Design and Operation of Multimode, Multiservice Logistics Systems

by

Karen Renee Smilowitz

B.S.Eng. (Princeton University) 1995

M.S (University of California, Berkeley) 1998

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering: Civil and Environmental Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Carlos Daganzo, Chair

Alper Atamtürk

Samer Madanat

Fall 2001

Abstract

Design and Operation of Multimode, Multiservice Logistics Systems

by

Karen Renee Smilowitz

Doctor of Philosophy in Engineering: Civil and Environmental Engineering

University of California, Berkeley

Professor Carlos Daganzo, Chair

This thesis introduces design strategies and operational planning techniques for multimode, multiservice networks for package delivery carriers where service levels are defined by the guaranteed delivery times of packages (e.g., overnight, two-day delivery, etc.). Large-scale transportation network design problems are typically challenging, due to the large number of interdependent decision variables and constraints. These problems are even more complex with multiple service levels. Conventional network design and routing models cannot sufficiently capture the complexity of multimode, multiservice networks. This thesis discusses two principal design and routing approaches employed in the literature, and shows how the two approaches can be integrated. One approach utilizes detailed mixed-integer programming formulations and numerical methods. The other employs less detailed models based on continuous approximations. While the first approach provides a much higher level of detail, the second is more revealing of "the big picture". Therefore, numerical methods are well suited for operational control, while continuous approximation methods are particularly effective for strategic planning and design, especially under uncertainty. An approach based on the complementary use of analytical approximation models and numerical optimization is developed to design, test and evaluate integrated strategies. This is the first application of hybrid continuous approximation/numerical optimization models to large-scale integrated networks with shipment choice. As such, advancements in both continuous approximation and numerical optimization, and the integration of the two, are required. Continuous approximation cost functions are shown to be capable of realistically modeling complex distribution systems with multiple transshipments and peddling tours. This research also demonstrates the application of solution techniques to reduce complex cost models to a series of subproblems that can be solved with common spreadsheet technology. Cost components are shown to accurately model costs using independent cost validation. A variety of integration scenarios are analyzed and the advantages of integrated operations are presented. Qualitative conclusions suggest that benefits of integration are greater when deferred demand exceeds express demand. This insight helps to explain the different business strategies of package delivery firms in industry today. This research demonstrates how hybrid modeling approaches can be used to better understand and better plan operating strategies for distribution companies.

In loving memory of my grandfather Louis Neffson.

I think you could have read this one twice.

Contents

List of Tables

List of Figures

Acknowledgements

First I would like to thank my advisor Carlos Daganzo for reminding me day after day why I began down this path. No one could have prepared me better for the next step.

I would like to thank my dissertation committee. I am grateful to Alper Atamtürk for giving me so much of his time, especially when the learning curve was steepest, and Samer Madanat for assistance with my dissertation and throughout my time at Berkeley.

I have been fortunate to have the opportunity to work closely with the faculty and students in the transportation operations group at Berkeley: Professors Mike Cassidy and the late Gordon Newell, and fellow students Alan Erera, Mike Mauch, Juan Carlos Muñoz, Yuwei Li, and Jorge Laval. You have provided such a good environment for conducting research.

I am grateful as well to the Air Operations Group at UPS, in particular Keith Ware and Alysia Wilson who allowed me to learn firsthand about distribution systems.

I would like to thank Susan Bolton, Mari Cook, Catherine Cortelyou, Steve Owen and Cindy Staffa and the entire staffs of both the Civil and Environmental Engineering Department and Institute of Transportation Studies. This research was conducted with the financial support of the Eisenhower Transportation Fellowship program and the University of California Transportation Center.

I wish to thank my friends both at Berkeley and far away, Laurie Baise, David Corr, Aaron Golub and Fenella Long, as well as Beth Flynn, Laura Rinzel, Sara Roberts and Emma Taylor, for offering support, much-needed diversions and occasional proofreading.

Finally, I could not imagine the past four years without the support of my family. I want to thank my parents and my sister Jennifer for always being a phone call away. I am so lucky to have a sister who understands the "complexity" of the PhD experience.

Chapter 1

Introduction

1.1 Background

The package delivery industry has grown quickly in the last decade. Carriers have begun to offer a wider range of service levels, defined by guaranteed time delivery window, to capture a larger share of the package delivery market and to utilize resources more efficiently. With these changes, effectively designing and operating distribution networks to accommodate multiple service levels becomes more challenging. New network configurations and routing strategies are possible when one considers integrating the operation of various service levels and transportation modes. This thesis considers both the merits of integration, as well as the challenges of designing and managing these complex multimode, multiservice networks.

Perhaps the best example of a multimode, multiservice delivery firm is United Parcel Service (UPS), which offers four overnight package services and four deferred services (2-day, 3-day, etc) domestically with an integrated air and ground network.[1] As demand for package delivery expands, offering both deferred and express services allows UPS to reach a wider range of customers in the market. Furthermore, with an integrated delivery

[1] United Parcel Service (2000c)

network, UPS can achieve higher utilization of sorting facilities, aircraft, and ground vehicles. For example, UPS overnight packages are sorted at the Louisville hub from 10:00 p.m. to 2:20 a.m.. Deferred service packages can be sorted during periods when sorting facilities are underutilized.[2] The cost of transporting some deferred packages on aircraft is marginal when excess capacity exists. According to company literature, "[UPS's] integrated air and ground network enhances pickup and delivery density and provides [UPS] with the flexibility to transport packages using the most efficient mode or combination of modes."[3]

Unlike UPS, Federal Express has chosen not to integrate its air network with the ground network recently acquired with the purchase of less-than-truckload carrier RPS. Analysts cite this decision as the reason Federal Express is losing to UPS in the e-commerce delivery market, typically composed of two- to five- day delivery.[4] Federal Express, however, believes that operations of the two networks are so different that integration is not feasible, arguing that "the optimal way to serve very distinct market segments, such as express and ground, is to operate highly efficient, independent networks."[5]

This research examines the questions of what conditions make multimode, multiservice networks most attractive, and how integration of networks can best be achieved. The advantages of integration cited by UPS (increased customer density, flexible mode choice) are examined in depth, along with additional benefits and costs from integration. Various levels of integration are studied, from the simplest case where only facilities are shared to the most complex where routing is fully integrated and the network configuration is redesigned for integration.

Conventional network design and routing models can not sufficiently capture the com-

[2] United Parcel Service (2000d)

[3] United Parcel Service (2000a)

[4] Rocks (2000), O'Reilly (2000)

[5] FedEx Corporation (2000)

plexity of multimode, multiservice networks. Network designs and routing must comply with the various time constraints for each service level. Unlike air passenger networks, shipments in freight networks can be routed in more circuitous ways to achieve economies of scale and density, provided time constraints are not violated. For deferred service shipments, these time constraints are somewhat relaxed and more cost efficient routings are possible. For these shipments, the distribution network should be designed optimally to exploit the advantages of various transportation modes. However, with the increased number of routing options and service levels, finding an optimal network design and distribution strategy becomes more difficult. Therefore, a new approach is needed to study these networks.

1.2 Related literature

The design and operation of large-scale transportation networks is a difficult task due to the large number of interdependent decisions variables and constraints. Many researchers have addressed components of this problem, utilizing a variety of approaches. Two principal approaches have been employed in the literature. In one approach, detailed mixed-integer programming formulations and numerical methods are utilized. In the other, less detailed models based on continuous approximations are used. While the former provide a much higher level of detail, the latter are more revealing of "the big picture". It is generally accepted (see below) that numerical methods are well-suited for operational control, and that continuous approximation methods are particularly effective for planning and design, especially under uncertainty.

1.2.1 Numerical optimization methods

Numerical optimization approaches to network modeling have been studied extensively, (see Magnanti and Wong (1984), Ahuja *et al.* (1993), and Ball *et al.* (1995)). The general network design model, formulated as a mixed integer program, selects integer network capacities (often in the form of transportation vehicles) to be included in the network from a discrete set of potential values and assigns continuous commodity flows over the chosen links. This model can be extended to solve facility location problems that include terminal location decisions.

As discussed in Magnanti and Wong (1984), Ahuja *et al.* (1993), and Ball *et al.* (1995), and other references on the subject (see Nemhauser and Wolsey (1999)), these mixed integer programs can be solved to optimality for small network problems. In some special cases, it is possible to solve large problems. In general, however, as the network size increases, problems become more difficult to solve, and it may be necessary to use heuristic approaches. Furthermore, collecting the necessary demand and cost data can be time-intensive and, at times, impossible.

Numerical optimization models have been successful in solving tactical and operational problems for multimode, multicommodity networks, offering detailed, cost-minimizing operating plans; see review in Crainic (2000). For example, Crainic and Rousseau (1986) determine the best use of resources for a given physical network, including vehicle routing, mode choice, and service frequency with a decomposition-based algorithm. Optimization-based approaches to network design and routing have been developed; see Powell and Sheffi (1983), Kuby and Gray (1993), and Popken (1994). A recent series of papers has looked at the service network design for United Parcel Service (Barnhart and Schneur (1996), Kim *et al.* (1999), and Armacost (2000)), and developed solution techniques to route aircraft and express items in the UPS network. While these papers study only transportation of a single service level (i.e., one time window), they provide good groundwork for studying operational issues for multiple service levels and are discussed further in Chapter 5.

1.2.2 Analytic methods

Analytic approaches use simple cost expressions, written in terms of few key parameters and decision variables to obtain near optimal solutions. Langevin *et al.* (1996) provide a comprehensive look at the most recent developments in continuous approximation models for various freight distribution systems, as well as a history of the continuous approximation approach. The earliest work on approximation methods (see Eilon *et al.* (1971), Newell (1973), Geoffrion (1976), and Daganzo and Newell (1986)) recognized that approximations can provide near optimal solutions, while offering valuable insight into operating strategies and network design. With simple cost expressions, it is easy to identify key trade-offs in network design. The most thorough treatment of the continuous approximation method and its applications is contained in Daganzo (1999). Results in this reference as well as others are discussed throughout the thesis as needed.

