Logistics Service Network Design: Models, Algorithms, and Applications

by

Su Shen

B.S., Tsinghua University (1997)
M.S.T., Massachusetts Institute of Technology (2000)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Transportation

at the

Massachusetts Institute of Technology

June 2004

Signature of Author ....
Department of Civil and Environmental Engineering
12 March 2004

Certified by ....
Cynthia Barnhat
Professor, Civil and Environmental Engineering and Engineering Systems
Thesis Co-supervisor

Certified by ....
Nigel H. M. Wilson
Professor, Civil and Environmental Engineering
Thesis Co-supervisor

Accepted by ....
Heidi M. Nepf
Chairman, Department Committee on Graduate Studies

Logistics Service Network Design: Models, Algorithms, and Applications

by

Su Shen

Submitted to the Department of Civil and Environmental Engineering
on 12 March 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Transportation

Abstract

Service network design is critical to the profitability of express shipment carriers. In this thesis, we consider two challenging problems associated with designing networks for express shipment service. The first problem is to design an integrated network for premium and deferred services simultaneously. Related existing models adapted to this problem are intractable for realistic instances of this problem: computer memory requirements and solution times are excessive. We introduce a disaggregated information-enhanced column generation approach for this problem that reduces the number of variables to be considered in the integer program from hundreds of thousands to only thousands, allowing us to solve previously unsolvable problem instances.

The second problem is to determine the express package service network design in its entirety, including aircraft routings, fleet assignments, and package flow routings, including hub assignments. Existing models applied to this problem have weak associated linear programming bounds and hence, fail to produce quality feasible solutions. For example, for a small network design problem instance it takes days to produce a feasible solution that is provably near-optimal using the best performing existing model. To overcome these tractability challenges, we introduce a new model, referred to as the *gateway cover and flow formulation*. Applying our new formulation to the same network design instance, it takes only minutes to find an optimal solution.

Applying our disaggregated information-enhanced column generation approach and gateway cover and flow formulation and solution approach to the network design problems of a large express package service provider, we demonstrate tens of millions of dollars in potential annual operating cost savings and reductions in the numbers of aircraft needed to perform the service. Moreover, we illustrate that, though designed for tactical planning, our new model and solution approach can provide insights for strategic decision-making, such as hub opening/closure, hub capacity expansion, and fleet composition and size.

Thesis Co-supervisor: Cynthia Barnhat
Title: Professor, Civil and Environmental Engineering and Engineering Systems

Thesis Co-supervisor: Nigel H. M. Wilson
Title: Professor, Civil and Environmental Engineering

To Lin and My Parents

Acknowledgements

I am deeply grateful to my advisor, Professor Cindy Barnhart, for her constant understanding and support, even at times when my own confidence started wondering. Her "can do" and "never give up" attitude always inspires me, and her intelligence is my unbounded resource. I learned a great deal from her both academically and personally.

My gratitude also goes to my co-supervisor, Professor Nigel Wilson. His insights contribute to the enhancement of this disseration, and the high standard he wants me to hold has always made me asking myself "Can you do better?"

I sincerely thank Professor Georgia Perakis for her invaluable advice. Her insightful comments and suggestions have contributed greatly to my dissertation.

Many thanks go to the people at United Parcel Service, Keith Ware, Alysia Wilson, Wynand Depuy, Lee Mitchell. Working on the UPS project has been an exciting experience for me.

I'd like to thank my colleagues at MIT. Andy Armacost, who has been the pioneer in this project and who offered a lot of help when I first started, Hong Jin and Manoj Lohatepanont, who have been supportive senior fellow students, Amr Frahat, Amy Mainville, Stephane Bratu, Shan Lan and many others. I also want to thank Maria Marangiellow for her great assistance.

This dissertation would not have been possible without the unwaivering support of my wife, Lin. As said in a Chinese proverb, one couldn't thank those closest to him/her. No word can be used to express my appreciation.

Finally, I want to thank my parents and elder brother. They encourage me to go higher and higher and never doubt my ability.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 1998, the two largest express shipment carriers in the United States, Federal Express (FedEx) and United Parcel Service (UPS) generated revenues of $9.4 billion and $7.1 billion, respectively, in domestic, air-express shipment service [40] and [89]. In 2002, while the revenue of FedEx's air-express shipment service was almost flat at $9.5 billion, the revenue of UPS' air-express shipment service grew by more than 15%, to about $8.2 billion [40] and [89]. Many Wall street analysts attributed UPS' revenue growth and gain in market share to its emphasis on operational efficiency [84]. Indeed, including international express shipment service, FedEx's *operating margin*, operating profit as a percentage of revenue, was 6.4% in 1998 and 4.8% in 2002. In contrast, the operating margins for UPS' package delivery business was 15.0% in 1998 and 13.6% in 2002. This includes the less profitable ground delivery service that accounts for 66% of UPS' domestic package delivery revenue. Clearly, efficient operations give carriers a decisive competitive advantage, allowing the carrier to price its service more aggressively and gain market share, or use the cash flow generated to make further investment to achieve an advantageous position.

Given the high-revenue and low-operating margins of air express shipment service, even a single-digit percentage improvement in operating costs translates to substantial savings. Because service network design is the first step in express shipment service planning, it is exceedingly important in improving operating efficiency. In this dissertation, our objective is to develop optimization models to help express shipment carriers design cost-minimizing service networks.

The difficulty of the Express Shipment Service Network Design (**ESSND**) problem comes from the need to model both integer aircraft routes, referred to as *design variables*, and continuous package flows. The LP relaxations of conventional network design formulations tend to produce fractional solutions that are difficult to transform into good-quality feasible solutions. Extended formulation techniques, as described in Armacost et al. [5] and Martin [74], embed package flow decisions within the design variables, resulting in improved linear programming (LP) bounds for special cases, though at the expense of greatly increased numbers of variables. In this research, we are interested in designing for the **ESSND** problem models and algorithms that improve tractability, typically through the provision of tight LP bounds.

1.1 Problem Description

We first describe the operations, modeling restrictions and planning process of a typical express shipment carrier and define the **ESSND** problem.

1.1.1 Express Shipment Delivery Operation

Express shipment carriers operate transportation equipment, including both aircraft and ground vehicles, and fixed facilities, such as gateways and hubs, to move shipments between customers within a small time window. Figure 1-1 depicts a partial express package delivery service network.

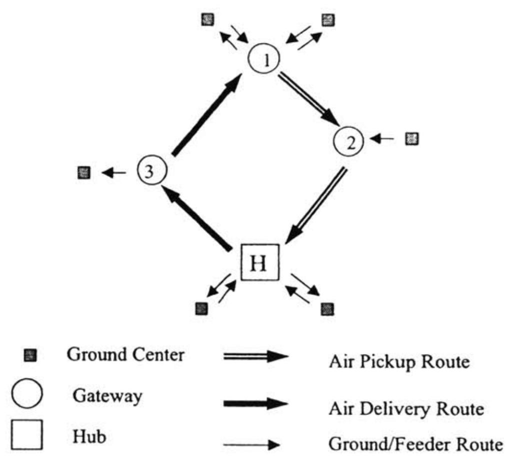


Figure 1-1: Express Package Service Operations

Packages are typically picked up by ground vehicles and first transported to *ground centers,* or more specifically, *origin ground centers.* A ground center can serve as both an origin and a destination ground center, depending on whether the operation is a pickup or delivery. A ground center is typically associated with a city, although there might be several ground centers for a large city. After packages arrive at the origin ground center, an *origin sort* is conducted to determine the routing for each package based on its destination and a pre-specified package service plan. Though there are exceptions, which we discuss later in this section, the shipment is typically transported to a *gateway* or an airport either by ground vehicles or small aircraft. Packages at gateways are then picked up by jets. A jet follows a *pickup route*, with arrival at a *hub.* Upon arrival at a hub, packages are sorted and consolidated by destination.

Next, each package is delivered via air along its *delivery route* to its destination gateway, and then transported to destination ground centers by ground vehicles or small aircraft. At the destination ground center, a *destination sort* is conducted, and packages are loaded onto ground vehicles for delivery.

Carriers typically offer different levels of service and charge higher premiums for higher levels of service. The level of service is characterized by the time from pick-up to delivery. For example, UPS offers both Next-Day and Second-Day services. For shipments picked up on a given day, Next-Day service guarantees delivery by the early morning of the next day, typically before 10 AM, and Second-Day service guarantees delivery by the end of the day after the next. For both services, the full premium is refunded to customers if delivery is not made on time (United Parcel Service [90]). Operations for different services are similar, with the same equipment and facilities used, though maybe at different times. For example, the same aircraft is used to deliver Next-Day shipments during the night and Second-Day shipments during the day. The equipment and facilities required for air operations, therefore, are largely determined by Next-Day service requirements because its time windows are the smallest.

1.1.2 Time Windows and Level of Service

To allow sufficient time for customers to prepare their packages, the carrier specifies the earliest time, typically after typical business hours, shipments can depart each ground center, denoted the *Earliest Pickup Time from Center* (*EPTC*). Similarly, to guarantee on-time delivery, the carrier specifies the latest time shipments may be delivered to a ground center, called the *Latest Delivery Time to Center* (*LDTC*). For Next-Day service, due to the small delivery time window, the closest gateway is generally the only feasible ground center option, and therefore, the earliest time a pickup aircraft can depart the gateway, denoted the *Earliest (Gateway) Pickup Time* (*EPT*), and the latest time a delivery aircraft may arrive at the gateway, denoted the *Latest (Gateway) Delivery Time* (*LDT*), can be determined by the *EPTC* and *LDTC* of the ground centers the gateway serves and the transportation time between the gateway and the ground centers. For Second-Day service, because more time is available for delivery, demands can be transported to a gateway farther away and decisions involve selecting which gateway to use. The *EPT* and *LDT* of a gateway in this case cannot be determined until decisions are made to determine which gateway serves each ground center.

Timing requirements at hubs are specified as the *Latest Hub Arrival Time* (*LHAT*) and the *Earliest Hub Departure Time,* (*EHDT*). Latest hub arrival time represents the latest time a pickup flight can arrive at the hub and still allow sufficient sort time for the packages it transports. Earliest hub departure time represents the earliest time a delivery flight may depart from the hub. To accommodate time zone differences, pickup flights originating from west coast gateways are typically allowed to arrive at non-west coast hubs later than those originating from non-west coast gateways.

1.1.3 Demand

Each ground center serves a designated service area. The actual customer-to-customer demands are therefore reflected as center-to-center demand, and demand estimates are reported at the center-to-center level. Given the specified timing requirements at the ground centers and hubs, the following demands can be handled *en route* by ground if:

- the demand can be transported from the origin ground center to the destination ground center by ground vehicles within the time window spanned by the *EPTC* of the origin ground center and the $LDTC$ of the destination ground center.

- the demand can be transported from the origin ground center to a hub by ground vehicle within the time window spanned by the $EPTC$ of the origin ground center and the $LHAT$ of the hub, and transported from the hub to the destination ground center by ground vehicle within the time window spanned by the $EHDT$ of the hub and the $LDTC$ of the destination ground center.

Because ground transportation is much less expensive than air transportation, demands are delivered via ground whenever *en route* ground delivery is possible, and such demands are not considered in the air service network work design problem.

For Next-Day service, center-to-center demands can be directly aggregated into origin gateway-destination gateway demands. For Second-Day service, gateway-to-gateway demands can be obtained after ground centers are assigned to gateways. We refer to origin gateway-destination gateway demands simply as *origin-destination (O-D) commodities or origin-destination volumes* hereafter. We assume all these demands are deterministic, whether expected values, conservative estimates, or another estimate.

1.1.4 Cost Elements for the Service Network Design Problem

Cost is incurred for aircraft operation, ground vehicle operation and package handling. Aircraft operating cost includes two components:

1. *Block time cost* based on the block time (that is, flying time plus taxi time) flown, includes mostly the crew and fuel costs;

2. *Fixed cycle cost* incurred on each flight leg, typically includes the landing fees and other one-time charges.

Ground vehicle operating costs, largely based on the distance traveled, are much smaller than aircraft operating costs, and hence, we considered them to be zero in this research.

Package handling cost also includes two components, a cost based on block time and a fixed handling cost. Block time cost is a proxy for the marginal fuel cost, and handling cost largely includes the package handling cost at ground centers and hubs. Package handling costs are also much smaller than the aircraft operating costs, and we again consider them to be zero in this research.

1.1.5 Other Restrictions

In addition to serving all demands within the specified timing windows, express shipment delivery service network design is subject to the following restrictions:

- Conservation of aircraft movement at both the gateways and hubs - if an aircraft departs a location, that aircraft must arrive at that location;
- Airport capacity - the number of aircraft arrivals at a hub cannot exceed the number of aircraft parking spots at the hub;
- Aircraft count - the number of aircraft of each fleet type used must not exceed the available number;
- Aircraft capacity - the packages assigned to each aircraft cannot exceed the aircraft capacity; and

- Hub sort capacity - the packages routed through a hub must not exceed its sort capacity.

1.1.6 Planning Process and Problem Definition

The carrier typically follows the process illustrated in Figure 1-2 for network planning. Focusing 5 years or more into the future, the strategic planning decisions to be made are: hub location, hub capacity expansion and aircraft purchases. The data needed include forecasted demands, hub set-up or capacity expansion costs and aircraft ownership costs. The objective is to minimize both fixed infrastructure and equipment investment costs and long-term variable operating costs.

Tactical planning for 6 month to 1 year into the future involves deciding hub assignments for O-D commodities, aircraft routes and fleet assignment, with location and capacity of hubs, and fleet composition and size all fixed. The objective is to minimize operating costs. Currently, the carrier develops these tactical plans using a two-step sequential process. First, using rules of thumb, O-D commodities are heuristically assigned to a specific hub. For example, if the origin gateway of an O-D commodity is close to a hub, and hence can be picked up by ground, the commodity is assigned to that hub. Some simple algorithms are also used to attempt to achieve commodity-hub assignments that minimize total package miles. Once the hub assignment decisions are made for every O-D commodity, commodities from the same gateway and assigned to the same hub are consolidated into *gateway-hub pickup demand.* Similarly, commodities to the same gateway and assigned to the same hub are consolidated into the *gateway-hub delivery demand.* (The first step of tactical planning for Second-Day service also includes selection of gateways to use and assignment of ground centers to gateways, before assigning origin gateway-destination gateway volumes to hubs.)

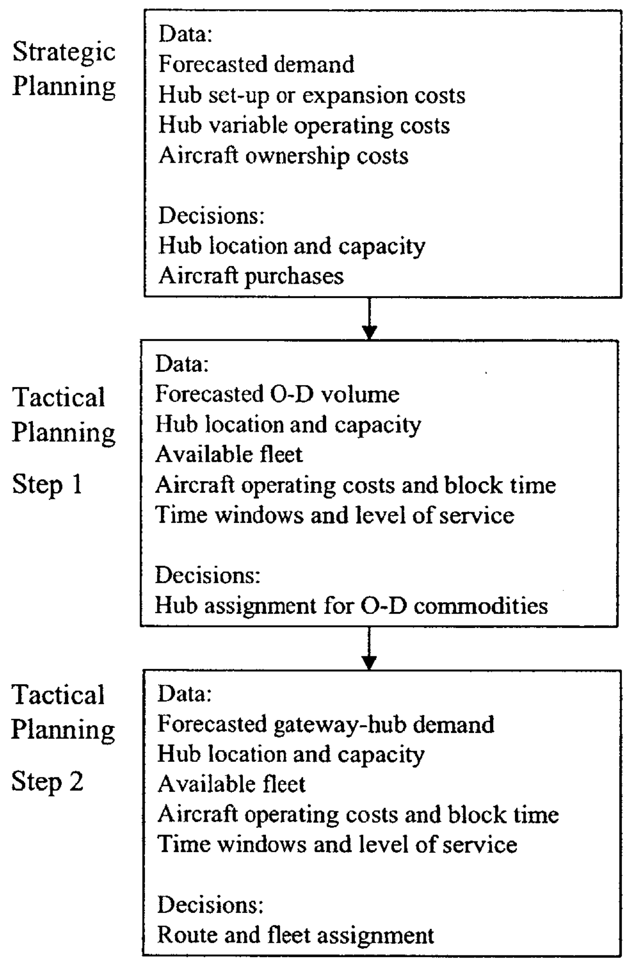


Figure 1-2: The Current Network Planning Process of a Typical Express Shipment Carrier

In the second step of the tactical planning process, (operating) cost-minimizing aircraft routes and fleet assignment decisions are made, given the set of gateway-hub demands. We define this step of the process as the *express shipment service network design problem with fixed hub assignment* or the *fixed hub assignment problem.* In contrast, we define the tactical planning process in its entirety, taking O-D commodities as input, determining hub assignments for each O-D commodity, selecting the set of routes and the corresponding fleet assignments that minimize total operating costs, as the *express shipment service network design problem with flexible hub assignment,* or the *flexible hub assignment problem.*

Although Next-Day and Second-Day operations are performed sequentially, the two services are interlinked. Fleet position as a result of the pickup and delivery operation of the Next-Day operation affects fleet position and costs associated with the Second-Day operation, and vice versa. In the carrier's current practice, because of problem size and complexity, tactical planning for Next-Day and Second-Day services is done sequentially, solving two independent **ESSND** problems, one for Next-Day and another for Second-Day service, with fleet position fixed in the second **ESSND** problem based on the results of the first. Planning Next-Day and Second-Day services simultaneously, that is, considering the *Integrated Next-Day and Second-Day problem*, can result in better fleet positioning and significant savings.

In the first part of this research, we investigate the Integrated Next-Day and Second-Day problem in which hub assignments for O-D commodities are determined *a priori.* We refer to this as the *Integrated Next–Day and Second-Day Problem with Fixed Hub Assignment.* In the second part of this research, we relax the fixed hub assignment assumption and consider the flexible hub assignment problem.

1.2 Contribution

In this research, our major contributions include:

- Designing a robust solution methodology to solve the integrated Next–Day and Second-Day problem with fixed hub assignment. Existing models that include all variables are intractable: computer memory requirements and solution times are excessive. We introduce a disaggregated information-enhanced column generation approach to reduce the number of variables in the integer program from hundreds of thousands to only thousands, allowing us to solve previously unsolvable **ESSND** problem instances. In addition to its relevance to express package delivery, disaggregated information-enhanced column generation can also be applied to other problem types, including multi-commodity flow problems and crew-scheduling problems, to reduce model size and improve solution speed.

- Developing an optimization model for the flexible hub assignment problem. Applying existing models to the flexible hub assignment problem proves to be ineffective: (1) LP relaxation bounds are weak; and (2) the numbers of variables are prohibitively large. Extending the composite variable concept described in Armacost et al. [5], we present the Gateway Cover and Flow model, significantly reducing the numbers of variables and improving LP bounds.

- Demonstrating the efficacy of both the information-enhanced column generation and the gateway cover and flow model on carrier-specific problem instances that could not be solved using existing approaches. We show that tens of millions of dollars in annual operating costs can potentially be saved in each case, with even greater potential savings in aircraft ownership costs and hub set-up and maintenance costs.

1.3 Thesis Overview

In Chapter 2, we review recent literature on a more general class of network design problems, namely the Network Design Problem, and the models and algorithms that have been designed for **ESSND** problems.

In Chapter 3, we describe the inter-connection of the Next-Day and Second-Day operations at a carrier and define the integrated Next-Day and Second-Day **ESSND** problem with fixed hub assignment. Next, we demonstrate that we can model the **ESSND** problem for the carrier's Second-Day operation as a daily problem, that is, the same operation is repeated daily. As a result, we can adapt the model designed for the Next-Day fixed hub assignment problem to the Integrated NDA-SDA Problem. This yields an intractable model because computer memory requirements and solution times are excessive. To overcome intractability issues, we design the disaggregated information-enhanced column generation approach, and demonstrate that it allows us to reduce dramatically the number of columns in the integer programming model and the solution time. We then apply disaggregated information-enhanced column generation to the UPS Integrated NDA-SDA problem and demonstrate that potential annual operating cost savings measure in the tens of millions of dollars. These savings result from: (1) reduced ferrying costs; and (2) better coordinated NDA and SDA fleet movements.

In Chapter 4, we consider the flexible hub assignment problem. We first review different variable definitions for modeling the flexible hub assignment problem, and present a formulation for the flexible hub assignment problem based on our best variable definition. Through a small service network design problem, we demonstrate the impact of variable definition on the associated LP bounds and solution time. Applying our formulation to the UPS Next-Day service network design problem, we demonstrate that potential operating cost savings measure in the tens of millions of dollars annually if hub assignment, route selection, and fleet assignment decisions are made simultaneously. Moreover, we demonstrate that, although designed as a tactical planning tool, our flexible hub assignment model can be used to provide insights for strategic planning, and to achieve even greater savings in aircraft ownership costs and hub set-up, capacity expansion and maintenance costs.

In the final chapter, we summarize the results and contributions of this thesis, and identify areas for future research.

Chapter 2

Literature Review

The express shipment service network design is a spacial case of a broader class of problems, that is, network design problems. Application of network design arises in distribution, transportation, telecommunications and many other areas. In network design problems, we have a set of demands, either single commodity or multi-commodity, and we want to install capacity in the network such that we can move all demands from their origins to their destinations with minimized flow and capacity-installation costs. In this chapter, we define the network design problem and review the recent literature on network design and express shipment service network design problems.

2.1 Classic Network Design Problems

Given a directed graph, $G = (N, A)$, a set of origin-destination commodities, K, and a set of facilities, F. Let c_{ij}^k denote the unit flow cost of commodity $k \in K$ flown on arc $(i,j) \in A$, d_{ij}^f denote the cost of installing one unit of facility type f on arc (i,j), and u_{ij}^f denote the capacity provided by one unit of facility f on arc (i,j). We denote the origin and destination of an O-D commodity as $O(k)$ and $D(k)$, respectively. Let y_{ij}^f be the integer decision variable indicating the number of facilities of type f installed on arc (i,j), denoted as the *design variable*, and x_{ij}^k be the continuous decision variable indicating amount of flow of commodity k on arc (i,j), also denoted as the *flow variable*. The network design problem is stated as:

$$\min \sum_{k \in K} \sum_{(i,j) \in A} c_{ij}^k x_{ij}^k + \sum_{f \in F} \sum_{(i,j) \in A} d_{ij}^f y_{ij}^f \tag{2.1}$$

subject to:

$$\sum_{k \in K} x_{ij}^k \leq \sum_{f \in F} u_{ij}^f y_{ij}^f \quad (i,j) \in A \tag{2.2}$$

$$\sum_{j:(i,j) \in A} x_{ij}^k - \sum_{j:(j,i) \in A} x_{ji}^k = \begin{cases} b^k & \text{if } i = O(k) \\ -b^k & \text{if } i = D(k) \\ 0 & \text{otherwise} \end{cases} \quad i \in N,\ k \in K \tag{2.3}$$

$$x_{ij}^k \geq 0 \quad (i,j) \in A,\ k \in K \tag{2.4}$$

$$y_{ij}^f \in \mathbb{Z}_+ \quad (i,j) \in A,\ f \in F. \tag{2.5}$$

Forcing constraints (2.2) ensure that the flow on any arc does not exceed the capacity installed on that arc. Constraints (2.3) ensure *conservation of flow* for the commodities.

If there is only a single facility type, the problem becomes the classic network design problem described in Magnanti and Wong [68] and Ahuja et al. [2]. If there are multiple facilities but the flow cost is zero, the problem is the network loading problem (**NLP**) presented in Magnanti and Mirchandani [70]. If there are multiple facilities and the flow cost is not zero, the problem is referred to as the multi-level network design problem described in Balakrishnan et al. [8].

Magnanti and Wong [68], and Minoux [75] provide surveys of network design models and applications. Magnanti and Wong [68] demonstrate that many of the combinatorial problems that arise in transportation planning are specializations or variations of network design, and summarize potential applications and solution strategies.

2.2 Valid Inequalities for Network Design

The forcing constraints (2.2) in the network design formulation tend to result in highly fractional solutions to the associated LP relaxations, and result in weak LP bounds. For this reason, significant research has been devoted to deriving valid inequalities and to characterizing network design polyhedra to strengthen the formulations.

Marchand et al. [73] present a survey on the use of valid inequalities for classes of problems, including network design. Van Roy and Wolsey [91] and Padberg et al. [82] present valid inequalities for fixed charge network design. Wolsey [94] derives a family of valid inequalities using submodularity. Magnanti et al. [71] develop families of facets and completely characterize the convex hull for two core subproblems of the network loading problem. Magnanti and Mirchandani [70] study a specialized version of the network design problem that is a generalization of the shortest-path problem and introduce two families of facets, providing geometric interpretations.

Magnanti et al. [72] study the two-facility capacitated network loading problem. They consider two solution approaches for solving the mixed integer program: a Lagrangian relaxation strategy and a cutting plane approach that uses three types of valid inequalities. They show that the linear programming formulation including the three types of inequalities provides a bound that is at least as good as that obtained from the Lagrangian relaxation, and demonstrate the effectiveness of these inequalities in improving the LP relaxation.

Bienstock and Günlük [27] study a capacity expansion problem and extend two types of the valid inequalities presented in Magnanti et al. [72]. Chopra et al. [29] consider the one-commodity, one-facility network design problem and present additional inequalities. Bienstock et al. [28] compare formulations for the single-facility-multi-commodity network design problem, describe two classes of valid inequalities and use them to characterize the corresponding polyhedron for a three-node graph. Mirchandani [76] considers an undirected network with two types of capacitated facilities for both the single-commodity case and multi-commodity case. *Equivalent* formulations in a lower-dimensional space are presented by projecting out the flow variables. Mirchandani's results strengthen existing results for multi-commodity flow problems. Atamtürk [6] gives a complete linear description of the cut-set polyhedron of the single commodity - single facility capacitated network design problem and extends the analysis to multi-commodity-multi-facility capacitated network design problems. Klabjan and Nemhauser [62] study the polyhedron of the single node capacitated network design model with integer variable upper bounds and give a characterization of valid inequalities that is useful in proving the validity of several classes of inequalities.

2.3 Solution Algorithms

Balakrishnan et al. [8] present a dual-based algorithm for the multi-level network design problem. Their algorithm first fixes certain design variables in a pre-processing step, and then applies a dual ascent procedure to generate lower and upper bounds on the optimal value. Balakrishnan et al. [9] investigate relationships between alternative formulations for the two-level network design problem and analyze the worst-case performance of a heuristic algorithm. Balakrishnan et al. [10] design a decomposition algorithm for local access telecommunications network expansion planning. They propose a Lagrangian relaxation scheme that solves an uncapacitated subproblem to generate upper and lower bounds. The uncapacitated subproblem is solved using a polynomial dynamic programming algorithm incorporating valid inequalities based on the problem-specific polyhedral structure. Balakrishnan et al. [11] present worst-case bounds for heuristics and LP relaxations of the overlay optimization problem and demonstrate worst-case bounds for the uncapacitated multi-commodity network design problem. Bienstock and Günlük [26] describe a cutting plane algorithm for the problem of network design to minimize the maximum load on any arc. Barahona [14] solves both the bifurcated and nonbifurcated versions of the network loading problem using a relaxation based on the cut condition for multi-commodity flows. Holmberg and Hellstrand [53] present a Lagrangian-based heuristic within a branch-and-bound framework for solving the uncapacitated network design problem. Günlük [50] present a branch-and-cut algorithm to solve capacitated network design problems using a knapsack branching rule. Stallaert [86] describes a simple procedure to derive network inequalities for capacitated fixed charge network problems by exploiting properties of fractional extreme point solutions to the LP relaxation. Chopra and Tsai [30] convert the multi-level network design problem to a Steiner tree problem. The resulting formulation contains fewer variables but exponential numbers of constraints. A branch-and-cut approach is used to solve the problem.

In addition to valid inequalities, Benders decomposition [24] is also used to solve the network design problem. Magnanti et al. [69] present Benders cuts for the uncapacitated network design problem and use a dual ascent procedure to accelerate the decomposition algorithm. Sridhar and Park [85] develop an implicit branch-and-bound algorithm for the fixed-charge capacitated network design problem, incorporating both Benders cuts and polyhedral cuts. They show that Benders cuts are more effective under heavy traffic loads, while cut set inequalities are more effective under light traffic loads.

Other algorithms used for solving the network design problem include approximation algorithms. Let Z_H be the objective value produced by an approximation heuristic H, and Z_{IP} be the integer programming optimum. We say H is a λ-*approximation* algorithm if $Z_H \leq \lambda Z_{IP}$. Agrawal et al. [1] present a polynomially solvable approximation algorithm for the general Steiner network problem. Goemans and Williamson [44] describe a general technique producing 2-approximation algorithms for a large class of graph problems, including the shortest path, minimum-cost spanning tree, minimum-weight perfect matching, traveling salesman and Steiner tree problems. Based on the work of Goemans and Williamson, Bertsimas and Teo [25] propose a primal-dual framework to design and analyze approximation algorithms for covering-type integer programming problems that uses valid inequalities in its design. Hochbaum and Naor [52] consider the network design problem with requirements specified for each subset of vertices of bounded size. They describe an approximation algorithm for the case of a proper requirement function. Jain [55] presents a 2-approximation for the generalized Steiner network problem using its linear programming relaxation and iteratively rounding-off the solution. Karger [58] uses random sampling-based approximation algorithms as a tool for solving undirected graph problems.

2.4 Related Problems

If the capacity installed on each arc is known, the network design problem (2.1)-(2.5) becomes the multi-commodity network flow (MCNF) problem. Ahuja et al. [2] describe general approaches for solving **MCNF** problems, including Lagrangian relaxation and Dantzig-Wolfe decomposition. Barnhart [15] develops dual ascent procedures for solving large-scale **MCNF** problems. Barnhart and Sheffi [16] develop primal-dual heuristics for **MCNF**s. Jones et al. [57] investigate the impact of problem formulation on Dantzig-Wolfe decomposition for the **MCNF** problem. Farvolden et al. [38] solve the **MCNF** problem using primal partitioning and Dantzig-Wolfe decomposition. Leighton et al. [66] develop polynomial-time approximation algorithms for **MCNF** problems. Barnhart et al. [17] solve large-scale **MCNF** problems with column generation methods. Barnhart et al. [19] and Barnhart et al. [22] use branch-and-price and branch-and-price-and-cut to solve large-scale *integer* **MCNF** problems in which the flow of each commodity is constrained to a single path between the commodity's origin and destination. Gabrel et al. [42] use Benders decomposition to solve **MCNF** problems with general step cost functions. This class of problems includes the multi-facility network loading problem as a special case.

Network reliability is important for many real-world applications. The survivable network design problem (SND) seeks a minimum-cost network configuration that provides a specified number of alternate edge-disjoint paths between the network. Goemans and Bertsimas [43] develop two heuristics for the survivable network design problem based on the parsimonious property. Williamson et al. [92] present a primal-dual based approximation algorithm for the SND problem. Gabow et al. [41] improve the efficiency of the Williamson et al. algorithm based on a combinatorial characterization of the "redundant" edges and Padberg and Rao's characterization of minimum odd cuts [81]. Balakrishnan et al. [12] introduce a multi-tier survivable network design problem and propose a solution procedure that solves the single-tier subproblems as matroids. Myung et al. [77] introduce network design models addressing survivability by specifying allowable loss of traffic during a network failure under prescribed conditions and develop an integer programming formulation solved by a heuristic procedure. Balakrishnan et al. [13] present connectivity-splitting models that provide tighter LP relaxation bounds for SND problems.

2.5 Network Design Applications and Service Network Design

Applications of network design arise in a wide range of areas. Aykin [7] presents an integer programming formulation for airline capacitated hub-and-spoke network design and describe an algorithm that first identifies a set of hub locations and then uses Lagrangian relaxation and subgradient optimization to select routes. Lederer and Nambimadom [65] analyze airline network and schedule choice using a model that permits derivation of analytical, closed form expressions for airline and passenger costs. They suggest that it is optimal for a profit maximizing airline to design its network and schedule to minimize the sum of airline and passenger costs.

Alevras et al. [3] develop cutting plane and heuristic approaches for solving the problem of installing capacity on arcs in a telecommunications network. Sung and Jin [87] consider a hub network design problem in which the network service areas are partitioned into predetermined zones (represented by node clusters). The objective is to determine the required routes and hub locations in the predetermined zones such that the total network cost (hub construction and transportation costs) is minimized. A dual-based solution approach is proposed to solve the problem. Riis and Andersen [83] study a capacity-expansion problem in telecommunication network design. They take uncertain demand into consideration and develop a two-stage stochastic integer programming formulation and propose an L-shaped solution procedure with additional facet-defining inequalities. Holmberg and Yuan [54] present a mixed integer programming model for internet protocol traffic network design and two heuristic solution procedures.

Barnhart et al. [23] formulate the railroad blocking problem as a network design problem with maximum degree and flow constraints on the nodes. They propose a heuristic Lagrangian relaxation approach to decompose the problem into a flow subproblem and a block subproblem, and use subgradient optimization to solve the Lagrangian dual.

For service network design, Farvolden and Powell [37] consider the service network design problem in the motor carrier industry and solve the problem using local-improvement heuristics based on subgradients derived from the optimal dual variables of the shipment routing subproblem. Gorman [45] and [46] uses a *tabu-enhanced* genetic search to solve the joint train-scheduling and demand-flow problem in railroad planning. Crainic [34] classifies various decision and management policies in freight transportation into three planning levels: strategic, tactical and operational, and surveys work on the service network design problem, which is increasingly used in tactical planning.

For express shipment service network design, Grünert and Sebastian [48] identify planning tasks faced by postal and express shipment companies and define corresponding optimization models. Leung and Cheung [67] propose models for the ground distribution network design problem. Kuby and Gray [64] consider the single-hub capacitated problem and use the conventional network design formulation to model the problem. They apply the formulation to the Federal Express west-coast hub case, which includes about 20 aircraft and a limited set of routes. Barnhart and Schneur [18] present a formulation for the uncapacitated single-hub problem and use column generation to obtain near-optimal solutions. Kim et al. [61] consider the multi-hub capacitated problem with flexible hub assignment and use a heuristic solution strategy. Building on the efforts of Kim et al., Krishnan et al. [63] present a heuristic algorithm to select routes. Armacost et al. [5] consider the multi-hub capacitated problem with fixed hub assignment. Under a restricted version of the problem, they transform conventional formulations to a new formulation using *composite variables.* The resulting linear programming relaxation gives stronger lower bounds than conventional approaches. They apply the formulation to the UPS next-day air delivery network and report substantial savings.

