of operating modes and European network designs were evaluated using the DELTA supply chain model. Various scenarios were analysed (e.g., separating parcel and freight volumes in Europe, reducing stop-time in PUD, altering the available time between PUD and network operations by varying the cut-off times, and investigating the robustness of our road and air networks). The insights gained from the DELTA supply chain analysis strongly contributed to the development of our strategic vision. Combining volumes (i.e., parcel and freight) resulted in a cost avoidance of 4 percent on supply chain costs. Although this insight was initially counter-intuitive, analysing the cost details resulting from the model convinced us of the correctness of the outcomes.

The DELTA supply chain model allows us to test our supply chain operation improvement ideas without having to experiment in practice. For example, when we used this model to calculate the impact of changing the cut-off times, it provided insights on the trade-offs between network costs and PUD costs that we would have to make. As a result, we initiated follow-up projects to analyse and change the cut-off times for potential new locations in our network. The model also indicated that substantial benefits could be achieved by increasing the capacity of certain depots and hubs. The reported benefits did not include the benefits of the Vision 2015 projects.

Finally, after observing the potential improvements that could be achieved by optimising the balance between the PUD and network processes, we started to develop models to optimise the depot and hub infrastructure. Initial results show that the savings potential is enormous; in the coming years, these savings will be achieved gradually because changing the infrastructure quickly is impossible. We will continue to use the DELTA supply chain solution regularly to search for large improvements.

B.6 General deployment challenges

Implementing tools such as TRANS and SHORTREC takes time and effort. Firstly, it takes time to convince analysts, network managers, and depot managers, most of whom are unfamiliar with optimisation, to adopt new tools that they might initially see as reducing their control of the analysis. Second, the available data, although numerous, were spread across many local information technology (IT) systems, thus reducing the quality and quick availability of the data required. Some data (e.g., the delivery or pickup address) must be detailed, and the data received are often either incomplete or incorrect. This problem was partially solved by introducing GO-Data Management, a data cleansing and conversion tool that makes the data retrieval from our IT systems repeatable. In particular, this tool describes a set of business rules to map the source data onto the GO-Data Structures; this is an evolving process because business rules or data structures sometimes change. Furthermore, we realised that significant effort is required to create a user-friendly and fast model that supports the analysts in evaluating various planning scenarios and understanding the differences between them.

When developing decision support solutions, we knew that we would be setting the standard for all countries in which TNT Express operates. This proved challenging because these countries have many ways of working, variances in volume profiles, and local regulations, any of which could impact our creation of a standard model. For example, TNT Express uses loose-loaded trucks in Italy, pallets in France, and cages in the European road network. TRANS had to support all these requirements; in addition, we had to find the right balance between generic and country-specific requirements in each instance.

Another challenge we faced was objectively tracking results. Managers who successfully meet their optimisation targets often receive a budget decrease in the subsequent year. This could affect the quality of the submitted results, especially because the managerial bonus system is based on the budget targets. Even a slight deviation from the optimal solution in practice may increase cost, making it difficult to attain the bonus targets. Therefore, to measure the results objectively, we introduced a benefits tracking system that uses three levels of savings to monitor the benefits: identified expected savings, agreed savings, and implemented (realised) savings. The savings presented in this chapter are the implemented (realised) savings; cost avoidance is not included.

B.7 GO-Academy

Most new users were resistant to change, even when we could demonstrate successes in nearby operating units, because they felt that their business was not comparable. Furthermore, the task of explaining the general optimisation principles of the tools was time-consuming. These two challenges led us to implement the GO-Academy, a unique learning concept in optimisation.

The main objective of the GO-Academy is to teach optimisation principles to TNT Express employees and to acquaint them (at a high level) with the available optimisation tools, without turning them into mathematicians. We discovered that one of the most important lessons we can teach is that investing a little in one part of the supply chain can result in large benefits in other parts. Another principle we taught is that to strategically optimise a supply chain, considering all details is unnecessary. Third, we teach that the combinatorial explosion forms the basis of many frequently encountered planning problems. We use various methods to teach these principles; these range from conceptual explanations and practical assignments to simple but powerful computer games. An example of the latter is the GO-Game, in which a tactical and strategic solution for an express network must be constructed; (Meuffels, et al., 2010) contain a description of this game. To date, over 500 managers and staff, including the TNT Express board of directors, have successfully completed this game.

The GO-Academy training program consists of six three-day modules, conducted over a period of two years, interspersed with small group assignments (see Table 21). After a group completes each module, the group composition is changed to promote networking—a key benefit of the academy. After completing the fifth module, students are given two days per week for six months to complete their final assignment, a master case study. The case study, sponsored by one or more senior managers and guided by an academic supervisor, is based on an actual challenge at TNT Express. Students present the master case results on graduation day. Within the operations arena of TNT Express, this day has become a big networking event, which most of our senior managers and our CEO, at times, attend.

In addition to optimisation skills, the GO-Academy training program focuses on the development of personal and interpersonal skills: presentation skills, debating ability, working on camera, and elevator pitches.

Graduating employees are designated as supply chain masters, an internal title that remains in effect as long as the employee completes at least one optimisation project per year. Project results, suggestions, and ideas are published on an internal website, Collaborate. These projects are judged as part of an annual supply chain master competition; on graduation day, an award is given for each category (i.e., networks, PUD, and supply chain).

Module	Name	Topics
1	Introduction module	Customers and their supply chains
2	Strategic optimisation	Infrastructure design (DELTA)
3	Networks and PUD	Planning in networks and PUD (TRANS & SHORTREC)
4	Hubs and depots	Bottleneck theory, mechanisation principles
5	Implementation	Change management techniques
6	Graduation	Presenting for impact, elevator pitches

Table 21: The six GO-Academy modules train participants in optimisation principles and practices used by TNT Express. The table lists the modules and the topics included in each module.

Since its inception, the GO-Academy has successfully met its main objective—to teach the principles of optimisation. We frequently find that it is unnecessary to convince people of the benefits of optimisation because they are coming to us for support and advice. This is a definite turn of events and one that proves the effectiveness of the training academy concept.

Furthermore, the supply chain master case studies have delivered a number of significant benefits in several areas. In addition to cost savings of approximately 5.7 million euros, networking within TNT Express has improved dramatically. After collaborating on assignments during their two years in the GO-Academy program, employees build strong relationships with each other; they also feel more empowered to ask for support from colleagues in other operating units or even in other parts of the world. Employees now use the same business language and have the same understanding of definitions and terminology (e.g., cut-off time, μ-zone). At TNT Express, it has become apparent that platforms such as the GO-Academy are ideal arenas for discussing and explaining the operational implication of strategy changes. The GO-Academy has exceeded our expectations.

B.8 Transportability

The above lessons are applicable and transportable to any organisation that wants to apply and embed OR on a large scale. To illustrate the transportability of our approach, we outline our contribution to the World Food Program (WFP), the world's largest humanitarian aid organisation, which feeds more than 90 million of the poorest people on earth. In conjunction with the WFP, we developed a simple hub-and-spoke network for food distribution for Ethiopia. The feeding of children in more than 2,000 schools in Liberia has been optimised with SHORTREC and has yielded 10 percent savings on transport costs. We are proud that the WFP is actively involved in our CoPs and GO-Academy. We currently are sharing our core ideas of supply chain optimisation with the WFP strategic logistic team in Rome.