Whereas numerical optimization models perform better on smaller problems, the opposite is true with continuous approximations. The larger the problem, the more accurate the large-scale approximations become, (see Daskin (1985), Campbell (1993), and Daganzo (1999)).

A recent review of continuous approximation models for freight distribution (Langevin *et al.* (1996)) identifies several critical gaps in continuous approximation research, focusing particularly on areas related to complex distribution systems. Current multiple origin/multiple destination distribution models do not adequately incorporate multiple transshipments and multistop peddling tours. Such activities should be included in models in order to realistically model complex systems. Additional operating costs beyond transportation and inventory are also missing from most current continuous approximation models. For the most part, previous continuous approximation models have not considered the cost of repositioning (except perhaps Jordan and Burns (1984) and Hall (1991)). As observed in Langevin *et al.* (1996), distribution network design involves more complicated trade-offs than the simple one between inventory and transportation costs.

Several studies have extended earlier work on analytic models to consider distribution of time sensitive items (see Han (1984), Daganzo (1987a) and Daganzo (1987b), and Kiesling (1995)); however, multiple time windows have not been addressed in these papers. Hall (1989a) addressed the problem of designing networks for express packages, finding network solutions that are nearly optimal without complex optimization techniques. Rather than focusing on route-specific details, this reference considers a few main parameters, and analyzes how these parameters impact network design for overnight air delivery in a region with multiple time zones. Multiple transportation modes have been included in a limited number of continuous models, see Hall (1989b).

In addition, few applications to real systems are provided in the literature. Examples can be found in the continuous approximation literature, see Blumenfeld *et al.* (1987) and Rosenfield *et al.* (1992); however, more applications are found in the numerical optimization literature (for example, Crainic and Rousseau (1986), Kuby and Gray (1993) and Kim *et al.* (1999)).

The review by Langevin et al. also cites the need for further integration of continuous approximation and discrete models. Previous work on hybrid continuous approximation and numerical optimization models can be found in Hall (1986), Robuste *et al.* (1990), and Campbell (1993). Clearly there are benefits and disadvantages to both approaches. Continuous approximation models are well-suited for large-scale design problems where decisions are made for long planning horizons in the presence of uncertainty. For operation (control) problems over shorter planning horizons with more information available, numerical optimization methods are better suited. Using the two approaches together can provide a complete design and operating methodology for complex logistics systems. Integration of the two approaches can also be helpful in model validation. Numerical optimization techniques can be used to validate continuous cost approximations, and vice versa. This research introduces a combination of the different methodologies for design and operation problems for multimode, multiservice networks, and addresses the above-mentioned gaps in the literature.

1.3 Scope of research

In this research, both strategic and operational issues related to multimode, multiservice transportation networks are analyzed. Long term strategic issues include the number, location, and hierarchy of terminals in the network, the modes serving each service level, and sensitivity of solutions to changes in inputs. Shorter term tactical and operation decisions include the routing of vehicles and items in integrated networks. Application of the analysis to other short term decisions, including the scheduling of sorting facilities, is left to future work. The following contributions are made in this research.

- Development of a hybrid solution methodology for a class of large-scale complex logistics systems, consisting of both continuous approximation and numerical optimization modules. This research demonstrates how continuous approximation and numerical optimization models complement each other.

- Development of formulation and optimization techniques as part of this methodology. This research utilizes a two-stage approach to both design and control decisions. Continuous approximation models are well-suited for the design of large-scale systems and facilitate the study of various "what-if" scenarios. Once network designs have been chosen with approximation models, detailed operating plans can be developed with numerical optimization models.

- Design and evaluation of distribution networks and integration strategies with these models.

- Exploration of solution approaches to large-scale multicommodity network flow problems with vehicle and item balancing constraints.
- Validation techniques for cost model components.

1.3.1 Thesis organization

This research begins with an overview of basic network configurations and routing principles for single mode, single service networks and multimode, multiservice logistics systems in Chapter 2. Chapter 3 presents the continuous approximation models used for the development of network design guidelines, strategy analysis, and cost estimation. This chapter also includes solution techniques to minimize total operating costs. The application of these models to network design and scenario analysis for a variety of case studies is presented in Chapter 4. This chapter also includes cost component validation. In Chapter 5, the focus shifts to the specific shorter term problem of routing deferred items and ground vehicles. This chapter presents the numerical optimization techniques to solve the deferred item and vehicle routing problem, and discusses the integration of continuous and discrete models. Finally, in Chapter 6, the findings of this research are summarized and future areas of research are discussed.

Chapter 2

System overview: concepts and definitions

2.1 Introduction

This chapter provides an introduction to the basic concepts of network design and routing for both traditional single mode, single service logistics networks and more complex multimode, multiservice logistics networks. In this research, two levels of service are studied: express and deferred. Express products are highly time sensitive items, whereas time is less critical for deferred packages. In the networks studied here, all local and regional transportation is conducted by ground vehicles (delivery vans, trucks, etc.), and two different modes are possible for longhaul transportation. These are ground vehicles (tractor-trailers) and aircraft. In traditional single mode, single service delivery networks, express items are transported, for the most part, by air due to restrictive time constraints. Deferred items are typically sent over separate, less expensive, ground longhaul networks.

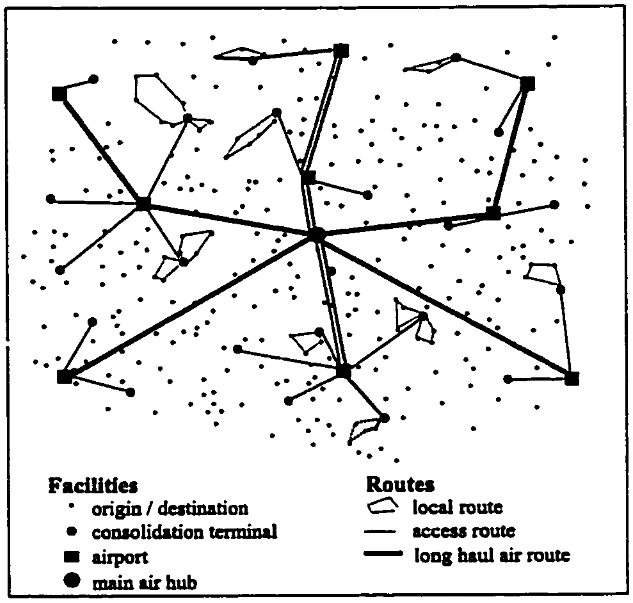


Figure 2.1: Single mode, single service: express air network

2.2 Single mode, single service express air network

Figure 2.1 displays a complete distribution network for an express air carrier. All packages are assumed to have express deadlines and all longhaul transportation is conducted by air. Items[1] first travel in delivery vans or trucks from their origins[2] along local delivery tours to the nearest regional consolidation terminal. These consolidation terminals consolidate cargo within the region for efficient longhaul transportation. Items then travel along access routes from the consolidation terminals to the nearest airport, for an evening flight departure to the main hub. As the figure shows, multiple-stop peddling tours between airports and the hub may be introduced (as opposed to a pure hub-and-spoke structure) to allow airplanes to accumulate higher volumes and still operate on daily

[1]Items are defined as a fixed portion of a truck, e.g. one pallet.

[2]Real origins are typically some customer/company interface, such as a Federal Express drop box. While there exist many types of origins, the differences among origins are ignored here. For strategic design purposes, demand is aggregated to the nearest local service center and the transportation from interface to service center is not considered.

frequencies to meet time deadlines. Items arrive at the main hub between 10 pm and 2 am, where the items are sorted by destination airport and then loaded onto aircraft for a morning departure. After arriving at the destination airport, items travel to a regional consolidation terminal and then to their final destination. These activities must be performed within the one-day express delivery window.

2.3 Single mode, single service: deferred ground network

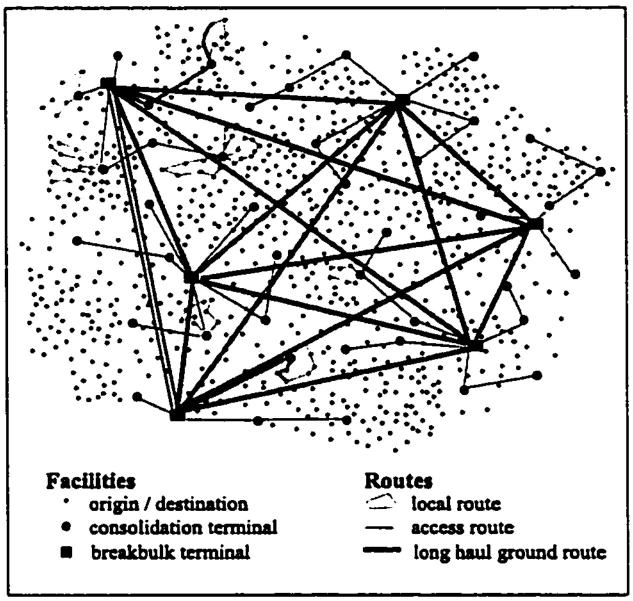


Figure 2.2: Single mode, single service: deferred ground network

Figure 2.2 displays a simple network for deferred items with ground transportation only. Again, only a single delivery time window is considered, assumed to be three to five days for deferred items. The local and access portions of the network are the same as in the express air network. The main differences exist in longhaul transportation. The ground network consists of several breakbulk terminals, which, like the airports, act as gateways to the longhaul network. However, unlike the air network, there is no single main hub. All breakbulk terminals serve as hubs, albeit for smaller percentages of the total network volume. Items are routed from originating consolidation terminal to destination consolidation terminal through at most two breakbulk terminals, depending on the location of the two consolidation terminals.

2.4 Multimode, multiservice networks

In this research, the two single mode, single service networks are integrated to form multimode, multiservice networks, as shown in Figure 2.3. Several levels of service and mode integration are considered and discussed in this section.