2.6 Models for Express Shipment Service Network Design

In this section, we first review a formulation, the **RF** formulation, introduced by Kim et al. [61] and Kim [59] for the **ESSND** problem with Flexible Hub Assignment. Next, we present the concept of composite variables and their application in the **ARM** formulation, introduced in Armacost et al. [5] and Armacost [4] for the **ESSND** problem with Fixed Hub Assignment.

To facilitate our discussion of these formulations, we first provide the following definitions:

Definition 1 *Two formulations are* ***equivalent*** *if for any feasible solution to one, there exists a corresponding feasible solution to the other with equal cost (and vice versa).*

Definition 2 *Let A and B be two equivalent (mixed) integer programming formulations, and* A_{LP} *and* B_{LP} *be their respective LP relaxations. We say A is* ***at least as strong as*** *B if, for* *any feasible solution to A_{LP}, we can find a feasible solution to B_{LP} with equal cost.*

2.6.1 Route and Flow Model for the Flexible Hub Assignment Problem

We define a *route*, denoted r, to be a sequence of connected flight legs. A *pickup route* visits one or more gateways and is destined to a hub, and *a delivery route* originates from a hub and visits one or more gateways. We define an *aircraft route*, denoted by (f, r), to be a route flown by fleet type $f \in F$. In the formulations introduced by Kim et al. [61], package flow variables can be represented as flows on arcs, paths or trees. The particular form of the package flow variables affects the number of variables and constraints, but does not affect the LP relaxation bound. Here, we represent package flow decisions using path variables. We define a *package pickup path*, or simply *pickup path*, as a sequence of flight legs from an origin gateway to a hub. Similarly, a *package delivery path*, or simply *delivery path*, is a sequence of flight legs from a hub to a destination gateway. We present the following notation and corresponding Route and Flow (**RF**) formulation introduced in Kim et al. [61] for the flexible hub assignment problem, assuming the package handling costs are negligible.

Sets

$\mathcal{A}$	set of pickup arcs.
$\mathcal{B}$	set of delivery arcs.
K	set of O-D commodities.
F	set of fleet types.
H	set of hubs.
$\mathcal{N}$	set of gateways.

R^f set of routes that can be flown by fleet type f, $f \in F$.

R_P^f set of pickup routes that can be flown by fleet type f, $f \in F$.

R_D^f set of delivery routes that can be flown by fleet type f, $f \in F$.

$\mathcal{P}_k^P$ set of pickup paths for O-D commodity k, $k \in K$.

$\mathcal{P}_k^D$ set of delivery paths for O-D commodity k, $k \in K$.

$\mathcal{P}_k^{P,h}$ set of pickup paths from the origin of O-D commodity k to hub h, $k \in K$, $h \in H^k$.

$\mathcal{P}_k^{D,h}$ set of delivery paths from hub h to the destination of O-D commodity k, $k \in K$, $h \in H^k$.

Data

a_h number of aircraft parking spots at hub $h \in H$.

b_k volume of O-D commodity $k \in K$.

d_r^f cost of flying route r with fleet type f.

e_h sorting capacity of hub $h \in H$.

n_f number of available aircraft of type $f \in F$.

u_r^f capacity of an aircraft of type $f \in F$ assigned to route $r \in R$.

δ_p^{ij} 1 if arc (i,j) is included in path p, and 0 otherwise.

δ_r^{ij} 1 if arc (i,j) is included in route r, and 0 otherwise.

$O(r)$ origin of route r.

$D(r)$ destination of route r.

Decision Variables

y_r^f number of aircraft of fleet type $f \in F$ assigned to route $r \in R^f$.

x_p^k flow of commodity $k \in K$ on pickup (or delivery) path $p \in \mathcal{P}_k^P$ (or $\mathcal{P}_k^D$).

The **RF** formulation is then:

$$\min \sum_{f \in F} \sum_{r \in R_P^f \cup R_D^f} d_r^f y_r^f \tag{2.6}$$

subject to:

$$\sum_{r \in R_P^f : O(r)=i} y_r^f - \sum_{r \in R_D^f : D(r)=i} y_r^f = 0 \quad i \in \mathcal{N},\ f \in F \tag{2.7}$$

$$\sum_{r \in R_D^f : O(r)=h} y_r^f - \sum_{r \in R_P^f : D(r)=h} y_r^f = 0 \quad h \in H,\ f \in F \tag{2.8}$$

$$\sum_{r \in R_P^f} y_r^f \leq n_f \quad f \in F \tag{2.9}$$

$$\sum_{f \in F} \sum_{r \in R_P^f : D(r)=h} y_r^f \leq a_h \quad h \in H \tag{2.10}$$

$$\sum_{f \in F} \sum_{r \in R_P^f} \delta_r^{ij} u_r^f y_r^f - \sum_{k \in K} \sum_{p \in \mathcal{P}_k^P} \delta_p^{ij} x_p^k \geq 0 \qquad (i,j) \in \mathcal{A} \tag{2.11}$$

$$\sum_{f \in F} \sum_{r \in R_D^f} \delta_r^{ij} u_r^f y_r^f - \sum_{k \in K} \sum_{p \in \mathcal{P}_k^D} \delta_p^{ij} x_p^k \geq 0 \qquad (i,j) \in \mathcal{B} \tag{2.12}$$

$$\sum_{p \in \mathcal{P}_k^{P,h}} x_p^k - \sum_{p \in \mathcal{P}_k^{D,h}} x_p^k = 0 \qquad k \in K,\ h \in H^k \tag{2.13}$$

$$\sum_{h \in H^k} \sum_{p \in \mathcal{P}_k^{P,h}} x_p^k = b_k \qquad k \in K \tag{2.14}$$

$$\sum_{k \in K} \sum_{p \in \mathcal{P}_k^{P,h}} x_p^k \leq e_h, \qquad h \in H \tag{2.15}$$

$$y_r^f \in \mathbb{Z}_+,\ r \in R^f,\ f \in F,\ x_p^k \geq 0,\ k \in K,\ p \in \mathcal{P}_k^P \cup \mathcal{P}_k^D$$

Constraints (2.7) and constraints (2.8), defined as *gateway aircraft balance* and *hub aircraft* *balance* constraints, respectively, ensure conservation of flow of aircraft by type at each gateway and each hub. Constraints (2.9), defined as *count* constraints, limit the number of aircraft of each fleet type selected in the solution to be no more than the number available. We need only to specify these constraints for pickup routes because conservation of flow constraints ensure that aircraft count will also be satisfied for delivery. Constraints (2.10), defined as *landing* constraints, ensure that the number of aircraft arriving at a hub does not exceed the parking spots available. We similarly only specify the landing constraints for pickup routes because aircraft conservation of flow ensures satisfaction for delivery. Forcing constraints (2.11) and (2.12) ensure that the package flow on an arc does not exceed the aircraft capacity provided. Constraints (2.13) guarantee the conservation of package flow at hubs. Commodity cover constraints (2.14) require that every commodity is served. Each hub has a *sorting rate*, the number of packages that can be sorted in a unit of time. Given the sort time of a hub, we can calculate the total number of packages that can be sorted at each hub, defined as the *hub sort capacity*. Constraints (2.15), defined as *hub sort* constraints, ensure the total number of packages sorted at a hub does not exceed the hub sorting capacity. In Chapter 4, we show how this set of constraints can be represented dynamically to ensures that all packages can be sorted based on the package arrival time.

The LP relaxation of the **RF** formulation provides a weak and ineffective bound and leads to highly fractional solutions for realistic problem instances (see Kim et al. [61]). To improve the LP bound, Kim et al. [61] apply cut-set inequalities, introduced in Magnanti, Mirchandani, and Vanchani [72], in which the set of nodes $\mathcal{N}$ is partitioned into two subsets, S and T, such that $S \cup T = \mathcal{N}$ and $S \cap T = \emptyset$. The $[S,T]$ cut denotes the set of arcs from S to T, and $D_{S,T}$ is the total demand originating in S and destined for T. Any feasible solution to the network design problem satisfies the following inequalities:

$$\sum_{(i,j)\in[S,T]} \sum_{f\in F} \sum_{r\in R^f} \delta_r^{ij} u_r^f y_r^f \geq D_{S,T} \quad \text{for any } [S,T] \text{ cut.} \tag{2.16}$$

Kim et al. [61] also strengthen the cut set inequalities with Chvátal-Gomory rounding (see Nemhauser and Wolsey [79]). Due to the large number of inequalities, cut-set inequalities are specified only for $|S| \leq 3$ or $|T| \leq 3$.

Even with the addition of the cut-set inequalities, the bound provided by the **RF** LP relaxation improves little. Moreover, the cut set inequalities increase the problem size significantly, resulting in intractability due to insufficient memory, even on high-end workstations. To reduce problem size, Kim et al. [61] present a modified **RF** model containing only route variables and constraints (2.7) through (2.10). In the absence of package flow variables, Kim et al. [61] uses the cut-set inequalities (2.16) as approximation of constraints (2.11)-(2.15). This approximate model does not guarantee a feasible solution to the flexible hub assignment problem. It does, however, provide guidance in potential route selection.

Krishnan et al. [63], building on the earlier efforts of Kim et al. [61], present an iterative process to select routes using the exact and approximate **RF** model. At each iteration, the approximate **RF** model is first solved to select a set of promising routes. With these routes selected, a variant of the multi-commodity flow problem is solved to move all shipments from or to as many gateways as possible. Next, a set of routes is "fixed" and shipments that can be transported by these fixed routes are "eliminated" from the network. This results in a smaller service network design problem. If it can be solved using the exact **RF** model, the exact **RF** model is used to generate the optimal routes for transporting the remaining shipments.

Otherwise, the iterative process repeats, fixing more routes and eliminating more shipments.

2.6.2 Carrier-Specific Route Construction

Although the route structure can be general, the carrier on which this research is based requires that the number of legs on a pickup (or a delivery) route to be no more than two. This restriction allows us to simplify the route construction process as detailed in Kim et al. [61] and Kim [59].

In Figure 2-1, we present an example depicting the time-space network used to generate all feasible routes for a single fleet type. In this example, there are two gateways and a single hub. The network contains a node for each gateway, corresponding to the Earliest Pickup Time (EPT) (e.g., node a), representing the starting node for pickup routes originating from the gateway. The network also contains a node for each hub. Based on the Latest Hub Arrival Time ($LHAT$) at a hub h, we determine for each gateway the latest time an aircraft can depart that gateway and arrive at hub h by $LHAT$ (that is, $LHAT$ minus block time). Then, a node corresponding to that gateway and that time is placed in the time-space network (e.g. node c), and a flight arc is added connecting this node and the hub node (e.g. node d). From the EPT node of gateway i, we consider all possible movements to other gateways at which the aircraft can land. The earliest arrival time at the second gateway j for an aircraft flying from gateway i is the EPT of gateway i plus the aircraft's block time. A node n is placed at the corresponding time and gateway j (e.g., node b) and a flight arc is added, connecting the EPT node of gateway i with node n. Similarly, based on the Latest Delivery Time (LDT) of gateways and Earliest Hub Departure Time ($EHDT$) of hubs, the nodes and flight arcs for the delivery operation are constructed. Ground arcs between successive nodes are then added (e.g. arcs (a, b) and (b, c)), and finally, wrap-around arcs connecting the last node with the *EPT* node are included at each gateway (e.g. arc (k, a)).

The time-space network for each fleet type allows us to construct the feasible set of pickup routes and delivery routes for each fleet type by enumerating the paths from the *EPT* nodes to the hub nodes and the hub nodes to the *LDT* nodes, with the condition that the time interval between two flight arcs on a path must exceed the *minimum turn time* for the fleet type. The minimum turn time for fleet type f is defined as the minimum time required for an aircraft of type f to stay on the ground for loading, unloading, and refueling between flight legs. In the example shown in Figure 2-1, the network contains the path a-b-c-d representing the pickup route from gateway 1 to the hub, path e-f-d representing the pickup route from gateway 2 to the hub, and path e-b-c-d representing the double-leg pickup route 2-1-H. Similarly, path d-g-h represents the delivery route from the hub to gateway 1, path d-l-p-q represents the delivery route from the hub to gateway 2, and path d-g-p-q represents the double-leg delivery route H-1-2.

We can similarly enumerate package pickup or delivery paths using the merged time-space networks for all fleet types. Unlike pickup and delivery routes, package pickup or delivery paths may contain more than two flight arcs, that is, packages might be transferred from one aircraft to another before arrival at hubs in the pickup operation or arrival at destination gateway in the delivery operation. We discuss this in more detail in Chapter 4.

2.6.3 Aircraft Routing Model for the Fixed Hub Assignment Problem

Observing the tractability issues associated with the **RF** model, Armacost et al. [5] developed a new model for express shipment service network design using *composite variables* to reduce

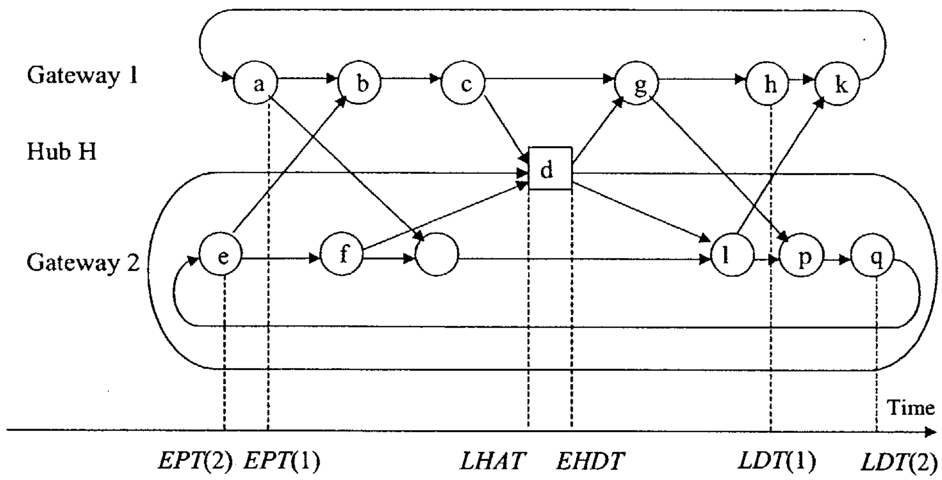


Figure 2-1: Time-space Network for a Single Fleet Type

fractionality of the LP relaxation and enhance tractability. Their composite variable model, however, is designed for **ESSND** problems in which each O-D commodity is assigned *a priori* to a hub. That is, Armacost et al. addresses the fixed hub assignment problem.

Given the fixed hub assignments, O-D commodities originating from the same gateway and assigned to the same hub can be consolidated as a single gateway-hub demand, defined as the *pickup demand* between the gateway-hub pair. Similarly, O-D commodities destined to the same gateway and assigned to the same hub can also be consolidated as a single gateway-hub demand, defined as the *delivery demand* between the gateway-hub pair.

Definition 3 *A* ***composite*** *is a set of aircraft routes that provide sufficient capacity for a set of pickup or delivery demands.*

Example 1 *We illustrate the idea of composite with the example in Figure 2-2. In this simple example, we have a demand of 3 units to be picked up from i to hub h. There is a single fleet* *type, and an aircraft of this type has 2 units of capacity. The operating cost of each aircraft on the route i to H is 10 units.*

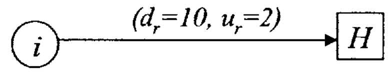


Figure 2-2: An Example of Composite Variable

In the ***RF*** *model introduced by Kim et al. [61], to ensure that the 3 units of demand are served, we specify a constraint*

$$2\,y \geq 3$$

with variable y representing the number of aircraft selected. The optimal solution to the LP relaxation is then 1.5 aircraft, with 15 units of operating cost.

In contrast, with the composite variable approach developed by Armacost et al., we design a composite c representing two aircraft going from i to h and providing sufficient capacity to serve the 3 units of demand. The condition that the demand must be served can then be specified as

$$c \geq 1$$

In the optimal solution to the LP relaxation using composite variables, c equals one, implying that two aircraft are selected to serve the demand, with a total operating cost of 20 units. This small example illustrates the improved LP relaxation bound achievable with composite variables.

We denote composites built with gateway-hub pickup and delivery demands, given the fixed hub assignment, as *demand composites.*

Notation for the Aircraft Routing Model (**ARM**), given the fixed hub assignments, is as follows.

Set

$\mathcal{C}_P$ set of pickup demand composite.

$\mathcal{C}_D$ set of delivery demand composite.

Data

b_P^{ih} pickup demand between gateway i and hub h.

b_D^{ih} delivery demand between gateway i and hub h.

γ_c^r number of aircraft routes r in composite c.

d_c cost of composite c, $d_c = \sum_{r \in R} \gamma_c^r d_r$.

γ_c^f number of aircraft of fleet type f in composite c.

$\gamma_c^f(\bar{i})$ number of aircraft of fleet type f originating at gateway i (or hub h) in demand composite c.

$\gamma_c^f(\underline{i})$ number of aircraft of fleet type f destined to gateway i (or hub h) in demand composite c.

$\delta_c^{i,h,P}$ 1 if demand composite c covers the pickup demand between gateway i and hub h, and 0 otherwise.

$\delta_c^{i,h,D}$ 1 if demand composite c covers the delivery demand between gateway i and hub h, and 0 otherwise.

Decision Variable

v_c equals 1 if composite c is selected, and 0 otherwise.

The Aircraft Routing Model (**ARM**) using demand composite variables is cast as:

$$\min \sum_{c \in \mathcal{C}_P \cup \mathcal{C}_D} d_c v_c \tag{2.17}$$

subject to

$$\sum_{c \in \mathcal{C}_P} \gamma_c^f(\bar{i}) v_c - \sum_{c \in \mathcal{C}_D} \gamma_c^f(\underline{i}) v_c = 0 \qquad i \in \mathcal{N},\ f \in F \tag{2.18}$$

$$\sum_{c \in \mathcal{C}_P} \gamma_c^f(\underline{h}) v_c - \sum_{c \in \mathcal{C}_D} \gamma_c^f(\bar{h}) v_c = 0 \qquad h \in H,\ f \in F \tag{2.19}$$

$$\sum_{C \in \mathcal{C}_P} \gamma_c^f v_c \leq n_f \qquad f \in F \tag{2.20}$$

$$\sum_{f \in F} \sum_{v \in \mathcal{C}_P} \gamma_c^f(\underline{h}) v_c \leq a_h \quad h \in H \tag{2.21}$$

$$\sum_{c \in \mathcal{C}_P} \delta_c^{i,h,P} v_c = 1 \quad (i,h) : b_P^{ih} > 0,\ i \in \mathcal{N},\ h \in H \tag{2.22}$$

$$\sum_{c \in \mathcal{C}_D} \delta_c^{i,h,D} v_c = 1 \quad (i,h) : b_D^{ih} > 0,\ i \in \mathcal{N},\ h \in H \tag{2.23}$$

$$v_c \in \{0,1\},\ c \in \mathcal{C}_P \cup \mathcal{C}_D \tag{2.24}$$

Constraints (2.18) through (2.21) are gateway balance, hub balance, count and landing constraints, which are similar to constraints (2.7), (2.8), (2.9), and (2.10) in the **RF** model. The package flow variables, forcing constraints (2.11) and (2.12), and commodity cover constraints (2.14) in **RF** are replaced by constraints (2.22) and (2.23), denoted the *cover constraints*, in **ARM**. These cover constraints ensure that at least one composite is selected to cover each nonzero gateway-hub demand. Because each demand composite is guaranteed to serve the associated gateway-hub demands fully, the cover constraints also ensure satisfaction of the aircraft capacity constraints.

Armacost [4] and Armacost et al. [5] show that the **ARM** model is at least as strong as the **RF** model. In fact, for realistic problem instances such as the UPS Next-Day fixed hub assignment problem, the LP relaxation of the **ARM** model provides much stronger bounds than the LP relaxation of the **RF** model.

Chapter 3

Integrated Next-Day and Second-Day Air Express Package Delivery

Express shipment carriers typically offer different levels of services, using the same equipment and facilities to provide the services, though maybe at different times. For example, the carrier we consider offers both Next-Day and Second-Day express services, using the aircraft for Next-Day Air (NDA) service during the night and Second-Day Air (SDA) service during the day. Currently, at the carrier, the NDA and SDA operations are designed in a sequential manner, with the design of the first operation defining the location of aircraft at the interface between the NDA and SDA operations. If, however, the NDA and SDA operations are designed simultaneously, aircraft movements can be better coordinated and costs significantly reduced.

Another shortcoming of the carrier's current practice is that the service network planning problem is modeled as a daily problem, that is, it is assumed that the same schedule is repeated everyday. In practice, demands vary by day of the week, and carriers adapt their schedules accordingly. For example, the demand for Saturday delivery is typically small, and thus, the carrier might have a different schedule for Saturday. At the carrier we consider, the design of different schedules are also treated as isolated problems with interface conditions.

Under the fixed hub assignment assumption, we can extend the **ARM** model to consider the service network design problems for multiple services or across multiple days. Unfortunately, the **ARM** model is already on the edge of solvability; a high end workstation with 2GB RAM had insufficient memory to produce even a feasible solution for the extended **ARM** model integrating the Next-Day and Second-Day problems with Fixed Hub Assignment.

In this chapter, we first present an extended **ARM** model to consider the Integrated Next-Day and Second-Day problem with Fixed Hub Assignment (Integrated NDA–SDA problem). We then present a decomposition solution approach breaking the problem into smaller hub pickup and delivery sub-problems. Applying our decomposition approach to the UPS Integrated NDA-SDA problem, we demonstrate the scalability of our approach and its practical significance.

In Section 3.1, we describe the interaction of Next-Day and Second-Day operations and define the Integrated NDA-SDA problem. In Section 3.2, we present a formulation for the Integrated NDA-SDA problem. In Section 3.3, we introduce a new approach we call *information-enhanced column generation.* In Section 3.4, we apply information-enhanced column generation to the UPS Integrated NDA-SDA problem and compare the results with both the current UPS solution and with results obtained from sequential approaches. In Section 3.5, we summarize our work.

3.1 Problem Description

We describe the Next-Day and Second-Day express shipment delivery operation of a large carrier and define the Integrated NDA-SDA problem in this section.

3.1.1 Next-Day Operation

Figure 3-1 depicts the service timeline for the Next-Day operation, assuming packages are collected on Day 1. Carriers typically schedule pickup of packages from customers as late as possible to allow customers sufficient time to prepare their packages. Hence, packages typically arrive at origin ground centers in the late afternoon or early evening. After the origin sort in the evening, packages to be picked up by air service are transported to origin gateways at night and loaded onto aircraft. From origin gateways, aircraft follow NDA pickup routes, arriving at hubs in the late night or early morning of the next day. The hub sort for NDA packages typically starts around midnight and lasts for 2-3 hours. After the hub sort, aircraft follow NDA delivery routes, delivering packages to their destination gateways. From destination gateways, packages are transported to destination ground centers, arriving in the early morning of the next day. After the destination sort, packages are loaded onto ground vehicles and delivered to customers in the morning. The same Next-Day air operation, starting with air pickup and ending with air delivery, is repeated each day except Sunday.

3.1.2 Second-Day Operation

The Second-Day operation is similar to the Next-Day operation except for an expanded service time. Figure 3-2 depicts the service timeline for the Second-Day operation, assuming packages are collected on Day 1. At the origin ground center, the origin sort for Second-Day packages

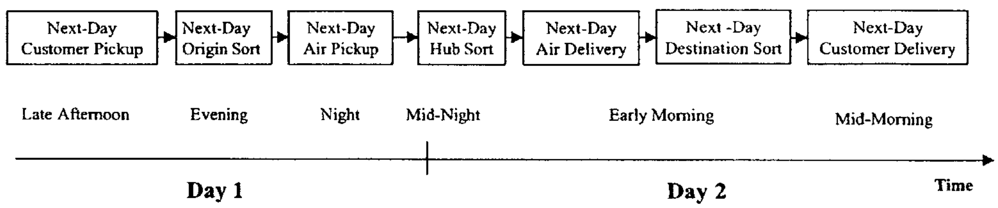


Figure 3-1: Next-Day Air Operations

begins at night after the origin sort for Next-Day packages is completed. Then, packages to be transported via air service (SDA packages) stay at the origin ground center overnight, while others are transported to destination ground centers or hubs via ground service. On the morning of the next day, SDA packages at origin ground centers are transported to gateways and loaded onto aircraft that have just completed their NDA delivery routes. Aircraft then follow SDA pickup routes, arriving at hubs before noon. After the hub sort, packages are delivered either to destination ground centers via ground service or to destination gateways via air service. In the case of air delivery, aircraft carrying SDA packages arrive at destination gateways in the evening of Day 2. After SDA packages are unloaded, aircraft are available to begin their NDA pickup routes. The unloaded SDA packages are transported to destination ground centers, where they wait overnight for other Second-Day packages transported via ground. On the morning of Day 3, the destination sort for Second-Day packages begins after the destination sort for NDA packages collected on Day 2 is completed. Packages are then delivered to customers in the afternoon. Note that compared with the Next-Day service, the extended service time allows more extensive use of ground transport.

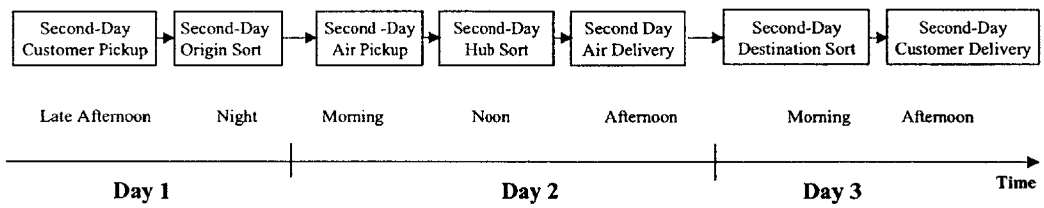


Figure 3-2: Second-Day Air Operation

Even though the complete Second-Day operation spans three days as depicted in Figure 3-2, we can model the SDA operation as a *daily problem,* that is, the same operation is repeated daily, because a new Second-Day operation initiates each day. We illustrate this concept as follows. In Figure 3-3, we depict Second-Day operations over three days. The number in the parenthesis at the upper left corner of each box indicates the starting day of the corresponding SDA operation. We refer to a Second-Day operation starting on Day n as *Second-Day Operation* n. On any given day n, there are three sets of Second-Day activities underway, one set for packages entering the system on Day $(n-2)$, one set for those entering on Day $(n-1)$, and finally, one for those on Day n. In the morning of Day n, the destination sort for Second-Day Operation $(n-2)$ is conducted and packages of Second-Day Operation $(n-1)$ are transported by air to hubs. Around noon, packages of Second-Day Operation $(n-1)$ are sorted at hubs, and then, in the afternoon, packages of Second-Day Operation $(n-2)$ are delivered to customers. Next, in the late afternoon, packages of Second-Day Operation $(n-1)$ are delivered by air to destination gateways, and packages of Second-Day Operation n are collected from customers. Finally in the night, the origin sort for the Second-Day operation n is conducted. As is evident in Figure 3-3, the same air operation is repeated daily in Second-Day operations.

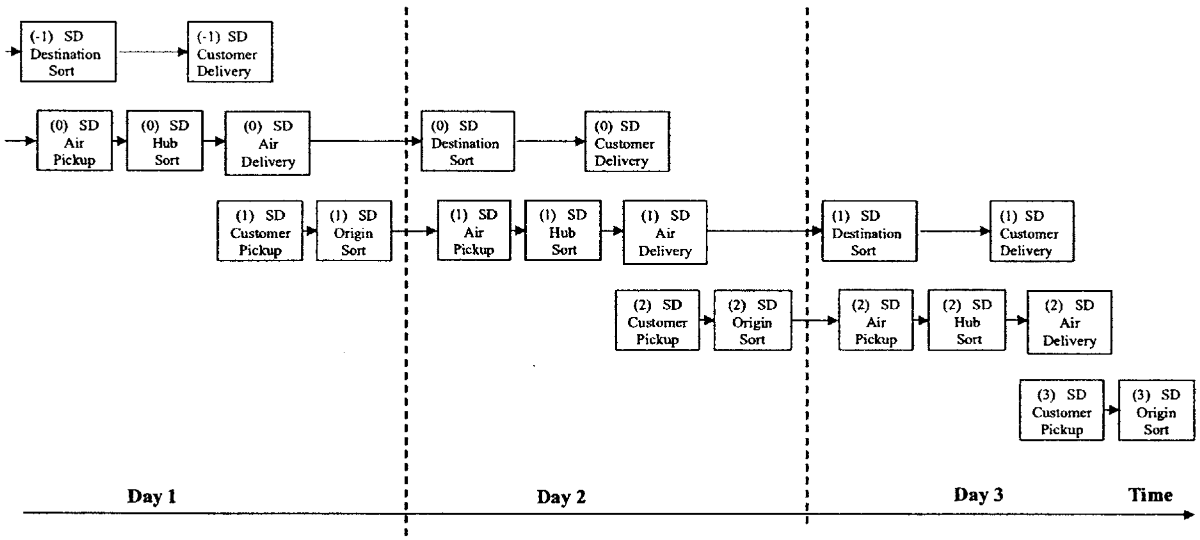


Figure 3-3: Daily Second-Day Air Operation

3.1.3 Integrated Next-Day Air and Second-Day Air Problem

The facility and equipment requirements for the express shipment carriers are largely driven by the demand for Next-Day service due to its tight delivery time windows, with the same resources used for both NDA and SDA services. The Next-Day and Second-Day operations are linked through the positioning of equipment. We illustrate this in the next example.

Example 2 *Figure 3-4 depicts a flight schedule, starting in the early morning with NDA delivery flights and ending at night with NDA pickup flights. There are two NDA delivery flights* *from hub A to gateway B and one from hub A to gateway C. Thus, two aircraft are available at gateway B for SDA pickup operations, and one available at gateway C, after the NDA operation. According to the schedule, however, two aircraft are required for SDA pickup flights from gateway C. To satisfy this, one aircraft is **ferried** from gateway B to gateway C before the SDA pickup operation. Ferry flights represent empty aircraft movement, and the cost structure of ferry flights is similar to that of pickup and delivery flights, consisting of both fixed cycle charges and block time costs.*

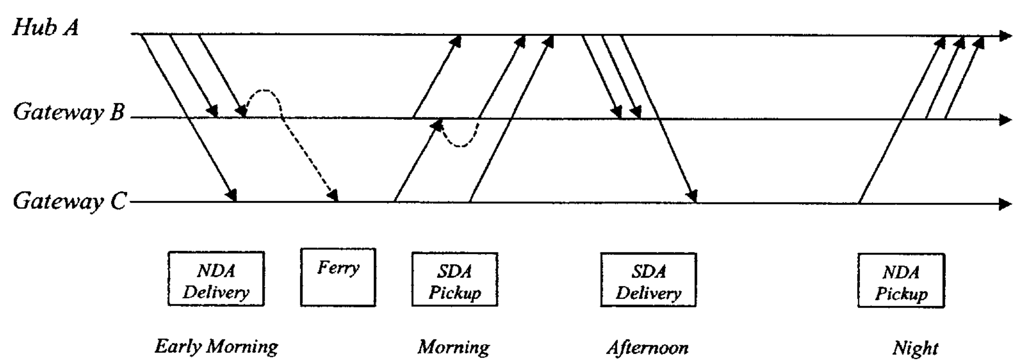


Figure 3-4: Aircraft Cycle for NDA and SDA Operations

In the carrier's current practice, the Next-Day and Second-Day service schedules are designed in a sequential manner. Once the schedule for one service is designed, the aircraft positions at the beginning of pickup operations and end of delivery operations are fixed as the boundary conditions for the other problem. In the above example, suppose we consider the NDA operations first and obtain the NDA schedule by ignoring the SDA operations. In a sequential approach, we next design the SDA network, but with the following conditions: (1) two aircraft are available at gateway B and one at gateway C at the beginning of the SDA pickup operation; and (2) two aircraft must be sent to gateway B and one to gateway C at the end of the SDA delivery operation.

Such a sequential approach results in sub-optimal solutions to the integrated problem. The next example illustrates such a case.

Example 3 *Consider the network depicted in Figure 3-5. Suppose there is a single fleet type, and an aircraft of this type has 2 units of capacity. We follow a sequential approach and consider the NDA network first. We balance the aircraft needed for the NDA pickup operations with those available after the NDA delivery and ferry operations. The optimal NDA solution includes pickup routes 1-H_1 and 2-H_1, delivery routes H_1-2 and H_1-2-1, and has an objective value 48.*

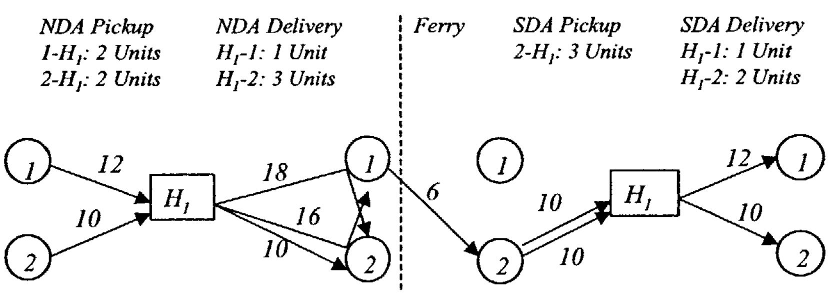


Figure 3-5: Sub-optimal Solution from the Sequential Approach

Using the number of aircraft available at gateways after the NDA delivery operations and the number of aircraft required for the NDA pickup operations as the fleet positioning conditions, we design the SDA network. Because there is no SDA pickup demand from gateway 1 to the hub and two aircraft are needed to pickup the SDA demand from gateway 2, the aircraft at gateway 1 has to be ferried to gateway 2 before the SDA pickup operations. The SDA solution *then includes pickup route 2-H_1 twice, delivery routes H_1-1 and H_1-2, and ferry route 1-2. The corresponding objective value of the SDA solution is* 48. *The total NDA and SDA cost of this sequentially generated solution is* 96.

In contrast, if we consider the integrated NDA and SDA problem, we obtain the optimal solution by replacing route H_1-2-1 with route H_1-1-2 in the NDA delivery operation. This allows us to eliminate the ferrying, and the total cost is 92.

Given the Next-Day and Second-Day demands, the Integrated Next-Day and Second-Day Problem seeks to determine simultaneously the cost-minimizing service network design for the Next-Day and Second-Day operations. Assuming fixed hub assignment, demand inputs are in the form of gateway-hub pickup or delivery demands, and all demands that can be transported by ground service are not considered in the air service network design problem. The resulting problem is the Integrated Next-Day Air (NDA) and Second-Day Air (SDA) problem.