B.9 Challenges

The introduction of optimisation to TNT Express has been and will continue to be an intensive process with many challenges. Of course, data availability and quality are always an issue in an OR project. In addition, more decision makers need to become familiar with optimisation principles, especially in cross-functional areas such as marketing, sales, and finance. Applying OR principles at the board level brings another set of challenges. Given the involvement of people from varying backgrounds, we required a great deal of discussion to determine the right level of detail for strategic modelling. Gaining acceptance of our approach required a high level of didactical skills and salesmanship. The most significant challenge related to the time pressure because of strict time lines (e.g., shareholder meetings, executive board meetings) and short decision time frames.

B.10 Concluding remarks

The best way to apply OR in a business that is unfamiliar with the concept of optimisation is to start simple and follow the maturity level of the business in applying more advanced methods. Forcing the use of OR—if it is not well understood—will only increase resistance and decrease user acceptance. The ability to visualise the business challenge contributes significantly to lowering resistance because users are able to more easily recognise the challenges. Using basic OR techniques from the start makes possible the building of trust and understanding, and improves data quality. Most of the savings and CO_2 reductions discussed in this chapter can be attributed to these basic OR solutions.

As soon as people become more familiar with OR modelling, advanced techniques (e.g., scenario analysis, simulation, and mixed-integer programming) can be introduced. The development of OR models and tools in close collaboration with the business leads to increased trust and acceptance by end users. We facilitated model and tool development by introducing CoPs, in which our subject matter experts and external OR experts derived and honed the requirements of the solutions to be developed.

Parallel to developing tools, we realised the importance of teaching employees the fundamentals of optimisation to encourage the dissemination of knowledge and allow for fast implementations. As a result of the GO-Academy, a huge network of people, who recognise optimisation possibilities and whose knowledge the company can easily tap into, exists within TNT Express. We initially set up our central GO-team with 5 people, and currently have about 30 people whose full-time jobs involve applying OR at TNT Express; senior management and over 200 supply chain masters support them. Optimisation has become part of the core values at TNT Express.

Glossary of terms

Term	Description
AIMMS	*AIMMS* is an optimisation modelling system with integrated GUI-builder.
Allocation	The term *allocation* or *assignment* is used to refer to the interaction between depots and hubs in network design; the term *single allocation* denotes that a depot is served by a single hub in the network while the term *multiple allocation* is used to denote that the transport of packages from and to the depot is arranged by a multiple of hubs.
Arc	See *tour* in the context of road transport and *flight* in the context of air transport.
Assignment	See *allocation.*
B2B	*B2B* refers to the *business-to-business* segment, which is originally the main segment in which express service providers operate.
B2C	*B2C* refers to the *business-to-consumer* segment, which is an upcoming industry in the CEP-market.
Balancing	See *repositioning.*
CAB-data	The well-known Civil Aeronautics Board (CAB) dataset, which is a dataset introduced by O'Kelly (1987).
CEP-market	The *Courier, Express, and Parcel-market,* abbreviated as *CEP-market* or *CEP*, is the industry that concerns the collection, transport and distribution of packages, and can be segmented along the dimensions of time and weight.
Community of Practice	See *GO-Program.*
Complete hub network	See *incomplete hub network*.
Connection	See *tour* in the context of road transport and *flight* in the context of air transport.
Consignment	The total shipment of a sender to a receiver that is handed to the express company under agreement of a service. A consignment may consist of multiple pieces of different shapes.

Cut-off time
A cut-off time is used to define the end of one process and the start of the next. In general, the cut-off time is used to make the distinction between the PUD process and the line-haul process. The *origin* or *pickup cut-off time* denotes the time at which the pickup process ends and the line-haul process starts; the *destination* or *destination cut-off time* denotes the time at which the line-haul process ends and the delivery process can start.

Delivery
See *pickup and delivery*.

Delivery cut-off time
See *cut-off time*.

Depot
Depots consolidate the packages from senders and take care of the delivery of packages to the receiving customers.

The depot that organises the pickup of consignments at the customer is referred as the *origin depot*. The depot that organises the delivery of consignments at the customer is referred as the *destination depot*.

Destination cut-off time
See *cut-off time*.

Destination depot
See *depot.*

Direct
A *direct* refers to the transport from depot to depot, bypassing all hub locations.

Document
Smallest type of shipment (<2 kilogram).

Economy service
See *service.*

Empty movement
See *repositioning.*

Express carrier
See e*xpress service provider.*

Express service provider
An *Express service provider* or *express carrier* is a company that takes care of the door-to-door transport of packages under service level agreements.

Express service
See *service.*

Flight
A *tour* describes the sequence of locations visited by an aircraft (and crew), including the times at which each location is visited. A *flightleg* connects two successive locations of a flight with no intermediary stop.

Characteristics of a flightleg are its departure and arrival time and corresponding vehicle type. A *leg* or *arc* or *connection* is the simplification of a flightleg without timing information, and in such describes the existence of a transport possibility between two consecutive locations.

Flightleg See *flight*.

Flow Term used to refer to a bundle of packages, for example the total of packages that is transported from one depot to another depot with a predefined service is known as the *ods-flow.*

Freight Heavy weight shipment (>30 kg).

GO-Academy See *GO-Program.*

GO-Program The *Global Optimisation-Program* is a program launched at TNT Express to assist the operating units of TNT Express to improve their package delivery in road and air networks. Within this program, *Communities of Practice (CoPs)* are established at which internal and external subject matter experts meet three times annually at an internal conference. Besides, two-year learning environment is created at TNT Express, which is known as the *GO-Academy*.

Hub Hubs are large sorting facilities used to consolidate the transport of packages between the depots.

Incomplete hub network A *complete hub network* refers to a network design in which transports between all pairs of hubs is arranged. In the situation of an *incomplete hub network*, some pairs of hubs have no direct interaction.

Leg See *flight*.

Line-haul *Line-haul network* or *network* or *long-haul*; the longer-distance transport between origin and destination depot locations of packages.

Long-haul See *line-haul*.

Milk-run See *stopover*.

Movement See *tour*.

Multiple allocation See *allocation.*

Multiple assignment See *allocation.*

Network See *line-haul*.

Normal service See *service.*

ODS-flow	See *flow.*
ODS-pair	The indicator for an origin, depot, and service combination.
Origin cut-off time	See *cut-off time*.
Origin depot	See *depot.*
ORTEC	ORTEC is an OR consulting and optimisation software provider that partners with CEP-related business on optimisation activities.
Package	A general reference for any type of consignment that is shipped in the network.
Parcel	Most common type of shipment (between 2 and 30 kilogram); this term is sometimes used as a general reference for any type of consignment that is shipped in the network.
Path	See *route.*
Pickup	See *pickup and delivery*.
Pickup and delivery	*Pickup and delivery (PUD)* or *collection and distribution (C&D)* refer to the short-distance transport between the customer and the depot location. The pickup or collection concerns the first transport action in the overall transport service of a consignment, while the delivery or distribution refers to the last part of the service offered. The latter part is also known as the *last mile*.
Pickup cut-off time	See *cut-off time*.
Piece	Single unit of a shipment.
Premium service	See *service.*
Repositioning	*Repositioning* or *balancing* concerns the rearrangement of vehicles or aircrafts to enable execution of a schedule. Note that these vehicles or aircrafts do not carry any packages and this operation is therefore often referred as an *empty movement*.
Route	The transportation *route* of an ods-pair defines the sequence of hubs to be visited by a package, from the depot of origin through to the destination depot, with scheduled times of arrival and departure of the hubs. A *path* is the simplification of a route, denoting only the sequence of hubs and excluding timing information. Related: *direct* is used to denote a route from depot to depot.