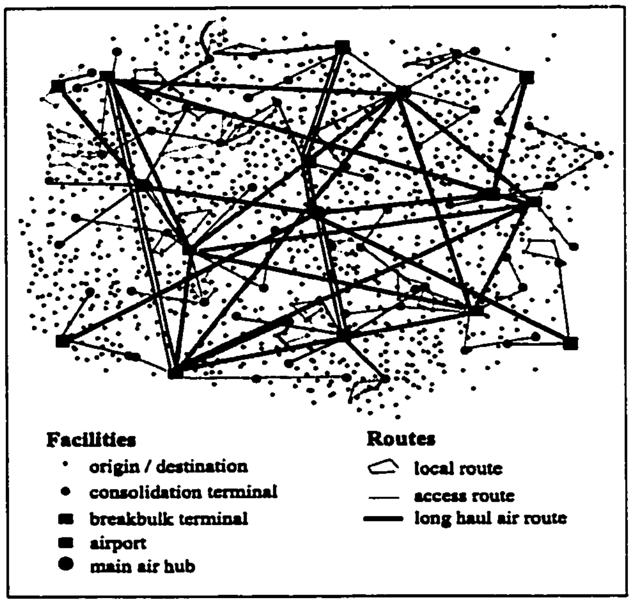


Figure 2.3: Multimode, multiservice network

Base case: no integration

This is the base case for comparison. Express and deferred items travel on two distinct networks, each designed for efficient *separate* operation. The two service levels have unique terminal sets and separate routing. Express items are routed over the longhaul air network since travel times by ground would be too long for on-time delivery. Deferred items are transported over the longhaul ground network.

Facility-integrated networks: existing infrastructure

Here the networks of the base case are merged without any change in network infrastructure. All consolidation terminals from both networks remain open after integration and may be used by all service levels. Routing, however, is performed separately for the service levels. This scenario allows customers to be served through closer consolidation terminals.

Facility-integrated networks: reduction in existing infrastructure

In a variation of the above, all consolidation terminals of the smaller network are removed to avoid an overcapacitated network. Routing is still performed separately by service level in this scenario.

Fully integrated networks: possible reduction in existing infrastructure

Here the networks of the base case are merged, and routing is integrated. The consolidation terminals of the smaller network are eliminated if such action reduces total cost. The integration of local pickup and delivery operations creates potential for more savings from consolidation. At consolidation terminals it is decided whether a deferred item travels by ground or air to the destination consolidation terminal where services are again joined. One of the benefits of operating a fully integrated multimode, multiservice network comes from the opportunities to fill excess capacity on aircraft with deferred items. For longhaul trips, it is assumed that express items are routed as in the air-only network; however, it may be possible to send some deferred packages by air as well. By filling excess capacity on aircraft with deferred items, all longhaul vehicles can be utilized better.

Fully integrated networks: reoptimized infrastructure

In this scenario, the entire network, with the exception of airports, is reoptimized for integrated routing. This can be beneficial since operating an integrated network with existing facilities can lead to an improperly capacitated network. Often facilities are expensive to relocate. Thus, comparing these strategies should help with facility relocation decisions.

In the following chapters, the above integration scenarios are analyzed at both the short term (control) and long term (design) levels. Control decisions, covered in Chapter 5, involve the efficient allocation of aircraft excess capacity and routing of ground vehicles. Design decisions, covered in Chapter 3, influence the structure of the integrated network such that control decisions can be made most effectively.

Chapter 3

System structure: modeling and optimization

3.1 Introduction

The costs and benefits of the integration strategies introduced in the previous chapter are quantified with a complete design methodology that facilitates the choice of integration level and the network design. In this chapter, analytic models are developed using idealizations of network geometries, operating costs, demand and customer distributions, and routing patterns. The goal is to find simple, yet realistic, guidelines to determine how to best design and operate a network. In Section 3.2, the decision variables and parameters that define the problem are presented. Section 3.3 details the development of models to approximate the total cost of operation as a function of these variables. The approach to cost minimization is presented in Section 3.4. The first four sections of this chapter consider spatially heterogenous (but time independent), deterministic demand for both deferred and express service levels. Models are expanded in Section 3.5 to incorporate uncertainty and demand variations over time.

3.2 Formulation

Continuous approximations can be used to facilitate decision making at an aggregate level. Rather than considering highly detailed and discrete operational data, continuous approximations use smooth functions to describe the data, such as a demand density function that varies with location. Such functions are used for decision variables as well, e.g. as in the case of spatially varying terminal densities. As explained in the literature, knowledge of these decision functions gives enough information to develop a network configuration and an operating plan with a predictable cost. The structure of the distribution network over a service area $\mathcal{A}$ and its cost performance are defined by a set of decision functions (variables) and data functions (parameters), which are described below. Location enters the problem through the coordinates, x, of points on the plane.

Network sets

$\mathcal{N}$ Set of network types, $\mathcal{N} = \{A, G\}$ for air and ground networks. Air and ground networks differ in the mode used for longhaul transportation; however, these sets define the entire distribution network regardless of local or access mode; i.e., local and access trips served by ground but ultimately feeding into the longhaul air network are considered part of the air network.

$\mathcal{S}$ Set of service levels, $\mathcal{S} = \{E, D\}$ for express and deferred items. In non-integrated networks, there is a one-to-one correlation between the service level and the network type; express items travel exclusively on the air network and deferred items travel on the ground network.

$\mathcal{L}$ Set of distribution levels, $\mathcal{L} = \{0, 1, 2\}$: local distribution (level 0), access (level 1) and longhaul (level 2).

$\mathcal{B}$ Set of route directions, $\mathcal{B} = \{i, o\}$ for trips inbound to and outbound from a terminal.

$\mathcal{V}$ Set of vehicle types, consisting of local delivery vans, larger trucks, longhaul tractor trailers, and longhaul aircraft. For simplicity of illustration, $\mathcal{V} = \{a, t\}$ for aircraft and truck in this chapter, although the analysis in Chapter 4 includes a variety of ground vehicle types.

$\mathcal{T}$ Set of terminal types, $\mathcal{T} = \{C, B, P, H\}$. Consolidation terminals (C) aggregate items from local origins. The longhaul network consists of a series of breakbulk terminals (B), airports (P), and one main air hub (H)

In scenarios where routing is not integrated, all distribution levels are further designated as part of the air or ground network. In fully integrated scenarios, local distribution is shared and only longhaul and access are designated by air or ground. The distribution levels are further subdivided by direction for trips inbound to and outbound from a terminal.

Demand Parameters

$\delta^s(x)$ spatial customer densities for service level $s \in \mathcal{S}$ *(customers/unit area)*

$\lambda^s(x^o, x^i)$ temporal demand rate from a region of unit area about x^o to a region of unit area about x^i for service level $s \in \mathcal{S}$ *(items/time*area^2)*

$\lambda_i^s(x^i)$ trip attraction rate in a region of unit area about x^i (independent of origin) *(items/unit area*time)*; $\lambda_i^s = \int_{x^o \in \mathcal{A}} \lambda^s(x^o, x^i) dx^o$

$\lambda_o^s(x^o)$ trip generation rate in a region of unit area about x^o (independent of destination) *(items/unit area*time)*; $\lambda_o^s = \int_{x^i \in \mathcal{A}} \lambda^s(x^o, x^i) dx^i$

Cost Parameters

c_d^u fixed costs of overcoming distance, for vehicle of type $u \in \mathcal{V}$(*\$/vehicle*distance*)

$c_d'^u$ marginal transportation cost per item, for vehicle of type $u \in \mathcal{V}(\$/item*trip)$

c_q^u cost of stopping a vehicle of type , $u \in \mathcal{V}$ at a terminal $(\$/stop)$[1]

c_f^y annualized fixed terminal cost of terminals of type $y \in \mathcal{T}(\$/terminal*time)$

$c_f'^y$ annualized variable terminal cost of terminals of type $y \in \mathcal{T}(\$/item*time)$

c_k cost of sorting $(\$/bit)$; i.e, the number of bits required to identify a sorting class, (2^n classes = n bits)

c_h storage (rent) cost for items $(\$/item*time)$

For simplicity of illustration, all facilities are assumed to have the same costs in this chapter and the superscript $y \in \mathcal{T}$ for c_f and c_f' is dropped. In the case studies presented in Chapter 4 this simplification is relaxed.

Decision variables

$\Delta_y(x)$ density of terminals of type $y \in \mathcal{T}$ *(terminals/unit area)*[2]

$h_l^{m,b}(x)$ headway of a route of type $l \in \mathcal{L}$ for network $m \in \mathcal{N}$ in direction $b \in \mathcal{B}$ *(time)*

$n_l^{m,b}(x)$ number of stops on a route of type $l \in \mathcal{L}$ for network $m \in \mathcal{N}$ in direction $b \in \mathcal{B}$

$v_l^{m,b}(x)$ shipment size per terminal on a route of type $l \in \mathcal{L}$ for network $m \in \mathcal{N}$ in direction $b \in \mathcal{B}$ *(items/terminal)*

$r_l^m(x)$ average linehaul distance [3] on a route of type $l \in \mathcal{L}$ for network $m \in \mathcal{N}$ *(distance)* Note linehaul distance is independent of direction.

[1] A second superscript can be added if the cost is terminal specific.

[2] As shown in Daganzo and Newell (1986), locating a terminal in the center of the region served by that terminal results in near optimal solutions. Therefore, one can assume terminals are located as such and determining $\Delta_y(x)$ is sufficient to obtain the number and approximate locations of terminals.

[3] As is conventional, this is the average distance from the terminal to the points served.

$\omega_b(x)$ fraction of deferred items sent by air for longhaul transportation in direction $b \in \mathcal{B}$

3.3 Logistic cost functions

This section describes the logistic cost functions used to approximate the total system cost. Here costs are developed as the average cost of operation per item. Therefore, to describe operations, the path of a typical item from origin to destination is illustrated in Figure 3.1, capturing all components of operation including transportation, handling, sorting, etc.