3.2 Integrated NDA-SDA Problem Formulation

With the fixed hub assignment assumption, we can apply the demand composite concept to the integrated problem. We first introduce the following notation. Let T indicate the type of service, Next-Day (denoted N) or Second-Day (denoted S); and O indicate the operation, pickup (denoted P) or delivery (denoted D). We then define the following sets and variables

Sets

F set of fleet types.

H set of hubs.

$\mathcal{N}$ set of gateways.

$\mathcal{C}^T$ set of demand composites for the NDA ($T = N$) or SDA ($T = S$) network.

$\mathcal{C}^T_O$ $\begin{cases} \text{set of pickup } (O = P) \text{ or delivery } (O = D) \text{ demand composites for the} \\ \text{NDA } (T = N) \text{ or SDA } (T = S) \text{ network.} \end{cases}$

Data

a^T_h $\begin{cases} \text{number of aircraft parking spots at hub } h \text{ for NDA } (T = N) \\ \text{or SDA } (T = S) \text{ network.} \end{cases}$

$b^{ih}_{T,O}$ $\begin{cases} \text{pickup } (O = P) \text{ or delivery } (O = D) \text{ demand between gateway } i \text{ and hub } h \\ \text{for NDA } (T = N) \text{ or SDA } (T = S) \text{ network.} \end{cases}$

γ^r_c number of aircraft routes r in demand composite c.

d_c cost of demand composite c, $d_c = \sum_{r \in c} \gamma^r_c d_r$.

d^f_{ij} ferrying cost for an aircraft of type f ferried from gateway i to j.

n^T_f number of aircraft of type f available for NDA ($T = N$) or SDA ($T = S$) network.

γ^f_c number of aircraft of type f in demand composite c.

$\gamma^f_c(\bar{i})$ number of aircraft of fleet type f originating at gateway i (or hub h) in demand composite c.

$\gamma^f_c(\underline{i})$ number of aircraft of fleet type f destined to gateway i (or hub h) in demand composite c.

$$\delta^{ih}_{T,O,c} = \begin{cases} 1 & \text{if demand composite } c \text{ covers NDA } (T = N) \text{ or SDA } (T = S) \text{ pickup} \\ & (O = P) \text{ or delivery } (O = D) \text{ demand between gateway } i \text{ and hub } h\text{, and} \\ 0 & \text{otherwise.} \end{cases}$$

Decision Variables

v_c equals 1 if demand composite c is selected, and 0 otherwise.

$\varpi_{f,i}^{T,O}$ $\left\{ \begin{array}{l} \text{number of aircraft of type } f \text{ on the ground at gateway (hub) } i \text{ during} \\ \text{NDA } (T = N) \text{ or SDA } (T = S) \text{ pickup } (O = P) \text{ or delivery} \\ (O = D) \text{ operation. } \varpi_{f,i}^{T,P} = \varpi_{f,i}^{T,D}, \text{ if } i \notin H. \end{array} \right.$

$\phi_{ij}^{T,f}$ $\left\{ \begin{array}{l} \text{number of aircraft of type } f \text{ ferried from gateway (hub) } i \text{ to } j \text{ after} \\ \text{the NDA } (T = N) \text{ or SDA } (T = S) \text{ operation.} \end{array} \right.$

We present the following formulation (**INS**) for the Integrated NDA-SDA problem:

$$\min \sum_{T=\{N,S\}} \sum_{c \in \mathcal{C}^T} d_c v_c + \sum_{T=\{N,S\}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} d_{ij}^f \phi_{ij}^{T,f} \tag{3.1}$$

subject to

$$\sum_{c \in \mathcal{C}_D^S} \gamma_c^f(\underline{i}) v_c - \sum_{c \in \mathcal{C}_P^N} \gamma_c^f(\overline{i}) v_c - \varpi_{f,i}^{N,P} + \varpi_{f,i}^{S,D} + \sum_{j \in \mathcal{N}, j \neq i} \phi_{ji}^{S,f} - \sum_{j \in \mathcal{N}, j \neq i} \phi_{ij}^{S,f} = 0, \; i \in \mathcal{N}, \; f \in F \tag{3.2}$$

$$\sum_{c \in \mathcal{C}_D^N} \gamma_c^f(\underline{i}) v_c - \sum_{c \in \mathcal{C}_P^S} \gamma_c^f(\overline{i}) v_c + \varpi_{f,i}^{N,D} - \varpi_{f,i}^{S,P} + \sum_{j \in \mathcal{N}, j \neq i} \phi_{ji}^{N,f} - \sum_{j \in \mathcal{N}, j \neq i} \phi_{ij}^{N,f} = 0, \; i \in \mathcal{N}, \; f \in F \tag{3.3}$$

$$\sum_{c \in \mathcal{C}_P^T} \gamma_c^f(\underline{h}) v_c + \varpi_{f,h}^{T,P} - \sum_{c \in \mathcal{C}_D^T} \gamma_c^f(\overline{h}) v_c - \varpi_{f,h}^{T,D} = 0, \qquad h \in H, \; f \in F, \; T = \{N, S\} \tag{3.4}$$

$$\sum_{c \in \mathcal{C}_P^T} \gamma_c^f v_c \leq n_f, \qquad f \in F, \; T = \{N, S\} \tag{3.5}$$

$$\sum_{f \in F} \sum_{c \in \mathcal{C}_P^T} \gamma_c^f(\underline{h}) v_c \leq a_h, \quad h \in H, \ T = \{N, S\} \tag{3.6}$$

$$\sum_{c \in \mathcal{C}_O^T} \delta_{T,O,c}^{ih} v_c \geq 1, \quad (i,h) : b_{T,O}^{ih} > 0, \ T = \{N, S\}, \ O = \{P, D\}, \ i \in \mathcal{N}, \ h \in H \tag{3.7}$$

$$v_c \in \{0,1\} \text{ for all } c \in \mathcal{C}^N \cup \mathcal{C}^S, \ \varpi_{f,i}^{T,O} \in \mathbb{Z}_+ \text{ for } T = \{N, S\}, \ O = \{P, D\}, \ i \in \mathcal{N}$$

$$\phi_{ij}^{N,f}, \phi_{ij}^{S,f} \in \mathbb{Z}_+ \ \text{ for } i,j \in \mathcal{N}, i \neq j, \ f \in F, \ T = \{N, S\} \tag{3.8}$$

The objective is to minimize the sum of the total NDA and SDA operating costs and the ferry costs between the operations. Constraints (3.2) and (3.3), called *boundary balance constraints*, ensure that aircraft at gateways are balanced between NDA and SDA operations. Constraints (3.2) require that the number of aircraft of type f originating and staying on the ground at a gateway (hub) i in the NDA pickup operation equals the number of aircraft of type f destined to and staying on the ground at gateway (hub) i in the SDA delivery operation, adjusted by the number of aircraft of type f ferried into and out of gateway (hub) i at the end of the SDA operation. Constraints (3.3) similarly require that the number of aircraft of type f destined to and staying at a gateway (hub) i in the NDA delivery operation equals the number of aircraft of type f originating and staying at gateway (hub) i in the SDA pickup operation, adjusted by the number of aircraft of type f ferried into and out of gateway (hub) i at the end of the NDA operations. Note that if an aircraft stays on the ground at a gateway location, it has to stay on the ground throughout NDA or SDA operation. In contrast, aircraft can stay on the ground during only pickup or delivery operations at a hub location. Constraints (3.4), (3.5), (3.6) and (3.7) are hub balance, count, landing and cover constraints similar to those in the **ARM** model, specified for both the NDA and SDA network.

3.3 Column Generation for the Integrated NDA-SDA Formulation

Populating the **INS** formulation with all possible variables results in an intractable model: computer memory requirements and solution times are excessive. To address this issue, we use column generation to reduce the number of columns considered in solving the IP.

In column generation, we maintain a restricted version of the original model, called the Restricted Master Problem (**RMP**), which includes only a limited set of columns. At each so called master iteration, we solve the **RMP** to obtain a set of dual prices. Using this set of dual prices, we can either compute the reduced cost of each column explicitly, or solve a pricing sub-problem, as in Dantzig-Wolfe decomposition (Dantzig and Wolfe [36]), to identify columns that can potentially improve the objective value of the **RMP**. If a problem has a diagonal block structure as shown in Figure 3-6, pricing sub-problems can be specified for each block, resulting in simpler sub-problems. We repeat the process until no column is generated in one master iteration.

In this section, we explore different solution approaches for the **INS** formulation. We refer to the first approach as *naive column generation*; an approach in which demand composite variables with negative-reduced cost are generated as needed, with restrictions on the number of variables generated per iteration. In the second approach, referred to as *aggregate information-enhanced column generation*, smaller hub pickup or delivery sub-problems are solved to generate

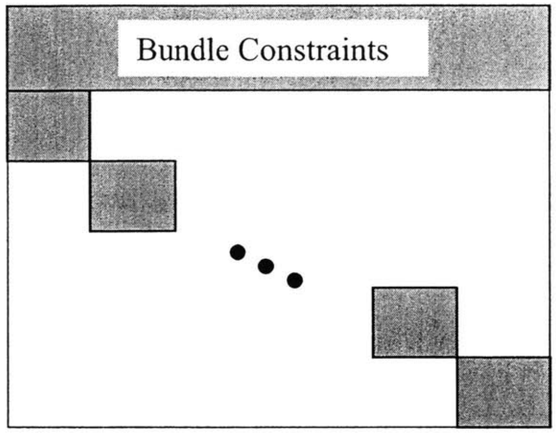


Figure 3-6: Block-Angular Matrix Structure

the necessary variables, and a master column represents the network design for the pickup or delivery operation of a hub. In the third approach, referred to as *disaggregate information-enhanced column generation*, we similarly solve hub pickup or delivery sub-problems, but each master column represents a demand composite variable, and we partition the hub sub-problem solution, that is, the solution to the pricing problem, into individual demand composites when adding columns to the **RMP**. In each of these solution approaches, we limit column generation to the root node LP relaxation, and consider only columns generated in solving the root node LP in branch-and-bound.

We apply these solution approaches to the UPS NDA problem to investigate their respective effectiveness. The UPS NDA network includes 101 gateways, 7 hubs, 9 fleet types, 198 pickup and 195 delivery gateway-hub demands. Formulation statistics are reported in Table 3.1.

Columns	195,009
Rows	3,302
Nonzeros	2,062,466

Table 3.1: UPS Next-Day Air Network Design Problem Statistics

All computations were performed on an HP C3000 workstation with 400MHz CPU and 2GB RAM, running HPUX 10.20. The models and column generation processes were compiled using HP's aCC compiler with calls to the ILOG CPLEX 6.5 Callable Library [32]. CPLEX MIP Solver settings are reported in Table 3.2. For parameters not indicated, the CPLEX default values were used.

Parameter	Setting
Backtrack	0.85
Branching Direction	Up direction selected first
Node Selection	Best estimate search
Variable Selection	Based on strong branching
Relative Best IP-Best Bound Gap Tolerance	0.0001

Table 3.2: Settings for CPLEX 6.5 MIP Solver

3.3.1 Naive Column Generation

In *naive column generation*, we evaluate the cost of demand composite variables explicitly using the dual prices obtained from solving the **RMP**. Denote the objective coefficient vector for demand composite variables as $\mathbf{d}$, and the constraint matrix for demand composite variables in constraints (3.2), (3.3), (3.4), (3.5), (3.6) and (3.7) as $\mathbf{B}_1, \mathbf{B}_2, \mathbf{H}, \mathbf{N}, \mathbf{A}$ and $\mathbf{C}$, respectively, and let the dual vector of the corresponding constraints be denoted $\pi^{\mathbf{B}_1}$, $\pi^{\mathbf{B}_2}$, $\pi^{\mathbf{H}}$, $\pi^{\mathbf{N}}$, $\pi^{\mathbf{A}}$ and $\pi^{\mathbf{C}}$. The reduced cost vector of demand composite variables is given by

$$\mathbf{d}' - (\pi^{\mathbf{B}_1})'\mathbf{B}_1 - (\pi^{\mathbf{B}_2})'\mathbf{B}_2 - (\pi^{\mathbf{H}})'\mathbf{H} - (\pi^{\mathbf{N}})'\mathbf{N} - (\pi^{\mathbf{A}})'\mathbf{A} - (\pi^{\mathbf{C}})'\mathbf{C}$$

Demand composite variables with negative reduced cost are generated when solving the LP relaxation. In order to limit the size of the integer programming model, we evaluate the effect of limiting the number of columns generated in one iteration to at most 100, 500, 1000, 2000, and 4000, respectively, and determine that generating at most 1000 columns in an iteration results in the fewest number of columns generated.

Our results for the naive column generation approach, limiting the number of columns generated in one iteration to at most 1000, are reported in Table 3.3. For comparison, we also solve the problem with *all* demand composite variables present, referred to as the *all-column approach*. "AC" represents the all-column approach, and "NCG" represents the naive column generation approach. In both approaches, the optimal LP value is the same. The objective value of the best IP solution using the naive column generation approach is 0.01% higher than that obtained with the all-column approach. This difference is explained by the fact that we generate columns only at the root node of the branch-and-bound tree, and hence, we do not consider certain demand composite variables whose reduced cost becomes negative as we branch in the branch-and-bound solution algorithm. This small degradation of the objective value is compensated for by the reduction in algorithmic complexity resulting from limiting column generation to the root node. In comparing running times, the naive column generation approach takes less than one fifth of the time required by the all-column approach.

Solution Approach		AC	NCG
Columns. Generated		-	16259
IP Objective Value		-	+0.01%
Run Time (sec.)	Root Node LP	28	23
	IP	8692	1550

Table 3.3: All-Column and Naive Column Generation Results for the UPS NDA Problem

3.3.2 Aggregate Information-Enhanced Column Generation

Instead of generating *individual* demand composite variables with negative reduced cost, we generate a *set* of demand composite variables that is both feasible and has, summing over the demand composites in the set, a negative reduced cost.

We define a *set* of pickup (or delivery) demand composites to be a *hub pickup* (or *delivery*) *composite* if it: (1) includes integer number of aircraft routes; and (2) satisfies the count constraints, and the landing constraints and cover constraints specified for the pickup (or delivery) gateway-hub demands, at a set of hubs. We introduce the following additional notation.

Sets and Data

$\mathcal{H}^T$ set of hub composites for the NDA ($T = N$) or SDA ($T = S$) network.

$\mathcal{H}_O^T$ set of pickup ($O = P$) or delivery ($O = D$) hub composites for the NDA ($T = N$) or SDA ($T = S$) network.

d_Θ cost of hub composite Θ, $d_\Theta = \sum_{c \in \Theta} d_c$.

γ_Θ^f number of aircraft of type f in hub composite Θ.

$\gamma_\Theta^f(\bar{i})$ number of aircraft of type f originating at gateway (hub) i in hub composite Θ.

$\gamma_\Theta^f(\underline{i})$ number of aircraft of type f destined to gateway (hub) i in hub composite Θ.

$$\delta_{T,O,\Theta}^{ih} = \begin{cases} 1 & \text{if hub composite } \Theta \text{ covers NDA } (T = N) \text{ or SDA } (T = S) \text{ pickup} \\ & (O = P) \text{ or delivery } (O = D) \text{ demand between gateway } i \text{ and hub } h\text{, and} \\ 0 & \text{otherwise.} \end{cases}$$

Decision Variables

v_Θ equals 1 if hub composite Θ is selected, and 0 otherwise.

We re-write the **INS** formulation with hub composite variables (**INS-H**) as follow.

$$\min \sum_{T=\{N,S\}} \sum_{\Theta \in \mathcal{H}^T} d_\Theta v_\Theta + \sum_{T=\{N,S\}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} d^f_{ij} \phi^{T,f}_{ij} \tag{3.9}$$

subject to

$$\sum_{\Theta \in \mathcal{H}^S_D} \gamma^f_\Theta(\underline{i}) v_\Theta - \sum_{\Theta \in \mathcal{H}^N_P} \gamma^f_\Theta(\bar{i}) v_\Theta - \varpi^{N,P}_{f,i} + \varpi^{S,D}_{f,i} + \sum_{j \in \mathcal{N}, j \neq i} \phi^{S,f}_{ji} - \sum_{j \in \mathcal{N}, j \neq i} \phi^{S,f}_{ij} = 0, \; i \in \mathcal{N}, \; f \in F \tag{3.10}$$

$$\sum_{\Theta \in \mathcal{H}^N_D} \gamma^f_\Theta(\underline{i}) v_\Theta - \sum_{\Theta \in \mathcal{H}^S_P} \gamma^f_\Theta(\bar{i}) v_\Theta + \varpi^{N,D}_{f,i} - \varpi^{S,P}_{f,i} + \sum_{j \in \mathcal{N}, j \neq i} \phi^{N,f}_{ji} - \sum_{j \in \mathcal{N}, j \neq i} \phi^{N,f}_{ij} = 0, \; i \in \mathcal{N}, \; f \in F \tag{3.11}$$

$$\sum_{\Theta \in \mathcal{H}^T_P} \gamma^f_\Theta(\underline{h}) v_\Theta + \varpi^{T,P}_{f,h} - \sum_{\Theta \in \mathcal{H}^T_D} \gamma^f_\Theta(\overline{h}) v_\Theta - \varpi^{T,D}_{f,h} = 0, \qquad h \in H, \; f \in F, \; T = \{N,S\} \tag{3.12}$$

$$\sum_{\Theta \in \mathcal{H}^T_P} \gamma^f_\Theta v_\Theta \leq n_f, \qquad f \in F, \; T = \{N,S\} \tag{3.13}$$

$$\sum_{f \in F} \sum_{c \in \mathcal{H}^T_P} \gamma^f_\Theta(\underline{h}) v_c \leq a_h, \quad h \in H, \; T = \{N,S\} \tag{3.14}$$

$$\sum_{\Theta \in \mathcal{H}^T_O} \delta^{ih}_{T,O,\Theta} v_\Theta \geq 1, \quad (i,h) : b^{ih}_{T,O} > 0, \; T = \{N,S\}, \; O = \{P,D\}, \; i \in \mathcal{N}, \; h \in H \tag{3.15}$$

$$v_\Theta \in \{0,1\} \text{ for all } \Theta \in \mathcal{H}^N \cup \mathcal{H}^S, \; \varpi^{T,O}_{f,i} \in \mathbb{Z}_+ \text{ for } T = \{N,S\}, \; O = \{P,D\}, \; i \in \mathcal{N}$$

$$\phi^{N,f}_{ij}, \phi^{S,f}_{ij} \in \mathbb{Z}_+ \text{ for } i,j \in \mathcal{N}, \; i \neq j, \; f \in F, \; T = \{N,S\} \tag{3.16}$$

The formulation is the same as the **INS** formulation except that demand composite variables are replaced with hub composite variables. We can prove the following lemma:

Lemma 1 *The **INS-H** formulation is at least as strong as the **INS** formulation.*

Proof. We first prove that the two formulations are equivalent integer programming formulations. Given any integer feasible solution to the **INS-H** formulation, we can partition hub composite variables into demand composite variables. The resulting demand composite variable solution is equivalent to the hub composite variable solution, and thus, satisfies all the constraints in the **INS** formulation.

Conversely, given any integer feasible solution to the **INS** formulation, all the selected pickup and delivery demand composite variables form trivial pickup and delivery hub composite variables for the complete hub set, respectively, satisfying all the constraints specified for the **INS-H** formulation.

The first part of the proof holds even if the integrality condition is relaxed. Hence, the **INS-H** formulation is at least as strong as the **INS** formulation. ■

In the next example, we describe a case in which the **INS-H** formulation is strictly stronger than the **INS** formulation.

Example 4 *Consider the example in Figure 3-7. There is a single fleet type with 2 units of capacity. We want to cover all gateway-hub demands in the example. We only consider the pickup operation for simplicity in the examples presented hereafter, but we can easily expand the examples to include delivery operations and aircraft balance without affecting formulation strength.*

*We consider only the demand composite variables and hub composite variables in the figure. (Other demand and hub composite variables do not affect the optimal integer or LP relaxation solution to the **INS** and **INS-H** formulation.) Excluding the balance, count, and landing*

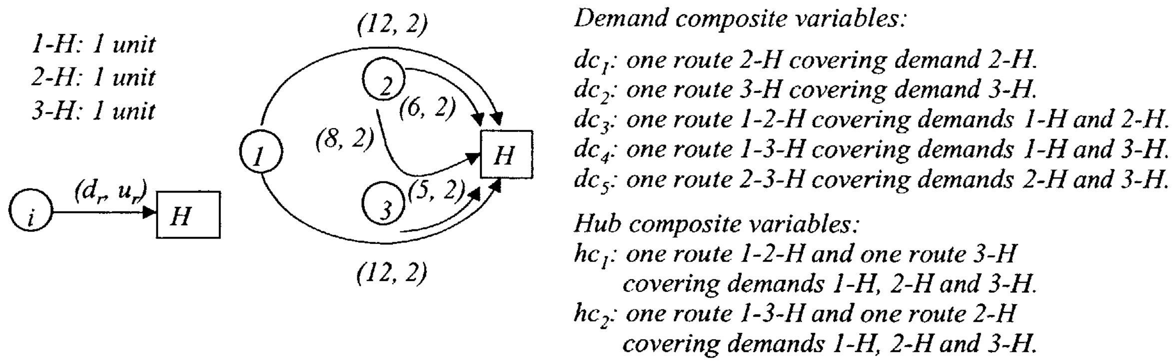


Figure 3-7: Example of Hub Composite Variable

*constraints, the **INS** formulation is*

$$dc_3 + dc_4 = 1$$

$$dc_1 + dc_3 + dc_5 = 1$$

$$dc_2 + dc_4 + dc_5 = 1$$

$$dc_i \in \{0,1\},\ i = 1,2,...,5.$$

The resulting optimal solution to the LP relaxation is $\{dc_1 = 0,\ dc_2 = 0,\ dc_3 = 0.5,\ dc_4 = 0.5,\ dc_5 = 0.5\}$, *with objective value* 16.

*Excluding the balance, count, and landing constraints, the **INS-H** formulation is*

$$hc_1 + hc_2 = 1$$

The optimal solution to the LP relaxation is $\{hc_1 = 1,\ hc_2 = 0\}$, *with objective value* 17.

We denote the dual vector for constraints (3.10), (3.11), (3.12), (3.13), (3.14) and (3.15) as

$\pi^{\mathbf{B}_1}$, $\pi^{\mathbf{B}_2}$, $\pi^{\mathbf{H}}$, $\pi^{\mathbf{N}}$, $\pi^{\mathbf{A}}$ and $\pi^{\mathbf{C}}$, respectively. Let $\mathcal{C}^{T,O,h}$ be the subset of NDA ($T = N$) or SDA ($T = S$) demand composite variables covering subsets of gateway-hub demands to hub h in the case of pickup ($O = P$) or subsets of gateway-hub demands from hub h in the case of delivery ($O = D$), and $\mathbf{v}^{T,O,h}$ the vector indicating the selection of those demand composite variables. Following the matrix notation introduced in describing naive column generation, we denote the constraint matrix for demand composite variables in $\mathcal{C}^{T,O,h}$ in constraints (3.2), (3.3), (3.4), (3.5), (3.6) and (3.7) as $\mathbf{B}_1^{T,O,h}$, $\mathbf{B}_2^{T,O,h}$, $\mathbf{H}^{T,O,h}$, $\mathbf{N}^{T,O,h}$, $\mathbf{A}^{T,O,h}$, $\mathbf{C}^{T,O,h}$, respectively. Denote the right-hand-side vector of constraints (3.5) and (3.6) for $T = \{N, S\}$ as $\mathbf{n}_T$ and $\mathbf{a}_T$. Denote the right-hand-side vector of constraints (3.7) for gateway-hub demands for $T = \{N, S\}$, $O = \{P, D\}$ and $h \in H$ as $\mathbf{I}^{T,O,h}$. We define the following sub-problem for $T = \{N, S\}$, $O = \{P, D\}$, and $h \in H$:

$$\min\ [\mathbf{d}' - (\pi^{\mathbf{B}_1})'\mathbf{B}_1^{T,O,h} - (\pi^{\mathbf{B}_2})'\mathbf{B}_2^{T,O,h} - (\pi^{\mathbf{H}})'\mathbf{H}^{T,O,h} - (\pi^{\mathbf{N}})'\mathbf{N}^{T,O,h} - (\pi^{\mathbf{A}})'\mathbf{A}^{T,O,h} - (\pi^{\mathbf{C}})'\mathbf{C}^{T,O,h}]\mathbf{v}^{T,O,h} \tag{3.17}$$

subject to

$$\mathbf{A}^{T,O,h}\ \mathbf{v}^{T,O,h} \leq \mathbf{a}_T \tag{3.18}$$

$$\mathbf{N}^{T,O,h}\ \mathbf{v}^{T,O,h} \leq \mathbf{n}_T \tag{3.19}$$

$$\mathbf{C}^{T,O,h}\ \mathbf{v}^{T,O,h} \geq \mathbf{I}^{T,O,h} \tag{3.20}$$

$$v_c \in \{0, 1\},\ \ c \in \mathcal{C}^{T,O,h} \tag{3.21}$$

Constraints (3.18) ensure that the selected demand composite variables at hub h satisfy the landing constraints at *all* hubs. We consider all hubs because the demand composite variables in $\mathcal{C}^{T,O,h}$ might include routes entering or departing hubs other than h. Constraints (3.19) are the count constraints, specified for each fleet type, and constraints (3.20) are the cover constraints specified for each gateway-hub demand to or from hub h.

The solution to a sub-problem is a hub composite, and the objective value is its reduced cost. If the objective value of a solution is negative, we add the corresponding hub composite variable to the **RMP**. The process terminates if after solving all sub-problems, one for the pickup operation and one for the delivery operation at each hub, and for NDA and SDA, not one sub-problem solution has a negative objective value. Because we ensure the set of columns generated are feasible and the sum of their reduced cost is negative, we call this approach *information-enhanced column generation.* We refer to the information-enhanced column generation approach in which a sub-problem solution is introduced into the **RMP** in its *aggregate* form, that is, as a hub composite variable, *aggregate information-enhanced column generation.*

We apply aggregate information-enhanced column generation to the same UPS NDA problem instance that we solved with the naive column generation and the all-column approaches. Our results are reported in Table 3.4. Compared with the **INS** formulation, the optimal LP objective value increases by 0.001%, as proved in Lemma 1. The MIP solver, however, runs out of memory and fails to find a feasible integer solution after 20 hours with the set of columns generated. The best bound achieved at that point is 2.4% higher than the true IP optimal objective value.

Columns. Generated		7101
Master Iterations		270
Objective Value	Root Node LP	+0.001%
	IP	N/A
Run Time (sec.)	Root Node LP	2842
	IP	N/A

Table 3.4: Aggregate Information-Enhanced Column Generation Results for the UPS NDA Problem

3.3.3 Disaggregate Information-Enhanced Column Generation

The columns generated by the aggregate information-enhanced column generation at the root node of the branch-and-bound tree fail to provide a feasible solution. The issue is that too many decisions are embedded in a column of the **RMP**. To overcome this issue, we introduce *disaggregate information-enhanced column generation.*

We replace the **INS-H** formulation with the **INS** formulation as the **RMP**. At each master iteration, we similarly solve the pricing problem (3.17)-(3.21) for the pickup and delivery operation of each hub and for NDA and SDA. If the objective value of a sub-problem is negative, instead of adding to the **RMP** a single column representing all the demand composite variables in the sub-problem solution, we partition the solution into individual demand composite variables and add to the **RMP** those that are not currently included. (Some demand composite variables might have been included in the **RMP** in earlier iterations.)

We apply disaggregate information-enhanced column generation to the same UPS NDA problem instance and report our results in Table 3.5.

Using disaggregate information-enhanced column generation, we generate less than 1% of all possible columns, and less than 10% of the number of columns generated using naive column generation. This indicates that the hub-based sub-problems are more effective than naive

Columns Generated		1535
Master Iterations		34
IP Objective Value		+0.11%
Run Time (sec.)	Root Node LP	307
	IP	185

Table 3.5: Disaggregate Information-Enhanced Column Generation Results for the UPS NDA Problem

column generation in identifying columns that can be used in an optimal solution. The root node LP converges to the true objective value, but the IP objective value is somewhat worse than that obtained with naive column generation, because columns are again generated only at the root node of the branch-and-bound tree. Compared with the naive column generation and the all-column approaches, the root node LP relaxation takes longer to solve, but the IP solution time is significantly reduced using disaggregate information-enhanced column generation. Overall, disaggregate information-enhanced column generation achieves a 70% reduction in total solution time compared with naive column generation, and a 95% overall reduction compared with the all-column approach.

Compared with aggregate information-enhanced column generation, disaggregate information-enhanced column generation not only produces fewer columns, but also converges in fewer master iterations. Most importantly, solutions with objective values close to the optimal value can be identified with the set of columns generated. Intuitively, it is easier to construct a feasible solution using "small components", that is, demand composite variables, than using "large components", that is, hub composite variables.

3.4 Case Study

We apply the disaggregate information-enhanced column generation approach to the integrated UPS NDA-SDA problem, with the objective to minimize daily operating costs. Problem statistics are reported in Table 3.6. The SDA network is relatively small compared to the NDA network. Using disaggregate information-enhanced column generation, we generate only 3113 columns, or 1.4% of all variables, in solving the LP relaxation.

Composite Variables	NDA	SDA
	168372	59969
Ferry and Ground Variables	76215	
Rows	4623	
Master Iterations	33	
Generated Demand Composite Variables	3113	

Table 3.6: UPS Integrated NDA-SDA Problem Statistics

In Table 3.7, we compare the results of the sequential and integrated approaches with the solution generated by planners at UPS. Costs are reported as the percentage difference from those of the UPS solution. In the UPS solution, the SDA network is designed manually, while the design of the NDA network is accomplished using **ARM**.

In the first scenario, *unconstrained NDA and SDA problem*, boundary balance conditions are not enforced between the NDA and SDA operations, and the two problems are solved independently, without aircraft balance constraints. Their combined solution value provides an upper bound on the potential savings achievable through integration of the NDA and SDA problems. In the second scenario, the NDA problem is first solved without aircraft balance constraints. Then the SDA problem is solved with balance constraints ensuring that the NDA operations can be executed as planned. The resulting total cost is slightly better than that of the UPS solution. Notably in this case, ferry costs increase significantly because many ferry flights are required to re-position aircraft before or after the NDA operation to perform the SDA operations. These ferry costs more than offset the savings achieved in the NDA solution. In the third scenario, a reverse sequence is followed, the SDA problem, without aircraft balance condition, is first solved, and the NDA problem, with balance constraints ensuring the execution of the SDA operations, is then solved. The resulting operating costs of the NDA solution are greater than those of the UPS solution, but the daily total cost is much lower. This sequential approach produces less expensive solutions than the previous one for the following reasons:

- Because the SDA operation uses only about one third of the fleet used in the NDA operation, there is sufficient flexibility to position the unused aircraft in the SDA operation to match the needs of the NDA operation; and

- Most aircraft re-positioning for the SDA operation can be accomplished with revenue flight movements in the NDA operation, given the large number of NDA gateway-hub demands to be served.

Scenario		Daily Revenue Flight Cost	Daily Ferry Flight Cost	Total Daily Cost	Fleet Usage
1	Unconstrained SDA	-23.4%	-100%	-15.9%	
	Unconstrained NDA	-7.3%			
2	Unconstrained NDA	-7.3%	+903.6%	-0.3%	-4
	Constrained SDA	-17.5%			
3	Unconstrained SDA	-23.4%	+218.5%	-5.9%	-3
	Constrained NDA	+1.9%			
4	Integrated SDA	-19.5%	+140.7	-8.1%	-5
	Integrated NDA	-1.2%			

Table 3.7: Sequential and Integrated Approach Results for the UPS NDA-SDA Problem

In the last scenario, we solve the integrated NDA-SDA problem with disaggregate information-enhanced column generation. Even though the ferrying costs are more than double those in the UPS solution, the NDA and SDA operating costs are both reduced, reflecting the better coordinated aircraft movements. The daily operating cost savings of the integrated approach translates into tens of millions of dollars annually. Compared with the best sequential approach, the savings from the integrated approach come from: (1) reduced ferry cost; and (2) better coordinated NDA and SDA fleet movements. Beyond the tens of millions of dollars in operating cost savings, two fewer aircraft are needed in the integrated solution than in the sequential solution. This is significant because annual ownership costs for aircraft measure in the millions of dollars.

In all scenarios, savings attributable to the NDA operation are small or nonexistent, whereas savings attributable to the SDA operation are large, reflecting the carrier's use of ARM to improve their NDA network design, but not the SDA network design.

We acknowledge that some operating requirements are not considered explicitly in our models. The staging of package arrivals at hubs is one example. Hence, the savings reported here might not be fully realized.

3.5 Summary

In this chapter, we adapt the **ARM** model to solve the integrated Next-Day and Second-Day **ESSND** problem with fixed hub assignment, and present a novel solution approach that can be used to solve large-scale **ESSND** problems. The disaggregate information-enhanced column generation approach is shown to generate many fewer columns and help reduce IP solution time significantly. Potential savings of tens of millions of dollars from solving the integrated

NDA-SDA problem is demonstrated.

We make the following observations about column generation approaches. First, high quality columns, that is, columns that are likely to be present in the optimal solution, and fewer generated columns can be achieved if interactions among columns are considered. To illustrate, compare naive column generation and disaggregate information-enhanced column generation. While we consider individual columns in naive column generation, we consider a *set* of columns in disaggregate column generation. Second, better convergence and fewer generated columns can be achieved if a column in the restricted master problem includes fewer decisions. This point is made by comparing disaggregate and aggregate information-enhanced column generation. In disaggregate information-enhanced column generation, each column in the **RMP** represents a single demand composite variable, indicating the selection of a small number of aircraft and routes. In contrast, each column in the **RMP** in aggregate information-enhanced column generation represents decisions for all aircraft routes at a hub.

In the service network design problem, using the disaggregate information-enhanced column generation approach, we decompose the original problem into hub pickup and delivery subproblems. Solving a subproblem, we generate a set of columns representing a solution to a *sub-network* of the overall network design problem. This approach greatly reduces the total number of columns generated, and is efficient in identifying columns that are likely to be in an optimal solution. We can extend the idea to other problems. For example, in the multicommodity network flow problem (**MCNF**), we can establish a feasible flow in part of the network at each iteration instead of a single commodity flow.

Chapter 4

Express Shipment Service Network Design with Flexible Hub Assignment

The models and solution algorithms designed for the Express Shipment Service Network Design (**ESSND**) problem with Fixed Hub Assignment have been successful in helping carriers design and improve their aircraft schedules. There are still limitations in the models, however. Most notably, the models require inputs specifying gateway-hub demands. Hence, the sorting hub for origin-destination commodities must be fixed before the models can be used. It is believed by the carrier that substantial further savings are achievable if the hub assignment decisions for O-D commodities are made simultaneously with route selection and fleet assignment. We denote this augmented problem as the **ESSND** problem with Flexible Hub Assignment.