Sector	See *tour*.
Service (level agreement)	A *service level agreement* specifies the delivery data and time that is offered to a customer. *Premium* or *premium service* refers to the fastest and most expensive service offered. *Normal* or *normal service* refers to the more commonly used services being express and economy services. *Express* or *express service* refers to the fastest variant of the common services. *Economy* or *economy service* is the more economically beneficial variant of the common services. The difference between express and premium is in general a difference in number of hours (e.g. premium service delivers before noon, express before end of business day), while the difference between express and economy is in general a difference in number of days (e.g. express provides a next day service while economy services deliver in two days).
Shipment	This term is used to refer to the total transportation process of packages that are moved from a sender to a receiver.
Single allocation	See *allocation.*
Single assignment	See a*llocation.*
Stopover	A *stopover* or *milk-run* denotes that the transport of multiple depots is combined from and/or to a hub location.
Tour	A *tour* describes the sequence of locations visited by a vehicle (and driver), including the times at which each location is visited. A *movement* connects two successive locations of a tour with no intermediary stop. Characteristics of a movement are its departure and arrival time and corresponding vehicle type. A *sector* or *arc* or *connection* is the simplification of a movement without timing information, and in such describes the existence of a transport possibility between two consecutive locations.
Unit Loading Device	A *Unit Loading Device (ULD)* is a container used to load packages in single units that can be loaded at once in an aircraft.

Bibliography

A.T. Kearney, Inc, 2012. *Europe's CEP Market - Growth on New Terms,* Berlin: A.T. Kearney.

Aarts, E. & Lenstra, J., 2003. *Local search in combinatorial optimization.* Princeton: Princeton Unitveristy Press.

Ahuja, R., Magnanti, T. & Orlin, J., 1993. *Network flows: theory, algorithms, and applications.* NJ, USA: Prentice-Hall, Inc..

Alumur, S. & Kara, B., 2008. Network hub location problems: the state of the art. *European Journal of Operational Research,* 190(1), pp. 1-21.

Alumur, S. & Kara, B., 2009. A hub covering network design problem for cargo applications in Turkey. *Journal of Operational Research Society,* 60(10), p. 21.

Armacost, A., Barnhart, C. & Ware, K., 2002. Composite variable formulations for express shipment service network design. *Transportation Science,* 36(1), pp. 1-20.

Armacost, A., Barnhart, C., Ware, K. & Wilson, A., 2004. UPS optimizes its air network. *Interfaces,* 34(1), pp. 15-25.

Aykin, T., 1994. Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *European Journal of Operational Research,* Volume 79, pp. 501-523.

Aykin, T., 1995. The hub location and routing problem. *European Journal of Operational Research,* Volume 83, pp. 852-863.

Baltz, A., Dubhashi, D., Srivastav, A. & Werth, S., 2007. Probabilistic analysis for a multiple depot vehicle routing problem. *Random Structure Algorithms,* 30(1-2), pp. 206-225.

Barnhart, C., Johnson, E., Nemhauser, G., Savelsbergh, M., Vance, P., 1998. Branch-and-price: column generation for solving huge integer programs. *Operations Research,* 46(3), pp. 316-329.

Barnhart, C. & Schneur, R., 1996. Air network design for express shipment service. *Operations Research,* 44(6), pp. 852-863.

Barnhart, C. & Shen, S., 2004. Logistics service network design for time-critical delivery. In: *Practice and Theory of Automated Timetabling V.* s.l.:Springer Berlin Heidelberg, pp. 86-105.

Campbell, J., 1994. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research,* Volume 72, pp. 387-405.

Campbell, J., Ernst, A. & Krishnamoorthy, M., 2005. Hub arc location problems: Part II - formulations and optimal algorithms. *Management Science,* 51(10), pp. 1556-1571.

Campbell, J. & Krishnamoorthy, M., 2005. Hub arc location problems: Part I - introduction and results. *Management Science,* 51(10), pp. 1540-1555.

Campbell, J. & O'Kelly, M., 2012. Twenty-five years of hub location research. *Transportation Science,* 46(2), pp. 153-169.

Collins, N., 1988. Simulated annealing - an anooted bibliography. *American Journal of Mathematical and Management Sciences,* 8(3-4), pp. 209-307.

Cordeau, J., Gendreau, M. & Laporte, G., 1997. A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks,* 30(2), pp. 105-119.

Cormen, T., Leiserson, C., Rivest, R. & Stein, C., 2001. *Introduction to algorithms.* Cambridge: The MIT University Press.

Crainic, T., 2000. Service network design in freight transportation. *European Journal of Operations Research,* 22(2), pp. 272-288.

Crainic, T., 2002. *A survey of optimization models for long-haul freight transportation. Handbook of transportation science.* 2nd ed. : Kluwer.

Dantzig, G. & Ramser, J., 1959. The truck dispatching problem. *Management Science,* 6(1), pp. 80-91.

Dashkin, M., 1995. *Network and Discrete Location.* 2nd ed. New Jersey: John Wiley and Sons, Inc..

Dejax, P. & Crainic, T., 1987. A review of empty flows and fleet management models in freight transportation. *Transportation Science,* 21(4), pp. 227-248.

DHL, 2012. *DHL Annual Report.*

Eiselt, H. & Marianov, V., 2013. *Foundations of Location Analysis.* 1st ed. New York: Springer-Verlag New York Inc..

Essen, J. v., Meuffels, W., Aardal, K. & Fleuren, H., 2014. Heuristic for the uncapacitated hub location and network design problem with a mixed vehicle fleet and regional differentiation. *Submitted to a scientific journal.*

EUR-lex, 2006. *Regulcations (EC) No 561/2006 of the European Parliament on the Council of 15 March 2006 on the harmonisation of certain social legislation to road transport.*

Farahani, R., Hekmatfar, M., Arabani, A. & Nikbakhsh, E., 2013. Hub location problems: a review of models, classification, techniques and application. *Computers & Industrial Engineering,* 64(4), pp. 1096-1109.

Farvolden, J. & Powell, W., 1994. Subgradient methods for the service network design problem. *Transportation Science,* 28(3), pp. 256-272.

Fleuren, H., Goossens, C., Hendriks, M., Lombard, M.-C., Meuffels, I., Poppelaars, J., 2013. Supply Chain-Wide Optimization at TNT Express. *Interfaces,* 43(1), pp. 5-20.

Gelareh, S. & Nickel, S., 2011. Hub location problems in transportation networks. *Transportation Research Part E,* Volume 47, pp. 1092-1111.

Goldberg, D., 1989. *Genetic algorithms in search, optimization and machine learning.* s.l.:Addison-Wesley Longman Publishing Co., Inc.

Golden, B., Raghaven, S., Wasil, E. & (Eds), 2008. *The Vehicle Routing Problem: Latest Advances and New Challenges,* New York: Springer.

Holmberg, K. & Hellstrand, J., 1998. Solving the uncapacitated network design problem by a lagrangean heuristic and branch-and-bound. *Operations Research,* 24(4), pp. 247-259.

Kant, G., Jacks, M. & Aantjes, C., 2008. Coca-Cola Enterprises optimizes vehicle routes for efficient product delivery. *Interfaces,* 38(1), pp. 40-50.

Kara, B. & Tansel, B., 2001. The latest arrival hub location problem. *Management Science,* 47(10), pp. 1408-1420.

Kara, B. & Tansel, B., 2003. The single-assignment hub covering problem: models and linearizations. *Journal of the Operations Research Society,* Volume 54, pp. 59-64.