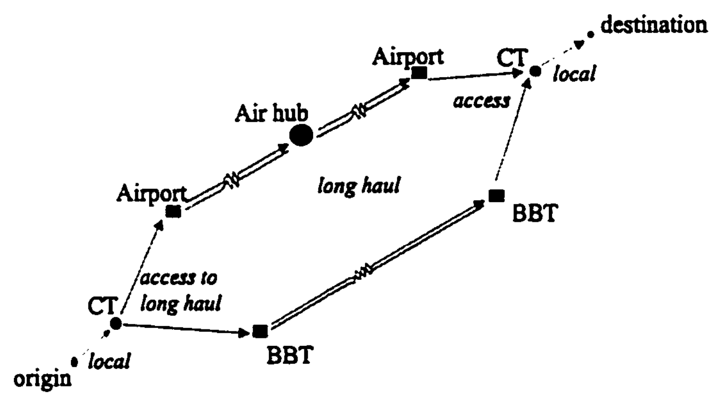


Figure 3.1: Distribution from origin to destination

As described in Daganzo (1999) and illustrated in the figures in Chapter 2, a hierarchical network structure is assumed. All items travel from an origin, to the closest consolidation terminal, to the longhaul network via the nearest airport or breakbulk terminal, and no intermediate step is skipped. Therefore, the cost of operations can be defined by the levels of distribution and the terminals visited. The logistic cost function for the origin-destination specific average cost per item, z, is comprised of the following components:

$$z = z_{local} + z_{access} + z_{longhaul} + z_{reposition} + z_{CT} + z_{airport} + z_{BBT} + z_{hub} \tag{3.1}$$

Equation (3.1) contains transportation costs for each distribution level: z_{local}, z_{access}, and $z_{longhaul}$; terminal costs: z_{CT}, $z_{airport}$, z_{BBT}, and z_{hub}; and empty vehicle repositioning costs: $z_{reposition}$. The functional form of each component is presented in the following subsections. The terms are defined for both separate and integrated networks, and for inbound and outbound operations. Further, each section includes a discussion of how these components change with various levels of integration and how these changes can be quantified. In Section 3.3.6, components are integrated over all items (deferred and express) and all distribution levels.

To aid in the development of cost models, auxiliary variables are defined. Recall that there are two service levels (E,D) and two networks (A,G). Therefore, we can economize on notation and define the auxiliary variables as follows:

$\lambda_b^A(x)$ Directional air network demand, $\lambda_b^A(x) = \lambda_b^E(x) + \omega_b(x)\lambda_b^D(x)$, for $b = i, o$

$\lambda_b^G(x)$ Directional ground network demand, $\lambda_b^G(x) = (1 - \omega_b(x))\lambda_b^D(x)$, for $b = i, o$

$\lambda_T^m(x)$ Bidirectional network-specific demand, $\lambda_T^m(x) = \sum_{b=i,o} \lambda_b^m(x)$, for $m = A, G$

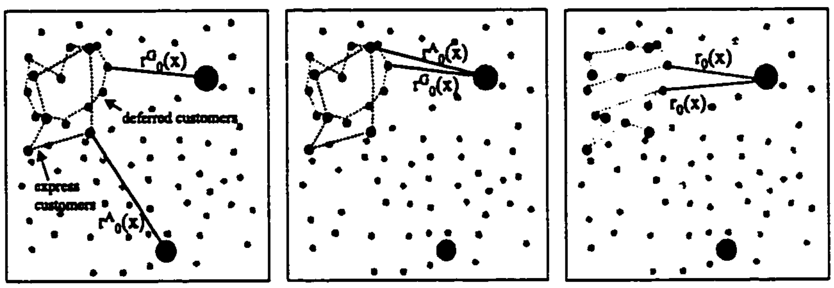
$\lambda_T(x)$ Bidirectional demand for combined networks, $\lambda_T(x) = \sum_{b=i,o} \sum_{m=A,G} \lambda_b^m(x)$

$\lambda_b(x)$ Directional demand for combined networks, $\lambda_b(x) = \sum_{m=A,G} \lambda_b^m(x)$, for $b = i, o$

$\delta(x)$ Total customer density for combined networks, $\delta(x) = \sum_{s=E,D} \delta^s(x)$

3.3.1 Local transportation costs

The local transportation costs cover pickup and delivery costs between origins/ destinations and consolidation terminals. Vehicles depart from a consolidation terminal each morning to make deliveries. After completing a tour of customers, vehicles do not return to the consolidation terminal. Rather, vehicles remain close to customers and, in the afternoon, begin a second tour of customers, this time collecting items for transportation to the consolidation terminal. Because these two activities occur at distinct times, it is assumed that pickup and delivery routes are designed independently, ignoring the repositioning of empty vehicles performed in the middle of the day caused by demand imbalances. These repositioning costs are covered in Section 3.3.4. Figure 3.2 illustrates the local distribution process for (a) non-integrated networks, (b) facility-integrated networks, and (c) fully integrated networks.



a b c

Figure 3.2: Local distribution

In non-integrated networks (Figure 3.2(a)) air and ground network operations are run independently. When integrating facilities, as in Figure 3.2(b), customers can be served through a closer consolidation terminal, although routing remains independent. Finally, when routing is integrated, the pickup and delivery tour distances decrease as the density of customers increases; see Figure 3.2(c). However, operating headways for all service levels must now meet the more stringent time restrictions of express items.[4]

The functional form of the cost model for all levels of integration is the same for both inbound and outbound trips, although parameters and decision variables change. The models presented below are based on earlier work on estimating the distances and costs of the vehicle routing problem (VRP), see Daganzo (1999). The VRP cost per item is approximated by the following:

$$f(r, v, n, \delta) = c_d' + \frac{rc_d + c_q}{nv} + \left(\frac{n-1}{n}\right)\frac{c_d k(\delta)^{-\frac{1}{2}} + c_q}{v} \tag{3.2}$$

[4]More intelligent strategies to deliver express items first and then deferred could be designed.

where k is a constant dependent on the distance metric; $k \approx 0.8$ for grids.

The first term of the VRP approximation represents the marginal handling cost per item, c'_d. The linehaul component that links the region about origins (or destinations) to the terminal and the detour component connecting individual origins (or destinations) on a route appear in the second and third terms, respectively. The average linehaul distance r is a function of the density of terminals, $r \approx \frac{2}{3}(\pi\Delta_t)^{-\frac{1}{2}}$. Clearly, this distance decreases as the density of terminals increases since customers will be located, on average, closer to a terminal. The average detour distance is a function of the average distance between customers,$(\delta)^{-\frac{1}{2}}$, and the constant, k. The detour distance decreases as the density of customers increases and each customer is then located, on average, closer to its nearest neighbor. Linehaul and detour costs also include a cost of stopping, c_q, at a customer or terminal.

Using expression (3.2), the local transportation costs can be expressed on a cost per item basis as a function of decision variables $r_0^m(x)$, $v_0^{m,b}(x)$, and $n_0^{m,b}(x)$ and parameter $\delta^s(x)$ for each routing direction, b and network type m.

$$z_{local}^{m,b}(x) = f(r_0^m(x), v_0^{m,b}(x), n_0^{m,b}(x), \delta^s(x)) =$$

$$c_d'^t + \frac{r_0^m(x)c_d^t + c_q^t}{n_0^{m,b}(x)v_0^{m,b}(x)} + \left(\frac{n_0^{m,b}(x) - 1}{n_0^{m,b}(x)}\right)\frac{c_d^t k(\delta^s(x))^{-\frac{1}{2}} + c_q^t}{v_0^{m,b}(x)} \quad (3.3a)$$

subject to:

$$n_0^{m,b}(x)v_0^{m,b}(x) \leq V_0 \quad (3.3b)$$

$$1 \leq n_0^{m,b}(x) \leq N_0 \quad (3.3c)$$

$$h_0^{m,b}(x) \leq H_0 \quad (3.3d)$$

$$v_0^{m,b}(x) = \frac{\lambda_b^s(x)}{\delta^s(x)}h_0^{m,b}(x) \quad (3.3e)$$

$$r_0^m(x) = \frac{2}{3}(\pi\Delta_C^m(x))^{-\frac{1}{2}} \quad (3.3f)$$

$$r_0^m(x), v_0^{m,b}(x), h_0^{m,b}(x) > 0 \quad (3.3g)$$

The above constraints have the following physical meaning. The total number of items on a route must not exceed the vehicle capacity, V_0 (3.3b). In local distribution vehicle capacity is often defined not by a physical volume or weight capacity, but rather by time constraints when the time required for item processing and paperwork is an issue. In these cases, the parameter V_0 can be redefined in terms of number of items that can be realistically delivered or collected in a certain time window. The maximum number of stops, N_0, is based on delivery time windows (which may be service/network specific) and the minimum number of stops is set to be one so that each customer is visited (3.3c). Constraints on operating headways (3.3d) require daily visits to customers ($H_0 = 1$), although this constraint may be relaxed for deferred items. A headway less than one day means that multiple trips are made each day, although this is rarely the case in local distribution. The average shipment size is a function of the headway between delivery tours, the customer density and the trip generation or attraction rate (3.3e). If, as an approximation, it is assumed that all consolidation terminals serve a circular area of uniform customer demand, the average distance to a customer is two-thirds of the radius of that area, and $r_0^m(x)$ can be defined as in (3.3f). The density of consolidation terminals $\Delta_C^m(x)$ enters the local costs through $r_0^m(x)$. Shipment sizes, headways, and linehaul distances must be positive (3.3g), and, since $r_0^m(x) > 0$, the density of consolidation terminals must be positive as well.

Expressions (3.3) are sufficient to quantify costs for all levels of integration. In non-integrated systems, four separate equations for $z_{local}^{m,b}(x)$ appear in the complete cost model (inbound and outbound for each network) since express items flow on the air network and deferred items on the ground network. Variables and parameters must be indexed by service level, direction, and network type as shown above. When integrating facilities, as in Figure 3.2(b), four equations are maintained and routing variables remain separate; however, the terminal densities are now combined (i.e., a single variable, $\Delta_C(x)$ rather than $\Delta_C^A(x)$ and $\Delta_C^G(x)$ is used for all four equations). As the figure shows, the linehaul distance for both networks ($r_0^A(x) = r_0^G(x)$) decreases as the density of consolidation terminals increases. When routing is integrated, only two equations for $z_{local}^b(x)$ (one for each direction) are needed. These equations capture operations for both service levels with one set of decision variables and parameters.