In this chapter, we first investigate different variable definitions to model the flexible hub assignment problem, and discuss the impact on the effectiveness of the formulation and size of the model. We then present a formulation for the **ESSND** problem with Flexible Hub Assignment based on the best variable definition and demonstrate its superiority through the case study of a small service network design problem. Applying the formulation to the UPS Next-Day service network design problem, we demonstrate that operating cost savings could measure in the tens of millions of dollars annually if hub assignment, route selection, and fleet assignment decisions are made simultaneously. Moreover, we demonstrate that, although designed as a tactical planning tool, our flexible hub assignment model can be used to provide insights for strategic planning.

4.1 Variable Definitions

In this section, we discuss the relative strengths and weaknesses of different variable definitions for modeling the **ESSND** problem with Flexible Hub Assignment. First, we introduce the following definitions:

Definition 4 *A hub h, $h \in H$, is a* ***candidate hub*** *of commodity k, $k \in K$, if there is a valid pickup aircraft route from the origin gateway of k to hub h, and a valid delivery aircraft route from hub h to the destination gateway of k.*

We define H^k as the set of candidate hubs for commodity k, $k \in K$.

Definition 5 *A commodity k, $k \in K$, is a* ***flexible commodity*** *if it has more than one candidate hub, that is, $|H^k| > 1$, and a* ***fixed commodity*** *if it has only one candidate hub, that is, $|H^k| = 1$.*

4.1.1 Route Variables

The route and flow based variables in the **RF** formulation, reviewed in Chapter 2, are capable of handling flexible hub assignment for origin-destination commodities. We review the notations and relevant constraints below.

Sets

H	set of hubs.
H^k	set of candidate hubs for commodity k, $k \in K$.
K	set of O-D commodities.
R	sets of aircraft routes.
R_P	sets of pickup aircraft routes.
R_D	sets of delivery aircraft routes.
$\mathcal{A}$	set of pickup arcs.
$\mathcal{B}$	set of delivery arcs.
$\mathcal{P}_k^P$	set of pickup paths for an O-D commodity k, $k \in K$.
$\mathcal{P}_k^D$	set of delivery paths for an O-D commodity k, $k \in K$.
$\mathcal{P}_k^{P,h}$	set of pickup paths from the origin of commodity k to hub h, $k \in K$, $h \in H^k$.
$\mathcal{P}_k^{D,h}$	set of delivery paths from hub h to the destination of commodity k, $k \in K$, $h \in H^k$.

Data

e_h	sort capacity of hub $h \in H$.
u_r	capacity of aircraft route r, $r \in R$.
b_k	volume of O-D commodity k, $k \in K$.
δ_p^{ij}	1 if arc (i, j) is included in path p, and 0 otherwise.

δ_r^{ij} 1 if arc (i,j) is included in aircraft route r, and 0 otherwise.

Decision Variables

y_r number of aircraft routes r, $r \in R$, selected.

x_p^k flow of commodity $k \in K$ on pickup (or delivery) path $p \in \mathcal{P}_k^P$ (or $\mathcal{P}_k^D$).

For simplicity, we only present the constraints modeling the package flows below

$$\sum_{r \in R_P} \delta_r^{ij} u_r y_r - \sum_{k \in K} \sum_{p \in \mathcal{P}_k^P} \delta_p^{ij} x_p^k \geq 0 \qquad (i,j) \in \mathcal{A} \tag{4.1}$$

$$\sum_{r \in R_D} \delta_r^{ij} u_r y_r - \sum_{k \in K} \sum_{p \in \mathcal{P}_k^D} \delta_p^{ij} x_p^k \geq 0 \qquad (i,j) \in \mathcal{B} \tag{4.2}$$

$$\sum_{p \in \mathcal{P}_k^{P,h}} x_p^k - \sum_{p \in \mathcal{P}_k^{D,h}} x_p^k = 0 \qquad k \in K,\ h \in H^k \tag{4.3}$$

$$\sum_{h \in H^k} \sum_{p \in \mathcal{P}_k^{P,h}} x_p^k = b_k \qquad k \in K \tag{4.4}$$

$$\sum_{k \in K} \sum_{p \in \mathcal{P}_k^{P,h}} x_p^k \leq e_h, \qquad h \in H \tag{4.5}$$

Forcing constraints (4.1) and (4.2) ensure the flow on an arc does not exceed the capacity provided. Constraints (4.3) guarantee the conservation of flow at hubs. Constraints (4.4) ensure that every commodity is covered. Constraints (4.5) limit the commodity flow through a hub to be no greater than the hub sort capacity.

Because the flow of an O-D commodity k can be assigned to any path between its origin and a candidate hub, and similarly, between a candidate hub and its destination, the commodity can be assigned to any of its candidate hubs. Hence, the **RF** formulation can be used to address the **ESSND** problem with flexible hub assignment. Unfortunately, though the **RF** formulation has relatively few variables, forcing constraints (4.1) and (4.2) result in fractional values of the route variables and render the **RF** formulation practically useless for large-scale service network design problems.

To compare with later formulations, we present a variant of the **RF** formulation, called the **RH** formulation. We introduce the following additional notation.

Sets

$\mathcal{P}^P$	sets of pickup paths.
$\mathcal{P}^D$	sets of delivery paths.
$\mathcal{N}$	set of gateways.

Data

d_r	cost of flying aircraft route r.
δ_p^l	1 if path p visits gateway l , and 0 otherwise.
$\varphi_p^{l,h}$	1 if path p connects gateway l with hub h, and 0 otherwise.
$\beta_k^{l,P}$	1 if gateway l is the origin of commodity k, and 0 otherwise.
$\beta_k^{l,D}$	1 if gateway l is the destination of commodity k, and 0 otherwise.

Decision Variables

y_r	number of aircraft routes r, $r \in R$, selected.
$x_{l,p}$	package flow from (or to) gateway l, $l \in \mathcal{N}$, on pickup (or delivery) path p, $p \in \mathcal{P}^P$ (or $\mathcal{P}^D$)
z_h^k	percentage of commodity k, $k \in K$, assigned to hub h, $h \in H^k$.

The route variables in the **RH** formulation are the same as those in the **RF** formulation. For simplicity, we denote the matrix corresponding to the gateway balance, hub balance, landing and count constraints for route variables as $\mathring{\mathbf{B}}$, $\mathring{\mathbf{H}}$, $\mathring{\mathbf{A}}$, $\mathring{\mathbf{N}}$, respectively, and the right-hand-side vector corresponding to the landing and count constraints as $\mathbf{a}$ and $\mathbf{n}$. The **RH** formulation is stated as

$$\min \sum_{r \in R} d_r y_r \tag{4.6}$$

subject to

$$\mathring{\mathbf{B}}\ \mathbf{y} = 0 \tag{4.7}$$

$$\mathring{\mathbf{H}}\ \mathbf{y} = 0 \tag{4.8}$$

$$\mathring{\mathbf{A}}\ \mathbf{y} \leq \mathbf{a} \tag{4.9}$$

$$\mathring{\mathbf{N}}\ \mathbf{y} \leq \mathbf{n} \tag{4.10}$$

$$\sum_{r \in R_P} \delta_r^{ij} u_r y_r - \sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^P} \delta_p^{ij} \delta_p^l x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{A} \tag{4.11}$$

$$\sum_{r \in R_D} \delta_r^{ij} u_r y_r - \sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^D} \delta_p^{ij} \delta_p^l x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{B} \tag{4.12}$$

$$\sum_{p \in \mathcal{P}^P} \varphi_p^{l,h} x_{l,p} - \sum_{k \in K} \beta_k^{l,P} b_k z_h^k = 0, \qquad l \in \mathcal{N},\ h \in H \tag{4.13}$$

$$\sum_{p \in \mathcal{P}^D} \varphi_p^{l,h} x_{l,p} - \sum_{k \in K} \beta_k^{l,D} b_k z_h^k = 0, \qquad l \in \mathcal{N},\ h \in H \tag{4.14}$$

$$\sum_{h \in H^k} z_h^k = 1, \quad k \in K \tag{4.15}$$

$$\sum_{k \in K} b_k z_h^k \leq s_h, \quad h \in H \tag{4.16}$$

$$y_r \in \mathbb{Z}_+,\ r \in R,\ z_k^h \geq 0,\ k \in K,\ h \in H,\ x_{l,p} \geq 0,\ l \in \mathcal{N},\ p \in \mathcal{P}^P \cup \mathcal{P}^D$$

The gateway balance, hub balance, landing and count constraints (4.7)-(4.10) in the **RH** formulation are the same as those in the **RF** formulation. Forcing constraints (4.11) and (4.12) specify that the capacity provided must exceed the total gateway path flow on any arc. Constraints (4.13) and (4.14) ensure that the amount of flow on all the pickup paths between a gateway-hub pair must equal the volume of commodities assigned to that hub originating at (or destined to) the gateway. Constraints (4.15) guarantee that every commodity is served.

We can prove the following lemma (See details in Appendix A):

Lemma 2 ***RF*** *and* ***RH*** *are equivalent formulations.*

4.1.2 Demand and Commodity Composite Variable

With the fixed hub assignment assumption, origin-destination commodities are consolidated into gateway-hub pickup and delivery demands. Taking gateway-hub demands as inputs, Armacost et al. [5] replace the route and flow variables with demand composite variables and introduce the **ARM** formulation for the fixed hub assignment problem. We introduce the following notation.

Sets and Data

$\mathcal{C}_P$	set of pickup demand composites.
$\mathcal{C}_D$	set of delivery demand composites.
$\mathcal{C}$	set of demand composites.
b_P^{ih}	pickup demand between gateway i and hub h, $i \in \mathcal{N}$, $h \in H$.
b_D^{ih}	delivery demand between gateway i and hub h, $i \in \mathcal{N}$, $h \in H$.

$\delta_c^{i,h,P}$ 1 if demand composite c covers the pickup demand between gateway i and hub h, and 0 otherwise.

$\delta_c^{i,h,D}$ 1 if demand composite c covers the delivery demand between gateway i and hub h, and 0 otherwise.

γ_c^r number of aircraft routes r in demand composite c.

Decision Variables

v_c equals 1 if demand composite c, $c \in \mathcal{C}$, is selected, and 0 otherwise.

The fixed hub assignment satisfies the hub sort constraints (4.5) in the **RF** formulation by design, and the demand composite definition allows constraints (4.1)-(4.4) in the **RF** formulation to be replaced by cover constraints (4.17) and (4.18).

$$\sum_{c \in \mathcal{C}_P} \delta_c^{i,h,P} v_c \geq 1, \qquad b_P^{ih} > 0 : \; i \in \mathcal{N}, \; h \in H. \tag{4.17}$$

$$\sum_{c \in \mathcal{C}_D} \delta_c^{i,h,D} v_c \geq 1, \qquad b_D^{ih} > 0 : i \in \mathcal{N}, \; h \in H. \tag{4.18}$$

$$v_c \in \{0,1\}, c \in \mathcal{C}_P \cup \mathcal{C}_D.$$

The route variables in the **RF** formulation correspond to the demand composite variables as follows:

$$\sum_{c \in \mathcal{C}} \gamma_c^r v_c = y_r \tag{4.19}$$

Armacost et al. [5], show that the **ARM** and **RF** are equivalent integer programming formulations for the **ESSND** problem with Fixed Hub Assignment. Moreover, Armacost et al. show that the **ARM** formulation is at least as strong as the **RF** formulation. In practice, the **ARM** formulation performs far better than the **RF** formulation.

By definition, a demand composite provides coverage for a *set* of O-D commodities, using a single pre-specified hub for each commodity. To extend the concept to the flexible hub assignment problem in which origin-destination commodities can be potentially sorted at multiple hubs, we introduce the following definitions.

Definition 6 *We say a vector ψ^k, $\psi^k \in \mathbb{R}^{|H|}$, is a hub assignment for a O-D commodity k, $k \in K$, if and only if $\sum_{h \in H^k} \psi^{k,h} = 1$.*

An entry of vector ψ^k, $\psi^{k,h}$, indicates the percentage of commodity k assigned to hub h.

Definition 7 *A **commodity composite**, c, is a set of aircraft routes R_c that provide sufficient capacity to pick up or deliver a set of O-D commodities K_c for a pre-specified hub assignment, ψ_c^k, for each O-D commodity k, $k \in K_c$.*

Let $\mathcal{P}_c$ be the set of (pickup or delivery) paths corresponding to R_c; $\mathcal{A}_c$ the set of arcs used by R_c; and $\mathcal{N}_c$ the set of gateways visited by R_c. Let ψ_c^k denote a hub assignment for an O-D commodity k, $k \in K_c$. Following previous notation, the following must be true for a pickup commodity composite c:

$$\sum_{p \in \mathcal{P}_c} \varphi_p^{l,h} x_{l,p} - \sum_{k \in K_c} \beta_k^{l,P} b_k \psi_c^{k,h} = 0, \qquad l \in \mathcal{N}_c,\ h \in H \tag{4.20}$$

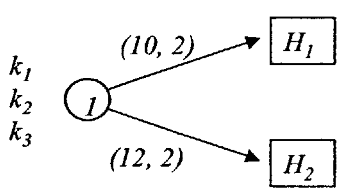
$$\sum_{r \in R_c} \gamma_c^r \delta_r^{ij} u_r - \sum_{p \in \mathcal{P}_c} \sum_{l \in \mathcal{N}_c} \delta_p^{ij} \delta_p^l x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{A}_c \tag{4.21}$$

Equalities (4.20) distribute the commodity flows to the paths between a gateway-hub pair.

The first term is the sum of the path flows from gateway l to hub h, and the second term is the commodity volume from gateway l and assigned to hub h. Inequalities (4.21) specify that the flow on each arc must not exceed the available capacity. Similar conditions can be specified for a delivery commodity composite.

We illustrate the commodity composite concept with the following example:

Example 5 *In Figure 4-1, suppose there are three commodities at gateway 1. Each commodity has one unit of volume and can be sorted at either of the two hubs. There is only one fleet type, and an aircraft of this type has two-units of capacity. In the figure, we list some examples of commodity composite variables .*



Commodity composite variables:
cc_1: *one route 1-H_1 picking up k_1 and k_2 to hub 1.*
cc_2: *one route 1-H_1 picking up k_1 and k_3 to hub 1.*
cc_3: *one route 1-H_1 picking up k_2 and k_3 to hub 1.*
cc_4: *one route 1-H_2 picking up k_1 and k_2 to hub 2.*
cc_5: *one route 1-H_2 picking up k_1 and k_3 to hub 2.*
cc_6: *one route 1-H_2 picking up k_2 and k_3 to hub 2.*
cc_7: *two routes 1-H_1, picking up k_1, k_2 and k_3 to hub 1.*
cc_8: *one route 1-H_1 and one route 1-H_2 picking up k_1 to hub 1, and k_2 and k_3 to hub 2.*

Figure 4-1: Examples of Commodity Composite

With fixed hub assignment in Example 1, 3 units of volume must be moved from gateway 1 to hub H_1. There is only one demand composite, namely two routes from gateway 1 to hub H_1, our commodity composite variable denoted cc_7. When building demand composites, we only need to consider one specific hub assignment for each of the commodities. When building commodity composites, however, we must consider every possible combination of hub assignment for *sets* of commodities. Furthermore, route sets previously providing insufficient capacity for the fixed gateway-hub demands can provide sufficient capacity for either a smaller set of commodities or a different set of hub assignments for the same set of commodities. Thus, the number of composite variables increases dramatically for the flexible hub assignment case.

We introduce the following additional notation:

Sets and Data

$\mathcal{C}_P$	set of pickup commodity composites.
$\mathcal{C}_D$	set of delivery commodity composites.
$\mathcal{C}$	set of commodity composites.
γ_c^r	number of aircraft routes r in commodity composite c.
d_c	cost for commodity composite c, $d_c = \sum_{r \in R} \gamma_c^r d_r$.
$\psi_c^{k,h}$	percentage of commodity k assigned to hub h in commodity composite c.

Decision Variables

v_c	equals 1 if commodity composite c is selected, and 0 otherwise.
z_h^k	percentage of commodity k assigned to hub h.

Similar to demand composite variables, commodity composite variables correspond to the route variables in the **RH** formulation by (4.19). Thus, the gateway balance, hub balance, landing and count constraints (4.7)-(4.10) specified with route variables in the **RH** formulation can be similarly specified with commodity composite variables. For simplicity, we denote the matrix corresponding to the gateway balance, hub balance, landing and count constraints for commodity composite variables as $\hat{\mathbf{B}}$, $\hat{\mathbf{H}}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{N}}$, respectively. We introduce the following Commodity Hub Assignment (**CHA**) formulation for the **ESSND** problem with Flexible Hub Assignment:

$$\min \sum_{c \in C} d_c v_c \tag{4.22}$$

subject to

$$\hat{\mathbf{B}}\ \mathbf{v} = 0 \tag{4.23}$$

$$\hat{\mathbf{H}}\ \mathbf{v} = 0 \tag{4.24}$$

$$\hat{\mathbf{A}}\ \mathbf{v} \leq \mathbf{a} \tag{4.25}$$

$$\hat{\mathbf{N}}\ \mathbf{v} \leq \mathbf{n} \tag{4.26}$$

$$\sum_{c \in \mathcal{C}_P} \psi_c^{k,h} v_c - z_h^k \geq 0,\ \ k \in K,\ \ h \in H \tag{4.27}$$

$$\sum_{c \in \mathcal{C}_D} \psi_c^{k,h} v_c - z_h^k \geq 0,\ \ k \in K,\ \ h \in H \tag{4.28}$$

$$\sum_{h \in H^k} z_h^k = 1, \quad k \in K \tag{4.29}$$

$$\sum_{k \in K} b_k z_h^k \leq e_h, \quad h \in H \tag{4.30}$$

$$v_c \in \{0,1\}, c \in \mathcal{C}_P \cup \mathcal{C}_D,\ \ z_h^k \geq 0,\ \ k \in K,\ \ h \in H$$

For the flexible hub assignment case, we must replace cover constraints (4.17) and (4.18) in the **ARM** formulation with constraints (4.27)-(4.30). Constraints (4.27)-(4.28) specify that the percentage of commodity k assigned to a hub must not exceed the percentage of k picked up and brought to hub h and the percentage of k delivered from hub h. Constraints (4.29) ensure that every commodity must be served in its entirety. Constraints (4.30) limit the commodities assigned a hub to be no more than the hub sort capacity.

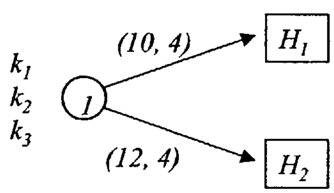
We can prove the following lemmas:

Lemma 3 *$\mathbf{CHA}$ and $\mathbf{RH}$ are equivalent integer programming formulations.*

Lemma 4 *The $\mathbf{CHA}$ formulation is at least as strong as the $\mathbf{RH}$ formulation.*

Detailed proof are included in Appendix A. In the next example, we describe a case in which the **CHA** formulation is strictly stronger than the **RH** formulation.

Example 6 *Consider the network depicted in Figure 4-2, each of the three commodities at gateway 1 has one unit of volume and can be sorted at either of the two hubs, H_1 and H_2. There is a single fleet type, and an aircraft of this type has 4 units of capacity. We have two single-leg routes, one from gateway 1 to hub H_1, and the other from gateway 1 to H_2, costing 10 and 12, respectively. We want to move the commodities to either of the two hubs.*



Commodity composite variables:
cc_1: one route 1-H_1 picking up k_1, k_2, and k_3 to hub 1.
cc_2: one route 1-H_2 picking up k_1, k_2, and k_3 to hub 2.

Figure 4-2: Example Demonstrating CHA Formulation Strength

Let y_1 and y_2 indicate the number of routes 1-H_1 and 1-H_2 selected in the solution, respectively. Let x_1 indicate the flow from gateway 1 on path 1-H_1, x_2 the flow from gateway 1 on path 1-H_2, and z_h^k the percentage of commodity k assigned to hub h. Excluding the balance, count, landing and hub sort constraints, the $\mathbf{RH}$ formulation is

$$\min\ 10y_1 + 12y_2$$

subject to:

$$4y_1 - x_1 \geq 0$$

$$4y_2 - x_2 \geq 0$$

$$x_1 - z_1^1 - z_1^2 - z_1^3 \geq 0$$

$$x_2 - z_2^1 - z_2^2 - z_2^3 \geq 0$$

$$z_1^1 + z_2^1 = 1$$

$$z_1^2 + z_2^2 = 1$$

$$z_1^3 + z_2^3 = 1$$

$$y_1, y_2 \in \mathbb{Z}_+, x_1, x_2 \geq 0, z_h^k \geq 0, k = 1, 2, 3, \text{ and } h = 1, 2.$$

The optimal route solution to the ***RH*** *LP relaxation is* $\{y_1 = 0.75,\ y_2 = 0\}$, *with objective value* 7.5, *and the integer optimal solution is* $\{y_1 = 1,\ y_2 = 0\}$, *with objective value* 10.

Next, we construct the ***CHA*** *formulation. For simplicity, we only consider the two commodity composites including one route listed in Figure 4-2. There are in fact more commodity composites, for example, those including both routes and assigning commodities to different hubs. Such composites, however, will not be in the optimal LP relaxation solution or integer solution because of the higher route cost needed to serve the commodities. Excluding the balance, count, landing and hub sort constraints, the* ***CHA*** *formulation is*

$$\min \ 10cc_1 + 12cc_2$$

subject to:

$$cc_1 - z_1^1 \geq 0$$

$$cc_1 - z_1^2 \geq 0$$

$$cc_1 - z_1^3 \geq 0$$

$$cc_2 - z_2^1 \geq 0$$

$$cc_2 - z_2^2 \geq 0$$

$$cc_2 - z_2^3 \geq 0$$

$$z_1^1 + z_2^1 = 1$$

$$z_1^2 + z_2^2 = 1$$

$$z_1^3 + z_2^3 = 1$$

$$cc_1, cc_2 \in \{0,1\}, \ z_h^k \geq 0, k = 1,2,3, \ \textit{and } h = 1,2.$$

The optimal commodity composite solution to the LP relaxation of the ***CHA*** *formulation is* $\{cc_1 = 1, \ cc_2 = 0\}$*, with objective value* 10*, and it is also the optimal integer solution.*

In the above example, the LP relaxation of the **CHA** formulation provides a better bound to the integer programming problem than that of the **RH** formulation. The route variables, together with the forcing constraints in the **RH** formulation, allow the excess route capacity to be eliminated from the LP solution, resulting in fractional aircraft usage. In contrast, the commodity composite definition allows the removal of the forcing constraints and coefficient reduction. This effectively forces excess capacity associated with whole planes to be included in the LP relaxation. The commodity composite concept, however, is not effective in this manner when multiple aircraft are needed to transport volume. The next example illustrates this point.

Example 7 *Consider the example in Figure 4-1, suppose the objective is still to move the three commodities to either of the two hubs. The route solution to the LP relaxation of the **RH** formulation is* $\{y_1 = 1.5, y_2 = 0\}$, *with objective value* 15, *while the integer optimal solution includes two routes 1-*H_1, *with objective value* 20.

*We only consider the commodity composites listed in the figure. (Other commodity composites including both routes and different hub assignments do not affect the integer solution to the **CHA** formulation or the solution to its LP relaxation.) Excluding the balance, count, landing and hub sort constraints, the **CHA** formulation is*

$$\min \ 10cc_1 + 10cc_2 + 10cc_3 + 12cc_4 + 12cc_5 + 12cc_6 + 20cc_7 + 22cc_8$$

subject to:

$$cc_1 + cc_2 + cc_7 + cc_8 - z_1^1 \geq 0$$

$$cc_1 + cc_3 + cc_7 - z_1^2 \geq 0$$

$$cc_2 + cc_3 + cc_7 - z_1^3 \geq 0$$

$$cc_4 + cc_5 - z_2^1 \geq 0$$

$$cc_4 + cc_6 + cc_8 - z_2^2 \geq 0$$

$$cc_5 + cc_6 + cc_8 - z_2^3 \geq 0$$

$$z_1^1 + z_2^1 = 1$$

$$z_1^2 + z_2^2 = 1$$

$$z_1^3 + z_2^3 = 1$$

$$cc_i \in \{0,1\},\ i = 1,2,...,7.\ z_h^k \geq 0, k = 1,2,3, \text{ and } h = 1,2.$$

The optimal composite solution to the **CHA** *LP relaxation is* $\{cc_1 = 0.5,\ cc_2 = 0.5,\ cc_3 = 0.5\}$ *(The values of variables not indicated are zero) with objective value* 15. *One of the optimal integer solutions is* $\{cc_7 = 1\}$*, with objective value* 20.

We see that if multiple aircraft routes are involved, the commodity composite concept allows excess capacity to be eliminated from the LP relaxation, even though this results in fractional aircraft usage. Hence, extending the demand composite concept to commodity composites results in some improvement of the formulation strength. This added strength, however, is insufficient to justify the huge number of variables created.

4.1.3 Gateway Composite Variable

Some of the fractionality in the LP relaxation of **CHA** results from the fact that commodity composites deal potentially, with only a subset of the commodities at a gateway. As shown in Example 7, at least two aircraft, totaling 4 units of capacity, are required to serve the 3 units of demand at the gateway, resulting in one unit of excess capacity. Each of the commodity composites cc_1, cc_2, and cc_3 serves two of the three commodities at gateway 1. Using one half of each of the commodity composites cc_1, cc_2, and cc_3, the **CHA** LP relaxation is able to provide sufficient capacity to cover all three commodities and eliminate the excess capacity, resulting in one and half aircraft.

To overcome the weakness of commodity composites, we introduce *gateway composite variables.*

Definition 8 *A* ***gateway composite****, g, is a set of aircraft routes R_g that provide sufficient capacity to pick up or deliver all the commodities at the set of gateways $\mathcal{N}_g$ visited by the routes.*

Similar to the notation used when defining the commodity composite, let γ_g^r be the number of aircraft routes r included in gateway composite g, $\mathcal{P}_g$ the set of paths based on R_g, $\mathcal{A}_g$ the set of arcs used by R_g, $x_{l,p}$ the flow of gateway l, $l \in \mathcal{N}_g$, on path p, $p \in \mathcal{P}_g$. We define $b^{l,P}$ and $b^{l,D}$ as the total pickup and delivery volume at gateway l, respectively. That is

$$\begin{aligned} b^{l,P} &= \sum_{k \in K} \beta_k^{l,P} b_k, \qquad l \in \mathcal{N}_g, \text{ and} \\ b^{l,D} &= \sum_{k \in K} \beta_k^{l,D} b_k, \qquad l \in \mathcal{N}_g. \end{aligned}$$

The following must be true for a pickup gateway composite g (similar conditions hold for a delivery gateway composite):

$$\sum_{r \in R_g} \gamma_g^r \delta_r^{ij} u_r - \sum_{p \in \mathcal{P}_g} \sum_{l \in \mathcal{N}_g} \delta_p^{ij} x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{A}_g, \text{ and} \tag{4.31}$$

$$\sum_{p \in \mathcal{P}_g} x_{l,p} = b^{l,P}, \qquad l \in \mathcal{N}_g. \tag{4.32}$$

Inequalities (4.31) specify that the flow on a path must not exceed the capacity provided. Equalities (4.32) ensure that all the volume at a gateway is distributed to the pickup paths. We illustrate the gateway composite concept by extending Example 7.

Example 8 *Consider the network in Figure 4-3. Each of the three commodities at gateway 1 has one unit of volume and can be sorted at either of the two hubs. There is one fleet type, and an aircraft of this type has two units of capacity. We list the gateway composites in the figure. Each of the gateway composites provide sufficient capacity to pickup all the volume at gateway 1.*

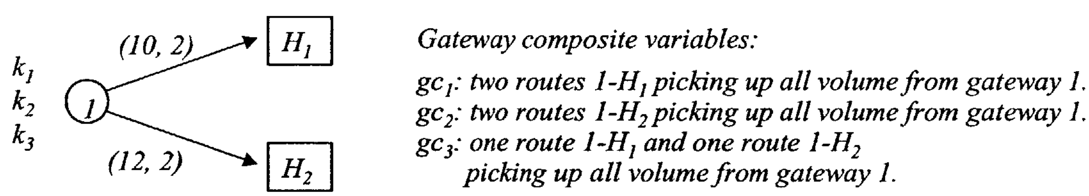


Figure 4-3: Example of Gateway Composite

To ensure there is sufficient capacity to transport the commodities, we need to include only the following constraint:

$$gc_1 + gc_2 + gc_3 = 1,$$

which results in an integer number of aircraft.

Compared to the commodity composite, the number of gateway composites is greatly reduced because a gateway composite does not specify the hub assignment for the commodities covered. For example, gateway composite gc_3 represents *all* possible ways to flow the three commodities through both hubs. To ensure flow conservation at hubs in an expanded example including both pickup and delivery, we would have to include flow variables.

4.2 Gateway Cover Flow Formulation

In this section, we introduce the Gateway Cover Flow formulation (**GCF**) based on gateway composite variables and compare the **GCF** formulation strength with that of the **RH** and **CHA** formulations. We also discuss ways to limit model size without eliminating optimal solutions.

4.2.1 Formulation

We begin by introducing additional notation.

Sets and Data

- $\mathcal{G}_P$ set of pickup gateway composites.
- $\mathcal{G}_D$ set of delivery gateway composites.
- $\mathcal{G}$ set of gateway composites.
- γ_g^r number of aircraft routes r in gateway composite g.
- d_g cost of gateway composite g, $d_g = \sum_{r \in R} \gamma_g^r d_r$.
- $\alpha_g^{l,P}$ 1 if gateway composite g picks up all the volume from gateway l, and 0 otherwise.
- $\alpha_g^{l,D}$ 1 if gateway composite g delivers all the volume to gateway l, and 0 otherwise.

Decision Variables

- w_g equals 1 if gateway composite g is selected, and 0 otherwise.
- z_h^k percentage of commodity k, $k \in K$, assigned to hub h, $h \in H^k$.

$x_{l,p}$ flow from (or to) gateway l on path p.

Similar to commodity composite variables, gateway composite variables correspond to the route variables in the **RH** formulation as follows:

$$\sum_{g \in \mathcal{G}} \gamma_g^r w_g = y_r. \tag{4.33}$$

The gateway balance, hub balance, landing and count constraints specified with route variables in the **RH** formulation can be similarly specified with gateway composite variables. For simplicity, we denote the matrix corresponding to the gateway balance, hub balance, landing and count constraints for gateway composite variables as $\bar{\mathbf{B}}$, $\bar{\mathbf{H}}$, $\bar{\mathbf{A}}$, $\bar{\mathbf{N}}$, respectively. We present the following Gateway Cover Flow formulation (**GCF**) for the **ESSND** problem with flexible hub assignment:

$$\min \sum_{g \in \mathcal{G}} d_g w_g \tag{4.34}$$

subject to

$$\bar{\mathbf{B}}\ \mathbf{w} = 0 \tag{4.35}$$

$$\bar{\mathbf{H}}\ \mathbf{w} = 0 \tag{4.36}$$

$$\bar{\mathbf{A}}\ \mathbf{w} \leq \mathbf{a} \tag{4.37}$$

$$\bar{\mathbf{N}}\ \mathbf{w} \leq \mathbf{n} \tag{4.38}$$

$$\sum_{g \in \mathcal{G}_P} \alpha_g^{l,P} w_g = 1, \qquad l \in \mathcal{N} \tag{4.39}$$

$$\sum_{g \in \mathcal{G}_D} \alpha_g^{l,D} w_g = 1, \qquad l \in \mathcal{N} \tag{4.40}$$

$$\sum_{g\in\mathcal{G}_P}\sum_{r\in R}\gamma_g^r\delta_r^{ij}u_r w_g - \sum_{p\in\mathcal{P}^P}\sum_{l\in\mathcal{N}}\delta_p^{ij}x_{l,p} \geq 0, \qquad (i,j)\in\mathcal{A} \tag{4.41}$$

$$\sum_{g\in\mathcal{G}_D}\sum_{r\in R}\gamma_g^r\delta_r^{ij}u_r w_g - \sum_{p\in\mathcal{P}^D}\sum_{l\in\mathcal{N}}\delta_p^{ij}x_{l,p} \geq 0, \qquad (i,j)\in\mathcal{B} \tag{4.42}$$

$$\sum_{p\in\mathcal{P}^P}\varphi_p^{l,h}x_{l,p} - \sum_{k\in K}\beta_k^{l,P}b_k z_h^k = 0, \qquad l\in\mathcal{N},\ h\in H \tag{4.43}$$

$$\sum_{p\in\mathcal{P}^D}\varphi_p^{l,h}x_{l,p} - \sum_{k\in K}\beta_k^{l,D}b_k z_h^k = 0, \qquad l\in\mathcal{N},\ h\in H \tag{4.44}$$

$$\sum_{h\in H^k} z_h^k = 1, \quad k\in K \tag{4.45}$$

$$\sum_{k\in K} b_k z_h^k \leq e_h, \quad h\in H \tag{4.46}$$

$$w_g\in\{0,1\},\ g\in\mathcal{G},\ z_h^k\geq 0, k\in K,\ h\in H,\ x_{l,p}\geq 0,\ l\in\mathcal{N},\ p\in\mathcal{P}^P\cup\mathcal{P}^D$$

Constraints (4.39) and (4.40) are cover constraints ensuring that at least one gateway composite is selected to cover the pickup and delivery volume of each gateway, respectively. It is not sufficient to specify that all the commodities at each gateway can be picked up or delivered. We also need to ensure there is a feasible path for every commodity from its origin to its destination. Hence, we introduce flow variables and hub assignment variables into the formulation. Constraints (4.41) and (4.42) specify that the capacity provided on each arc must exceed its flow. Constraints (4.43) and (4.44) distribute the commodity flows to the paths between gateway-hub pairs. Constraints (4.45) guarantee that every commodity is fully served. Constraints (4.46) ensure that hub sort capacity is not violated.

4.2.2 GCF Formulation Strength

We can prove the following lemmas:

Lemma 5 *The* ***GCF*** *formulation and the* ***RH*** *formulation are equivalent IP formulations for the* ***ESSND*** *problem with flexible hub assignment.*

Proof. We first show that, given a feasible solution to the **GCF** formulation, we can construct a feasible solution to the **RH** formulation, with the same cost.