Kärkkäinen, I., 2006. *Methods for Fast and Reliable Clustering,* Joensuu: University of Joensuu (Dissertation).

Kim, D., Barnhart, C., Ware, K. & Reinhardt, G., 1999. Multimodal express package delivery: a service network design application. *Transportation Science,* 33(4), pp. 391-407.

Krajewska, M. & Kopfer, H., 2009. Transportation planning in freight forwarding companies - tabu search algorithm for the integrated operational transportation planning problem. *European Journal of Operational Research,* Volume 197, pp. 741-751.

Kuby, M. & Gray, R., 1993. The hub network design problem with stopovers and feeders: the case of Federal Express. *Transportation Research Part A,* 27A(1), pp. 1-12.

Leung, J., Magnanti, T. & Singhal, V., 1990. Routing in point-to-point delivery systems: formulations and solution heuristics. *Transportation Science,* 24(4), pp. 245-260.

Lin C-C, C. S.-H., 2008. An integral constrained generalized hub-and-spoke network design problem. *Transportation Research Part E,* Volume 44, pp. 817-848.

Lin, C., 2001. The freight routing problem of time-definite freight delivery common carriers. *Transportation Research Part B,* Volume 35, pp. 525-547.

Lin, C. & Chen, S., 2004. The hierarchical network design problem for time-definite express common carriers. *Transportation Research Part B,* Volume 38, pp. 271-283.

Louwerse, I., Mijnarends, J., Meuffels, I., Huisman, D., Fleuren, H., 2014. Scheduling movements in the network of an express service provider. *Flexible Services Manufacturing Journal*, Volume 26(4), pp. 565-584.

Lübbecke, M. & Desrosiers, J., 2005. Selected topics in column generation. *Operations Research,* 53(6), pp. 1007-1023.

Magnanti, T. & Mirchandani, P., 1993. Shortest paths, single origin-destination network deisgn, and associated polyhedra. *Networsk,* 23(2), pp. 103-121.

Magnanti, T. & Wong, R., 1984. Network design and transportation planning: models and algorithms. *Transportation Science,* 18(1), pp. 1-55.

Melkote, S. & Dashkin, M., 2001. Capacitated facility location/ network design problem. *Transportation Research Part A,* Volume 35, pp. 515-538.

Melo, M.-T., S, N. & Saldanha da Gama, F., 2009. Facility location and supply chain managament - a review. *Europan Journal of Operational Research,* Volume 196, pp. 401-412.

Meuffels, I., Fleuren, H., Cruijssen, F. & Van Dam, E., 2010. Enriching the tactical network design of express service carriers with fleet scheduling characteristics. *Flexible Services and Manufacturing Journal,* 22(1-2), pp. 3-35.

Meuffels, I., Fleuren, H., Poppelaars, J., Hoornenborg, H., De Rooij, F., 2010. The design of express networks in a nutshell - playing the global optimisation game (GO-Game). *OR News,* Volume 39, pp. 6-8.

Meuffels, I., Van Dijk, T. & Fleuren, H., 2014. Reduced Hub Handling in Multiple Hub Air Network Design for Express Providers by the Introduction of Pre-Sorted Unit Loading Devices. *Submitted to a scientific journal.*

Mirchandani, P., 1989. *Polyhedral Structure of a Capacitated Network Design Problem.* Dissertation ed. Massachusetts: Massachusetts Institute of Technology.

Mirchandani, P. & Francis, R., 1990. The Uncapacitated Facility Location Problem . In: *Discrete location theory.* s.l.:Wiley, p. 127.

Mitroff, I., Betz, F., Pondy, L. & Sagasti, F., 1974. On managing science in the systems age: two schemas for the study of science as a whole systems phenomenon. *Interfaces,* 4(3), pp. 46-58.

Mitrović-Minić, S. & Laporte, G., 2006. The pickup and delivery problem with time windows and transshipment. *INFORMS,* 44(3), pp. 217-228.

Ngamchai, S., 2007. *Air express network design with hub sorting,* College Park, USA: University of Maryland - Dissertation.

Nickel, S., Schobel, A. & Sonneborn, T., 2001. *Hub location problems in urban traffic networks. Mathematics methods and optimization in transportation systems.* s.l.:Kluwer Academic Publishers.

O'Kelly, M., 1986. The location of interacting hub facilities. *Transportation Science,* 20(2), pp. 92-106.

O'Kelly, M., 1987. A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research,* Volume 32, pp. 393-404.

O'Kelly, M., 1992. Hub facility location with fixed costs. *Pap Regional Science RSAI,* 71(3), pp. 293-306.

O'Kelly, M. & Bryan, D., 1998. Hub location with flow economies of scale. *Transportation Research Part B,* 32(8), pp. 605-616.

O'Kelly, M., Bryan, D., Skorin-Kapov, D. & Skorin-Kapov, J., 1996. Hub network design with single and multiple allocation: a computational study. *Location Science,* 4(3), pp. 125-138.

O'Kelly, M. & Miller, J., 1994. The hub network design problem. *Journal of Transport Geography,* 2(1), pp. 31-40.

Pedersen, M., Crainic, T. & Madson, O., 2009. Models and tabu search metaheuristics for service network design with asset-balance requirements. *Transportation Science,* 43(2), pp. 158-177.

Podnar, H., Skorin-Kapov, J. & Skorin-Kapov, D., 2002. Network cost minimization using threshold-based discounting. *European Journal of Operational Research,* Volume 137, pp. 371-386.

Revelle, C. & Eiselt, H. D. M., 2008. A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research,* 4(1), pp. 1264-1277.

Root, S. & Cohn, A., 2008. *A novel modeling approach for express package carrier planning.* [Online] Available at: https://deepblue.lib.umich.edu/bitstreem/2027.42/60965/1/20310_ftp.pdf
[Accessed 16 April 2012].

Schultes, D., 2008. *Route Planning in Road Networks,* Karlsruhe: University Fridericiana (Dissertation).

Simchi-Levi, D., Kaminsky, P. & Simchi-Levi, E., 2003. *Designing and managing the supply chain: concepts, strategies, and case studies.* 2nd ed. Boston: McGraw-Hill/Irwin.

Skorin-Kapov, D., Skorin-Kapov, J. & O'Kelly, M., 1996. Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research,* Volume 94, pp. 582-593.

Tan, P., Kara, B. & Tansel, B., 2007. The latest arrival hub location problem for cargo delivery systems with stopovers. *Transportation Research Part B,* Volume 41, pp. 906-919.

TNT Express, 2010. *TNT Annual Report.*

Van Krieken, M., 2006. *Solving set partitioning problems using Lagrangian relaxation.* [Online] Available at: https://arno.uvt.nl/show.cgi?fid=47051
[Accessed 16 April 2012].

Wasner, M. & Günter, Z., 2004. An integrated multi-depot hub-location vehicle routing model for network planning of parcel service. *International Journal Production Economics,* Volume 90, pp. 403-419.

Wieberneit, N., 2008. Service network design for freight transportation: a review. *OR Spectrum,* 30(1), pp. 77-112.

Yaman, H., 2011. Allocation strategies in hub networks. *European Journal of Operational Research,* Volume 211, pp. 442-451.

Yaman, H., Kara, B. & Tansel, B., 2007. The latest arrival hub location problem for cargo delivery stystems with stopovers. *Transportation Research Part B,* Volume 41, pp. 906-919.