3.3.2 Access transportation costs

The next level of distribution is the transportation of items from consolidation terminals to the longhaul network. It is shown in Daganzo and Newell (1986) that many-to-many logistic networks can be designed optimally with a hierarchical structure (i.e., consolidation terminals feed into an assigned regional breakbulk terminal or airport, shown in Figure 3.3 for the three levels of integration). As with local distribution, the non-integrated networks (Figure 3.3(a)) are run independently; therefore, access tours are separate. In Figure 3.3(b), vehicles must stop at all consolidation terminals to collect/ distribute items for both service levels, since local collection and distribution tours for both service levels use the entire set of consolidation terminals. It is easy to see from the figure how this would increase the detour portion of access costs. This is true as well when routing is integrated (Figure 3.3(c)). A portion of deferred items travel with express items to/from airports rather than breakbulk terminals when routing is integrated.

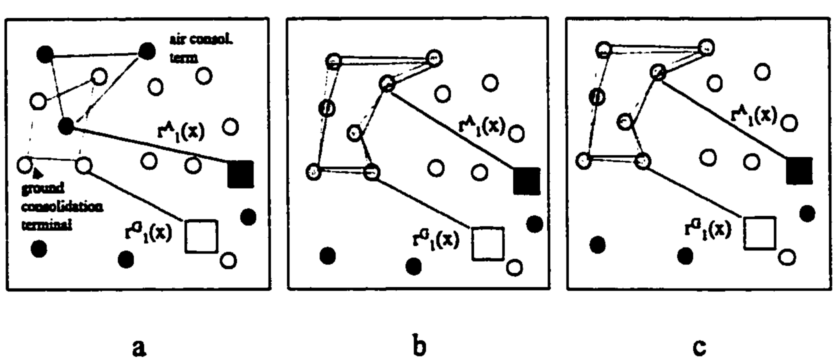


Figure 3.3: Access to longhaul transportation

With this hierarchical structure, the operation of access tours is quite similar to that of local distribution tours. Therefore, the VRP approximation can be used, and access costs can be modeled in a similar way to local costs. Again, access costs are expressed on an average cost per item basis, as a function of $r_1^m(x)$, $v_1^{m,b}(x)$, $n_1^{m,b}(x)$, and $\Delta_C^m(x)$.

$$z_{access}^{m,b}(x) = f(r_1^m(x), v_1^{m,b}(x), n_1^{m,b}(x), \Delta_C^m(x))$$
$$= c_d'^t + \frac{r_1^m(x)c_d^t + c_q^t}{n_1^{m,b}(x)v_1^{m,b}(x)} + \left(\frac{n_1^{m,b}(x) - 1}{n_1^{m,b}(x)}\right)\frac{c_d^t k(\Delta_C^m(x))^{-\frac{1}{2}} + c_q^t}{v_1^{m,b}(x)} \quad (3.4a)$$

subject to:

$$n_1^{m,b}(x)v_1^{m,b}(x) \le V_1 \quad (3.4b)$$

$$1 \le n_1^{m,b}(x) \le N_1 \quad (3.4c)$$

$$h_1^{m,b}(x) \le H_1 \quad (3.4d)$$

$$v_1^{m,b}(x) = \frac{\lambda_b^m(x)}{\Delta_C^m(x)} h_1^{m,b}(x) \quad (3.4e)$$

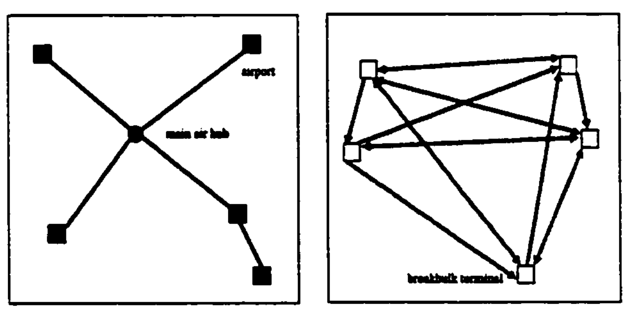
$$r_1^G(x) = \frac{2}{3}(\pi\Delta_B(x))^{-\frac{1}{2}} \qquad r_1^A(x) = \frac{2}{3}(\pi\Delta_P(x))^{-\frac{1}{2}} \quad (3.4f)$$

$$r_1^m(x), v_1^{m,b}(x), h_1^{m,b}(x) > 0 \quad (3.4g)$$

The operational constraints are the same for access costs. In the non-integrated networks (Figure 3.3(a)) all variables, including $\Delta_C^m(x)$, are indexed by network. In facility-integrated networks, this is true as well, except for the consolidation terminals which are not indexed by network. This is also true for fully integrated networks, except in this case one must also specify the network demand rates that appear in (3.4e) with another equation since service level demand no longer directly correlates to network demand. Recall that $\lambda_b^A(x) = \lambda_b^E(x) + \omega_b(x)\lambda_b^D(x)$ and $\lambda_b^G(x) = (1-\omega_b(x))\lambda_b^D(x)$ for $b = i, o$; i.e., network demand depends on the decision variable, $\omega_b(x)$, which is used to represent the fraction of deferred items shifted to the longhaul air network, and must be specified separately for inbound and outbound trips. The available capacity on aircraft determines the values of $\omega_b(x)$ and this is discussed in the next section.

3.3.3 Longhaul transportation costs

The longhaul operations are shown in Figure 3.4 for both air and ground networks. In the air network, all packages are routed through the main hub; see Figure 3.4(a). Packages must arrive before a certain sorting cut-off time in the evening and depart only after all packages have been sorted the next morning. The system operates on a daily headway. It is assumed that daily demand at any airport is always less than the capacity of an aircraft, V_2^A. Otherwise, one can always reduce access costs by using another airport (assuming reasonable airport infrastructure exists and can accommodate demand). As the figure indicates, multi-stop peddling tours between small airports and the main hub may be introduced (as opposed to a pure hub-and-spoke structure) to operate the network with a smaller fleet and maintain the daily frequency needed to meet time deadlines. However, one must be careful not to design the network too tightly; there must be slack in operations to ensure on-time delivery. Therefore, no more than two airports can be visited on a route. More details on the design of air networks can be found in Appendix A.



a b

Figure 3.4: Longhaul distribution: (a) air network (b) ground network

Figure 3.4(b) depicts longhaul ground operations. Here there is no single hub; items travel from the breakbulk terminal closest to the origin to the breakbulk terminal closest to the destination. Since time constraints are more flexible, routing adjustments such as increased headways are possible to allow vehicles to travel full. Further, either peddling or collecting may occur at one end of the longhaul distribution to fill vehicles, but not both.

Since the longhaul transportation differs significantly in the air and ground networks, two separate cost functions are developed. In the air network, all items are served through one main hub; the problem decomposes into a many-to-one distribution problem inbound to the hub and a one-to-many distribution problem outbound for which the VRP approximation holds. Due to time and operating constraints in longhaul air transportation, there are fewer decision variables. Operating headways are restricted to one day ($h_2^A = \tilde{h} = 1$ day), and they are not decision variables here. In addition, the average linehaul distance, $r_2^A(x)$ is simply the distance from x to the hub which depends on the location of the main hub. For discussion on the location of the main hub see Appendix A. The number of stops, although significantly constrained, is still a decision variable, as are the density of airports and the shipment size. As shown in Appendix A, it may be inefficient to operate a symmetric air network (inbound trips to a region mirror outbound trips from that region). Therefore, inbound and outbound longhaul trips are modeled separately.

$$
\begin{aligned}
z_{longhaul}^{A,b}(x) &= f(r_2^A(x), v_2^{A,b}(x), n_2^{A,b}(x), \Delta_P(x)) \\
&= c_d'^a + \frac{r_2^A(x)c_d^a + c_q^a}{n_2^{A,b}(x)v_2^{A,b}(x)} + \left(\frac{n_2^{A,b}(x) - 1}{n_2^{A,b}(x)}\right)\frac{c_d^a k(\Delta_P(x))^{-\frac{1}{2}} + c_q^a}{v_2^{A,b}(x)} \qquad (3.5a)
\end{aligned}
$$

subject to:

$$n_2^{A,b}(x)v_2^{A,b}(x) \leq V_2^A \qquad (3.5b)$$

$$1 \leq n_2^{A,b}(x) \leq N_2^A \qquad (3.5c)$$

$$v_2^{A,b}(x) = \frac{\lambda_b^A(x)}{\Delta_P(x)}\tilde{h} \qquad (3.5d)$$

$$v_2^{A,b}(x) > 0 \qquad (3.5e)$$

$$\Delta_P(x) \geq \frac{1}{\rho^2\pi} \qquad (3.5f)$$

As mentioned earlier, $N_2^A = 2$. A new constraint on the density of airports (3.5f) is added to ensure that there is an adequate number of airports to operate the express network. The service radius from an airport is limited to guarantee consolidation terminals are close enough to airports to complete access and local distribution on time. The parameter ρ is the maximum service radius that would allow on-time delivery.

For non-integrated networks and facility integrated networks, the above expression is used only for express items. However, when routing is integrated, a fraction of the deferred items may be added to the air network, provided excess capacity exists. Therefore, a sixth constraint is added to (3.5a) to determine the amount shifted $\omega_b(x)$.

$$\frac{n_2^{A,b}(x)}{\Delta_P(x)}\left(\omega_b(x)\lambda_b^D(x) + \lambda_b^E(x)\right) \leq \nu V_2^A, \qquad \text{for } \nu < 1 \tag{3.5g}$$

The right-hand side of (3.5g) is less than the physical capacity V_2^A by a shift factor ν, recognizing that it is not economical to fill aircraft to the same capacity level as with more profitable express items. Constraint (3.5b) still appears in integrated cost models since an aircraft may be filled to true physical capacity with express items. However, because there are a limited number of stops and $\Delta_P(x)$ is restricted by service region, it is likely that $\frac{n_2^{A,b}(x)}{\Delta_P(x)}\lambda_b^E(x) < V_2^A$ and some amount of excess capacity will exist. If results from the non-integrated case show that the left-hand side of (3.5b) is greater than νV_2^A, then $\omega_b(x) = 0$ and the additional constraint (3.5g) is not included in the model.