We construct the route solution $\mathbf{y}$ with a gateway composite solution $\mathbf{w}$ based on mapping (4.33). For the objective function, we have

$$\sum_{g\in\mathcal{G}} d_g w_g = \sum_{g\in\mathcal{G}}\sum_{r\in R} \gamma_g^r d_r w_g = \sum_{r\in R} d_r \sum_{g\in\mathcal{G}} \gamma_g^r w_g = \sum_{r\in R} d_r y_r.$$

Thus, the resulting route solution has the same objective value as the gateway composite solution. Furthermore, constraints (4.7)-(4.10) in the **RH** formulation are satisfied as a result of the mapping. We use the count constraints as an illustrating example. In the **GCF** formulation, for a fleet type f, we have

$$\sum_{g\in\mathcal{G}}\sum_{r\in R} \delta_r^f \gamma_g^r w_g \le n_f,$$

where δ_r^f is 1 if fleet type f is assigned to aircraft route r, and 0 otherwise. With mapping (4.33), we can re-write the constraints as

$$\sum_{r\in R} \delta_r^f \sum_{g\in\mathcal{G}} \gamma_g^r w_g = \sum_{r\in R} \delta_r^f y_r \le n_f.$$

We can similarly show that the route solution satisfies constraints (4.7), (4.8) and (4.9).

We can directly use the gateway path flow and hub assignment solution to **GCF** to satisfy constraints (4.13)-(4.16). The remaining forcing constraints (4.11) and (4.12) are also satisfied because

$$\sum_{r \in R_P} \delta_r^{ij} u_r y_r = \sum_{g \in \mathcal{G}_P} \sum_{r \in R_P} \delta_r^{ij} u_r \gamma_g^r w_g \geq \sum_{p \in \mathcal{P}^P} \sum_{l \in \mathcal{N}} \delta_p^{ij} x_{l,p}, \ (i,j) \in \mathcal{A}, \text{ and} \tag{4.47}$$

$$\sum_{r \in R_D} \delta_r^{ij} u_r y_r = \sum_{g \in \mathcal{G}_D} \sum_{r \in R_D} \delta_r^{ij} u_r \gamma_g^r w_g \geq \sum_{p \in \mathcal{P}^D} \sum_{l \in \mathcal{N}} \delta_p^{ij} x_{l,p}, \ (i,j) \in \mathcal{B}. \tag{4.48}$$

Conversely, given an integer feasible solution to the **RH** formulation, the pickup routes and delivery routes form a pickup and a delivery gateway composite covering the volume at all the gateways, respectively, each including integer numbers of routes. These two composites along with the gateway path flow and hub assignment solution to the **RH** formulation satisfy all the constraints in the **GCF** formulation and form an integer feasible solution.

In summary, the **GCF** and **RH** are equivalent integer formulations for the **ESSND** problem with Flexible Hub Assignment. ∎

Lemma 6 *The **GCF** formulation is at least as strong as the **RH** formulation.*

Proof. The first part of the proof of Lemma 5 shows that we can find a feasible integer solution to **RH** with the same cost given a feasible integer solution to **GCF**. The proof holds even if integrality constraints are relaxed. That is, given a feasible solution to the **GCF** LP relaxation, we can find a feasible solution to the **RH** LP relaxation with the same cost. ∎

Following is an example in which this strength is strict.

Example 9 *Consider the pickup network in Figure 4-3. Let x_1 indicate the flow from gateway*

1 on path 1-H_1, x_2 the flow from gateway 1 on path 1-H_2, and z_h^k the percentage of commodity k assigned to hub h. Excluding the balance, count, landing and hub sort constraints, the ***GCF*** *formulation is*

$$\min\ 20gc_1 + 24gc_2 + 22gc_3$$

$$\textit{subject to:}\quad gc_1 + gc_2 + gc_3 = 1$$

$$4gc_1 + 2gc_3 - x_1 \geq 0$$

$$4gc_2 + 2gc_3 - x_2 \geq 0$$

$$x_1 - z_1^1 - z_1^2 - z_1^3 \geq 0$$

$$x_2 - z_2^1 - z_2^2 - z_2^3 \geq 0$$

$$z_1^1 + z_2^1 = 1$$

$$z_1^2 + z_2^2 = 1$$

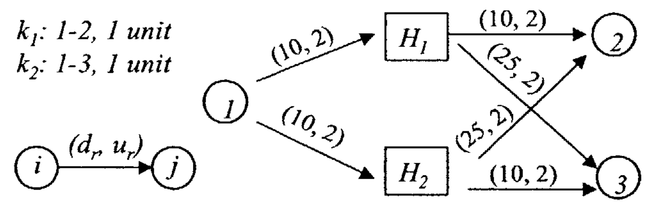
$$z_1^3 + z_2^3 = 1$$

$$gc_i \in \{0,1\},\ i = 1,2,3.\ z_h^k \geq 0, k = 1,2,3,\ \textit{and}\ h = 1,2.$$

The optimal solution to the ***RH*** *LP relaxation is* $\{gc_1 = 1,\ gc_2 = 0,\ gc_3 = 0,\ z_1^1 = 1,\ z_1^2 = 1,\ z_1^3 = 1\}$, *with objective value* 20.

In the above example, the **GCF** formulation is stronger than both the **RH** and **CHA** formulations. (The optimal LP relaxation objective values of the **RH** and **CHA** formulations are the same, 15, according to Example 7.) Following is an example, however, in which the **CHA** formulation is stronger than the **GCF** formulation.

Example 10 *Consider the network in Figure 4-4. There are two O-D commodities with one unit of volume each, and one fleet type. An aircraft of this fleet type has 2 units of capacity. The commodities can be sorted at either of the two hubs.*



Commodity composite:
cc_1: one route 1-H_1 picking up k_1 and k_2 to hub 1.
cc_2: one route 1-H_2 picking up k_1 and k_2 to hub 2.
cc_3: one route H_1-2 delivering k_1 from hub 1.
cc_4: one route H_1-3 delivering k_2 from hub 1.
cc_5: one route H_2-2 delivering k_1 from hub 2.
cc_6: one route H_2-3 delivering k_2 from hub 2.

Gateway composites:
gc_1: one route 1-H_1 picking up all volume from gateway 1.
gc_2: one route 1-H_2.picking up all volume from gateway 1.
gc_3: one route 1-H_1 and one route 1-H_2.
picking up all volume from gateway 1
gc_4: one route H_1-2 delivering all volume to gateway 2.
gc_5: one route H_1-3 delivering all volume to gateway 3.
gc_6: one route H_2-2 delivering all volume to gateway 2.
gc_7: one route H_2-3 delivering all volume to gateway 3.

Figure 4-4: Strength of ***CHA*** *and* ***GCF*** *Formulation*

Let z_h^k be the percentage of commodity k assigned to hub h. Excluding the balance, count, landing and hub sort constraints, the ***CHA*** *formulation is*

$$\min 10cc_1 + 10cc_2 + 10cc_3 + 25cc_4 + 25cc_5 + 10cc_6$$

subject to:

$$cc_1 - z_1^1 \geq 0$$

$$cc_3 - z_1^1 \geq 0$$

$$cc_2 - z_2^1 \geq 0$$

$$cc_5 - z_2^1 \geq 0$$

$$cc_1 - z_1^2 \geq 0$$

$$cc_4 - z_1^2 \geq 0$$

$$cc_2 - z_2^2 \geq 0$$

$$cc_6 - z_2^2 \geq 0$$

$$z_1^1 + z_2^1 = 1$$

$$z_1^2 + z_2^2 = 1$$

$$cc_i \in \{0,1\},\ i = 1,2,...6.\ z_h^k \geq 0, k = 1,2,\ \text{and}\ h = 1,2.$$

*The optimal solution to the **CHA** LP relaxation is* $\{cc_1 = 1,\ cc_2 = 1,\ cc_3 = 1,\ cc_6 = 1,\ z_1^1 = 1,\ z_2^2 = 1\}$, *with objective value* 40.

Let $x_{i,h}$ *and* x_{hi} *indicate the flow from gateway i on path i-h and the flow to gateway i on path h-i, respectively. Excluding the balance, count, landing and hub sort constraints, the **GCF** formulation is*

$$\min\ 10gc_1 + 10gc_2 + 20gc_3 + 10gc_4 + 25gc_5 + 25gc_6 + 10gc_7$$

subject to:

$$gc_1 + gc_2 + gc_3 = 1$$

$$gc_4 + gc_6 = 1$$

$$gc_5 + gc_7 = 1$$

$$2gc_1 + 2gc_3 - x_{1H_1} \geq 0$$

$$2gc_2 + 2gc_3 - x_{1H_2} \geq 0$$

$$2gc_4 - x_{H_1 2} \geq 0$$

$$2gc_5 - x_{H_1 3} \geq 0$$

$$2gc_6 - x_{H_2 2} \geq 0$$

$$2gc_7 - x_{H_2 3} \geq 0$$

$$x_{1H_1} - z_1^1 - z_1^2 \geq 0$$

$$x_{1H_2} - z_2^1 - z_2^2 \geq 0$$

$$x_{H_1 2} - z_1^1 \geq 0$$

$$x_{H_2 2} - z_2^1 \geq 0$$

$$x_{H_1 3} - z_1^2 \geq 0$$

$$x_{H_2 3} - z_2^2 \geq 0$$

$$z_1^1 + z_2^1 = 1$$

$$z_1^2 + z_2^2 = 1$$

$gc_i \in \{0, 1\},\ i = 1, 2, \ldots 7,\ z_h^k \geq 0, k = 1, 2,$ *and* $h = 1, 2,\ x_{1H_1}, x_{1H_2}, x_{H_1 2}, x_{H_1 3}, x_{H_2 2}, x_{H_2 3} \geq 0.$

The optimal solution to the **GCF** *LP relaxation is* $\{gc_1 = 0.5,\ gc_2 = 0.5,\ gc_4 = 1,\ gc_7 = 1,$ $z_1^1 = 1, z_2^2 = 1\}$, *with objective value* 30.

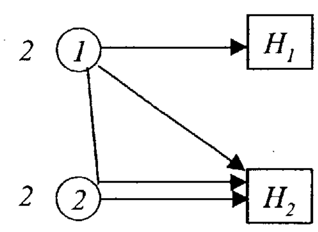
Though the **GCF** formulation is not always stronger than the **CHA** formulation, for realistic problem instances in which the number of commodities is large and multiple aircraft routes are required to cover the volume at gateways, the **GCF** formulation is likely to be stronger than the **CHA** formulation.

4.2.3 Restricting Composite Variable Set without Affecting Optimality

Our definition of gateway composites allows *redundant* gateway composites. In this section, we define redundancy in the context of gateway composites and introduce two concepts that help eliminate redundant composites.

We say a gateway composite variable is *redundant* if the exclusion of the variable does not eliminate the optimal solution.

Example 11 *Consider the network depicted in Figure 4-5. Gateway 1 and 2 each has two-units of demand. There is a single fleet type, and an aircraft of this type has three units of capacity. We depict four valid aircraft routes in the figure and list some gateway composites. Consider gateway composites* gc_1 *and* gc_3*. Both gateway composites provide sufficient capacity to pickup the demand at gateway 1, but there is a redundant aircraft route in gateway composite* gc_3*. In the case of gateway composites* gc_4 *and* gc_5*, both composites provide sufficient capacity to pick up the demand at both gateways to hub* H_2*. There is a redundant stop on the double-leg route 1-2-*H_2*, however. Replacing route 1-2-*H_2 *with route 1-*H_2 *will not affect any hub assignment feasible previously. In contrast, consider gateway composite* gc_6 *and* gc_7*. If*



Gateway composites:
gc_1: one routes 1-H_1 picking up all volume from gateway 1.
gc_2: one routes 2-H_2 picking up all volume from gateway 2.
gc_3: two routes 1-H_1 picking up all volume from gateway 1.
gc_4: one route 1-2-H_2 and one route 2-H_2 picking up all volume from gateway 1 and 2.
gc_5: one route 1-H_2 and one route 2-H_2 picking up all volume from gateway 1 and 2.
gc_6: one route 1-H_1 and one route 1-2-H_2 picking up all volume from gateway 1 and 2.
gc7: one route 1-H_1 and one route 2-H_2 picking up all volume from gateway 1 and 2.

Figure 4-5: Examples of Redundant Gateway Composites

gateway composites gc_6 is selected, some commodities at gateway 1 can be assigned to hub H_2. If we replace route 1-2-H_2 with route 2-H_2, commodities at gateway 1 all have to be assigned to hub H_1.

We define a gateway composite to be *minimal* if the removal of an aircraft route or the skipping of a gateway in a double-leg aircraft route included in the gateway composite renders at least one hub assignment feasible in the original gateway composite infeasible. We can prove the following lemma:

Lemma 7 *If ferrying is possible, any optimal solution to the **GCF** formulation can be represented by minimal gateway composites and ferry routes.*

Proof. Suppose a gateway composite in the optimal solution is not minimal. If we remove the redundant routes or flight legs in the gateway composite, the resulting gateway composite will become minimal and only aircraft balance is likely to be affected. Because the redundant routes and flight legs must be feasible for ferrying, we can restore the aircraft balance by introducing ferrying variables. The set of ferry flights needed is at most all the routes and flight legs removed. The resulting total cost of the minimal gateway composite and the ferry flights is therefore no more than the cost of the original gateway composite. Thus, the original optimal solution can be represented by minimal composites and ferrying routes. ■

If we consider gateway composites gc_1, gc_2 and gc_7 in Figure 4-5, all three gateway composites are minimal. If gc_7 is included in the optimal solution, we can replace it with gateway composites $gc_1$and gc_2, without violating any constraints. We define a gateway composite to be *connected* if the routes included in the gateway composite form a connected graph, for example, gateway composites gc_4 and gc_6 in Figure 4-5. We can prove the following lemma:

Lemma 8 *Any optimal solution to the GCF formulation can be represented by connected gateway composites.*

Proof. If a non-connected gateway composite is included in the optimal solution, it can be broken into a set of connected gateway composites. Because the set of aircraft routes represented by the composites is the same, the new solution satisfies every constraints and has the same cost. Thus, the original optimal solution can also be represented by connected composites. ■

4.3 Sub-network Column Generation

In ARM, with fixed hub assignment, we are able to enumerate a tractable set of composite variables with certain restrictions (Armacost [4]) and solve the corresponding model. With flexible hub assignment, the size of the composite set is much larger. In this section, we introduce a column generation approach for generating gateway composites. The column generation approach for generating commodity composites is similarly derived.

Definition 9 *A **sub-network** is a triplet ($\mathcal{N}_G$, H_G, R_G) including a subset of gateways, $\mathcal{N}_G$, a set of hubs, H_G, and the complete set of aircraft routes, R_G, visiting only the set of gateways $\mathcal{N}_G$.*

The tractability issue arises because many gateway composites can be formed on a single sub-network. The column generation approach to circumvent this issue is to first solve a restricted master problem (**RMP**), that is, the **GCF** formulation with a limited set of gateway composites. Then, use the dual prices obtained in solving the **RMP** to price-out and identify only those gateway composites that can potentially improve the current solution to the sub-networks. These new composites are added to the **RMP** to create an augmented model, and the process repeats until no gateway composite can be identified.

Definition 10 *The **size of a gateway composite** is the number of gateways the composite covers.*

To control the size of the model and the number of sub-problems, we assume that we will only consider minimal and connected gateway composites with size up to M. Thus, we only need to consider sub-networks including up to M gateways.

Consider the example network in Figure 4-6, we construct the 1-gateway, 2-gateway and 3-gateway sub-networks. Note that we only need to consider connected sub-networks. Hence, in the example, the subset including gateway 2 and 3 is not considered because there is no connected graph containing only gateway 2 and 3.

We then define a pricing problem for each subnetwork to generate gateway composites. Any aircraft route included in a gateway composite affects constraints (4.35) through (4.38) and constraints (4.41) or (4.42). We define π_r as the sum of the products of the dual price

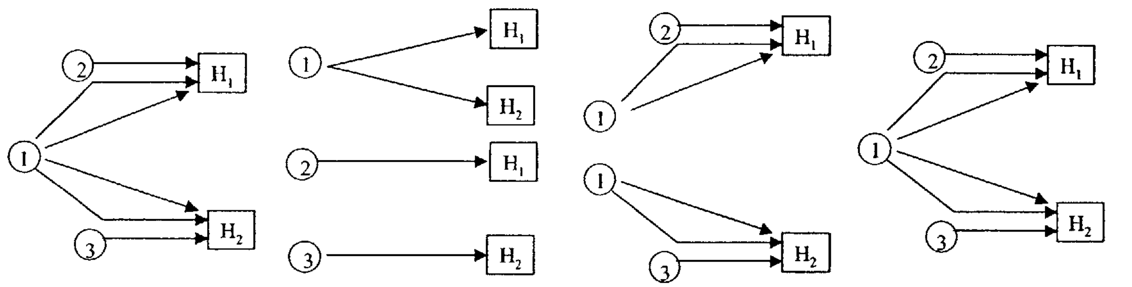


Figure 4-6: Sub-networks

of a constraint and the corresponding coefficient of aircraft route r in that constraint, for each constraint containing aircraft route r. We define π_G as the sum of the dual prices associated with cover constraints (4.39) or (4.40) corresponding to the gateways in $\mathcal{N}_G$. Because each gateway composite formed on subnetwork ($\mathcal{N}_G$, H_G, R_G) must cover all the gateways in $\mathcal{N}_G$, π_G is a constant for all gateway composites covering $\mathcal{N}_G$. We define $\mathcal{P}_G$ as the set of *gateway flow pickup* (or *delivery*) *paths* and $\mathcal{A}_G$ as the set of flight arcs corresponding to the aircraft routes in R_G, and b^l as the total pickup (or delivery) volume at gateway l. We can then specify the following pricing problem to generate gateway composites:

$$\min \sum_{r \in R_G} (d_r - \pi_r) y_r - \pi_G \tag{4.49}$$

subject to

$$\sum_{r \in R_G} \delta_r^{ij} u_r y_r - \sum_{p \in \mathcal{P}_G} \sum_{l \in \mathcal{N}_G} \delta_p^{ij} x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{A}_G \tag{4.50}$$

$$\sum_{h \in H_G} \sum_{p \in \mathcal{P}_G} \varphi_p^{l,h} x_{l,p} = b^l, \qquad l \in \mathcal{N}_G \tag{4.51}$$

$$y_r \in \mathbb{Z}_+,\ r \in R_G,\ x_{l,p} \geq 0,\ l \in \mathcal{N}_G,\ p \in \mathcal{P}_G$$

Problem (4.49)-(4.51) is itself a network design problem, with the objective function coefficients modified by the dual prices obtained from solving the **RMP**. Constraints (4.50) specify that the total flow on each path must not exceed the capacity provided by the selected routes. Constraints (4.51) require that all the demand from (or to) each of the gateways in $\mathcal{N}_G$ are assigned to the pickup (or delivery) paths.

We include all the ferry routes in the **RMP**, and reduce their costs by an infinitesimal amount. Thus, the pricing problems will generate only minimal gateway composites. (Suppose the reduced cost of a non-minimal gateway composite g' is negative, we can replace the gateway composite with a minimal gateway composite g and a set of redundant routes. The redundant routes can be equivalently replaced by the ferry flights in the **RMP**. Because the reduced cost of ferry flights is non-negative, the reduced of the redundant routes must be positive as a result of our adjustment. Hence, the reduced cost of gateway composite g is smaller than that of g'.)

Because every gateway in $\mathcal{N}_G$ must be covered, the size of the gateway composite is equal to the number of gateways included in the sub-network. To improve efficiency, at each column generation iteration, we first consider sub-networks including one gateway, then sub-networks including two gateways, and so on. If the objective value of a pricing problem is negative, the reduced cost of the corresponding gateway composite is negative. If we also find the routes selected form a connected graph, we generate the corresponding gateway composite and include it in the **RMP**. If the gateway composite is not connected, it could be broken into smaller connected gateway composites. We do not need to change the **RMP**, however, because those smaller connected gateway composites must have been generated when we consider sub-networks containing smaller numbers of gateways. Overall, we can ensure the sub-network column generation does not produce redundant gateway composites.

We can similarly specify the pricing problem to generate commodity composites. We define K_l^P and K_l^D to be the set of commodities to be picked up from and delivered to gateway l, respectively, and K_G to be the complete set of commodities on sub-network G. We define π_r as the sum of the products of the dual price of a constraint and the corresponding coefficient of aircraft route r in that constraint, for each constraint containing aircraft route r (constraints (4.23)-(4.26)), and $\pi_h^{k,P}$ and $\pi_h^{k,D}$ as the dual prices of pickup and delivery cover constraints (4.27) and (4.28), respectively. Let μ_k be the binary decision variable which equals 1 if commodity k is covered and 0 otherwise. For generating pickup commodity composites, we have

$$\min \sum_{r\in R_G} (d_r - \pi_r) y_r - \sum_{k\in K_G} \sum_{h\in H^k} \pi_h^{k,P} \psi^{k,h} \tag{4.52}$$

subject to

$$\sum_{h\in H^k} \psi^{k,h} = \mu_k, \qquad k \in K_G \tag{4.53}$$

$$\sum_{r\in R_G} \delta_r^{ij} u_r y_r - \sum_{p\in \mathcal{P}_G} \sum_{l\in \mathcal{N}_G} \delta_p^{ij} x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{A}_G \tag{4.54}$$

$$\sum_{h\in H} \sum_{p\in \mathcal{P}_G} \varphi_p^{l,h} x_{l,p} - \sum_{k\in K_l^P} \beta_k^{l,P} b_k \psi^{k,h} = 0, \qquad l \in \mathcal{N}_G \tag{4.55}$$

$$y_r \in \mathbb{Z}_+,\ r\in R_G,\ \mu_k \in \{0,1\},\ k\in K_G,\ z_h^k \geq 0,\ k\in K_G,\ h\in H^k,\ x_{l,p}\geq 0,\ l\in \mathcal{N}_G,\ p\in \mathcal{P}_G$$

The pricing problem to generate delivery commodity composites can be defined similarly. In the objective function (4.52), the first term is the route costs modified by the dual prices, and the second term is the sum of the values for covering commodities. Note that if a commodity is selected to be covered, by constraints (4.53), we have $\sum_{h \in H^k} \psi^{k,h} = 1$, resulting in a hub assignment for commodity k. Constraints (4.54) specify that the flow on a path must not exceed the capacity provided by the selected routes. Constraints (4.55) assign commodity flows to the paths. Unlike the pricing problem for generating gateway composites, the problem above does not require each commodity to be served. Whether a commodity is served by the set of routes selected is determined by the pricing problem. Because the pricing problem for generating commodity composites involves individual commodity hub assignment variables, its size is much larger and the problem is more difficult to solve.

4.4 Case Study 1: Small Service Network Design Problem

In this section, we apply the **CHA** formulation and the **GCF** formulation to a small service network design problem to investigate the effectiveness of the two formulations. The small service network design problem, shown in Figure 4-7, contains nine gateways, two hubs and 70 O-D commodities. Among the gateways, three of them are located to the north of the northern hub, another three to the south of the southern hub, and the remaining three between the two hubs. The hub sort capacity is assumed to be infinite. There are two fleet types, and there is no limit on the available number of aircraft of each fleet type. To match the capacity of aircraft, the carrier's origin-destination volume associated with the 70 selected O-D commodities is scaled by 9. (Without scaling, the problem becomes an uncapacitated problem.) The operating cost of the aircraft is equal to the carrier's actual cost, and the schedule requirements match those of the carrier.

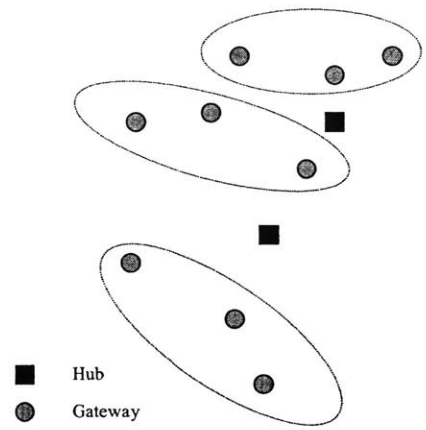


Figure 4-7: Small Network Design Problem

All computations for this case study were performed on an HP9000 Model D370 workstation. The models and procedures were compiled with HP's aCC compiler with calls to the ILOG CPLEX 6.5 Callable Library. We first solved the problem with fixed hub assignments matching those in the UPS solution. The resulting optimal solution obtained with the **ARM** formulation is shown in Figure 4-8. With fixed hub assignment, both hubs are used and 4 of the 9 gateways send or receive volume from both hubs.

We then applied the **CHA** formulation and the **GCF** formulation to the problem, thus enabling flexible hub assignment. We allowed each commodity to be assigned to either hub, and we used sub-network column generation to generate the relevant composite variables at the root node. We generated composites with up to three gateways. After column generation at the root node, the **RMP** was solved with the CPLEX Mixed Integer Program (MIP) solver. The results reported in Table 4.1 indicate that the LP relaxation of the **GCF** formulation provides a much tighter bound than the **CHA** formulation. Moreover, because of this strength, the

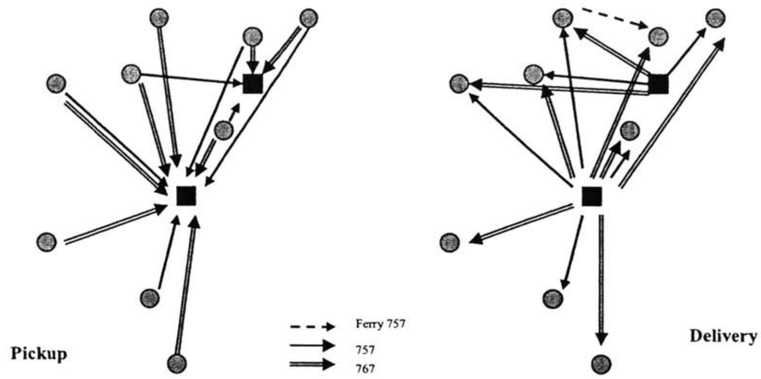


Figure 4-8: Solution to the Small Network Design Problem with Fixed Hub Assignment

GCF formulation obtained the best IP solution within one minute, and achieved a *final best bound-best IP gap* within 0.01%. For the **CHA** formulation, the CPLEX MIP solver could not close the gap to within 10%, even after 24 hours of computing time.

The best integer solution found to the **GCF** model is illustrated in Figure 4-9. Compared with the fixed hub assignment solution, operating cost savings are about 20%. Moreover, one fewer hub and one fewer aircraft are used.

The computational results of this case study indicate that the **GCF** formulation is potentially far more tractable than the **CHA** formulation for the service network design problem with flexible hub assignment.

Formulation		CHA	GCF
Problem Statistics	Sub-problems	90	90
	Master Iterations	31	32
	Columns Generated	1075	1094
Objective Value ($)	LP Relaxation	69422	82339
	First IP	94350	85189
	Best IP	85305	85189
	LP Relaxation-First IP Gap	35.9%	3.5%
	LP Relaxation-Best IP Gap	17.9%	3.5%
	Best LP Bound - Best IP Gap	>10%	<0.01%
Run Time	LP Relaxation	572 sec.	923 sec.
	First IP	1789 sec.	1 sec.
	Best IP	5.3 hr.	41 sec.
	Total IP	24 hr.	42 sec.
	Total Solution	24.2 hr.	0.27 hr.

Table 4.1: Case 1 Results

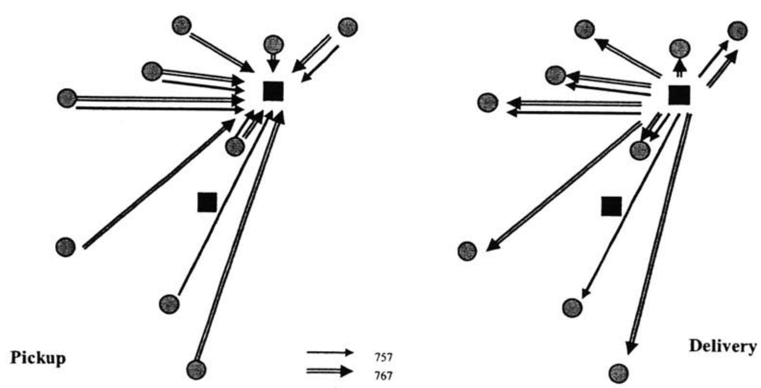


Figure 4-9: Solution to the Small Network Design Problem with Flexible Hub Assignment

4.5 Case Study 2: UPS Next-Day Air Problem

In this section, we apply the **GCF** formulation to the UPS Next-Day Air (NDA) service network design problem. First, we investigate the effectiveness of the sub-network column generation approach when applied to UPS' large-scale NDA service network design problem. Second, we enumerate a restricted set of gateway composite variables and solve the resulting UPS NDA problem. Finally, we use the model to explore the impact of potential strategic changes to the air network, such as operating with different numbers of hubs or managing increased demand.

The models and procedures in this case study were compiled with HP's aCC compiler with calls to the ILOG CPLEX 6.5 Callable Library. To build the MIP model, the column generation procedures were run on the HP 9000 Model D370 workstation. The MIP model was then exported to a Dell Precision M50 workstation with 2.5 GHz Pentium4 CPU and 2GB RAM, and was solved using ILOG CPLEX 8.5.

The UPS domestic air network we consider consists of 85 gateways, six of which are hubs. Distances between gateways are measured by *block time*, that is, flying time plus taxi time. *Earliest Pickup Times* and *Latest Delivery Times* are specified for each gateway, and a route-specific *Latest Hub Arrival Time* and a *Earliest Hub Departure Time* are specified for each hub. Pickup routes originating from west coast gateways to non-west coast hubs may arrive later than routes originating from non-west coast gateways. There are four fleet types, each with a specified flying range and an allowable set of gateway and hub locations for landing. Aircraft capacity is specified in number of containers. Carrier policy requires that we reduce by 30% the capacity of the fleet assigned to pickup routes from west-coast gateways to non-west coast hubs to compensate for the tighter time windows for loading and unloading.

It is the carrier's current practice to develop two aircraft routing plans each year. The first one specifies aircraft movements in the ten months spanning January to October, and the second specifies aircraft movements for the peak retail season that occurs during November and December. Each plan is typically designed for the highest demand in the planning period because system reliability is highly important. Due to the time and complexity involved in creating aircraft plans with the carrier's existing tools, considering other demand scenarios and designing more demand-specific plans has not been possible.

Demands are specified by origin gateway and destination gateway and are measured in number of packages. At the gateway locations, packages are packed into containers. The number of packages that can be packed into a container is generally a constant, though there can be small variations depending on the time available before the departure of the pickup aircraft, which is largely determined by the distance from the gateway to the destination hub. For example, there is typically more time to pack a container to be sent to a nearby hub, and thus, a larger number of packages can be packed into the container. To be consistent with our units of aircraft capacity, we must convert all demands into container units. In this case study, we assume a constant number of packages can be packed into a container, but a gateway-hub-specific conversion factor can be used if necessary to translate package demand into container demand.

If a gateway is located close to a hub, the volume assigned to the hub from that gateway can be transported by ground vehicles. For this reason, both ground and aircraft services must be considered. Because the ground fleet is much cheaper to operate than the air fleet, ground service (with assumed unlimited capacity) is always used if time allows. In some cases, there might not be enough time to transport by ground *all* the volume from the gateway to the hub.

For example, if all the volume does not arrive at the gateway before the ground vehicle has to leave the gateway to arrive at the hub sort on time, only a portion of the volume is assigned to ground transport. A *maximum pickup* (or *delivery*) *ground percentage*, based on historical data is computed and used as a capacity limit on the corresponding ground route.

We also consider direct flights *by-passing* hubs and moving O-D commodities directly from their respective origins to their respective destinations. Based on the carrier's preferences, if the potential load factor is greater than 50%, a direct flight is viable. Direct flights can be included in a gateway composite. In our implementation, however, we create a variable for each direct flight. The direct flight variables are included in constraints (4.35)-(4.38), but not constraints (4.39)-(4.44). Instead, it provides one cover for the corresponding O-D commodity in constraint (4.45) specified for that commodity. Because each gateway composite is still required to cover all the volume at a gateway, if direct flights are selected to transport some O-D commodities, we might over-restrict the solution. In the UPS case, however, there are only tens of O-D commodities qualifying for direct flights and they account for only a very small percentage of the total volume at their respective origin or destination gateways. Thus, the above approach to handle direct flights has little impact on the solution.

4.5.1 Column Generation

We use the column generation approach described in Section 4.3 to solve the LP relaxation of the **GCF** formulation. The best feasible solution with the existing UPS fixed hub assignments, referred to as the *fixed hub assignment solution*, is used as the starting set of columns for column generation. We solve 65,786 sub-network subproblems in total, with 4000 subproblems solved at each master iteration. The column generation process converges very slowly. We terminate it after about 100 hours when the LP relaxation objective value improvement from 20 consecutive master iterations is less than 0.1%. At that point, 206 master iterations are executed and 16,896 columns generated. The gap between the objective value of the LP relaxation and the fixed hub assignment solution is about 12.0%. The MIP model, including the generated columns, is then solved with the fixed hub assignment solution as the advanced starting solution, but the CPLEX MIP solver fails to find any improved integer solution after 100 hours. The gap between the best LP bound and the fixed hub assignment solution is still 11.5% at that point.

Network size is a primary reason for the failure of the column generation approach. Opportunities to combine routes to form gateway composites leading to low cost, but fractional, solutions are greatly increased on large-scale networks. Though the column generation at the root node can produce columns that lead to a good solution to the LP relaxation, those columns might not produce a good IP solution. We illustrate this point in the next example.

Example 12 *Consider Example 10 in Figure 4-4. Suppose at some point during column generation, gateway composites* gc_1, gc_2, gc_4, gc_5, gc_6 *and* gc_7 *are included in the RMP, but not* gc_3. *When solving the LP relaxation, the reduced cost of* gc_3 *is non-negative, so* gc_3 *will not be generated. Without* gc_3, *the solution to the problem is sub-optimal.*

The tractability issue is further exacerbated by the presence of ground routes. The ground routes can form zero-cost gateway composites and fractional values of them can be combined with other fractional gateway composites to satisfy gateway cover constraints (4.39) and (4.40).

4.5.2 Variable Enumeration

Instead of using column generation to generate gateway composites, we define the characteristics of gateway composites that are likely to be included in the optimal solution, and then enumerate gateway composites based on these characteristics.

The carrier maintains a major hub in their network. To ensure reliability, the carrier specifies the following *major hub connectivity condition*:

Each gateway must be visited by a pickup route destined to the major hub, and a delivery route originating at the major hub.

We say an **ESSND** problem is *uncapacitated* if the capacity of the smallest aircraft that can be assigned to each pickup (or delivery) route is sufficient to serve all the volume from (or to) the gateways the route contains. We can prove the following lemma.