Yaman, H., Karasan, O. & Kara, B., 2008. *Release time scheduling in cargo delivery.* [Online] Available at: www.bilkent.edu.tr/~hyaman/release.htm
[Accessed 01 06 2009].

Yan, X., Jinghui, W. & Liqun, Z., 2013. An Express Transportation Model of Hub-and-Spoke Network with Distribution Center Group. In: *Proceedings of the 2012 International Conferene of Modern Computer Science and Applications Advances in Intelligent Systems and Computing.* s.l.:Springer Berlin Heidelberg, pp. 467-472.

Yoon, M. & Current, J., 2008. The hub location and network design problem with fixed and variable arc costs: formulation and dual-based solution heuristic. *Journal of the Operational Research Society,* Volume 59, pp. 80-89.

Zäpfel, G. & Wasner, M., 2002. Planning and optimization of hub-and-spoke transportation networks of cooperative third-party logistics providers. *International Journal Production Economics,* Volume 78, pp. 207-220.

Summary

Express service providers move packages (i.e., parcels, documents, or pieces of freight) from a sender to a receiver under various but guaranteed service level agreements. Each service level agreement consists of collecting packages at a sender, transporting them generally via a road and or air network, and delivering them to a receiver within a specific delivery date and time. The European courier and express market in which express service providers operate is dominated by four major players, i.e. DHL, UPS, TNT Express, and FedEx. To retain their leading position in Europe, express service providers focus on cost cutting strategies while maintaining their service levels to remain competitive. Meanwhile, they have to respond on trends that shape the market by investing in innovative solution methods. This research is a clear example of the continuous search for improvement in the express market.

Research in this dissertation is dedicated to the design of so-called express line-haul transports from the first consolidation point, the origin depot, to the last consolidation point, the destination depot. The line-haul transport is commonly organised either via road or via air. Hub locations are used to sort packages of incoming line-hauls, and to consolidate and load these packages on the outgoing line-hauls. The strategic network design problem is concerned with decisions on these hub locations in the network: how many hubs are needed and where should these be located, and which line-hauls are unloaded and loaded at each consolidation point? These decisions are typically of a long-term planning nature. The tactical planning level at express service providers concerns decisions at a medium-term planning horizon, and contains decisions on package routings from origin to destination and asset usages. At the operational planning level, short-term decisions are made on the detailed use of vehicles, facilities and activities. This dissertation focusses on the strategic and tactical network design of the line-haul part in express networks, which provides input to the operational level.

Network design problems occur in various industries, like the transportation industry in general, but also for example in the production-distribution industry. However, the design of express networks is one of the most complex network design problems due to its multi-commodity nature and resulting size: different types of packages with varying service level agreements are shipped from many senders to many receivers. Moreover, network connectivity has to be recognised sufficiently during the design, as tight service agreements may restrict consolidation possibilities.

In this dissertation, we design express networks that are organised either via road or air, and present methodologies that suit the characteristics of each network. Firstly, we discuss the strategic and tactical network design for road networks. Afterwards, we discuss similar designs for air networks. We conclude by providing a general discussion on similarities and differences in the design of road and air networks, and then provide overall conclusions and recommendations.

Strategic road network design

Firstly, we consider the strategic road network design as we observe that scientific research to this challenging problem can be improved. One of the general assumptions seen in network design literature is that cost is represented by a linear function with economies of scale on inter-hub

transport, which represents the level of consolidation that is assumed on these transports. As depot to depot flows in express networks are in general low and consolidation is not always possible due to tight service agreements, the classical cost function does not suit to the cost made at the operational level in express networks. A second improvement in strategic road network design concerns the decision to transport packages between a pair of locations, the arc decisions. The complexity of the arc decisions in express networks is insufficiently recognised in research on strategic road network design, particularly when assumptions are drawn on the completeness of the hub network or restrictions are posed on the arcs between depots and hubs. In Chapter 3 we present two heuristics, simulated annealing and genetic algorithms, for the strategic road network design problem; these heuristics include a refined cost representation and a better reflection of the arc selection to fit to the operation of an express network. Both heuristics are tested on benchmark data and on modified instances of real-world express networks, and showed cost reductions and regional differences in the hub location choices.

The strategic road network design provides answers to long term decisions on the hub locations to be operated. The opening of a new hub location or the extension of existing buildings initiates investment and development processes that may take years. However, at some stage, the new building or extended equipment becomes available for use, and at that moment details on the changes in the operation need to be known.

Tactical road network design

Tactical road network design should give answers to flow routing decisions and scheduling decisions as soon as strategic changes in equipment become available. In the existing literature answers on routing and fleet scheduling are provided in the design of air networks, but the scheduling problem has almost gone unremarked in the design of long-distance road networks. As the characteristics of road networks differ from air networks, particular in the sizes of the networks that are to be designed, tactical network design approaches for road network design have to be enriched. That is the purpose of our work in Chapter 4, where we design a method that first decides on the flow routes and estimates the operational cost of the fleet, and afterwards a fleet scheduling heuristic is presented that supports in the design of a basic fleet schedule. The model is tested on benchmark data based on realistic instances, resulting in significant cost reductions.

The tactical road network design and fleet scheduling heuristic as presented in Chapter 4 has also been used to test sensitivities in the design of express networks. Firstly, a classical cost function with economies of scale has been compared to a more detailed cost function. The fleet schedules that resulted from the more detailed cost function showed significant cost reductions compared to the fleet schedules generated with the classical cost function. Further comparison of the different cost components showed that road network design is driven by the balance between transport cost, i.e. the cost per kilometre, and variable hub handling cost, and is most sensitive to the latter. Based on this observation, we recommend that both practice and science should pay more attention to hub handling cost estimates in road network design.

Our last work in the area of tactical road network design makes refinements on the fleet scheduling heuristic as presented in our work in Chapter 5 and makes a first attempt to consider hub operations in the generation of a basic schedule. The fleet scheduling extensions regard the use of multiple vehicle types and the consideration of stop-overs, which both were motivated by the observation of low-utilised depot to hub transports and the resulting opportunity for further consolidation. The extension of hub operations resulted from the observation that the number of transports that need to be handled at a hub is limited, e.g. there are only a limited number of doors available to unload or load vehicles, which might cause connectivity issues at the hubs. In Chapter 5 we presented a column generation approach and a local search algorithm in an iterative approach, and tested our methodology again on modified instances of express networks with good results. However, the test instances also showed that further research is desired to deal with hub operations and refined fleet scheduling characteristics for large-scale networks.

Strategic air network design

Air networks at European scale are in general operated via single hub networks. As air hubs are operated at airports and the number of airports at continental scale is limited, also due to night regulations, the strategic air network design problem can be answered via scenario analyses of tactical air network designs.

Tactical air network design

The tactical network design solutions that we suggested for the design of road networks do not suit the design of air networks, for reasons that will be discussed in detail below. In particular, air network design desires a higher level of detail at the scheduling level, and hence requires a different approach.

Some researchers have tackled the tactical network design problem in the design of air networks for express service providers. However, all research so far has been focussed on the design of single hub air networks. In practice, some drawbacks of single hub networks were observed: firstly, changing package flows had put pressure on hub handling capabilities, and secondly, single hub networks are undesirable from a contingency perspective. The idea arose to design a network with pre-sorted unit loading devices, which are units that can be loaded at once in an airplane. In single hub networks, pre-sorted unit loading devices (ULDs) would reduce the necessary handling capacity; and in multiple hub networks, pre-sorted unit loading devices would reduce hub handling time. The drawback of using pre-sorted ULDs is a possible loss of capacity, if not enough packages are available to create fully utilised ULDs.