The ground network contains multiple breakbulk terminals; therefore, the problem cannot be decomposed in the same manner. Fortunately, continuous approximation models for many-to-many single mode, single service distribution systems with breakbulk terminals have been developed already. It has been shown (see Daganzo (1999)) that, without specifying the exact routing of items, the linehaul component between breakbulk terminals can be easily estimated when it can be assumed that ground vehicles travel full. As $\Delta_B(x) \to \infty$, the linehaul term is very close to the ratio of the total demanded item-miles $\psi\bar{d}$ (a constant obtained from the total number of items generated per unit time ψ

and the average distance between customers $\bar{d}$) and the vehicle capacity. Assuming the distance function between two points is written $\Gamma(x_1, x_2)$, the total linehaul distance is:

$$\psi \bar{d} = \int_{x^o \in \mathcal{A}} \int_{x^i \in \mathcal{A}} \lambda^G(x^o, x^i) \Gamma(x^o, x^i) dx^i dx^o$$

It is possible for ground vehicles to stop at multiple breakbulk terminals to collect items before traveling to distant breakbulk terminals. To estimate this detour distance, the average deferred demand rate for the entire service region across all origins and destinations, $\bar{\lambda}^G$ (items/area^2), is defined as

$$\bar{\lambda}^G = \int_{x^o \in \mathcal{A}} \int_{x^i \in \mathcal{A}} \frac{\lambda^G(x^o, x^i)}{|\mathcal{A}|^2} dx^i dx^o$$

The average breakbulk terminal density, $\bar{\Delta}_B$, is defined as

$$\bar{\Delta}_B = \int_{x \in \mathcal{A}} \frac{\Delta_B(x)}{|\mathcal{A}|} dx$$

Then the average shipment size collected from a breakbulk terminal at x for destination x^i is $\frac{\lambda^G(x,x^i) h_2^G(x)}{\Delta_B(x) \Delta_B(x^i)}$. Averaged across destinations, this may be approximated by $v_2^G(x) \approx \frac{\bar{\lambda}^G h_2^G(x)}{\bar{\Delta}_B^2}$. The average number of collecting stops per vehicle trip at or around x is then $n_2^G(x) \approx \frac{V_2^G}{v_2^G(x)} = \frac{V_2^G \bar{\Delta}_B^2}{\bar{\lambda}^G h_2^G(x)}$. The detour distance component is a function of the density of breakbulk terminals, $\left(\frac{n_2^G(x)-1}{n_2^G(x)}\right) \frac{c_d k(\Delta_B(x))^{-\frac{1}{2}} + c_q}{v_2^G(x)}$. It is assumed that the headway for all routes leaving a breakbulk terminal is the same. Further, a marginal transportation cost c_d' is charged to each item traveling by ground.

3.3.4 Vehicle repositioning costs

When demand inbound to a region is not balanced with outbound demand, it may be necessary to reposition empty vehicles. In local distribution, the number of vehicles leaving a consolidation terminal for morning deliveries may be insufficient to cover afternoon collection. In this case, extra empty vehicles must be deployed from the consolidation terminal. Conversely, vehicles may return empty to the consolidation terminal after morning distribution if inbound demand exceeds outbound demand. The same is true for access trips, if there is an imbalance in inbound and outbound demand between consolidation terminals and a breakbulk terminal or an airport. It is assumed that in the case of deterministic demand each terminal at both the local and access levels of distribution is self-sufficient (i.e. vehicles are not shared between terminals). However, demand imbalances between breakbulk terminals require the repositioning of empty tractor trailers among terminals.

The repositioning costs of the additional trips required between origin/demand points and the main terminal at the local and access levels can be modeled as a function of the demand rate imbalance, $|\lambda_o^m(x) - \lambda_i^m(x)|$ in the region. The number of empty vehicle movements (of capacity V) required in a region served by a terminal with density $\Delta(x)$ would be $\frac{|\lambda_o^m(x)-\lambda_i^m(x)|}{\Delta(x)V}$.

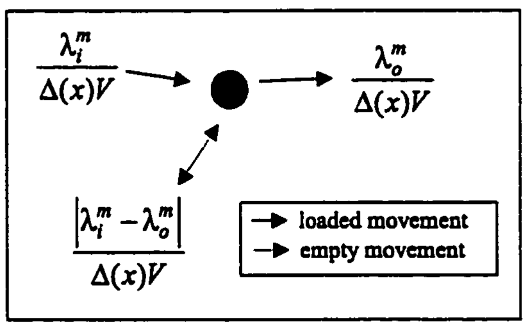


Figure 3.5: Repositioning empty vehicles at a terminal

The repositioning cost in that region would then be $\frac{|\lambda_o^m(x)-\lambda_i^m(x)|}{\Delta(x)V} c_d r(x)$. For local vehicles, the value of $r(x) = r_0(x) = \frac{2}{3}(\pi\Delta_C(x))^{-\frac{1}{2}}$, and for access vehicles, $r(x) = r_1^G(x) = \frac{2}{3}(\pi\Delta_B(x))^{-\frac{1}{2}}$ for the ground network and $r(x) = r_1^A(x) = \frac{2}{3}(\pi\Delta_P(x))^{-\frac{1}{2}}$ for the air network. Including these definitions, and prorating the cost to a cost per item basis, with $\frac{\lambda_T^m(x)}{\Delta(x)}$ items per terminal, we obtain

$$\frac{\frac{2}{3}c_d|\lambda_o^m(x) - \lambda_i^m(x)|}{\lambda_T^m(x)V}(\pi\Delta(x))^{-\frac{1}{2}}$$

The repositioning of empty tractor trailers between breakbulk terminals is more difficult to model. An upper bound to this cost can be obtained by multiplying the net supply or demand at a breakbulk terminal by the average distance between breakbulk terminals and summing these costs across all breakbulk terminals. However, it would be more efficient to reposition vehicles between nearby terminals.[5] This repositioning operation is essentially a transportation problem of linear programming, where some breakbulk terminals serve as sinks for needed vehicles, and others as supplies of extra vehicles. In Appendix C an approximation of the transportation problem of linear programming is developed that can easily be included in continuous approximation models. The following repositioning cost per item is derived for a service area of size $|\mathcal{A}|$ comprised of N terminals:

$$\frac{\sigma_N}{V_2^G}\left(\frac{N}{|\mathcal{A}|}\right)^{-\frac{1}{2}} c_d(0.42 + 0.031 * \log_2(N))$$

The standard deviation in net supply or demand across all terminals, expressed in units of "items", is σ_N. It is determined by the demand data and the allocation of demand to the N terminals.

The above expression, derived in the appendix, assumes independence of demand across terminals. However, net supplies or demands for breakbulk terminals within a region of common demand parameters are not likely to be independent and using the terminals as nodes would underestimate costs. Therefore, breakbulk terminals are grouped by region, and the repositioning is approximated as a cost of repositioning among regions of terminals, rather than the terminals themselves, plus a lower level cost that is a function of the aggregation. Thus, the upper level cost is independent of decision variables and can be neglected in the optimization phase, although this cost is included in the case studies in Chapter 4. Additional repositioning costs due to demand uncertainty are introduced in Section 3.5.1.

[5] It is assumed that this repositioning can be performed within the time frame necessary for on-time deferred delivery.

3.3.5 Terminal costs

Terminals within a large-scale distribution network serve a variety of purposes, as described in Chapter 2. The cost of operation at terminals is composed of handling costs, facility charges, sorting expenses, and storage fees. The value of the parameters and decision variables are different across terminal types, but the functional form of the terminal operating cost is the same. Therefore, this section first introduces the generic cost model, and then details specific costs for each terminal type.

Consider a terminal serving a total inbound and outbound flow of $Q = Q^i + Q^o$ items per unit time where two sorts are performed daily, one sort performed on items traveling outbound from the region served by the terminal and one sort performed on items inbound to that region. Each sort has a complexity of $K_b, b \in i, o$, which is a function of the possible destinations from a terminal. The cost per item for a generic terminal with both inbound and outbound sorts is

$$g(Q, K_i, K_o, h_o, h_i) = c_f' + \frac{c_f}{Q} + \sum_{b=i,o} c_k \frac{Q^b}{Q} \log(K_b) + \sum_{b=i,o} c_h h_b$$

Terminal costs include a marginal cost per item through the terminal, c_f', representing the cost of handling items at the terminal. The fixed cost per terminal per unit time c_f is prorated to all the items flowing through in a time unit. The flow through a terminal can be determined by the trip attraction to and trip generation from that terminal, and may be network specific, $Q = \frac{\lambda_i^m(x)}{\Delta_y(x)} + \frac{\lambda_o^m(x)}{\Delta_y(x)}$. The sorting costs are assumed to increase logarithmically with the complexity of the sort, K_b. Finally, the storage costs, which depend on the length of time an items is held at a terminal, are included. The chosen expression assumes that the length of time an item is stored at a terminal is a linearly dependent on the routing headways, h_b. Although the proposed cost expression is general, the values of its inputs are terminal-specific. They may also depend on the integration level, as explained below.

Consolidation terminal costs

Following afternoon collection tours, items arrive at consolidation terminals for offloading from local vans and loading onto larger trucks for access trips. The flow of items (equal to $\frac{\lambda_o^m(x)}{\Delta_C^m(x)}$) is considered to be "outbound" since these items will ultimately leave the region for their destination. As items travel directly from a consolidation terminal to the breakbulk terminal or airport serving the area, no outbound sort occurs to determine access routes in non-integrated networks ($K_o = 1$). However, when terminals are integrated (both facility integrated and fully integrated scenarios) a binary outbound sort must be performed to determine outbound access route. The complexity of this sort is $K_o = 2$. For delivery routes, items arrive at consolidation terminals from breakbulk terminals or airports, at an inbound flow equal to $\frac{\lambda_i^m(x)}{\Delta_C^m(x)}$. Items must again be offloaded and then sorted by delivery van, a sort of complexity $\frac{\delta^m(x)}{\Delta_C^m(x)V_0}$, representing the number of customer locations per consolidation terminal, divided into separate delivery vans. Storage costs are a function of local pickup and delivery headways, $h_0^{m,b}(x)$.