Lemma 9 *For the uncapacitated **ESSND** problem, there is an optimal hub assignment under the major hub connectivity condition assigning every O-D commodity to the major hub.*

Proof. We prove the lemma by construction. Relaxing the requirements to serve all demands, we first obtain a network design solution satisfying the major hub connectivity condition and other constraints with the minimum cost. The cost of the solution to the relaxed problem provides a lower bound on the optimal cost. Because each pickup (or delivery) aircraft route connecting gateways and the major hub in this solution provides sufficient capacity to serve all the demands from (or to) the gateways it visits, the solution we obtain also satisfies the requirements to serve all demands, with an objective value equal to the lower bound on the optimal cost. Hence, we have obtained an optimal solution, and we can assign every O-D commodity to the major hub in this solution. ■

In the capacitated case, if the total volume of a gateway is larger than the capacity of the smallest aircraft but is smaller than the capacity of the largest aircraft, assigning the volume to multiple hubs would likely result in more fixed cycle charges and increased costs due to additional aircraft needed. Therefore, in the optimal solution, all the volume at such gateways would most likely be assigned to the major hub, served by a single aircraft. For this reason, we impose the following assumption, defined as the **small demand gateway rule**:

If the total volume from (or to) a gateway is smaller than the capacity of the largest aircraft that can serve that gateway, the gateway is categorized as a ***small demand gateway*** *and all the associated volume is assigned to the major hub.*

Because the commodities from or to small demand gateways are all assigned to the major hub, there is only one pickup and one delivery gateway-hub demand pair for each small demand gateway. A demand composite covering the pair is also a gateway composite. Therefore, we can use the **ARM** composite generator (Armacost [4]) to create gateway composites including small demand gateways exclusively. In addition, due to timing constraints, the west coast hub is the only feasible hub for commodities with both origins and destinations on the west coast. Hence, the hub assignments for these commodities are also fixed, resulting in fixed pickup and delivery demand between gateways on the west coast and the west coast hub. We therefore specify two sets of gateway pickup and delivery cover constraints (4.39) and (4.40) for gateways on the west coast. One corresponds to west coast-west coast volume, and the other corresponds to west coast to non-west coast or non-west coast to west coast volume. The small demand gateway rule is also applied to west coast to non-west coast or non-west coast to west coast volume. **ARM** composite generator can be used to build composites covering west coast-west coast volume and the volume satisfying the small demand rule.

For the remaining gateways, those not satisfying the small-demand rule after strictly west coast demands are removed, we need only to consider *minimal* and *connected* gateway composites (as proved in Lemma 7 and 8). Such gateway composites can be enumerated by combining aircraft routes to satisfy the capacity requirement.

The number of feasible aircraft route combinations fully serving a sub-set of gateways can still be huge, especially if the demand of the gateways is large. Many of the gateway composites however, might not be practical at all, and are unlikely to be included in the optimal solution. For example, instead of using a single large-capacity aircraft to serve a gateway with large volume, a feasible yet costly and impractical option (due to the resulting complexity of the operation) is to assign many small-capacity aircraft to serve that gateway. We exclude such options from our model by imposing the following restrictions when enumerating the gateway composites:

- Let $u_l^{\max}$ be the capacity of the largest aircraft that can serve gateway l, and b_l be the total volume of gateway l. The number of planes serving gateway g in a gateway composite must be no more than $\lceil \frac{b_l}{u_l^{\max}} \rceil + 1$; and the number of single-leg aircraft routes with capacity $u_l^{\max}$, must be at least $\lceil \frac{b_l}{u_l^{\max}} \rceil - 2$.

- There can be at most two double-leg aircraft routes included in a gateway composite. If both routes are from or to the same hub, they are not allowed to visit the same set of gateways.

- If a double-leg aircraft route visiting at least one small-demand gateway is included in a gateway composite, it must cover the volume of at least one small-demand gateway fully.

- A composite variable can cover at most one non-small-demand gateway.

The first restriction imposes practical limits on the fleet serving a gateway, based on the demand of that gateway. The second restriction imposes a limit on the use of double-leg routes, based on the following lemma:

Lemma 10 *Two double-leg aircraft routes visiting the same two gateways and going to or coming from the same hub do not need to be in an optimal solution.*

Proof. Suppose the optimal solution includes two double-leg pickup aircraft routes r_1 and r_2 visiting the same two gateways, a and b, and going to the same hub h (The proof for delivery routes follows similarly). Let x_a and x_b be the volumes from gateways a and b, respectively, transported on the two routes r_1 and r_2. Denote the capacity of aircraft route r_1 as u_1 and aircraft route r_2 as u_2. Without loss of generality, assume $x_a \geq x_b$ and $u_1 \geq u_2$. We can construct a solution with cost at least as good as the existing solution using route r_1 and replacing route r_2 with a single-leg route r_3 with capacity u_2 from gateway a to hub h. We know:

$$u_1 \geq u_2 \Longrightarrow u_1 \geq \frac{1}{2}(u_1 + u_2), \text{ and}$$

$$x_b \leq x_a \text{ and } (x_b + x_a) \leq (u_1 + u_2) \Longrightarrow x_b \leq \frac{1}{2}(u_1 + u_2), \text{ and hence,}$$

$$u_1 \geq x_b.$$

Thus, the capacity of route r_1 is sufficient to pick up all the demand at gateway b. Moreover, the remaining capacity of r_1 and the capacity of the single-leg route r_3 from gateway a to hub h is sufficient to pick up the demand at gateway a (because $(x_b + x_a) \leq (u_1 + u_2)$.)

If, in the original solution, route r_2 goes from gateway a to gateway b and then h, then in the new solution, route r_3, visiting only gateway a, does not change aircraft balance. Hence, the new solution serves all the demand at gateway a and b, with balance condition unchanged, and does so with less cost. If instead, route r_2 in the original solution goes from gateway b to gateway a and then hub h, the new solution will have to include a ferried aircraft with capacity u_2 from gateway b to gateway a prior to the pickup operation. The resulting new solution has same operating costs as the original solution, and hence, in this case, costs remain the same.

■

Similarly, if a gateway composites including many double-leg routes, it is likely that we can substitute one or more double-leg routes with single-leg routes to achieve lower costs.

The third restriction, based on an operational preference of the carrier, requires that each double-leg route serve at least one of the two gateway-hub demands fully. If a double-leg route visits two small-demand gateways, because the volume at the small-demand gateway is all assigned to the major hub, the route is required to cover fully the volume of either one of the gateways. If a double-leg route visits a small-demand gateway and a non-small demand gateway, it is possible that the route does not provide sufficient capacity for the demand between the non-small-demand gateway and the major hub. We therefore require that the route must cover the volume of a small demand gateway fully.

For the last restriction, if a minimal and connected gateway composite covers two non-small demand gateways, it must include a double-leg route visiting the two gateways. (Otherwise, these two gateways must be connected with the same small-demand gateway by two double-leg routes because there can be at most two double-leg routes by the second restriction. By the third restriction, the two double-leg routes must cover all the volume at the small-demand gateway, and thus, one stop on one of the two double-leg routes must be redundant. We have a contradiction.) The rationale for not allowing this is that it implies splitting the aircraft capacity between two gateways with relatively large demand, whereas it is usually best for a double-leg route to cover fully the demand at a small demand gateway and use the remaining capacity to cover a portion of the demand at the non-small-demand gateway. One implication of this restriction is that double-leg routes going to or from hubs other than the major hub need not to be considered when generating the gateway composites. Such routes must visit two non-small demand gateways, given that the volume of a small demand gateway is all sent to the major hub.

The procedures described in Figures B-1 and B-2 in Appendix B are used to generate pickup gateway composites. Procedures for generating delivery composites follow similarly. All the valid pickup aircraft routes are processed first with the procedure in Figure B-1 and a list of potential routes that can be combined to form gateway composites are maintained for each gateway. The combinations of the routes maintained at each gateway are then enumerated with the procedure described in Figure B-3. A valid gateway composite must provide sufficient capacity to:

1. cover the fixed demand to or from the major hub for each of the gateways it covers; and

2. cover the *minimum air pickup or delivery volume* for each of the gateway it covers.

For most gateways, all the volume must be transported by air, and the minimum air pickup or delivery volume is equal to the total volume. For some gateways, however, volume can be transported to one of the hubs by ground. In the **ESSND** problem with fixed hub assignment, the volume transported by ground is pre-determined and can be removed from the model.

In the flexible hub assignment problem, however, the volume transported by ground must be determined by the model. As a result, we cannot determine the exact air pickup capacity required for such gateways. Instead, we require that the capacity provided for such gateways exceeds the *minimum air pickup* or *delivery volume*, which is determined based on hub *service territory*. The service territory of a hub is a set of gateways that a hub can potentially serve based on the timing requirements and block time, usually those close to the hub. Obviously, the maximum volume a gateway might send to a hub by ground, defined as the *maximum pickup (*or *delivery) ground volume* hereafter, is the minimum of: (1) the total volume from (or to) that gateway to (or from) all other gateways in the service territory of the hub; and (2) the total volume that can be transported by ground to (or from) the hub from (or to) that gateway within its time window. Recall, this is specified by the total volume at the gateway multiplied by the *maximum pickup* (or *delivery*) *ground percentage*, as calculated from historical data. Once we determine the maximum ground volume, we can determine the minimum air pickup or delivery volume by subtracting the maximum ground volume from the total volume. Note that when enumerating gateway composite variables for gateways with a ground service option, we need to enumerate gateway composites with sufficient capacity to handle the minimum air volume up to the total volume.

Because the commodities from (or to) small demand gateways are all fixed to the major hub, there is a certain amount of fixed demand between a gateway and the major hub. To improve the solvability of our model, we remove the capacity used for transporting fixed demand with the following procedure (See Figure B-3 in Appendix B for details):

For gateway composites including no more than one double-leg route, we extract fixed demand from the single-leg routes first then from the double-leg route. For gateway composites with multiple double-leg routes and some *double-leg route junction gateways*, that is, gateways visited by more than one double-leg routes, the fixed demand at the junction gateways can be transported on multiple routes visiting the gateways, rendering the capacity available for the flexible commodities uncertain. For example, in Figure 4-10, the capacity of each aircraft route is 5 units and there are 2 units of fixed demand from each of the three gateways to the hub. The total demand at gateway 1 is 4 units, and at gateways 2 and 3 is 3 units. In Figure 4-10 (a), the fixed demand at gateway 1 can be transported on either of the two routes. If all the fixed demand of gateway 1 is transported on route 1-2-H, the available capacity for flexible commodities on route 1-2-H will be 1 unit, and the available capacity for flexible commodities on route 1-3-H will be 3 units. If all the fixed demand of gateway 1 is transported on route 1-3-H, the situation is reversed. Hence, the capacity available for the flexible commodities on respective routes cannot be determined.

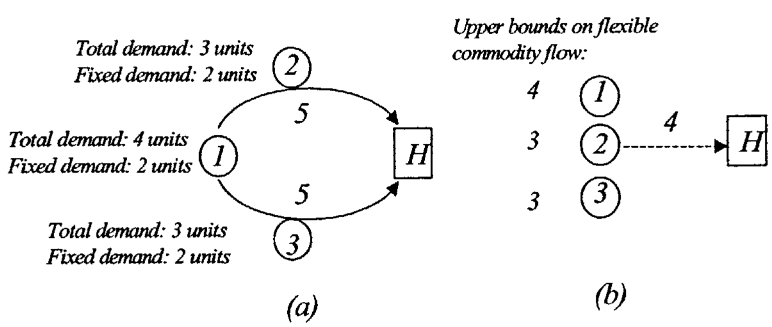


Figure 4-10: Artificial Path for Gateway Composite

For this reason, an "aggregate" path is created for this composite in Figure 4-10 (b). The capacity provided by the aggregate path equals the total capacity of all routes included in the composite minus the total fixed demand. At each gateway, we specify upper bounds on the available capacity for flexible commodities at the gateway. Because the capacity of each route is 5 units, and there are 2 units of fixed demand from each of the three gateways to the hub, the available capacity for flexible commodities at gateway 1 is 4 units, and at gateways 2 and 3 is 3 units.

4.5.3 UPS Next Day Air Problem Results

We enumerate the gateway composites for the UPS NDA problem, imposing the restrictions we detail above. After applying the small-demand gateway rule and fixing the hub assignment for west coast to west coast volume, the O-D commodities at 53 of the 85 gateways are all fixed commodities. In addition, based on the carrier's operational preferences, commodities from or to the major hub are assigned to the major hub. Overall, fixed commodities account for about 60% of the total volume. We then enumerate the gateway composites for the remaining gateways. After the gateway composites are enumerated, the MIP model is solved with CPLEX 8.5 interactive solver on a Dell Precision M50 laptop computer with Pentium 4 2.5GHz CPU and 2G RAM.

Table 4.2 details the computational results. The first section reports the number of variables. Note that only the composite variables are specified as integer variables. The second section reports the problem size after standard CPLEX preprocessing (ILOG [33]).

We compare the best network design obtained using **ARM** with the existing UPS hub assignment and the best integer solution obtained using the **GCF** formulation with flexible hub assignment. In both cases, NDA pickup aircraft are balanced against delivery aircraft.

Table 4.3 contains the information about packages assigned to different hubs under the existing UPS hub assignment and the hub assignment from the solution to the **GCF** formulation.

Variables	Path Variables	69062
	Ferry Variables	20807
	Composite Variables	33985
Problem Size	Columns	91825
	Rows	20962
	Nonzeros	678930
Run Time	LP Relaxation	317 sec.
	Best IP Solution	41 hr.
Gaps	LP Relaxation - Best IP Solution	2.5%
	Final Best LP Bound - Best IP Solution	$< 1\%$

Table 4.2: Computational Results of the UPS Next-Day Air Problem

Letters "A" through "F" represent the different hubs. The row "Network" represents the whole network. Column "P" is the number of packages processed as a percentage of the total number of packages, columns "GP" and "GD" represent the number of packages picked up or delivered, respectively, by ground to or from that hub as a percentage of the number of packages processed at that hub. Note, packages that can be picked up and delivered all by ground exclusively are excluded from the air service network design problem. Among the six hubs, the packages processed by the west coast hub are the same under both hub assignment plans, those with origins and destinations all on the west coast. In our hub assignment, two hubs are not used as air service hubs. We observe: (1) both hubs handled a very small percentage of the total packages in the existing hub assignment; and (2) both hubs provided service links to only a small set of gateways in the existing hub assignment. Though more than one quarter of the packages processed by the two hubs can be either picked up or delivered by ground, air service is still required to serve these packages. Hence, establishing more hubs to allow more packages to be transported by ground is not a particularly effective strategy.

In Table 4.4, we compare costs and network characteristics of the design obtained using **ARM** with the existing UPS hub assignment and that obtained using **GCF**. For confidentiality

	Existing Hub Assignment			Flexible Hub Assignment		
	P	GP	GD	P	GP	GD
A	0.044	0.299	0.314	N/A	N/A	N/A
B	0.059	0.283	0.254	N/A	N/A	N/A
C	0.077	0.287	0.259	0.077	0.287	0.259
D	0.182	0.334	0.259	0.166	0.385	0.280
E	0.190	0.194	0.120	0.165	0.223	0.157
F	0.448	0.030	0.020	0.592	0.022	0.015
Network	1.000	0.213	0.145	1.000	0.194	0.125

Table 4.3: Package Distribution

reasons, we do not report the exact cost but rather report percentage differences in the solutions, using the **ARM** solution as the baseline. About 7.5% savings in operating cost can be achieved through the flexible hub assignment, translating to tens of millions of dollars in annual operating cost savings. Reduced cycle costs account for a large portion of the savings. The reduction in cycle costs is a direct result of the reduced number of *air service links*. An air service link represents a non-zero air pickup or delivery demand for a gateway-hub pair, implying at least one aircraft route visiting the gateway. In the existing UPS hub assignment, a gateway sends volume to or receives volume from 2.08 hubs by air on average. In contrast, a gateway sends volume to or receives volume from 1.68 hubs by air on average in our solution. Furthermore, there are more double-leg routes used in the fixed hub assignment solution because there are more, and hence smaller, gateway-hub demands to serve.

In addition to operating cost savings, two aircraft and air service for two hubs can be eliminated from the existing network, implying tens of millions of dollars in annual savings. Interestingly, the above savings are reported based on tactical models, in which no credit is given for reduction in the number of aircraft or the number of hubs operated. If we adjust the formulation to include aircraft ownership and hub set-up and operating costs, the potential

	Existing Hub Assignment	Flexible Hub Assignment
Operating Cost	-	-7.5%
Cycle Cost	-	-11.4%
Flying Cost	-	-5.3%
Air Service Links	351	277
Block Time	590.4	558.5
Pickup Routes	133	129
Delivery Routes	133	131
Aircraft Needed	133	131
Double-leg Routes	79	46
Air Hubs	6	4

Table 4.4: Flexible Hub Assignment Improved Results for UPS NDA Problem

savings would likely be even greater.

Table 4.5 reports the average flight load factor and the number of air service links. Column "LF" is the average flight load factor, and column "SL" is the number of air service links. The average flight load factor of the network in our hub assignment increases by about 5%: the result of both the reduced number of aircraft used and a higher percentage of packages moved by air as indicated in row "Network" of Table 4.3.

	Existing Hub Assignment				Flexible Hub Assignment			
	Pickup		Delivery		Pickup		Delivery	
	LF	SL	LF	SL	LF	SL	LF	SL
A	0.773	7	0.675	9	N/A	N/A	N/A	N/A
B	0.815	13	0.766	15	N/A	N/A	N/A	N/A
C	0.766	15	0.740	14	0.774	15	0.745	14
D	0.754	25	0.841	28	0.814	17	0.905	20
E	0.845	28	0.813	25	0.900	19	0.862	20
F	0.806	86	0.829	86	0.859	86	0.863	86
Network	0.803		0.806		0.853		0.860	

Table 4.5: Average Flight Load Factor

4.5.4 Staged Volume Arrival at Hubs

The **GCF** formulation specifies the hub sort capacity as a single constraint for each hub. In a more realistic representation, each hub can handle a limited number of packages for each time period. Hence, we need to ensure that package arrivals are spaced over the hub sort period and not concentrated at the end of the sort.

Armacost [4] uses aircraft arrivals as a proxy of package arrivals and specify conditions to ensure spaced aircraft arrival. We can similarly specify the conditions for package arrivals. Suppose x_i is the number of packages arriving at the hub in interval τ, and the number of intervals is T. Let the hub sort capacity during interval τ be e_τ. Then, the following conditions, called *staggering constraints*, must be satisfied for each interval, τ:

$$\sum_{j=\tau}^{T} x_j \leq \sum_{j=\tau}^{T} e_j, \qquad \tau = 1, ..., T.$$

We can specify the staggering constraints using the path variables to represent the package flows in the **GCF** formulation. Let δ_p^τ be 1 if the earliest possible hub arrival time ($EHAT$) for pickup path p belongs to interval τ, and 0 otherwise. The staggering constraints are stated as:

$$\sum_{j=\tau}^{T} \sum_{p \in \mathcal{P}^P} \sum_{l \in \mathcal{N}} \delta_p^j \varphi_p^{l,h} x_{l,p} \leq \sum_{j=\tau}^{T} e_j, \qquad \tau = 1, ..., T, \; h \in H. \tag{4.56}$$

For a given pickup path, however, the $EHAT$'s of the aircraft routes covering that path vary by assigned fleet types because block times and turn times differ by fleet type. To extend constraints (4.56) to capture this variability is non-trivial, requiring specifying the time-space network with *EHAT*. Instead of further complicating our model, we obtain the solution to the flexible hub assignment problem *without* the staggering constraints. Because most routes selected are single-stop routes that generally arrive at hubs early, our solution is less likely to violate the staggering constraints. We check the feasibility of the staggering constraints for 15-minute intervals spanning the sort period nevertheless. We calculate the *EHAT* for the routes in our solution and assign the gateway-hub demand to routes. If there are multiple routes from a gateway to a hub, the earlier routes are filled up first. We then calculate the number of packages arriving at the hub in each 15-minute interval. We find that none of the hub staggering constraints is violated. In Figure 4-11, we depict the number of packages arriving at the major hub by the earliest possible hub arrival time. We calculate the minimum required sorting rate for each sorting interval, i.e., 4 to 7.5, 4.25 to 7.5, and so on. The maximum sorting rate required, at about 180,000 packages/hour, is for the 4 to 7.5 interval as a result of the relatively concentrated volume in the early period. Staggering constraints for other hubs are not violated, either. This suggests that the staggering constraints need not be included in our formulation, and a total sorting capacity constraint for each hub is sufficient.

The hub assignment obtained from our model was used as the fixed hub assignment for the **ARM** model. A 8.2% potential savings using the new hub assignments was reported by UPS for the same problem instance, with gateway-hub-specific conversion factors translating package demand into container demand.

4.5.5 Ramp-transfer Operations

Ramp-transfer operations allow packages to be transferred between aircraft visiting the same gateway. Packages are taken from one aircraft, and either placed onto another aircraft arriving

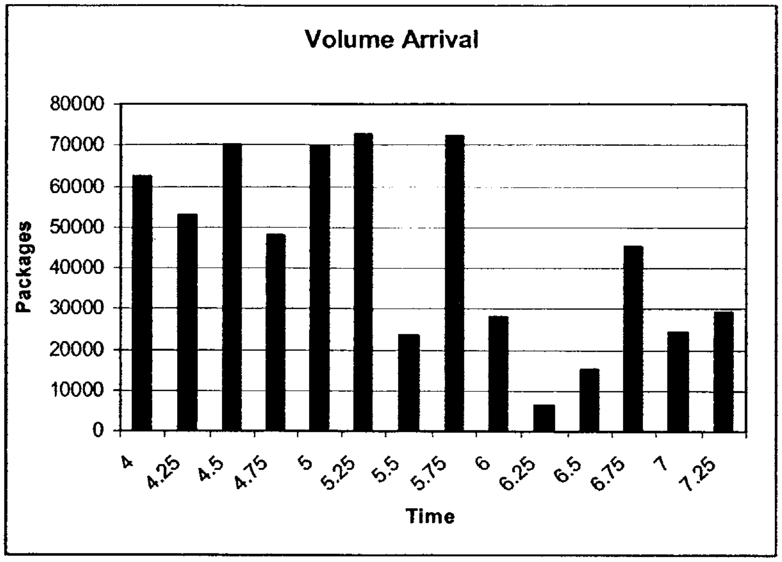


Figure 4-11: Volume Arriving at the Major Hub by E.H.A.

at the gateway earlier or left on the ramp for another aircraft to pickup at a later time. In the former case, ramp-transfer is possible only if the aircraft first arriving at the gateway can be held until the ramp-transfer is finished and still arrive at the hub on time. Figure 4-12 depicts an example of a ramp-transfer operation. Suppose there is only one fleet type and an aircraft of this type has 3 units of capacity. Each of the three gateways has two O-D commodities, and each commodity has one unit of demand. We assume the best hub assignments for these commodities are known, as depicted in the figure. With coordinated timing, we can use one aircraft flying route 1-3-H_1 to carry commodities k_1 and k_2 to gateway 3, and a second aircraft flying route 2-3-H_2 to carry commodities k_3 and k_4 to gateway 3. At gateway 3, we transfer commodity k_2 from the first aircraft to the second, and k_3 from the second to the first. Then, after picking up commodity k_5, the first aircraft continues to hub H_1 carrying commodities k_1, k_3, and k_5, and the second aircraft, after picking up commodity k_6, continues to hub H_2 carrying commodities k_2, k_4, and k_6. Commodities k_2 and k_3 are served by two different aircraft in this pickup operation.

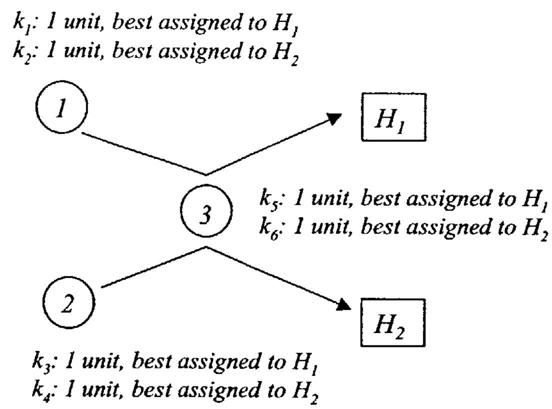


Figure 4-12: An Example of Ramp-transfer Operation

Ramp-transfer operations can effectively reduce aircraft movements, and hence, operating costs. In the above example, if ramp-transfer is not allowed, we need at least three aircraft to transport all commodities. A possible solution is route 1-3-H_1 transporting commodities k_1 and k_5 to hub H_1, route 1-3-H_2 transporting commodities k_2 and k_6 to hub H_1, and route 2-H_2-H_1 transporting commodities k_3 and k_4 to hub H_1 and H_2.

We can build gateway composites involving ramp-transfer operations for the **GCF** formulation. In the example depicted in Figure 4-12, routes 1-3-H_1 and 2-3-H_2 form a gateway composite covering gateways 1, 2 and 3. Similar to other gateway composites, ramp-transfer gateway composites provide capacity on flights arcs contained in constraints (4.41) and (4.42).

Although theoretically feasible, we do not consider ramp-transfer composites in our implementation because there are fewer ramp-transfer opportunities as a result of the small-demand gateway rule. If ramp-transfer is considered, the potential operating cost savings might be even greater.

4.5.6 Scenario Analysis

As we have seen, though the **GCF** formulation is primarily a tool for tactical planning purposes, its solution can provide insights to guide the strategic planning process. In this section, we investigate two scenarios. In the first scenario, we relax the hub sort capacity constraints and investigate the optimal set of hubs to use. In the second scenario, we look at the effects of maximum consolidation by allowing use only of the major hub and the west coast hub.

In the best network design solution we obtained, four hubs are used. We refer to the solution as the 4-hub solution. Because some of the hubs might be used due to limited hub sort capacity at other hubs, we relax hub sort capacity constraints, and obtain a new solution (reported in Table 4.6). In this solution, we use only 3 hubs and we refer to it as the 3-hub solution. We then compare the 3-hub and 4-hub solutions in Table 4.6. The percentages reflect the differences from the 4-hub solution.

	3-Hub Solution
Operating Cost	-2.4%
Cycle Cost	-3.7%
Flying Cost	-1.7%
Air Service Links	236
Block Time	546.8
Pickup Routes	127
Delivery Routes	127
Aircraft Needed	129
Double-leg Routes	40
Air Hubs	3

Table 4.6: Flexible Hub Assignment Solution with Relaxed Hub Sort Capacity Constraints

When hub sort capacity is not constraining, we can further cut operating costs by 2.4%, translating to tens of millions of dollars each year. Furthermore, we can save another two aircraft and eliminate one more air hub, implying even greater savings. The network maintains a hub on the east coast, a hub on the west coast and the major hub. The hub sort capacity is violated by about 10% at the major hub, and not violated at the other two hubs. The current hub capacity expansion project at the major hub, however, will easily accommodate the capacity needed in the 3-hub solution.

In the second scenario, we assign every commodity to the major hub except commodities with both origins and destinations on the west coast; these commodities must be handled by the west-coast hub. The network therefore includes only two hubs and all commodities are fixed. The hub sort capacity at the major hub only is relaxed. Table 4.7 reports the results, referred to as the 2-hub solution. The percentages again reflect the differences from the 4-hub solution.

	2-Hub Solution
Operating Cost	+1.6%
Cycle Cost	+1.9%
Flying Cost	+1.4%
Air Service Links	201
Block Time	571.4
Pickup Routes	133
Delivery Routes	132
Aircraft Needed	133
Double-leg Routes	43
Air Hubs	2

Table 4.7: Maximum Consolidation Results

The results show that the number of air service links is greatly reduced compared to the 4-hub operation, but the cycle cost is higher. The reason is that more volume has to be transported by air now, and therefore, more aircraft routes are required. Flying cost is also higher in the 2-hub solution than the 4-hub solution; many aircraft have to travel a longer distance in the 2-hub solution. Although not optimal, the 2-hub solution still has a lower operating cost than **ARM** solution.

Our analysis indicates that UPS could potentially reduce operating costs by increasing consolidation. In practice, however, other factors might affect the optimal hub configuration, for example, long-term demand growth and network robustness.

4.5.7 Demand Growth

The optimal set of hubs to be used depends on the demand of the network. The higher the demand, the more hubs the optimal hub set includes. Demand growth of express shipment is relatively steady. Because it takes a long time to build hubs or expand hub capacity, however, we need to take into consideration potential demand growth several years ahead when considering the set of hubs to use. In this section, we look at a 5-year planning horizon, in which we assume that domestic demand will grow by approximately 5.4% each year, yielding about a 30% demand increase in 5 years. We base these projections on the carrier's annual report [89] from 1995 to 2002. Their 5-year growth rates range from 27% to 35%. We assume the demand increase will be uniform, that is, volume of every O-D commodity will increase by 30%. We also increase the number of aircraft available, based on the carrier's current aircraft purchase plan. Finally, we relax the hub sort capacity.

Using the demand estimates for 5 years out, we solve **GCF** and report the results in Table 4.8. We compare operating costs of our solution with the 3-hub solution using current demand volume and relaxed hub sort capacity. The results suggest that the 4-hub network structure is more appropriate, taking into consideration long-term demand growth. In addition to achieving minimal operating costs in the future demand scenario, the 4-hub network structure is also more robust than the 3-hub network structure. Under the 3-hub network structure, the major hub is the only hub that can handle east coast-west coast and west coast-east coast demand. If something is wrong at the major hub, for example, weather conditions delay arrivals at the major hub and result in insufficient sort capacity, there is no alternative feasible operating strategy. In contrast, flights can be re-routed to the alternative hub under the 4-hub network structure to reduce the demand at the major hub.

	Future 4-Hub Solution
Operating Cost	+15.6%
Cycle Cost	+10.1%
Flying Cost	+18.5%
Air Service Links	273
Block Time	637.1
Pickup Routes	145
Delivery Routes	145
Aircraft Needed	145
Double-leg Routes	43
Air Hubs	4

Table 4.8: Results of a 30 Percent Demand Increase

Though the overall demand increases by 30%, the total operating cost increases by only 15.6%. This is achieved largely by replacing small old aircraft with large new aircraft, the available number of which is almost doubled in the carrier's new aircraft purchase plan. Figure 4-13 (a) shows the usage of each fleet type in the 3-hub solution given current demand volumes, and Figure 4-13 (b) shows the usage of each fleet type with a 30% increase in demand and new aircraft purchased under the carrier's purchase plan.

When the new large type D aircraft become more available, none of the old and smaller type A and C aircraft is used. The number of required aircraft increases by only 12%. Because the fixed cycle costs of new aircraft are lower than those of old ones, the total cycle cost increases by only 10.1%. Block hours increase by 15%, but the flying cost increases by 18.5%. The reason is that the block hour cost of type B aircraft is the lowest. Under the new network

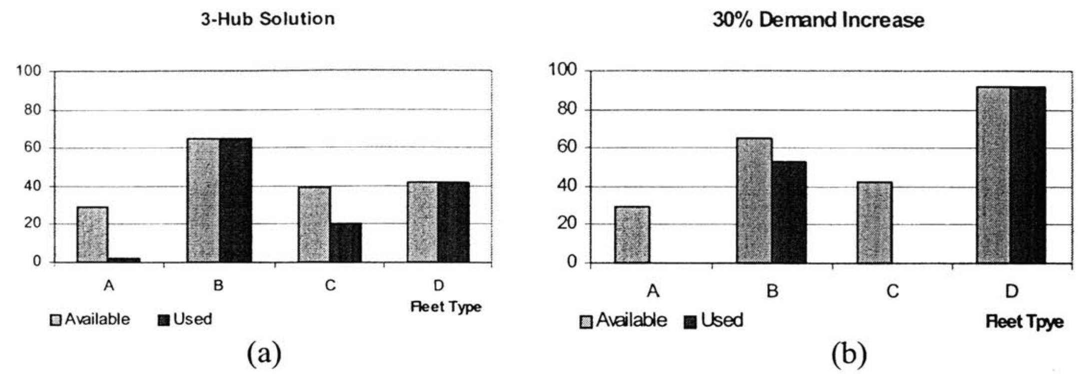


Figure 4-13: Available Fleet and Actual Usage

design, fewer type B aircraft are flown because more large type D aircraft become available.

To illustrate that the multiple-hub network structure becomes increasingly attractive when demand increases, we consider the network design in which we allow only the major hub and the west-coast hub to be used to handle the future demand. In Table 4.9, we compare the resulting operating costs with the 4-hub solution with future demand volumes and relaxed hub sort capacity. Operating costs increase by 4.9%, and another 6 aircraft are needed if only two hubs are allowed. In contrast, operating costs increase by only about 4.1% and only 4 more aircraft are needed changing from the best solution (the 3-hub solution) to the 2-hub operation with relaxed hub capacity and current demand volumes.

In summary, when demand increases, the solution is likely to increase the number of hubs used and its use of large efficient aircraft.

	Future 2-Hub Solution
Operating Cost	+4.9%
Cycle Cost	+4.8%
Flying Cost	+4.9%
Air Service Links	201
Block Time	663.27
Pickup Routes	151
Delivery Routes	151
Aircraft Needed	151
Double-leg Routes	49
Air Hubs	2

Table 4.9: Results of a 30 Percent Demand Increase with 2-Hub Operation

4.6 Summary

In this chapter, we present an effective model, the Gateway Cover Flow (**GCF**) model, for large-scale express shipment service network design with flexible hub assignment. We demonstrate that for the UPS domestic Next-Day Air problem, the **GCF** model generates solutions representing tens of millions of dollars in annual operating cost savings. Though designed for tactical planning purposes, the **GCF** model can be effective in providing insights for strategic decision-making, decisions involving the addition or elimination of hubs or hub capacity.

The composite variable concept can be a useful tool to capture difficult constraints implicitly. We demonstrate however, the type of decisions captured in a composite variable greatly affects the effectiveness of the model.

Chapter 5

Future Research

Each composite variable represents a *sub-assembly* of a possible solution. By modeling the express shipment service network design problem using sub-assemblies instead of *raw components*, that is, individual route variables, we are able to tighten the bounds provided by the associated LP relaxations and improve problem tractability. This achievement, however, requires both careful consideration of the variable definition and advanced algorithmic solution approaches. In applying our novel models and solution approaches, we demonstrate that express package delivery problem instances from a large carrier are solvable, with potential impacts including tens of millions of dollars in annual operating cost savings and even greater reductions in aircraft ownership, hub set-up, and maintenance costs.