Both multiple-hub routing and ULD-routing decisions posed questions to the network design stage that had not been dealt with in literature. In Chapter 6 we illustrate the methodology that we propose to design multiple hub air networks with pre-sorted ULDs. We use a three-stage iterative solution method, in which a pre-processing phase is used to make routing decisions and to specify hub sort windows; afterwards flight-optimisation occurs making decisions on number, types, and routings of aircrafts; and finally, a ULD-optimisation model is used to route the ULDs through the network. The method is illustrated on modified instances of an express service provider, and shows that a two hub network can be operated with hubs at half or their capacity in a single hub network.

Road versus air network design

So far, we summarised our work on several methodologies to design road and air networks for express service providers, at a strategic and tactical stage of the planning horizon. That also raises questions on the similarities and differences in the design of such networks. In fact, the major difference in the design of road and air networks concerns the level in which decisions on the assets are taken into account in each planning phase: asset decisions in road network design are part of the tactical/operational level, while asset decisions in air network design are part of the strategic/tactical decision level. The reason for this is two-fold: firstly, fixed cost of an airplane is the largest cost contributor in the design of air networks, while in road networks, transport cost in terms of the kilometres driven and hub handling cost are the main cost components; secondly, road transport can easily attract additional resources at the operational level, via third parties, while air transport has hardly any possibility for arranging additional resources at this stage.

As a result of the difference in which assets are considered in the network design phases for road and air networks, the output of the tactical network design is different: tactical road network design gives guidance in scheduling of vehicles, which can be refined based on subcontractor negotiations and may constitute a level of flexibility to react on operational fluctuations. Tactical air network design provides a final schedule of its assets and only relative small changes, like swapping aircraft types for certain flights, are applied at the operational level. Due to the lacking possibilities to attract other resources, the final schedule has to be robust for package flow fluctuations.

Lastly, note that when technological advances on computation power and mathematical techniques become available, we expect that road networks might be designed in a similar way as the current approaches in air network design. However, when air networks become larger in size, one may question if road network design approaches could be used to design these types of networks. It is our expectation that the lack of the inclusion of fixed asset cost in road network design would result in cost inefficient air operations, so that further research is desired when air networks become more complex.

Conclusions and recommendations

In conclusion, we extended strategic road network design by designing models that consider express network typologies in detail. We showed that the inclusion of these typologies result in different hub configurations with significant cost reductions as a result. Furthermore, we showed tactical road network design enrichments, and presented a methodology to create a basic schedule for the operation of vehicles in an express network. The methodology was also used to create valuable insights in the sensitivities of road network design. For smaller sized networks, we designed an approach that refines basic schedules, by the scheduling of multiple vehicle types and transports with stopovers; it is recommended to consider further research on these topics for larger sized networks as well as further research is needed on the subject of hub operations. Tactical air network design can be enriched by scheduling of pre-sorted unit loading devices which reduces the required handling capacities. Additionally, the scheduling of unit loading devices supports in the operation of multiple-hub networks, and a first methodology on the design of such multiple-hub air network has been presented in our work.

Some further recommendations are drawn from the limitations that we encountered during this period of research. Data availability is a topic that still is a constraining condition when applying operations research in practice. Luckily, businesses become more and more conscious of the power of data, and invest in its quality. Network design techniques can be enriched to become more robust, to for example volume fluctuations and network disturbances, when more data on these dynamics becomes available. We furthermore noticed that the express business is still a niche in the area of scientific research, while the growing e-commerce business increases the demand for time-guaranteed transportation. We expect that research in this segment will gain more attention and invite other researchers to compare their work to the research presented in this dissertation via the reference data sets that were made available. Furthermore, researchers and practitioners should understand the timelines and complexities of network design implementations and improvement measurement at the start of a project. Further research on the design of the gradual way of improvements might help practitioners to make network design solutions work in practice. Lastly, we would like to mention that the scope of our research concerned the line-haul express network, but that further opportunities are expected when pickup and delivery operations as well as depot and hub operations are considered simultaneously.

Summary (in Dutch)

Expressbedrijven vervoeren pakketten (pakjes, documenten en vrachtstukken) van verzenders naar ontvangers, waarbij ze service garanderen met betrekking tot de afleverdatum en de aflevertijd. De serviceafspraken gaan ook over het ophalen van een pakket bij een verzender op een afgesproken tijdstip en het transport, meestal via weg- of luchtvervoer. In de Europese koeriers-, express-, en pakkettenmarkt, waarin expressbedrijven opereren, zijn vier spelers dominant: DHL, UPS, TNT Express en FedEx. Om hun leidende positie binnen Europa te borgen, focussen deze bedrijven op kostenbesparende strategieën waarbij service-afspraken niet mogen verslechteren. Daarnaast zullen zij moeten reageren op de laatste ontwikkelingen in de markt door te investeren in innovatieve oplossingen. Binnen dit veld speelt ons onderzoek zich af.

In dit proefschrift staat het ontwerp van langeafstandtransport centraal. Dit is het transport, meestal via de weg of door de lucht, vanaf het eerste verzamelpunt in de keten, het depot van origine, tot het laatste verzamelpunt, het depot van bestemming. Hub-locaties worden gebruikt om pakketten te sorteren van inkomende transporten, waarna pakketten geconsolideerd en geladen worden op de uitgaande transporten. De vraag hoeveel hubs er nodig zijn, waar ze zich moeten bevinden en welke transporten bediend worden door elke hub is een strategisch netwerkontwerp-probleem. De hub-locatiekeuze is een typisch probleem met een lange termijn karakter. Tactische planning, de planning voor de middellange termijn, ondersteunt in beslissingen op het vlak van routeringen van pakketten en de inzet van transportmiddelen. Op het operationele niveau worden korte termijn beslissingen genomen met gedetailleerde weergaves van transportmiddelen, consolidatiepunten en activiteiten op de sorteercentra. Dit proefschrift is gericht op de lange en middellange planning en biedt daarmee input aan korte termijn beslissingen.

Netwerkontwerpproblemen bestaan bij verschillende industrieën, zoals de transport-industrie en de productiedistributie-industrie. Echter, het ontwerpen van expressnetwerken is een van de meest complexe vormen, vanwege de grote aantallen producten en daarmee de grootte van het planningsprobleem: verschillende soorten pakketten met variërende serviceafspraken moeten vervoerd worden tussen grote aantallen verzenders en ontvangers. Daarnaast moet ook de aansluiting in het netwerk voldoende worden ondersteund, omdat strakke serviceafspraken strikte beperkingen opleggen aan de mogelijkheden om geconsolideerd transport te organiseren.

In dit proefschrift bieden we methodes aan om wegtransport en luchttransport voor expressbedrijven te ontwerpen, op een manier die aansluit bij de specifieke karakteristieken van dergelijk transport. We zullen eerst een overzicht geven van de strategische en tactische planningsproblemen en oplossingsmethodes voor het wegtransport. Daarna bespreken we vergelijkbare problematiek en bijbehorende oplossingsmethodes voor het luchttransport. Tot slot zullen we de gelijkenissen en verschillen bespreken rondom de plannings- en oplossingsmethodes voor weg- en luchttransport, gevolgd door een algehele conclusie met aanbevelingen.