For non-integrated networks, separate cost models are used for ground consolidation terminals and air consolidation terminals since these terminals are not shared across networks.

$$
\begin{aligned}
z_{CT}^m(x) &= g\left(\frac{\lambda_o^m(x)+\lambda_i^m(x)}{\Delta_C^m(x)}, 1, \frac{\delta^m(x)}{\Delta_C^m(x)V_0}, h_0^{m,i}(x), h_0^{m,o}(x)\right) \\
&= c_f' + \frac{\Delta_C^m(x)}{\lambda_T^m(x)}c_f + \frac{\lambda_i^m(x)}{\lambda_T^m(x)}c_k' \log\left(\frac{\delta^m(x)}{\Delta_C^m(x)V_0}\right) + \sum_{b=i,o} c_h h_0^{m,b}(x) \qquad (3.6a)
\end{aligned}
$$

For facility integrated and fully integrated networks, there is no need to distinguish between network types or service level, and therefore the superscript, m, for network type is dropped. Outbound sorting complexity $K_o = 2$.

$$z_{CT}(x) = g\left(\frac{\lambda_o(x)+\lambda_i(x)}{\Delta_C(x)}, 2, \frac{\delta(x)}{\Delta_C(x)V_0}, h_0^i(x), h_0^o(x)\right)$$
$$= c_f' + \frac{\Delta_C(x)}{\lambda_T(x)}c_f + \frac{\lambda_o(x)}{\lambda_T(x)}c_k' \log(2) + \frac{\lambda_i(x)}{\lambda_T(x)}c_k' \log\left(\frac{\delta(x)}{\Delta_C(x)V_0}\right) + \sum_{b=i,o} c_h h_0^b(x) \tag{3.6b}$$

Breakbulk terminal costs

Upon arrival at a breakbulk terminal from a consolidation terminal, items are sorted by the next breakbulk terminal en route to the final destination. The outbound flow of items through the breakbulk terminal is equal to the flow of items leaving the region divided by the density of terminals in the region, $\frac{\lambda_o^G(x)}{\Delta_B(x)}$. Here, the complexity of the outbound sort can be approximated by a decision among all breakbulk terminals, $K_o = \bar{\Delta}_B|\mathcal{A}|$. Although ground vehicles may make multiple stops, items are assumed to be stowed/loaded in vehicles by destination terminal. Conversely, the flow inbound to a breakbulk terminal from another breakbulk terminal is $\frac{\lambda_i^G(x)}{\Delta_B(x)}$. Before transporting items to the consolidation terminal closest to their destination, another sort is performed and the complexity of this sort depends on the number of consolidation terminals in the area surrounding the breakbulk terminal, $K_i = \frac{\Delta_C^G(x)}{\Delta_B(x)}$. Therefore, the cost is written

$$z_{BBT}(x) = g\left(\frac{\lambda_o^G(x)+\lambda_i^G(x)}{\Delta_B(x)}, \frac{\Delta_C^G(x)}{\Delta_B(x)}, \bar{\Delta}_B|\mathcal{A}|, h_1^{G,i}(x), h_1^{G,o}(x), h_2^G(x)\right)$$
$$= c_f' + \frac{\Delta_B(x)}{\lambda_T^G(x)}c_f + \frac{\lambda_i^G(x)}{\lambda_T^G(x)}c_k \log\left(\frac{\Delta_C^G(x)}{\Delta_B(x)}\right) + \frac{\lambda_i^G(x)}{\lambda_T^G(x)}c_k \log\left(\bar{\Delta}_B|\mathcal{A}|\right)$$
$$+ \sum_{b=i,o} c_h h_1^{G,b}(x) + c_h h_2^G(x) \tag{3.7}$$

Note that storage costs arise from items held both on access routes and longhaul routes, since items visit two breakbulk terminals en route from origin to destination. In fully integrated networks, the flow through breakbulk terminals is reduced to $\lambda_o^G(x) = (1-\omega_o(x))\lambda_o^D(x)$ and $\lambda_i^G(x) = (1-\omega_i(x))\lambda_i^D(x)$, but the sorts performed are the same.

Airport and main hub costs

The cost for an airport is combined with the air hub costs for each item, and the fixed cost of the hub is neglected since it is independent of decision variables. Every evening, items arrive at an airport from consolidation terminals, at a daily rate of $\frac{\lambda_o^A(x)}{\Delta_P(x)}$. The items are loaded onto aircraft for a flight to the main hub. No sort occurs at that airport since there is only one hub in the network. At the main hub, items are sorted by destination airport, a sort complexity that can be approximated by the total number of airports $K_o = \int_{x\in\mathcal{A}} \Delta_P(x)dx = \bar{\Delta}_P|\mathcal{A}|$. Items are then loaded onto aircraft again and sent to another airport the next morning. When items arrive at an airport from the hub, the items are sorted by destination consolidation terminal; therefore, $K_i = \frac{\Delta_C^A(x)}{\Delta_P(x)}$. The flows inbound and outbound from the main hub are equal. The flow inbound to an airport from the main hub is $\frac{\lambda_i^A(x)}{\Delta_P(x)}$.

$$
\begin{aligned}
z_{airport}(x) &= g\left(\frac{\lambda_o^A(x)+\lambda_i^A(x)}{\Delta_P(x)}, \frac{\Delta_C^A(x)}{\Delta_P(x)}, \bar{\Delta}_P|\mathcal{A}|, h_1^{A,o}(x), h_1^{A,i}(x)\right) \\
&= c_f' + \frac{\Delta_P(x)}{\lambda_T^A(x)}c_f + \frac{\lambda_i^A(x)}{\lambda_T^A(x)}c_k' \log\left(\frac{\Delta_C^A(x)}{\Delta_P(x)}\right) + \frac{\lambda_i^A(x)}{\lambda_T^A(x)}c_k' \log(\bar{\Delta}_P|\mathcal{A}|) \\
&\qquad + \sum_{b=i,o}\left(c_h h_1^{A,b}(x) + c_h\tilde{h}\right) \qquad (3.8)
\end{aligned}
$$

As with breakbulk terminals, storage costs arise from both access route headways and longhaul headways ($h_2^A = \tilde{h} = 1$ day). For integrated networks, the sorting of deferred packages at main hub may be conducted during the day. There may be some changes in operating costs due to increased load factors, but larger fixed costs are shared across more items.

3.3.6 Complete model

A complete logistic cost function, containing all transportation and terminal costs in the region per unit time, is used to compare non-integrated networks with integrated networks and to evaluate the merits of integration and address the question of attractiveness/feasibility of integration. To obtain this function, the cost components described in the previous sections are integrated over all items in the service area.

The complete model is presented in its most complex form: a fully integrated network. By setting $\omega_b(x) = 0, \forall x \in \mathcal{A}, b \in \mathcal{B}$ and separating local costs by network type, the model represents costs for facility integrated networks. For non-integrated networks, $\Delta_C(x)$ should be indexed by network type. While the complete model may appear rather complex, Section 3.4 shows how the entire model can be reduced to a series of subproblems that can be easily programmed into a spreadsheet.

$$\min z = \int_{x^o \in \mathcal{A}} \Bigg\{ \int_{x^i \in \mathcal{A}} \Bigg\{$$

$$\sum_{b=i,o} \lambda_b(x) \left(c_d'^t + \frac{r_0(x) c_d^t + c_q^t}{n_0^b(x) v_0^b(x)} + \left(\frac{n_0^b(x) - 1}{n_0^b(x)} \right) \frac{c_d^t k (\delta(x))^{-\frac{1}{2}} + c_q^t}{v_0^b(x)} \right)$$

$$+ \sum_{b=i,o} \sum_{m=A,G} \lambda_b^m(x) \left(c_d'^t + \frac{r_1^m(x) c_d^t + c_q^t}{n_1^{m,b}(x) v_1^{m,b}(x)} + \left(\frac{n_1^{m,b}(x) - 1}{n_1^{m,b}(x)} \right) \frac{c_d^t k (\Delta_C(x))^{-\frac{1}{2}} + c_q^t}{v_1^{m,b}(x)} \right)$$

$$+ \sum_{b=i,o} \lambda_b^A(x) \left(c_d'^a + \frac{r_2^A(x) c_d^a + c_q^a}{n_2^{A,b}(x) v_2^{A,b}(x)} + \left(\frac{n_2^{A,b}(x) - 1}{n_2^{A,b}(x)} \right) \frac{c_d^a k (\Delta_P(x))^{-\frac{1}{2}} + c_q^a}{v_2^{A,b}(x)} \right)$$

$$+ \lambda_o^G(x) \left(c_d'^t + \left(\frac{n_2^G(x) - 1}{n_2^G(x)} \right) \frac{c_d^t k (\Delta_B(x))^{-\frac{1}{2}} + c_q^t}{v_2^G(x)} \right) + \frac{c_d \bar{d}}{V_2^G} (\bar{\lambda}^G)$$

$$+ \frac{\frac{2}{3} |\lambda_o(x) - \lambda_i(x)|}{V_0 \sqrt{\pi \Delta_C(x)}} c_d + \frac{\frac{2}{3} |\lambda_o^A(x) - \lambda_i^A(x)|}{V_1 \sqrt{\pi \Delta_P(x)}} c_d + \frac{\frac{2}{3} |\lambda_o^G(x) - \lambda_i^G(x)|}{V_1 \sqrt{\pi \Delta_B(x)}} c_d$$

$$+ \lambda_T(x) c_f' + \Delta_C(x) c_f + \lambda_i(x) c_k \log\left(\frac{\delta(x)}{\Delta_C(x) V_0} \right) + \lambda_o(x) c_k \log(2) + \sum_{b=i,o} c_h \lambda_b(x) h_0^b(x)$$