We suggest the following additional research on the topic of express shipment service network design.

- **Column generation combined with variable enumeration for the GCF model.**

  In our **GCF** implementation, gateway composite variables are enumerated. Many of these variables are not likely to be included in an optimal solution. Column generation with some restrictions might allow us to reduce the size of the integer programming model and reduce solution time.

- **Gateway selection and assignment of ground centers to gateways.** In the Second-Day problem, the volume from a ground center need not be assigned to the closest gateway because there is more time available for delivery. Thus, savings can be achieved if we optimally select a set of gateways to use and assign ground centers to gateways when considering the Second-Day problem. This embeds a facility-location problem within the flexible hub assignment problem. We can potentially adapt **GCF** to model this problem. The challenge is the likely dramatic increase in the numbers of variables.

- **Serving part of the Second-Day demand on Next-Day flights.** Demand for Second-Day service cannot be served by Next-Day flights, given current practice. It could prove worthwhile to investigate the feasibility and potential benefits for doing so. If benefits are evident, the models and algorithms we develop in this dissertation can be adapted to consider serving Second-Day demand on Next-Day flights. The challenge is again the increase in the numbers of variables.

- **Robust network design.** In this dissertation, our objective is to minimize operating costs. A more realistic representation of the system should take service reliability into consideration and allow the resulting network to accommodate certain unusual events, for example, loss of hub capacity due to inclement weather conditions.

- **Operations recovery.** When unusual events disrupt scheduled service, the carrier might need to re-route O-D commodities through alternative hubs to recover or achieve a feasible plan. The **GCF** model could be adapted to consider such cases, and restrictions in the operations recovery problem might allow us to improve the solution speed.

- **Incremental network design.** In practice, dramatic changes to the current network might be unacceptable to carriers. A more readily implementable approach might take allowable changes and switching costs into consideration. This could in fact help improve tractability because the model is more restrictive.

In addition to the express shipment service network design problem, we suggest the following areas of research related to the models and algorithms we design in this dissertation.

- **Information-enhanced column generation applied to other problems.** One possible application is the multi-commodity network flow problem, in which we can design the flows for the sets of commodities on a portion of the network.

- **The composite variable concept applied to other fixed charge network design problems.** One possible application is the facility-location problem. Although many facility-location problems contain significant flow costs, we can design models combining composite variables and flow variables. Doing so will not compromise the **GCF** formulation strength.

Appendix A

Proof of Formulation Strength

Lemma 2 ***RH** and **RF** are equivalent formulations.*

Proof. Given a feasible solution to the **RH** formulation, the routes selected satisfy constraints (2.7), (2.8), (2.9) and (2.10) in the **RF** formulation, because they are the same as those in the **RH** formulation. We only need to find a set of feasible commodity-based path flows for the **RF** formulation, given the set of feasible gateway-based path flows and hub assignments for the **RH** formulation. Consider the pickup operation between a gateway-hub pair (l, h). We denote the set of pickup paths between gateway l and hub h as $\mathcal{P}_l^{P,h}$, and the set of commodities that originate at gateway l and can be sorted at hub h as $K_l^{P,h}$. Given the set of feasible gateway-based pickup path flows, we can construct a set of commodity-based pickup path flows between the gateway-hub pair as follow.

- **Step 1**: select a path p from $\mathcal{P}_l^{P,h}$, and let $x_{l,p}$ be the flow from gateway l on path p in the solution to the **RH** formulation.
- **Step 2**: select a commodity k from $K_l^{P,h}$, and let x_h^k be the flow of commodity k through

hub h, $x_h^k = b_k z_h^k$.

- **Step 3**: If $x_{l,p} > x_h^k$, let $x_p^k = x_h^k$, $x_{l,p} = x_{l,p} - x_h^k$, $x_h^k = 0$, and repeat **Step 2**. If $x_{l,p} \leq x_h^k$, let $x_p^k = x_{l,p}$, $x_h^k = x_h^k - x_{l,p}$, $x_{l,p} = 0$, and select another path p from $\mathcal{P}_l^{P,h}$ and repeat **Step 3**.

The above process terminates when $x_{l,p} = 0$ for each $p \in \mathcal{P}_l^{P,h}$ and $x_h^k = 0$ for each $k \in K_{l,h}^P$. This is true because of constraints (4.13). Similarly, we can construct commodity-based delivery path flows between gateway l and hub h and pickup and delivery path flows between other gateway-hub pairs. Let $O(k)$ and $D(k)$ denote the origin and destination gateway of commodity k, respectively. Because an O-D commodity k can be transported on any pickup path between its origin and one of its candidate hubs and on any delivery path between one of its candidate hubs and its destination, we have

$$\begin{aligned} \mathcal{P}_{O(k)}^{P,h} &= \mathcal{P}_k^{P,h}, \text{ and} \\ \mathcal{P}_{D(k)}^{D,h} &= \mathcal{P}_k^{D,h}. \end{aligned}$$

Thus,

$$\sum_{k \in K} \sum_{p \in \mathcal{P}_k^P} x_p^k = \sum_{l \in \mathcal{N}} \sum_{h \in H} \sum_{k \in K_l^{P,h}} \sum_{p \in \mathcal{P}_l^{P,h}} x_p^k = \sum_{l \in \mathcal{N}} \sum_{h \in H} \sum_{p \in \mathcal{P}^P} \varphi_p^{l,h} x_{l,p} = \sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^P} \delta_p^l x_{l,p}.$$

The first equality is the result of grouping commodities by their origin gateway, and the second equality is achieved by construction. By constraints (4.11),

$$\sum_{r \in R_P} \delta_r^{ij} u_r y_r - \sum_{k \in K} \sum_{p \in \mathcal{P}_k^P} \delta_p^{ij} x_p^k = \sum_{r \in R_P} \delta_r^{ij} u_r y_r - \sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^P} \delta_p^{ij} \delta_p^l x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{A}.$$

So the constructed commodity-based pickup path flows satisfy constraints (2.11). We can similarly prove that the constructed commodity-based delivery path flows satisfy constraints (2.12).

By construction,

$$\begin{aligned}
\sum_{h\in H^k}\sum_{p\in\mathcal{P}_k^{P,h}} x_p^k &= \sum_{h\in H^k}\sum_{p\in\mathcal{P}_{O(k)}^{P,h}} x_h^k = \sum_{h\in H^k} b_k z_h^k = b_k, \qquad k\in K, \text{ and} \\
\sum_{p\in\mathcal{P}_k^{P,h}} x_p^k - \sum_{p\in\mathcal{P}_k^{D,h}} x_p^k &= \sum_{p\in\mathcal{P}_{O(k)}^{P,h}} x_p^k - \sum_{p\in\mathcal{P}_{D(k)}^{P,h}} x_p^k = b_k z_h^k - b_k z_h^k = 0, \qquad k\in K,\ h\in H^k.
\end{aligned}$$

By constraints (4.16),

$$\sum_{k\in K}\sum_{p\in\mathcal{P}_k^{P,h}} x_p^k = \sum_{k\in K} x_h^k = \sum_{k\in K} b_k z_h^k \le e_h, \qquad h\in H.$$

Thus, the constructed set of commodity-based path flows satisfy constraints (2.11)-(2.15). Overall, for any feasible solution to the **RH** formulation, we can find a feasible solution to the **RF** formulation with equal cost.

Conversely, given a feasible solution to the **RF** formulation, the routes selected also satisfies constraints (4.7)-(4.10). We can construct a set of gateway-based path flows and hub assignment variables with equations (A.1)-(A.3).

$$x_{l,p} = \sum_{k \in K_l^{P,h}} x_p^k, \qquad l \in \mathcal{N},\ h \in H,\ p \in \mathcal{P}_l^{P,h} \tag{A.1}$$

$$x_{l,p} = \sum_{k \in K_l^{D,h}} x_p^k, \qquad l \in \mathcal{N},\ h \in H,\ p \in \mathcal{P}_l^{D,h} \tag{A.2}$$

$$z_h^k = \sum_{p \in \mathcal{P}_k^{P,h}} x_p^k / b_k, \qquad k \in K,\ h \in H^k \tag{A.3}$$

Reversing the reasoning in the first part of the proof, we can show that the set of constructed gateway-based path flows and hub assignments variables satisfy constraints (4.11)-(4.16). Thus, for any feasible solution to the **RF** formulation, we can also find a feasible solution to the **RH** formulation with equal cost. ■

Lemma 3 *CHA and RH are equivalent integer programming formulations.*

Proof. Given any integer commodity composite solution to the **CHA** formulation, we can construct an integer route solution to the **RH** formulation through the following mapping

$$\sum_{c \in C} \gamma_c^r v_c = y_r. \tag{A.4}$$

The constructed route solution has the same objective value as the commodity composite solution. We can show it satisfies constraints (4.7)-(4.10), using count constraints (4.10) as an illustrating example. In the **CHA** formulation, for a fleet type f, we have

$$\sum_{c \in C} \sum_{r \in R} \delta_r^f \gamma_c^r v_c \leq n_f,$$

where δ_r^f is 1 if fleet type f is assigned to route r, and 0 otherwise, and γ_c^r is the number of aircraft route r included in composite c. With mapping (A.4), we can re-write the constraint as

$$\sum_{r\in R}\delta_r^f\sum_{c\in C}\gamma_c^r v_c = \sum_{r\in R}\delta_r^f y_r \leq n_f.$$

So the route solution obtained by mapping (A.4) satisfies the count constraints, and we can similarly show that balance and landing constraints in the **RH** formulation are also satisfied.

Next, we construct a set of feasible gateway-based path flows for the **RH** formulation. Let ψ_c^k be the hub assignment for O-D commodity k and $x_{l,p}^c$ the flow from (or to) gateway l on pickup (or delivery) path p in commodity composite c. Summing equations (4.20) for the pickup commodity composites selected in the given solution, we have

$$\sum_{c\in\mathcal{C}_P}\sum_{p\in\mathcal{P}^P}\varphi_p^{l,h}x_{l,p}^c v_c - \sum_{c\in\mathcal{C}_P}\sum_{k\in K}\beta_k^{l,P}b_k\psi_c^{k,h}v_c = 0, \qquad l\in\mathcal{N},\ h\in H.$$

By constraints (4.27),

$$\sum_{c\in\mathcal{C}_P}\sum_{p\in\mathcal{P}^P}\varphi_p^{l,h}x_{l,p}^c v_c \geq \sum_{k\in K}\beta_k^{l,P}b_k z_h^k, \qquad l\in\mathcal{N},\ h\in H. \tag{A.5}$$

Following the notation used in the proof of **Lemma 2**, we construct a set of feasible gateway-based pickup path flows between a gateway-hub pair (l,h) by the following process:

- **Step 1**: select a path p from $\mathcal{P}_l^{P,h}$, and let $x_{l,p} = \sum_{c\in\mathcal{C}_P} x_{l,p}^c v_c$ be the flow from gateway l on path p.
- **Step 2**: select a commodity k from $K_l^{P,h}$, and let x_h^k be the flow of commodity k through hub h, $x_h^k = b_k z_h^k$.

- **Step 3**: If $x_{l,p} > x_h^k$, let $x_p^k = x_h^k$, $x_{l,p} = x_{l,p} - x_h^k$, $x_h^k = 0$, and repeat **Step 2**. If $x_{l,p} \leq x_h^k$, let $x_p^k = x_{l,p}$, $x_h^k = x_h^k - x_{l,p}$, $x_{l,p} = 0$, and select another path p from $\mathcal{P}_l^{P,h}$ and repeat **Step 3**.

The above process terminates when $x_h^k = 0$ for each $k \in K_l^{P,h}$ by equation (A.5). Similarly, we can construct gateway-based delivery path flows between gateway l and hub h and pickup and delivery path flows between other gateway-hub pairs.

We use this constructed set of gateway-based path flows as the solution for the **RH** formulation. By construction, we have

$$\sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^P} \delta_p^{ij} \delta_p^l x_{l,p} = \sum_{k \in K} \beta_k^{l,P} b_k z_h^k, \qquad l \in \mathcal{N},\ h \in H, \text{ and} \tag{A.6}$$

$$\sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^D} \delta_p^{ij} \delta_p^l x_{l,p} = \sum_{k \in K} \beta_k^{l,D} b_k z_h^k, \qquad l \in \mathcal{N},\ h \in H. \tag{A.7}$$

Therefore, the constructed set of gateway-based path flows satisfy constraints (4.13) and (4.14). In addition,

$$\sum_{r \in R_P} \delta_r^{ij} u_r y_r = \sum_{r \in R_P} \delta_r^{ij} u_r \sum_{c \in C_P} \gamma_c^r v_c \geq \sum_{c \in C_P} \sum_{p \in \mathcal{P}^P} \sum_{l \in \mathcal{N}} \delta_p^{ij} \delta_p^l x_{l,p}^c v_c \geq \sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^P} \delta_p^{ij} \delta_p^l x_{l,p}, \quad (i,j) \in \mathcal{A}.$$

The first inequality follows inequalities (4.21), and the second follows inequalities (A.5). The constructed gateway-based pickup path flows therefore satisfy constraints (4.11). Similarly, we can prove that the construct a set of gateway-based delivery path flows satisfy constraints (4.12). For hub assignment variables, we can use the hub assignment solution for the **CHA** formulation to satisfy constraints (4.15) and (4.16). Overall, given an integer feasible solution to the **CHA** formulation, we can find an integer feasible solution to the **RH** formulation with the same cost.

Conversely, given an integer feasible solution to the **RH** formulation, we can construct two trivial commodity composites with the set of routes selected in the solution. One includes all the pickup routes selected and picks up all the O-D commodities, and the other includes all the delivery routes selected and delivers all the O-D commodities, following the hub assignment in the given solution. These two commodity composites and the hub assignment solution satisfy all the constraints in the **CHA** formulation and have the same cost. ■

The first part of the above proof, (given any integer commodity composite solution to the **CHA** formulation, we can construct an integer route solution to the **RH** formulation with the same cost) holds when integrality constraints are relaxed. Thus, we can prove the following lemma.

Lemma 3 *The **CHA** formulation is at least as strong as the **RH** formulation.*

Appendix B

Procedures

```
Procedure ProcessRoutes(List of feasible pickup aircraft routes )
   while not End-of-List do
      select a route r  from the list of feasible pickup aircraft routes
      if (route r is a single-leg route) // suppose it visits gateway i and hub h
         add r to i.pickupCompositeRouteList
         r.availableCapacityFor[i]  ← r.capacity
      else // r is a double-leg route, suppose it visits gateway i, j, and hub h
         if (h is the major hub)
            if ( i is a small-demand gateway and
               r.capacity ≥ i.fixedPickupDemandtoMajorHub )
               add r to j.pickupCompositeRouteList
               r.availableCapacityFor[j]  ← r.capacity -
               i.fixedPickupDemandtoMajorHub
            endif
            if ( j is a small-demand gateway and
               r.capacity ≥ j.fixedPickupDemandtoMajorHub )
               add r to i.pickupCompositeRouteList
               r.availableCapacityFor[i]  ← r.capacity -
               j.fixedPickupDemandtoMajorHub
            endif
         endif
      endif
   end while
End Procedure
```

Figure B-1: Procedure for Processing Routes for Gateway Composite Variable Generation

```
Procedure GenerateGatewayCompositeVariables()
    select a gateway i from the list of gateways
    while not End-of-List do
        routeList ← 0
        CompositeRecursive(i, 0, 0, routeList, 0, 0)
    end while
End Procedure

Procedure CompositeRecursive (i, capacity, majorHubCapacity, routeList, numRoutes, k)
    while not End-of-List do
        get the kth route r from i.pickupCompositeRouteList
        if ( r is a single-leg route or r satisfies double-leg route conditions)
            if ( r goes to the major hub)
                tempMajorHubCapacity ← majorHubCapacity + r. availableCapacityFor[i]
            endif
            copy routeList to newRouteList
            add r to newRouteList
            if ( capacity + r. availableCapacityFor[i] < i.minAirPickupDemand )
                if ( numRoutes + 1 < r.compsiteRouteLimit )
                    CompositeRecursive( i, capacity + r. availableCapacityFor[i] ,
                        tempMajorHubCapacity, newRouteList, numRoutes +1, k)
                endif
            else
                if ( tempMajorHubCapacity ≥ i.fixedPickupDemandToMajorHub )
                    build gateway composite variable c from newRouteList
                    add c to r.gatewayCompositeList
                endif
                if ( numRoutes + 1 < r.compsiteRouteLimit and
                    capacity + r.availableCapacityFor[i] < i.totalPickupDemand )
                    CompositeRecursive( i, capacity + r. availableCapacityFor[i] ,
                        tempMajorHubCapacity, newRouteList, numRoutes +1, k)
                endif
            endif
        endif
        k ← k+1
    end while
    delete routeList
End Procedure
```

Figure B-2: Procedures for Generating Gateway Composite Variables

```
Procedure RemoveCapacityForFixedDemand( a gateway composite variable)
   while not End-of-List do
      select a route r from the set of routes included in the composite variable
      if (route r is a double-leg route)
         numberDoubleStopRoutes ← numberDoubleStopRoutes + 1
         add r to doubleStopRouteList
      else // r is a single-leg route, suppose it visits gateway i and hub h
         if the r.capacity ≥ i.fixedDemandFor[h] then
            r.capacity ← r.capacity - i.fixedDemandFor[h]
            i.fixedDemandFor[h] ← 0
         else
            r.capacity ← 0
            i.fixedDemandFor[h] ← i.fixedDemandFor[h] - r.capacity
         endif
      endif
   end while
   if ( numberDoubleStopRoutes = 1 ) then
      get route r from doubleStopRouteList
      // suppose r serves gateway i and j, and hub h
         r.capacity ← r.capacity - i.fixedDemandFor[h] - j.fixedDemandFor[h]
         i.fixedDemandFor[h] ← 0
         j.fixedDemandFor[h] ← 0
   else if ( numberDoubleStopRoutes = 2 ) then
      create an artificial path p
      select a route r from the set of routes included in the composite variable
      // suppose r serves gateway i and j, and hub h
      while not End-of-List do
         p.capacity ← p.capacity + r.capacity - i.fixedDemandFor[h]
                  - j.fixedDemandFor[h]
         p.flowUpperBoundFor[i] ← p.flowUpperBoundFor[i] + r.capacity
                                   - j.fixedDemandFor[h]
         p.flowUpperBoundFor[j] ← p.flowUpperBoundFor[j] + r.capacity
                                   - i.fixedDemandFor[h]
      end while
   end if
End Procedure
```

Figure B-3: Procedure for Removing the Capacity for the Fixed Demand

Appendix C

Glossary

Aircraft Route: Combination of an aircraft of a particular fleet type flying a particular route.

ARM: Aircraft Routing Model.

Block Time: Flying time plus taxi time.

Block Time Cost: Cost based on the block time flown, includes mostly the crew and fuel costs.

Candidate hub: A hub h is a candidate hub of an O-D commodity k if there is a valid pickup aircraft route from the origin gateway of k to hub h, and a valid delivery aircraft route from hub h to the destination gateway of k.

CHA: Commodity Hub Assignment Model.

Commodity Composite: A set of aircraft routes R_c that provide sufficient capacity to pick up or deliver a set of O-D commodities K_c for a pre-specified hub assignment, ψ_c^k, for each O-D commodity k, $k \in K_c$.

(Gateway-Hub) Delivery Demand: Volume consolidated from O-D commodities destined to the same gateway and assigned to the same hub.

(Package) Delivery Path: A sequence of flight legs from a hub to a destination gateway.

Delivery Route: A route originating at a hub and visiting one or two gateways.

Demand Composite: A set of aircraft routes that provide sufficient capacity for a set of pickup or delivery demands.

Destination Ground Center: A location where packages are sent before delivered to customers.

Destination Sort: Sort for packages delivered to a destination ground center.

Double-Leg Route Junction Gateways: A gateway visited by more than one double-leg routes.

Earliest Hub Arrival Time (EHAT): The earliest time a pickup aircraft can arrive at a hub.

Earliest Hub Departure Time (EHDT): The earliest time a delivery flight may depart from the hub.

Earliest Pickup Time (EPT): The earliest time a pickup aircraft can depart a gateway.

Earliest Pickup Time from Center (EPTC): The earliest time shipments can depart a ground center.

ESSND: Express Shipment Service Network Design.

Ferrying: Empty aircraft movement.

Fixed Commodity: A commodity that has only one candidate hub.

Fixed Cycle Cost: Cost incurred on each flight leg, typically includes the landing fees and other one-time charges.

Fixed Hub Assignment Problem: Express Shipment Service Network Design Problem with Fixed Hub Assignment.

Flexible Commodity: A commodity that has more than one candidate hub.

Gateway: An airport at which packages are transferred from ground vehicles and feeder aircraft to the airplanes that fly the pickup routes. Gateways are also the points at which packages are transferred from airplanes that fly the delivery routes to ground vehicles and feeder aircraft.

Gateway Composite: A set of aircraft routes R_g that provide sufficient capacity to pick up or deliver all the commodities at the set of gateways $\mathcal{N}_g$ visited by the routes.

Gateway Flow Pickup (or Delivery) Path: The amount of flow from a gateway on a pickup (or delivery) path.

GCF: Gateway Cover Flow Model.

Ground Center: A location where packages enter or depart the express shipment service network.

Hub: Airport at which packages are sorted. Hubs serves as the terminating location for pickup routes and the starting point for delivery routes.

Hub Pickup/Delivery Composite: A set of pickup (or demand) composites, if they satisfy the count constraints, the landing constraint for that hub (and other hubs), and the cover constraints for the pickup (or delivery) gateway-hub demands to (or from) that hub.

INS: Integrated NDA-SDA Formulation (with demand composite variables).

INS-H: Integrated NDA-SDA Formulation with Hub composite variables.

Integrated NDA-SDA Problem: The Integrated Next-Day and Second-Day **ESSND** Problem with fixed hub assignment.

Latest Delivery Time (LDT): The latest time a delivery aircraft may arrive at the gateway.

Latest Delivery Time to Center (LDTC): The latest time shipments may be delivered to a ground center.

Latest Hub Arrival Time (LHAT): The latest time a pickup flight can arrive at the hub and still allow sufficient sort time for the packages it transports.

Minimum Air Pickup (or Delivery) Volume: Total volume from (or to) a gateway subtracted by the maximum ground volume at that gateway.

Minimum Turn Time: The minimum time required for an aircraft of type f to stay on the ground for loading, unloading, and refueling between flight legs.

Maximum Pickup (or Delivery) Ground Volume: The maximum volume from (or to) a gateway that can be picked up (or delivered) by ground service.

Maximum Pickup (or Delivery) Ground Percentage: The maximum percentage of the pickup demand between a gateway-hub pair can be served by ground service, computed based on historical data.

MCNF: Multi-Commodity Network Flow Problem.

NDA: Next-Day Air.

NDP: The Network Design Problem.

Origin-Destination (O-D) Commodity/Origin-Destination Volume: Origin gateway-destination gateway demands.

Origin Ground Center: A location where packages collected from customers are first sent.

Origin sort: Sort for packages collected to a origin ground center.

(Gateway-Hub) Pickup Demand: Volume consolidated from O-D commodities originating from the same gateway and assigned to the same hub.

(Package) Pickup Path: A sequence of flight legs from an origin gateway to a hub.

Pickup Route: A route visiting one or two gateways and destined to a hub.

RF: Route and Flow Model.

RH: Route and Hub Assignment Model.

RMP: Restricted master problem.

Route: A sequence of connected flight legs.

SDA: Second-Day Air.

Small Demand Gateway/Small Demand Gateway Rule: If the total volume from (or to) a gateway is smaller than the capacity of the largest aircraft that can service the gateway, the gateway is categorized as a **small demand gateway** and all the associated volume is assigned to the major hub.

Size of a Gateway Composite: The number of gateways a gateway composite covers.

Subnetwork: a triplet ($\mathcal{N}_G$, H_G, R_G) including a subset of gateways, $\mathcal{N}_G$, a set of hubs, H_G, and the complete set of routes, R_G, visiting only the set of gateways $\mathcal{N}_G$.

Appendix D

Notation

Sets:

$\mathcal{A}$ set of pickup arcs.

$\mathcal{A}_c$ set of pickup arcs used by the aircraft routes included in commodity composite c.

$\mathcal{A}_g$ set of pickup arcs used by the aircraft routes included in gateway composite g.

$\mathcal{B}$ set of delivery arcs.

$\mathcal{C}$ set of demand (or commodity) composites.

$\mathcal{C}_P$ set of pickup demand (or commodity) composites.

$\mathcal{C}_D$ set of delivery demand (or commodity) composites.

$\mathcal{C}^T$ set of demand composites for the NDA ($T = N$) or SDA ($T = S$) network.

$\mathcal{C}_O^T$ set of pickup ($O = P$) or delivery ($O = D$) demand composites for the NDA ($T = N$) or SDA ($T = S$) network.

$\mathcal{C}^{T,O,h}$ $\begin{cases} \text{set of pickup } (O = P) \text{ or delivery } (O = D) \text{ demand composites} \\ \text{covering gateway-hub to } (O = P) \text{ or from } (O = D) \text{ hub } h \\ \text{in the NDA } (T = N) \text{ or SDA } (T = S) \text{ network.} \end{cases}$

F set of fleet types or facilities.

$\mathcal{G}_P$ set of pickup gateway composites.

$\mathcal{G}_D$ set of delivery gateway composites.

$\mathcal{G}$ set of gateway composites.

H set of hubs.

H_G set of hubs on sub-network G.

H^k set of candidate hubs for commodity k, $k \in K$.

$\mathcal{H}^T$ set of hub composites for the NDA $(T = N)$ or SDA $(T = S)$ network.

$\mathcal{H}_O^T$ set of pickup $(O = P)$ or delivery $(O = D)$ hub composites for the NDA $(T = N)$ or SDA $(T = S)$ network.

K set of O-D commodities.

K_c set of O-D commodities served by commodity composite c.

K_G set of O-D commodities on sub-network G.

K_l^P set of commodities to be picked up from gateway l.

K_l^D set of commodities to be delivered to gateway l.

$K_l^{P,h}$ set of commodities that originate at gateway l and can be sorted at hub h.

$K_l^{D,h}$ set of commodities that are destined to gateway l and can be sorted at hub h.

$\mathcal{N}$ set of gateways.

$\mathcal{N}_c$ set of gateways visited by the aircraft routes included in commodity composite c.

$\mathcal{N}_g$ set of gateways visited by the aircraft routes included in gateway composite g.

$\mathcal{N}_G$ set of gateways on sub-network G.

$\mathcal{P}_c$ sets of paths based on the aircraft routes included in commodity composite c.

$\mathcal{P}_g$ sets of paths based on the aircraft routes included in gateway composite g.

$\mathcal{P}_G$ sets of paths on sub-network G.

$\mathcal{P}^P$ sets of pickup paths.

$\mathcal{P}^D$ sets of delivery paths.

$\mathcal{P}_k^P$ set of pickup paths for an O-D commodity k, $k \in K$.

$\mathcal{P}_k^D$ set of delivery paths for an O-D commodity k, $k \in K$.

$\mathcal{P}_k^{P,h}$ set of pickup paths from the origin of commodity k to hub h, $k \in K$, $h \in H^k$.

$\mathcal{P}_k^{D,h}$ set of delivery paths from hub h to the destination of commodity k, $k \in K$, $h \in H^k$.

$\mathcal{P}_l^{P,h}$ set of pickup paths from gateway l to hub h, $l \in \mathcal{N}$, $h \in H$.

$\mathcal{P}_l^{D,h}$ set of delivery paths from hub h to gateway l, $l \in \mathcal{N}$, $h \in H$.

R sets of aircraft routes.

R_c sets of aircraft routes included in commodity composite c.

R_g sets of aircraft routes included in gateway composite g.

R_G sets of aircraft routes on sub-network G.

R_P sets of pickup aircraft routes.

R_D sets of delivery aircraft routes.

R^f set of routes that can be flown by fleet type f, $f \in F$.

R_P^f set of pickup routes that can be flown by fleet type f, $f \in F$.

R_D^f set of delivery routes that can be flown by fleet type f, $f \in F$.

$[S, T]$	set of arcs from $S \subset \mathcal{N}$ to $T \subset \mathcal{N}$, $S \cup T = \mathcal{N}$ and $S \cap T = \emptyset$.

Data:

a_h	number of aircraft parking spots at hub $h \in H$.
a_h^T	number of aircraft parking spots at hub h for NDA $(T = N)$ or SDA $(T = S)$ network.
b_k	volume of O-D commodity $k \in K$.
b_P^{ih}	the pickup demand between gateway i and hub h.
b_D^{ih}	the delivery demand between gateway i and hub h.
$b_{T,O}^{ih}$	pickup $(O = P)$ or delivery $(O = D)$ demand between gateway i and hub h for NDA $(T = N)$ or SDA $(T = S)$ network.
c_{ij}^k	unit flow cost of commodity k flown on arc (i, j).
d_{ij}^f	cost of installing one unit of facility type f on arc (i, j).
d_r^f	cost of flying route r with fleet type f.
d_r	cost of flying aircraft route r.
d_c	cost of demand (or commodity) composite c, $d_c = \sum_{r \in R} \gamma_c^r d_r$.
d_g	cost of gateway composite g, $d_g = \sum_{r \in R} \gamma_g^r d_r$.
d_Θ	cost of hub composite Θ, $d_\Theta = \sum_{c \in \Theta} d_c$.
d_{ij}^f	ferrying cost for an aircraft of type f ferried from gateway i to j.
e_h	sorting capacity of hub $h \in H$.
n_f	number of available aircraft of type $f \in F$.
n_f^T	number of aircraft of type f available for NDA $(T = N)$ or SDA $(T = S)$ network.
u_{ij}^f	capacity provided by one unit of facility f on arc (i, j).

u_r capacity of aircraft route r, $r \in R$.

u_r^f capacity of an aircraft of type $f \in F$ assigned to route $r \in R$.

$D_{S,T}$ total demand originating in set $S \subset \mathcal{N}$ and destined for $T \subset \mathcal{N}$.

Indicators:

$\alpha_g^{l,P}$ 1 if composite variable g picks up all the volume from gateway l, and 0 otherwise.

$\alpha_g^{l,D}$ 1 if composite variable g delivers all the volume to gateway l, and 0 otherwise.

$\beta_k^{l,P}$ 1 if gateway l is the origin of commodity k, and 0 otherwise.

$\beta_k^{l,D}$ 1 if gateway l is the destination of commodity k, and 0 otherwise.

γ_c^r number of aircraft routes r in demand (or commodity) composite c.

γ_g^r number of aircraft routes r in gateway composite g.

γ_Θ^f number of aircraft of type f in hub composite Θ.

γ_c^f number of aircraft of fleet type f in composite c.

$\gamma_c^f(\bar{i})$ number of aircraft of fleet type f originating at gateway i (or hub h) in demand composite c.

$\gamma_c^f(\underline{i})$ number of aircraft of fleet type f destined to gateway i (or hub h) in demand composite c.

$\gamma_\Theta^f(\bar{i})$ number of aircraft of type f originating at gateway (hub) i in hub composite Θ.

$\gamma_\Theta^f(\underline{i})$ number of aircraft of type f destined to gateway (hub) i in hub composite Θ.

δ_p^{ij} 1 if arc (i,j) is included in path p, and 0 otherwise.

δ_r^{ij} 1 if arc (i,j) is included in route r, and 0 otherwise.

δ_p^l 1 if path p visits gateway l , and 0 otherwise.

δ_r^f 1 if fleet type f is assigned to route r, and 0 otherwise.

$\delta_c^{i,h,P}$ 1 if demand composite c covers the pickup demand between gateway i and hub h, and 0 otherwise.

$\delta_c^{i,h,D}$ 1 if demand composite c covers the delivery demand between gateway i and hub h, and 0 otherwise.

$$\delta_{T,O,c}^{ih} = \begin{cases} 1 & \text{if demand composite } c \text{ covers NDA } (T = N) \text{ or SDA } (T = S) \text{ pickup} \\ & (O = P) \text{ or delivery } (O = D) \text{ demand between gateway } i \text{ and hub } h \text{, and} \\ 0 & \text{otherwise.} \end{cases}$$

$$\delta_{T,O,\Theta}^{ih} = \begin{cases} 1 & \text{if hub composite } \Theta \text{ covers NDA } (T = N) \text{ or SDA } (T = S) \text{ pickup} \\ & (O = P) \text{ or delivery } (O = D) \text{ demand between gateway } i \text{ and hub } h \text{, and} \\ 0 & \text{otherwise.} \end{cases}$$

$\varphi_p^{l,h}$ 1 if path p connects gateway l with hub h, and 0 otherwise.

$\psi_c^{k,h}$ percentage of commodity k assigned to hub h in commodity composite c.

$O(k)$ origin gateway of O-D commodity k.

$D(k)$ destination gateway of O-D commodity k.

$O(r)$ origin of route r.

$D(r)$ destination of route r.

Decision Variables:

v_c equals 1 if demand (or commodity) composite c, $c \in \mathcal{C}$, is selected, and 0 otherwise.

v_Θ equals 1 if hub composite Θ is selected, and 0 otherwise.

w_g equals 1 if gateway composite g is selected, and 0 otherwise.

y_{ij}^f number of facilities of type f installed on arc (i, j).

y_r^f number of aircraft of fleet type $f \in F$ assigned to route $r \in R^f$.

y_r number of aircraft routes r, $r \in R$, selected.

x_p^k amount of flow of commodity k on arc (i,j).

x_{ij}^k flow of commodity $k \in K$ on pickup (or delivery) path $p \in \mathcal{P}_k^P$ (or $\mathcal{P}_k^D$).

$x_{l,p}$ package flow from (or to) gateway l, $l \in \mathcal{N}$, on pickup (or delivery) path p, $p \in \mathcal{P}^P$ (or $\mathcal{P}^D$).

z_h^k percentage of commodity k, $k \in K$, assigned to hub h, $h \in H^k$.

μ^k equals 1 if commodity k is covered in a sub-problem, and 0 otherwise.

$\varpi_{f,i}^{T,O}$ number of aircraft of type f on the ground at gateway (hub) i during NDA ($T = N$) or SDA ($T = S$) pickup ($O = P$) or delivery ($O = D$) operation. $\varpi_{f,i}^{T,P} = \varpi_{f,i}^{T,D}$, if $i \notin H$.

$\phi_{ij}^{T,f}$ number of aircraft of type f ferried from gateway (hub) i to j after the NDA ($T = N$) or SDA ($T = S$) operation.

$\psi^{k,h}$ percentage of commodity k, $k \in K$, assigned to hub h, $h \in H^k$, in a sub-problem.