Strategische planning van wegtransport

Allereerst hebben we het strategische ontwerp van wegtransport bekeken, omdat bestaand wetenschappelijk onderzoek op dit vlak niet toereikend is. Eén van de gangbare aannames in de literatuur over netwerkontwerp betreft het gebruik van een lineaire kostenfunctie met lagere

kosten voor vervoer tussen hub locaties in tegenstelling tot transport tussen andere locaties in het netwerk. Deze kortingsschaal representeert de aangenomen efficiëntie die verwacht wordt door consolidatie van wegverkeer tussen de hubs. Maar omdat de pakketstromen tussen depots in expressnetwerken klein kunnen zijn en consolidatie niet altijd mogelijk is vanwege krappe tijdrestricties, volstaat een dergelijke kostenfunctie niet als weergave van de praktijk. Daarnaast kunnen oplossingsstrategieën voor netwerkontwerpproblemen verbeterd worden bij de vraag tussen welke locaties er getransporteerd gaat worden: de keuze voor de zogenaamde "takken" in netwerkontwerpterminologie. Deze complexiteit wordt over het algemeen onvoldoende erkend in onderzoek naar strategische netwerkontwerpen, in het bijzonder vanwege aannames die gemaakt worden over de volledigheid van het hubnetwerk en de beperkingen op de toewijzing van de transporten aan depots en hubs. In hoofdstuk 3 zullen we twee heuristieken presenteren voor het strategische netwerkontwerpprobleem bij expressbedrijven, één heuristiek gebaseerd op "Simulated Annealing" en één heuristiek gebaseerd op het "Genetic Algorithms"-principe. De gepresenteerde heuristieken zijn verrijkt met een representatieve kostenfunctie en verbeterde optimalisatie voor het opzetten van transporten. De methodes zijn getest op bestaande datasets binnen de literatuur en op realistische data uit de praktijk. De testresultaten laten duidelijke besparingen zien maar ook regionale verschillen in de uiteindelijke locatiekeuze voor hubs.

De strategische beslissingen binnen wegvervoer bieden langetermijnantwoorden rondom de inzet van hubs in een expressnetwerk. Het openen van een nieuwe hub maar ook de uitbreiding van een bestaande hub initieert een investerings- en ontwikkelingsproces dat jaren in beslag kan nemen. Echter, als er op een gegeven tijdstip een nieuwe hub of uitbreiding beschikbaar komt, moet op dat moment bekend zijn hoe de gegenereerde extra capaciteit het best benut kan worden. Dit is het onderwerp van tactisch netwerkontwerp.

Tactische planning van wegtransport

Tactische ontwerpen voor wegvervoer bieden antwoorden op routeringvraagstukken voor pakketten en de inzet en het inroosteren van voertuigen. In de bestaande literatuur worden antwoorden op dergelijke vraagstukken van tactische aard gegeven binnen de luchtvaart, maar het roosterprobleem voor langeafstandvervoer wordt zelden besproken binnen wegtransport. Aangezien wegvervoer en luchtvervoer karakteristiek verschillen, alleen al in de grootte van het planningsprobleem, waren wij genoodzaakt om de literatuur op dit gebied uit te breiden. Het resultaat daarvan wordt gepresenteerd in hoofdstuk 4, waar we een methode presenteren die allereerst beslissingen maakt rondom pakketrouteringen gevolgd door een roosterheuristiek die een basisrooster geeft voor de inzet van voertuigen. Deze aanpak is getest op representatieve data die beschikbaar is gesteld vanuit een expressbedrijf. Hierbij zijn significante besparingen aangetoond.

Dezelfde methode is gebruikt om de gevoeligheden in het ontwerp van expressnetwerken over de weg te analyseren. Allereerst hebben we gekeken naar de gevoeligheid van het netwerkontwerp voor de kostenfunctie. Dit is gebeurd door de roosteruitkomsten van een ontwerp dat gebaseerd is op de klassieke kostenrepresentatie met kortingen te vergelijken met de voorgestelde verfijnde kostenfunctie. We hebben geconcludeerd dat de roosters die gebaseerd zijn op de meer gedetailleerdere kostenfunctie significant lagere kosten geven dan de roosters gebaseerd op de klassieke kostenfunctie. Verder hebben we de gevoeligheid van het rooster getest voor de verschillende kostencomponenten in het wegverkeer. Hieruit hebben we geconcludeerd dat het tactische ontwerp voor het wegtransport van expressbedrijven wordt bepaald door de balans tussen transportkosten (de kosten per kilometer) en de variabele afhandelingskosten op een hub. Het ontwerp van het netwerk blijkt bijzonder gevoelig te zijn voor de kosten die gemaakt worden op de hub en zowel wetenschap als praktijk zouden meer aandacht mogen besteden aan het meenemen en correct inschatten van dergelijke kosten.

Ons laatste onderzoek op het gebied van het wegtransport richt zich op verfijningen van de roosterheuristiek gepresenteerd in hoofdstuk 5. In dit onderzoek bekijken we huboperaties tijdens het ontwerp van het basisrooster. De verfijningen in de heuristiek ontstaan door de toevoeging van verschillende voertuigtypes. Daarnaast hebben we verdere mogelijkheden tot verfijningen ontdekt door laag benutte transporten van en naar de depots te combineren. De uitbreidingen rondom de huboperatie werden vanuit de praktijk voorgesteld. Er werd opgemerkt dat het aantal aankomsten en vertrekken dat behandeld kan worden op een hub beperkingen heeft, bijvoorbeeld door de beschikbaarheid van het aantal deuren waaraan voertuigen ontladen of geladen kunnen worden. In hoofdstuk 5 bieden we een oplossing aan die gebaseerd is op een “Column Generation” –techniek in combinatie met een lokaal zoekalgoritme. We hebben die aanpak wederom getest op realistische datasets van expressnetwerken; dit met goede resultaten. Echter, de testresultaten toonden ook aan dat verder onderzoek nodig is op het vlak van huboperaties en bij grootschalige netwerken.

Strategische planning van luchttransport

Luchtvervoer op Europese schaal wordt in het algemeen georganiseerd via een netwerk met één hublocatie. De specifieke locatie van een dergelijke hub is gebonden aan de locatie van vliegvelden, en wordt nog strikter vanwege wetgeving rondom nachttransporten. Hierdoor kan strategische planning van luchtvervoer volgen uit scenarioanalyses op het vlak van tactische planning.

Tactische planning van luchttransport

De manier waarop tactische vragen rondom wegvervoer worden beantwoord, levert geen bruikbare antwoorden op in de luchtvaart, om redenen die we later zullen toelichten. De planning van luchttransporten op dit niveau vergt een hoger detailniveau dan de planning van wegtransporten en vraagt daarmee om een andere aanpak.

Enkele onderzoekers hebben tactische vraagstukken voor het netwerkontwerpprobleem binnen luchtnetwerken bekeken. Deze onderzoeken hebben gemeen dat ze oplossingen bieden voor één-hubnetwerken. In de praktijk hebben netwerken met één hub nadelen: allereerst omdat veranderende pakketstromen druk kunnen zetten op de hubcapaciteiten, maar ook omdat de operatie van een netwerk met één hub slecht bestand is tegen onvoorziene omstandigheden. Eén van de ideeën die ontstonden om dergelijke problemen op te vangen, is de introductie van voorgesorteerde laadeenheden (ULDs), die in één keer in een vliegtuig geladen kunnen worden. In netwerken met één hub, zouden dergelijke laadeenheden de benodigde sorteercapaciteit kunnen verlagen wanneer voorsortering elders georganiseerd kan worden; in netwerken met meer hubs kan een dergelijke oplossing tijd opleveren om de servicegaranties te voldoen. Nadeel van het gebruik van voorgesorteerde ULDs is een mogelijk verlies van laadcapaciteit, wanneer er te weinig pakketten beschikbaar zijn om volledig gevulde ULDs op te bouwen.

Het netwerkontwerpprobleem met meer hubs in een netwerk en de complexiteit van de inzet van voorgesorteerde laadeenheden, stellen eisen aan het ontwerp van het luchtnetwerk die niet eerder zijn behandeld in de literatuur. In hoofdstuk 6 schetsen wij onze aanpak van dit probleem. We maken daarin gebruik van een iteratieve methode die bestaat uit drie stappen. In een voorbereidingsstap maken we beslissingen rondom de routering van pakketroutes en specificeren we de hubsorteertijden. Daarna optimaliseren we vluchten door beslissingen te nemen over het aantal en type vliegtuigen dat gebruikt wordt en de bijbehorende route van elk specifiek vliegtuig. In de laatste fase worden ULDs gerouteerd door het netwerk en worden er beslissingen gemaakt over de voorsortering van die ULDs. Deze methode wordt geïllustreerd aan de hand van gemodificeerde data van een expressbedrijf en toont dat een netwerk met twee hubs kan worden onderhouden waarin elke hub minder dan de helft van zijn capaciteit gebruikt dan in het geval van een één-hub netwerk.

Planning van wegtransport versus luchttransport

Tot dusver hebben we verschillende methodes besproken om weg- en luchttransport te ontwerpen voor expressbedrijven, op de lange en middellange planningstermijn. Dit roept ook vragen op over de gelijkenissen en verschillen rondom het ontwerp van dergelijke netwerken. In feite is het grootste verschil in de aanpak van weg- en luchtvervoer de mate waarin beslissingen over de transportmiddelen meegenomen worden tijdens de verschillende planningsfases. Beslissingen over voertuigen worden bij wegvervoer genomen op tactisch/operationeel niveau, terwijl dergelijke beslissingen over vliegtuigen gemaakt worden op het strategisch/tactische niveau. De reden hiertoe is tweeledig: allereerst zijn de vaste kosten voor de aanschaf van een vliegtuig de belangrijkste kostenpost in het ontwerp van een luchtnetwerk, terwijl kosten in wegvervoer worden bepaald door de gereden kilometers en de benodigde afhandeling op de hubs. Bovendien bestaat er binnen het wegvervoer de mogelijkheid om op korte termijn extra transportmiddelen aan te trekken, via andere partijen, terwijl die optie nauwelijks beschikbaar is voor luchtvervoer.

Het resultaat van tactisch ontwerp van planningen voor wegverkeer en luchtverkeer verschilt dan ook. Voor wegverkeer biedt de uitkomst van een tactische planning een initieel rooster dat bijgeschaafd wordt aan de hand van onderhandelingen met derde partijen waar vervoer wordt ingekocht, met bepaalde vrijheden om bij te sturen op basis van de operatie. De tactische planning van luchtvervoer levert een definitief rooster op, waarin slechts wat kleine veranderingen volgen gedurende de operatie, zoals de wisseling van een type vliegtuig tussen vluchten. Omdat de bijsturingsmogelijkheden in luchttransport dusdanig laag zijn, moet het definitieve schema van vluchten robuust zijn voor mogelijke fluctuaties in volumestromen.

Tot slot willen we opmerken dat technologische ontwikkelingen in rekenkracht en wiskundige technieken er voor zullen zorgen dat het ontwerp van wegtransport meer en meer gaat lijken op het ontwerp van luchttransport. Wanneer luchttransport complexer wordt bijvoorbeeld door grotere netwerken, kunnen we ons afvragen of technieken voor het organiseren van wegtransport bruikbaar zijn in de luchtvaart. Het is onze verwachting dat verder onderzoek nodig is, wanneer luchttransport complexer wordt. Voornaamste reden hiertoe is dat de mate waarin de vaste kosten van het transportmiddel worden geschat in de tactische planning van het wegtransport een te sterke onderschatting geven van de operatie van vliegtuigen in een luchtnetwerk.

Conclusies en aanbevelingen

In ons onderzoek hebben we strategische netwerkontwerpen verfijnd door methodes te presenteren die gedetailleerde expressnetwerktypologieën kunnen ontwerpen. We hebben aangetoond dat de toevoeging van dergelijke typologieën resulteert in andere hub-locaties met lagere kosten. Daarnaast hebben we tactische modellen voor wegvervoer verfijnd en presenteren wij een methodologie die in staat is om een initieel rooster op te leveren voor de transportoperatie. Die methodologie heeft ook waardevolle inzichten opgeleverd over de gevoeligheden van het netwerkontwerpprobleem voor wegtransport. De gepresenteerde methode hebben we verder verfijnd voor de inzet van verschillende type voertuigen en het creëren van betere consolidatie van transporten tussen depots en hubs met tussenstops. Hoewel die verfijnde methode goed werkt voor kleine netwerken, is er verder onderzoek noodzakelijk naar toepassingen op grotere schaal; dit geldt ook voor verfijningen in het basisrooster die tot een betere aansluiting bij huboperaties zou moeten leiden. Het tactisch ontwerp van luchttransport kan worden verbeterd door voorgesorteerde laadeenheden in beschouwing te nemen, die de benodigde sortering op hubs vermijden. Daarnaast bieden voorgesorteerde laadeenheden de mogelijkheid om netwerken met meer hubs toe te passen. In dit proefschrift presenteren wij een methode om dergelijke meervoudige-hubnetwerken te ontwerpen, inclusief de beslissing rondom voorsortering van laadeenheden.

We sluiten af met aanbevelingen die volgen uit de beperkingen die we tegen zijn gekomen gedurende dit onderzoek. De beschikbaarheid van data is een onderwerp dat nog steeds een beperkende factor vormt voor de toepassing van besliskunde in de praktijk. Het is goed om te zien dat bedrijven meer en meer beseffen dat data een belangrijke bijdrage kunnen leveren en dat zij het belang zien van investeringen in kwaliteit. Netwerkontwerptechnieken zullen beter en robuuster worden wanneer databeschikbaarheid en kwaliteit verbetert, bijvoorbeeld door gebruik te maken van informatie rondom fluctuaties in pakketvolumes of informatie rondom netwerkverstoringen. Daarnaast zien we dat de expressmarkt nog steeds een niche vormt in de wetenschap, terwijl de groeiende e-commerce business steeds meer vraagt om transporten met strikte serviceafspraken. Het ligt dan ook in de verwachting dat dit segment meer aandacht zal krijgen. We nodigen onderzoekers uit om hun werk te vergelijken met onderzoek gepresenteerd in dit proefschrift door gebruik te maken van de beschikbaar gestelde datasets. Daarnaast zullen wetenschappers en praktijkdeskundigen begrijpen dat het toebrengen van veranderingen aan expressnetwerken tijd kost en met een behoorlijke complexiteit gepaard gaat mede doordat verbeteringen moeilijk te meten zijn. Verder onderzoek zou zich kunnen richten op de uitwerking van de transitiestappen en de meetbaarheid van dergelijke stappen zodat de daadwerkelijke toepassing vergemakkelijkt wordt. Tot slot, willen we benadrukken dat het onderwerp van dit onderzoek beperkt was tot het langeafstandtransport in expressnetwerken; verdere kansen verwachten we wanneer de planning van het collectie- en distributieproces en de operatie van depots en hubs gezamenlijk beschouwd worden.