Dual Prices:

π_r sum of the products of the dual price of a constraint and the corresponding coefficient of route r in that constraint, for each constraint containing route r.

π_G sum of the dual prices associated with gateway pickup (or delivery) cover constraints (4.39) (or (4.40)) corresponding to set of gateways on sub-network G.

$\pi_h^{k,P}$ dual prices of pickup cover constraints (4.27) specified for commodity k and hub h.

$\pi_h^{k,D}$ dual prices of delivery cover constraints (4.28) specified for commodity k and hub h.

Matrices:

$\mathbf{A}$	constraints matrix for demand composite variables in landing constraints (3.6).
$\mathring{\mathbf{A}}$	constraints matrix for route variables in landing constraints (4.9).
$\hat{\mathbf{A}}$	constraints matrix for commodity composite variables in landing constraints (4.25).
$\bar{\mathbf{A}}$	constraints matrix for gateway composite variables in landing constraints (4.37).
$\mathbf{A}^{T,O,h}$	constraints matrix for variables in set $\mathcal{C}^{T,O,h}$ in landing constraints (3.6).
$\mathbf{B}_1$	constraints matrix for demand composite variables in bounday balance constraints (3.2).
$\mathbf{B}_2$	constraints matrix for demand composite variables in bounday balance constraints (3.3).
$\mathbf{B}_1^{T,O,h}$	constraints matrix for variables in set $\mathcal{C}^{T,O,h}$ in bounday balance constraints (3.2).
$\mathbf{B}_2^{T,O,h}$	constraints matrix for variables in set $\mathcal{C}^{T,O,h}$ in bounday balance constraints (3.3).
$\mathring{\mathbf{B}}$	constraints matrix for route variables in gateway balance constraints (4.7).
$\hat{\mathbf{B}}$	constraints matrix for commodity composite variables in gateway balance constraints (4.23).
$\bar{\mathbf{B}}$	constraints matrix for gateway composite variables in gateway balance constraints (4.35).
$\mathbf{C}$	constraints matrix for demand composite variables in cover constraints (3.7).
$\mathbf{C}^{T,O,h}$	constraints matrix for variables in set $\mathcal{C}^{T,O,h}$ in cover constraints (3.7).
$\mathring{\mathbf{H}}$	constraints matrix for route variables in hub balance constraints (4.8).
$\hat{\mathbf{H}}$	constraints matrix for commodity composite variables in hub balance constraints (4.24).

$\mathbf{H}$ constraints matrix for demand composite variables in hub balance constraints (3.4).

$\bar{\mathbf{H}}$ constraints matrix for gateway composite variables in hub balance constraints (4.36).

$\mathbf{H}^{T,O,h}$ constraints matrix for variables in set $\mathcal{C}^{T,O,h}$ in hub balance constraints (3.4).

$\mathbf{N}$ constraints matrix for demand composite variables in count constraints (3.5).

$\mathring{\mathbf{N}}$ constraints matrix for route variables in count constraints (4.10).

$\hat{\mathbf{N}}$ constraints matrix for commodity composite variables in count constraints (4.26).

$\bar{\mathbf{N}}$ constraints matrix for gateway composite variables in count constraints (4.38).

$\mathbf{N}^{T,O,h}$ constraints matrix for variables in set $\mathcal{C}^{T,O,h}$ in count constraints (3.5).

Vectors

$\mathbf{a}$ parking spot vector for landing constraints (4.9), (4.25) or (4.37).

$\mathbf{d}$ objective coefficient vector for demand composite variables.

$\mathbf{a}_T$ parking spot vector for landing constraints (3.6) in NDA ($T = N$) or SDA ($T = S$) network.

$\mathbf{n}$ available aircraft number vector for count constraints (4.10), (4.26) or (4.38).

$\mathbf{n}_T$ available aircraft number vector for count constraints (3.5) in NDA ($T = N$) or SDA ($T = S$) network.

$\mathbf{I}^{T,O,h}$ right-hand-side vetor for pickup ($O = P$) or delivery ($O = D$) gateway-hub demands at hub h, $h \in H$, in NDA ($T = N$) or SDA ($T = S$) network.

$\mathbf{v}^{T,O,h}$ decision variable vector for demand composites in $\mathcal{C}^{T,O,h}$.

$\boldsymbol{\pi}^{\mathbf{A}}$ dual price vetor for landing constraints (3.6) or (3.14).

$\boldsymbol{\pi}^{\mathbf{B}_1}$ dual price vetor for bounday balance constraints (3.2) or (3.10).

$\boldsymbol{\pi}^{\mathbf{B}_2}$ dual price vetor for bounday balance constraints (3.3) or (3.11).

$\pi^{\mathbf{H}}$ dual price vetor for hub balance constraints (3.4) or (3.12).

$\pi^{\mathbf{C}}$ dual price vetor for cover constraints (3.7) or (3.15).

$\pi^{\mathbf{N}}$ dual price vetor for count constraints (3.5) or (3.13).

Appendix E

Formulations

Aircraft Routing Model (ARM)

$$\min \sum_{c \in \mathcal{C}_P \cup \mathcal{C}_D} d_c v_c$$

subject to

$$\sum_{c \in \mathcal{C}_P} \gamma_c^f(\bar{i}) v_c - \sum_{c \in \mathcal{C}_D} \gamma_c^f(\underline{i}) v_c = 0 \qquad i \in \mathcal{N},\ f \in F$$

$$\sum_{c \in \mathcal{C}_P} \gamma_c^f(\underline{h}) v_c - \sum_{c \in \mathcal{C}_D} \gamma_c^f(\overline{h}) v_c = 0 \qquad h \in H,\ f \in F$$

$$\sum_{C \in \mathcal{C}_P} \gamma_c^f v_c \leq n_f \qquad f \in F$$

$$\sum_{f \in F} \sum_{v \in \mathcal{C}_P} \gamma_c^f(\underline{h}) v_c \leq a_h \quad h \in H$$

$$\sum_{c \in \mathcal{C}_P} \delta_c^{i,h,P} v_c = 1 \quad (i,h) : b_P^{ih} > 0,\ i \in \mathcal{N},\ h \in H$$

$$\sum_{c \in \mathcal{C}_D} \delta_c^{i,h,D} v_c = 1 \quad (i,h) : b_D^{ih} > 0,\ i \in \mathcal{N},\ h \in H$$

$$v_c \in \{0,1\},\ c \in \mathcal{C}_P \cup \mathcal{C}_D$$

Commodity Hub Assignment Model (CHA)

$$\min \sum_{c \in \mathcal{C}} d_c v_c$$

subject to

$$\hat{\mathbf{B}}\ \mathbf{v} = 0$$

$$\hat{\mathbf{H}}\ \mathbf{v} = 0$$

$$\hat{\mathbf{A}}\ \mathbf{v} \leq \mathbf{a}$$

$$\hat{\mathbf{N}}\ \mathbf{v} \leq \mathbf{n}$$

$$\sum_{c \in \mathcal{C}_P} \psi_c^{k,h} v_c - z_h^k \geq 0,\ \ k \in K,\ \ h \in H$$

$$\sum_{c \in \mathcal{C}_D} \psi_c^{k,h} v_c - z_h^k \geq 0,\ \ k \in K,\ \ h \in H$$

$$\sum_{h \in H^k} z_h^k = 1, \quad k \in K$$

$$\sum_{k \in K} b_k z_h^k \leq e_h, \quad h \in H$$

$$v_c \in \{0,1\}, c \in \mathcal{C}_P \cup \mathcal{C}_D,\ \ z_h^k \geq 0,\ \ k \in K,\ \ h \in H$$

Gateway Cover and Flow Model (GCF)

$$\min \sum_{g \in \mathcal{G}} d_g w_g$$

subject to

$$\bar{\mathbf{B}}\ \mathbf{w} = 0$$

$$\bar{\mathbf{H}}\ \mathbf{w} = 0$$

$$\bar{\mathbf{A}}\ \mathbf{w} \leq \mathbf{a}$$

$$\bar{\mathbf{N}}\ \mathbf{w} \leq \mathbf{n}$$

$$\sum_{g \in \mathcal{G}_P} \alpha_g^{l,P} w_g = 1, \qquad l \in \mathcal{N}$$

$$\sum_{g \in \mathcal{G}_D} \alpha_g^{l,D} w_g = 1, \qquad l \in \mathcal{N}$$

$$\sum_{g \in \mathcal{G}_P} \sum_{r \in R} \gamma_g^r \delta_r^{ij} u_r w_g - \sum_{p \in \mathcal{P}^P} \sum_{l \in \mathcal{N}} \delta_p^{ij} x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{A}$$

$$\sum_{g \in \mathcal{G}_D} \sum_{r \in R} \gamma_g^r \delta_r^{ij} u_r w_g - \sum_{p \in \mathcal{P}^D} \sum_{l \in \mathcal{N}} \delta_p^{ij} x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{B}$$

$$\sum_{p \in \mathcal{P}^P} \varphi_p^{l,h} x_{l,p} - \sum_{k \in K} \beta_k^{l,P} b_k z_h^k = 0, \qquad l \in \mathcal{N},\ h \in H$$

$$\sum_{p \in \mathcal{P}^D} \varphi_p^{l,h} x_{l,p} - \sum_{k \in K} \beta_k^{l,D} b_k z_h^k = 0, \qquad l \in \mathcal{N},\ h \in H$$

$$\sum_{h \in H^k} z_h^k = 1, \quad k \in K$$

$$\sum_{k \in K} b_k z_h^k \leq e_h, \quad h \in H$$

$$w_g \in \{0,1\},\ g \in \mathcal{G},\ z_h^k \geq 0, k \in K,\ h \in H,\ x_{l,p} \geq 0,\ l \in \mathcal{N},\ p \in \mathcal{P}^P \cup \mathcal{P}^D$$

Integrated NDA-SDA Model with Demand Composite Variable (INS)

$$\min \sum_{T=\{N,S\}} \sum_{c \in \mathcal{C}^T} d_c v_c + \sum_{T=\{N,S\}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} d_{ij}^f \phi_{ij}^{T,f}$$

subject to

$$\sum_{c \in \mathcal{C}_D^S} \gamma_c^f(\underline{i}) v_c - \sum_{c \in \mathcal{C}_P^N} \gamma_c^f(\bar{i}) v_c - \varpi_{f,i}^{N,P} + \varpi_{f,i}^{S,D} + \sum_{j \in \mathcal{N}, j \neq i} \phi_{ji}^{S,f} - \sum_{j \in \mathcal{N}, j \neq i} \phi_{ij}^{S,f} = 0, \; i \in \mathcal{N}, \; f \in F$$

$$\sum_{c \in \mathcal{C}_D^N} \gamma_c^f(\underline{i}) v_c - \sum_{c \in \mathcal{C}_P^S} \gamma_c^f(\bar{i}) v_c + \varpi_{f,i}^{N,D} - \varpi_{f,i}^{S,P} + \sum_{j \in \mathcal{N}, j \neq i} \phi_{ji}^{N,f} - \sum_{j \in \mathcal{N}, j \neq i} \phi_{ij}^{N,f} = 0, \; i \in \mathcal{N}, \; f \in F$$

$$\sum_{c \in \mathcal{C}_P^T} \gamma_c^f(\underline{h}) v_c + \varpi_{f,h}^{T,P} - \sum_{c \in \mathcal{C}_D^T} \gamma_c^f(\overline{h}) v_c - \varpi_{f,h}^{T,D} = 0, \qquad h \in H, \; f \in F, \; T = \{N,S\}$$

$$\sum_{c \in \mathcal{C}_P^T} \gamma_c^f v_c \leq n_f, \qquad f \in F, \; T = \{N,S\}$$

$$\sum_{f \in F} \sum_{c \in \mathcal{C}_P^T} \gamma_c^f(\underline{h}) v_c \leq a_h, \quad h \in H, \; T = \{N,S\}$$

$$\sum_{c \in \mathcal{C}_O^T} \delta_{T,O,c}^{ih} v_c \geq 1, \quad (i,h) : b_{T,O}^{ih} > 0, \; T = \{N,S\}, \; O = \{P,D\}, \; i \in \mathcal{N}, \; h \in H$$

$$v_c \in \{0,1\} \text{ for all } c \in \mathcal{C}^N \cup \mathcal{C}^S, \; \varpi_{f,i}^{T,O} \in \mathbb{Z}_+ \text{ for } T = \{N,S\}, \; O = \{P,D\}, \; i \in \mathcal{N}$$

$$\phi_{ij}^{N,f}, \phi_{ij}^{S,f} \in \mathbb{Z}_+ \;\text{ for } i,j \in \mathcal{N}, i \neq j, \; f \in F, \; T = \{N,S\}$$

Integrated NDA-SDA Model with Hub Composite Variable (INS-H)

$$\min \sum_{T=\{N,S\}} \sum_{\Theta \in \mathcal{H}^T} d_\Theta v_\Theta + \sum_{T=\{N,S\}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} d_{ij}^f \phi_{ij}^{T,f}$$

subject to

$$\sum_{\Theta \in \mathcal{H}_D^S} \gamma_\Theta^f(\underline{i}) v_\Theta - \sum_{\Theta \in \mathcal{H}_P^N} \gamma_\Theta^f(\overline{i}) v_\Theta - \varpi_{f,i}^{N,P} + \varpi_{f,i}^{S,D} + \sum_{j \in \mathcal{N}, j \neq i} \phi_{ji}^{S,f} - \sum_{j \in \mathcal{N}, j \neq i} \phi_{ij}^{S,f} = 0, \; i \in \mathcal{N}, \; f \in F$$

$$\sum_{\Theta \in \mathcal{H}_D^N} \gamma_\Theta^f(\underline{i}) v_\Theta - \sum_{\Theta \in \mathcal{H}_P^S} \gamma_\Theta^f(\overline{i}) v_\Theta + \varpi_{f,i}^{N,D} - \varpi_{f,i}^{S,P} + \sum_{j \in \mathcal{N}, j \neq i} \phi_{ji}^{N,f} - \sum_{j \in \mathcal{N}, j \neq i} \phi_{ij}^{N,f} = 0, \; i \in \mathcal{N}, \; f \in F$$

$$\sum_{\Theta \in \mathcal{H}_P^T} \gamma_\Theta^f(\underline{h}) v_\Theta + \varpi_{f,h}^{T,P} - \sum_{\Theta \in \mathcal{H}_D^T} \gamma_\Theta^f(\overline{h}) v_\Theta - \varpi_{f,h}^{T,D} = 0, \qquad h \in H, \; f \in F, \; T = \{N, S\}$$

$$\sum_{\Theta \in \mathcal{H}_P^T} \gamma_\Theta^f v_\Theta \leq n_f, \qquad f \in F, \; T = \{N, S\}$$

$$\sum_{f \in F} \sum_{c \in \mathcal{H}_P^T} \gamma_\Theta^f(\underline{h}) v_c \leq a_h, \quad h \in H, \; T = \{N, S\}$$

$$\sum_{\Theta \in \mathcal{H}_O^T} \delta_{T,O,\Theta}^{ih} v_\Theta \geq 1, \quad (i,h) : b_{T,O}^{ih} > 0, \; T = \{N, S\}, \; O = \{P, D\}, \; i \in \mathcal{N}, \; h \in H$$

$$v_\Theta \in \{0,1\} \text{ for all } \Theta \in \mathcal{H}^N \cup \mathcal{H}^S, \; \varpi_{f,i}^{T,O} \in \mathbb{Z}_+ \text{ for } T = \{N, S\}, \; O = \{P, D\}, \; i \in \mathcal{N}$$

$$\phi_{ij}^{N,f}, \phi_{ij}^{S,f} \in \mathbb{Z}_+ \;\text{ for } i, j \in \mathcal{N}, i \neq j, \; f \in F, \; T = \{N, S\}$$

Route and Flow Model (RF)

$$\min \sum_{f \in F} \sum_{r \in R_P^f \cup R_D^f} d_r^f y_r^f$$

subject to:

$$\sum_{r \in R_P^f : O(r)=i} y_r^f - \sum_{r \in R_D^f : D(r)=i} y_r^f = 0 \quad i \in \mathcal{N},\ f \in F$$

$$\sum_{r \in R_D^f : O(r)=h} y_r^f - \sum_{r \in R_P^f : D(r)=h} y_r^f = 0 \quad h \in H,\ f \in F$$

$$\sum_{r \in R_P^f} y_r^f \leq n_f \quad f \in F$$

$$\sum_{f \in F} \sum_{r \in R_P^f : D(r)=h} y_r^f \leq a_h \quad h \in H$$

$$\sum_{f \in F} \sum_{r \in R_P^f} \delta_r^{ij} u_r^f y_r^f - \sum_{k \in K} \sum_{p \in \mathcal{P}_k^P} \delta_p^{ij} x_p^k \geq 0 \qquad (i,j) \in \mathcal{A}$$

$$\sum_{f \in F} \sum_{r \in R_D^f} \delta_r^{ij} u_r^f y_r^f - \sum_{k \in K} \sum_{p \in \mathcal{P}_k^D} \delta_p^{ij} x_p^k \geq 0 \qquad (i,j) \in \mathcal{B}$$

$$\sum_{p \in \mathcal{P}_k^{P,h}} x_p^k - \sum_{p \in \mathcal{P}_k^{D,h}} x_p^k = 0 \qquad k \in K,\ h \in H^k$$

$$\sum_{h \in H^k} \sum_{p \in \mathcal{P}_k^{P,h}} x_p^k = b_k \qquad k \in K$$

$$\sum_{k \in K} \sum_{p \in \mathcal{P}_k^{P,h}} x_p^k \leq e_h, \qquad h \in H$$

$$x_p^k \geq 0 \quad k \in K,\ p \in \mathcal{P}_k^P \cup \mathcal{P}_k^D$$

$$y_r^f \in \mathbb{Z}_+ \quad r \in R^f,\ f \in F$$

Route and Hub Assignment Model (RH)

$$\min \sum_{r \in R} d_r y_r$$

subject to

$$\mathring{\mathbf{B}}\ \mathbf{y} = 0$$

$$\mathring{\mathbf{H}}\ \mathbf{y} = 0$$

$$\mathring{\mathbf{A}}\ \mathbf{y} \leq \mathbf{a}$$

$$\mathring{\mathbf{N}}\ \mathbf{y} \leq \mathbf{n}$$

$$\sum_{r \in R_P} \delta_r^{ij} u_r y_r - \sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^P} \delta_p^{ij} \delta_p^l x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{A}$$

$$\sum_{r \in R_D} \delta_r^{ij} u_r y_r - \sum_{l \in \mathcal{N}} \sum_{p \in \mathcal{P}^D} \delta_p^{ij} \delta_p^l x_{l,p} \geq 0, \qquad (i,j) \in \mathcal{B}$$

$$\sum_{p \in \mathcal{P}^P} \varphi_p^{l,h} x_{l,p} - \sum_{k \in K} \beta_k^{l,P} b_k z_h^k = 0, \qquad l \in \mathcal{N},\ h \in H$$

$$\sum_{p \in \mathcal{P}^D} \varphi_p^{l,h} x_{l,p} - \sum_{k \in K} \beta_k^{l,D} b_k z_h^k = 0, \qquad l \in \mathcal{N},\ h \in H$$

$$\sum_{h \in H^k} z_h^k = 1, \quad k \in K$$

$$\sum_{k \in K} b_k z_h^k \leq s_h, \quad h \in H$$

$$y_r \in \mathbb{Z}_+,\ r \in R,\ z_k^h \geq 0,\ k \in K,\ h \in H,\ x_{l,p} \geq 0,\ l \in \mathcal{N},\ p \in \mathcal{P}^P \cup \mathcal{P}^D$$

Bibliography

[1] Agrawal, A., P. Klein, R. Ravi. 1995. When trees collide: an approximation algorithm for the generalized steiner problem on networks. *SIAM J. Comput.* **24**(3) 440-456.

[2] Ahuja, R. K., T. L. Magnanti, J. B. Orlin. 1993. *Network Flows: Theory, Algorithms and Applications.* Prentice Hall.

[3] Alevras, D., M. Grötschel, R. Wessaly. 1998. Cost-efficient network synthesis from leased lines. *Ann. Oper. Res.* **76** 1-20.

[4] Armacost, A. 2000. *Composite Variable Formulations for Express Shipment Service Network Design*, PhD dissertation, Operations Research Center, Massachusetts Institute of Technology.

[5] Armacost, A., C. Barnhart, K. Ware. 2002. Composite variable formulations for express shipment service network design. *Transportation Sci.* **36**(1) 1-20.

[6] Atamtürk A. 2002. On capacitated network design cut-set polyhedra. *Math. Programming* **92**(3) 425-437.

[7] Aykin T. 1994. Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *Eur. J. Oper. Res.* **79** 501-523.

[8] Balakrishnan, A., T. L. Magnanti, P. Mirchandani. 1994. A dual-based algorithm for multi-level network design. *Management Sci.* **40**(5) 567-581.

[9] Balakrishnan, A., T. L. Magnanti, P. Mirchandani. 1994. Modeling and heuristic worst-case performance analysis of the two-level network design problem. *Magagment Sci.* **40**(7) 846-867.

[10] Balakrishnan, A., T. L. Magnanti, R. T. Wong. 1995. A Decomposition algorithm for local access telecommunications network expansion planning. *Oper. Res.* **43**(1) 58-76.

[11] Balakrishnan, A., T. L. Magnanti, P. Mirchandani. 1996. Heuristics, LPs, and trees on trees: network design analysis. *Oper. Res.* **44**(3) 478-496.

[12] Balakrishnan, A., T. L. Magnanti, P. Mirchandani. 1998. Designing hierarchical survivable networks. *Oper. Res.* **46**(1) 116-136.

[13] Balakrishnan A., T. L. Magnanti, P. Mirchandani. 2004. Connectivity-splitting models for survivable network design. *Networks* **43**(1) 10-27.

[14] Barahona, F. 1996. Network design using cut inequalities. *SIAM J. Optim.* **6**(3) 823-837.

[15] Barnhart, C. 1993. Dual-ascent methods for large-scale multicommodity flow problems. *Nav. Res. Log.* **40**(3) 305-324.

[16] Barnhart, C., Y. Sheffi. 1993. A network based primal-dual heuristic for the solution of multicommodity network flow problems. *Transportation Sci.* **27**(2) 102-117.

[17] Barnhart, C., C. A. Hane, E. L. Johson, G. Sigismondi. 1995. A column generation and partitioning approach for multicommodity flow problems. *Telcommun. Syst.* **3**(3-4) 239-258.

[18] Barnhart, C., R. R. Schneur. 1996. Air network design for express shipment service. *Oper. Res.* **44**(6) 852-863.

[19] Barnhart, C., C. A. Hane, P. H. Vance. 1997. Integer multicommodity flow problems. *Lect. Notes. Econ. Math.* **450** 17-31.

[20] Barnhart, C. , E. L. Johnson, G. L. Nemhauser, M. W. F. Savelsbergh, P. H. Vance. 1998. Branch and price: column generation for solving huge integer programs. *Oper. Res.* **46**(3) 316-329.

[21] Barnhart, C., R. G. Shenoi. 1998. An approximation model and solution approach for the long-haul crew pairing problem. *Transportation Sci.* **32**(3) 221-231.

[22] Barnhart, C., C. A. Hane, P. H. Vance. 2000. Using branch-and-price-and-cut to solve origin-destination integer multcommodity flow problems. *Oper. Res.* **48**(2) 318-326.

[23] Barnhart, C., H. Jin, P. H. Vance. 2000. Railroad blocking: a network design application. *Oper. Res.* **48**(4) 603–614.

[24] Benders, J. F. 1962. Partitioning procedure for solving mixed variable programming problems. *Numer. Mathe.* **4** 238-252.

[25] Bertsimas, D., C. P. Teo. 1998. From valid inequalities to heuristics: a unified view of primal-dual approximation algorithms in covering problems. *Oper. Res.* **46**(4) 503-514.

[26] Bienstock, D., O. Günlük. 1995. Computational experience with a difficult mixed-integer multicommodity flow problem. *Math. Programming* **68**(2) 213-237.

[27] Bienstock, D., O. Günlük. 1996. Capacitated network design - polyhedral structure and computation. *INFORMS J. Comput.* 8(3) 243-259.

[28] Bienstock, D., S. Chopra, O. Günlük, C. Y. Tsai. 1998. Minimum cost capacity installation for multicommodity networks. *Math. Programming* **81**(2) 177-199.

[29] Chopra, S., L. Gilboa, S. T. Sastry. 1998. Source sink flows with capactiy installation in batches. *Discrete Appl. Math.* **85**(3) 165-192.

[30] Chopra S., C. Y. Tsai. 2002. A branch-and-cut approach for minimum cost Multi-level Network design. *Discrete Math.* **242**(1-3) 65-92.

[31] Cook, W. J., W. H. Cunningham, W. R. Pulleyblank, A. Schrijver. 1998. *Combinatorial Optimization.* New York: Wiley.

[32] ILOG CPLEX 6.5 User's Manual, 1999.

[33] ILOG CPLEX 8.0 User's Manual, 2002.

[34] Crainic, T. G. 2000. Service network design in freight transportation. *Eur. J. Oper. Res.* **122**(2) 272-288.

[35] Crowder, H., E. L. Johnson, M. W. Padberg. 1983. Solving large-scale zero-one linear programming problems. *Oper. Res.* **31**(5) 803-834.

[36] Dantzig, G. B., P. Wolfe. 1960. Decomposition principle for linear programs. *Oper. Res.* **8** 101-111.

[37] Farvolden, J. M., W. B. Powell. 1994. Subgradient methods for the service network design problem. *Transportation Sci.* **28**(3) 256-272.

[38] Farvolden, J. M., W. B. Powell, I. J. Lustig. 1993. A primal partitioning solution for the arc-chain formulation of a multicommodity network flow problem. *Oper. Res.* **41**(4) 669-693.

[39] FedEx Corporation 1998 Annual Report.

[40] FedEx Corporation 2002 Annual Report.

[41] Gabow, H. N., M. X. Goemans, D. P. Williamson. 1998. An efficient approximation algorithm for the survivable network design problem. *Math. Programming.* **82**(1-2) 13-40.

[42] Gabrel, V., A. Knippel, M. Minoux. 1999. Exact solution of multicommodity network optimiztation problems with general step cost functions. *Oper. Res. Lett.* **25**(1) 15-23.

[43] Goemans, M. X., D. J. Bertsimas. 1993. Survivable networks, linear programming relaxations and the parsimonious property. *Math. Programming.* **60**(2) 145-166.

[44] Goemans, M. X., D. P. Williamson. 1995. A general approximation technique for constrained forest problems. *SIAM J. Comput.* **24**(2) 296-317.

[45] Gorman, M. F. 1998. An application of genetic and tabu searches to freight railroad operating plan problem. *Ann. Oper. Res.* **78** 51-69.

[46] Gorman, M. F. 1998. Santa Fe railway uses an operating-plan model to improve its service design. *Interfaces* **28**(4) 1-12.

[47] Göthe-Lundgren, M., T. Larsson. 1994. A set covering reformulation of the pure fixed charge transportation problem. *Discrete Appl. Math.* **48**(3) 245-259.

[48] Grünert, T., H. J. Sebastian. 2000. Planning models for long-haul operations of postal and express shipment companies. *Eur. J. Oper. Res.* **122**(2) 289-309.

[49] Guignard, M., S. Kim. 1987. Lagrangean decomposition - a model yielding stronger Lagrangean bounds. *Math. Programming* **39**(2) 215-228.

[50] Günlük, O. 1999. A branch-and-cut algorithm for capacitated network design problems. *Math. Programming* **86**(1) 17-39.

[51] Hane, C. A., C. Barnhart, E. L. Johnson, R. E. Marsten, G. L. Nemhauser, G. Sigismondi. 1995. The fleet assignment problem: solving a large-scale integer program. *Math. Programming* **70**(2) 211-232.

[52] Hochbaum, D. S., J. S. Naor. 1996. Approximation algorithms for network design problems on bounded subsets. *J. Algorithms* **21**(2) 403-414.

[53] Holmberg, J., J. Hellstrand. 1998. Solving the uncapacitated network design problem by a Lagrangian heuristic and branch-and-bound. *Oper. Res.* **46**(2) 247-259.

[54] Holmberg K., D. Yuan. 2004. Optimization of internet protocol network design and routing. *Networks* **43**(1) 39-53.

[55] Jain, K. 1998. A factor 2 approximation algorithm for the generalized steiner network problem. *Proc. 39th Annual Sympos. on Foundations of Computer Science (FOCS '98).*

[56] Johnson, E., G. L. Neumhauser, M. Bazaraa, J. Sokol, T. K. Choon, T. C. Plaw, H. H. Chuen. 2000. *Air Cargo White Paper*, National University of Singapore.

[57] Jones, K. L., I. J. Lustig, J. M. Farvolden, W. B. Powell. 1993. Multicommodity network flows: the impact of formulation on decomposition. *Math. Program. Stud.* **62**(1) 95-117.

[58] Karger, D. R. 1999. Random sampling in cut, flow, and network design problems. *Math. Oper. Res.* **24**(2) 383-413.

[59] Kim, D. 1997. *Large Scale Transportation Service Network Design: Models, Algorithms and Applications.* PhD dissertation, Massachusetts Institute of Technology.

[60] Kim, D., C. Barnhart. 1999. Transportation service network design: models and algorithms. *Lect. Notes. Econ. Math. Computer-Aided Transit Scheduling* **471** 259-283

[61] Kim, D., C. Barnhart, K. Ware, G. Reinhardt. 1999. Multimodal express package delivery: a service network design application. *Transportation Sci.* **33**(4) 391-407.

[62] Klabjan D., G. L. Nemhauser. 2002. A polyhedral study of integer variable upper bounds. *Math. Oper. Res.* **27**(4) 711-739.

[63] Krishnan, N., C. Barnhart, D. Kim, K. Ware. 2002. Network design for express shipment delivery. *Comput. Optim. Appl.* **21**(3) 239-262.

[64] Kuby M., R. G. Gray. 1993. The hub network design problem with stopovers and feeders: the case of Federal Express. *Transportation Res. A* **27A**(1) 1-12.

[65] Lederer, P. J., R. S. Nambimadom. 1998. Airline network design. *Oper. Res.* **46**(6) 785-804.

[66] Leighton, T., F. Makedon, S. Plotkin, C. Stein, E. Tardos, S. Tragoudas. 1995. Fast approximation algorithms for multicommodity flow problems. *J. Comput. Syst. Sci.* **50**(2) 228-243.

[67] Leung L., Cheung W. 2000. An integrated decision methodology for desgining and operating an air express courier's distribution network. *Decision Science* **31**(1) 105-126.

[68] Magnanti, T. L., R. T. Wong. 1984. Network design and transportation plannning: models and algorithms. *Transportation Sci.* **18**(1) 1-55.

[69] Magnanti, T. L., P. Mireault, R. T. Wong. 1986. Tailoring benders decomposition for uncapacitated network design problems. *Math. Program. Stud.* **26** 112-154.

[70] Magnanti, T. L., P. Mirchandani. 1993. Shortest paths, single origin-destination network design and associated polyhedra. *Networks* **23**(2) 103-121.

[71] Magnanti, T. L., P. Mirchandani, R. Vachani. 1993. The convex hull of two core capacitated netwrok design problems. *Math. Programming* **26**(2) 233-250.

[72] Magnanti, T. L., P. Mirchandani, R. Vachani. 1995. Modelling and solving the two-facility capacitated network loading problem. *Oper. Res.* **43**(1) 142-157.

[73] Marchand H., A. Martin, R. Weismantel, L. Wolsey. 2002. Cutting planes in integer and mixed integer programming. *Discrete Appl. Math.* **123**(1-3) 397-446.

[74] Martin R. K. 1999. *Large Scale Linear and Integer Optimization: A Unified Approach.* Kluwer Academic Publishers.

[75] Minoux, M. 1989. Network synthesis and optimum network design problems: models, solution methods and applications. *Networks* **19**(3) 313-360.

[76] Mirchandani P. 2000. Projections of the capacitated network loading problem. *Eur. J. Oper. Res.* **122**(3) 534-560.

[77] Myung, Y. S., H. J. Kim, D. W. Tcha. 1999. Design of communication network with survivability constraints. *Management Sci.* **45**(2) 238-252.

[78] Nemhauser, G. L., M. W. P. Savelsbergh, G. S. Sigismondi. 1994. MINTO, a Mixed INTeger Optimizer. *Oper. Res. Lett.* **15**(1) 47-58.

[79] Nemhauser, G. L., L. A. Wolsey. 1988. *Integer and Combinatorial Optimization.* John Wiley and Sons, Inc.

[80] Newton, H. N., C. Barnhart, P. H. Vance. 1998. Constructing railroad blocking plans to minimize handling costs. *Transportation Sci.* **32**(4) 330-345.

[81] Padberg, M. W., M. R. RAO. 1982. Odd minimum cut-sets and B-matchings. *Math. Oper. Res.* **7**(1) 67-80.

[82] Padberg, M. W., T. J. VanRoy, L. A. Wolsey. 1985. Valid linear inequalities for fixed charge problems. *Oper. Res.* **33**(4) 842-861.

[83] Riis M., K. A. Andersen. 2002. Capacitated network design with uncertain demand. *INFORMS J. Comput.* **14**(3) 247-260.

[84] Salomon Smith Barney. 2000. *Equity Research: United Parcel Service Inc.*

[85] Sridhar V., J. S. Park. Benders-and-cut algorithm for fixed-charge capacitated network design problem. *Eur. J. Oper. Res.* **125**(5) 622-632.

[86] Stallaert, J. 2000. Valid inequalities and seperation for capacitated fixed charge flow problems. *Discrete Appl. Math.* **98**(3) 265-274.

[87] Sung C. S., H. W. Jin. 2001. Dual-based approach for a hub network design problem under non-restrictive policy. *Eur. J. Oper. Res.* **132**(1) 88-105.

[88] United Parcel Service 1998 Annual Report.

[89] United Parcel Service 2002 Annual Report.

[90] *Terms and Conditions of Service*, United Parcel Service, 2003.

[91] VanRoy, T. J., L. A. Wolsey. 1985. Valid inequalities and seperation for uncapacitated fixed charge networks. *Oper. Res. Lett.* **4**(3) 105-112.

[92] Williamson, D. P., M. X. Goemans, M. Mihail, V. V. Vazirani. 1995. A primal-dual approximation algorithm for generalized steiner network problems. *Combinatorica* **15**(3) 435-454.

[93] Wolsey, L. A. 1975. Faces of linear inequalities in 0-1 variables. *Math. Programming* **8** 165-178.

[94] Wolsey, L. A. 1989. Submodularity and valid inequalities in capacitated fixed charge networks. *Oper. Res. Lett.* **8** 119-124.