$$+ \lambda_T^G(x) c_f' + \Delta_B(x) c_f + \lambda_i^G(x) c_k \log\left(\frac{\Delta_C(x)}{\Delta_B(x)} \right) + \lambda_o^G(x) c_k \log(\bar{\Delta}_B |\mathcal{A}|)$$

$$+ \lambda_T^A(x) c_f' + \Delta_P(x) c_f + \lambda_i^A(x) c_k \log\left(\frac{\Delta_C(x)}{\Delta_P(x)} \right) + \lambda_i^A(x) c_k \log(\bar{\Delta}_P |\mathcal{A}|)$$

$$+ \sum_{b=i,o} c_h \lambda_b^G(x) h_1^{G,b}(x) + \lambda_o^G(x) c_h h_2^G(x) + \sum_{b=i,o} c_h \lambda_b^A(x) (\bar{h} + h_1^{A,b}(x))$$

$$\Bigg\} dx^i \Bigg\} dx^o \tag{3.9a}$$

The integrand begins with local transportation costs, summing both collection and delivery costs. The next line represents access costs for trips to and from airports and breakbulk terminals. The following two lines represent longhaul costs for ground transportation and air, respectively. The repositioning costs are included for local delivery vans and access costs. Since the repositioning of longhaul ground vehicles is independent of decision variables, that term does not appear. Terminal costs appear in the final lines of the integrand. The goal is to choose the decision functions of the logistic cost function subject to constraints defined in the previous section. These are:

$$n_l^{m,b}(x)v_l^{m,b}(x) \leq V_l \qquad \forall b \in \mathcal{B}; m \in \mathcal{N}; l \in \mathcal{L} \tag{3.9b}$$

$$1 \leq n_l^{m,b}(x) \leq N_l^m \qquad \forall b \in \mathcal{B}; m \in \mathcal{N}; l \in \mathcal{L} \tag{3.9c}$$

$$h_l^{m,b}(x) \leq H_l^m \qquad \forall b \in \mathcal{B}; m \in \mathcal{N}; l \in \mathcal{L} \tag{3.9d}$$

$$v_0^b(x) = \frac{\lambda_b(x)}{\delta(x)} h_0^b(x) \qquad \forall b \in \mathcal{B} \tag{3.9e}$$

$$v_1^{m,b}(x) = \frac{\lambda_b^s(x)}{\Delta_C(x)} h_1^{m,b}(x) \qquad \forall b \in \mathcal{B}; m \in \mathcal{N} \tag{3.9f}$$

$$v_2^{A,b}(x) = \frac{\lambda_b^A(x)}{\Delta_P(x)} \bar{h} \qquad \forall b \in \mathcal{B} \tag{3.9g}$$

$$v_2^G(x) = \frac{\bar{\lambda}^G h_2^G(x)}{E[\Delta_B(x)]^2} \qquad \forall b \in \mathcal{B} \tag{3.9h}$$

$$r_0(x) = \frac{2}{3}(\pi \Delta_C(x))^{-\frac{1}{2}} \tag{3.9i}$$

$$r_1^G(x) = \frac{2}{3}(\pi \Delta_B(x))^{-\frac{1}{2}} \qquad r_1^A(x) = \frac{2}{3}(\pi \Delta_P(x))^{-\frac{1}{2}} \tag{3.9j}$$

$$\frac{n_2^{A,b}(x)}{\Delta_P(x)} \lambda_b^A(x) \leq \nu V_2^A \qquad \forall b \in \mathcal{B} \tag{3.9k}$$

$$r_l^m(x), v_1^{m,b}(x), h_l^{m,b}(x) > 0 \qquad \forall b \in \mathcal{B}; m \in \mathcal{N}; l \in \mathcal{L} \tag{3.9l}$$

$$\Delta_P(x) \geq \frac{1}{\rho^2 \pi} \tag{3.9m}$$

3.4 Optimization

This section describes the reduction of problem (3.9) to a series of simple subproblems that can be solved in closed form. As a result, it is then easy to design integrated networks and analyze potential cost and demand scenarios.

3.4.1 Solution approach

As a first step, the number of variables appearing in the complete model is reduced. It is shown below how the set of number of stops, the shifted amounts $\omega_b(x)$, the shipment sizes from terminals, and the linehaul distances can be eliminated from the optimization. Consequently, the entire cost model can then be written as a function of the terminal densities and operating headways.

Since in-vehicle inventory costs are not considered, it should be intuitive that it is cost-efficient to use full vehicles wherever possible. See Daganzo (1999) for further discussion. Vehicles can be filled by holding vehicles for longer headways or increasing the number of stops on a route. There is more flexibility in headways for the ground network; one need not operate on a daily basis, although this may be desirable for scheduling drivers and performing vehicle maintenance. By setting constraints (3.9b) to equality ($n_l^{m,b}(x)v_l^{m,b}(x) = V_l^m$) for all distribution levels, route direction and service levels, this full vehicle conditions can be enforced. It then becomes possible to remove $n_l^{m,b}(x)$ from the model. Importantly, the full vehicle condition does not hold for longhaul air routes due to tight time restrictions.

In all scenarios, it is assumed that the longhaul portion of the air network is optimally configured for express items only (i.e., the location of airports is determined by express demand only and this, in turn, specifies the excess capacity available). It is further assumed that in fully integrated scenarios the maximum amount of deferred items is shifted to the air network. Therefore, the variable $\omega_b(x)$ can be considered input data to the model, once the non-integrated network has been optimized.

The shipment size from a terminal or a customer is a function of operation headway and demand parameters, and can be replaced in the final model, using constraints (3.9e-3.9h). Likewise, linehaul distance $r_l^m(x)$ is a function of terminal density and can be replaced as well, using constraints (3.9i) and (3.9j). The only remaining variables in the complete cost model are the terminal densities and operating headways, as claimed above.

In addition, the logarithmic sorting terms involving $\Delta_C(x)$ cancel each other out if one assumes that terminals have the same cost of sorting c_k. With a few additional assumptions, the logarithmic sorting terms involving $\Delta_B(x)$ and $\Delta_P(x)$ cancel out as well. Since the number of breakbulk terminals is not expected to change significantly between regions as a result of the optimization and the log function mutes any differences, $\bar{\Delta}_B$ can be accurately approximated by $\Delta_B(x)$ and $\bar{\Delta}_P$ by $\Delta_P(x)$. (This approximation will systematically overestimate costs by Jensen's inequality. Fortunately, as the area of the service region increases, the amount of this overestimation should become quite small.) Furthermore if $\lambda_i^G(x) \approx \lambda_o^G(x)$ then the log terms involving $\Delta_B(x)$ should cancel out; e.g., $\lambda_i^G(x)c_k \log\left(\frac{1}{\Delta_B(x)}\right) + \lambda_o^G(x)c_k \log(\Delta_B(x)) \approx 0$. Unlike the first approximation, we do not know a priori the direction of the error caused by this approximation. However, since $\bar{\lambda}_i^G(x) = \bar{\lambda}_o^G(x)$, we do know that it must be a second order effect and likely quite small. In the validation section of Chapter 4, it is shown that the effects of both approximations on the optimization results are minimal. The approximation $\bar{\Delta}_B \approx \Delta_B(x)$ is used in defining ground network longhaul shipment size and number of stops as well.

After making these changes, the complete model can be further simplified by grouping the terms independent of decision variables into one constant Π,

$$\begin{aligned}\Pi =& \lambda_T(x)\left(c_d^{t'} - \frac{c_d^t k(\delta(x))^{-\frac{1}{2}}}{V_0}\right) + \lambda_T(x)c_d^{t'} + \lambda_o^G(x)c_d^{t'} + \lambda_T^A(x)\left(c_d^{a'} + 2c_h\bar{h}\right) \\ &+ \lambda_i(x)c_k \log\left(\frac{\delta(x)}{V_0}\right) + \lambda_o(x)c_k \log(2) + 2\lambda_T(x)c_f' + \lambda_i(x)c_k \log(|\mathcal{A}|)\end{aligned}$$

where it is understood that Π is a function of x. Further economy of notation can be achieved with the following abbreviations (again the dependence on x has been omitted): Coefficients for local operating headways:

$$\alpha_1^b = \lambda_b c_h;\ b = i, o \qquad\qquad \alpha_2 = c_d^t k(\delta)^{\frac{1}{2}} + c_q^t \delta$$

Coefficients for consolidation terminal densities and access operating headways:

$$\beta_1 = \lambda_T\left(\frac{\frac{2}{3}c_d^t}{\sqrt{\pi}V_0} - \frac{c_d^t k}{V_1}\right) + \frac{\frac{2}{3}|\lambda_o - \lambda_i|}{\sqrt{\pi}V_0}c_d$$

$$\beta_2 = c_d^t k \qquad \beta_3 = c_q \qquad \beta_4^b = \lambda_b^A\frac{c_h}{2} \qquad \beta_5^b = \lambda_b^G c_h;\ b = i, o \qquad \beta_6 = c_f$$

Coefficients for airport densities:

$$\chi_1 = \lambda_T^A\left(\frac{\frac{2}{3}c_d^t}{\sqrt{\pi}V_1}\right) + \frac{\frac{2}{3}|\lambda_o^A - \lambda_i^A|}{\sqrt{\pi}V_1}c_d$$

$$\chi_2 = c_f + \sum_{b=i,o}\frac{r_2^A c_d + n_2^{A,b}c_q}{n_2^{A,b}} \qquad \chi_3 = \sum_{b=i,o}\frac{n_2^{A,b} - 1}{n_2^{A,b}}c_d k$$

Coefficients for breakbulk terminal densities and longhaul operating headways:

$$\kappa_1 = \lambda_o^G\left(\frac{\frac{2}{3}c_d^t}{\sqrt{\pi}V_1}\right) + \frac{\frac{2}{3}|\lambda_o^G - \lambda_i^G|}{\sqrt{\pi}V_1}c_d - \bar{\lambda}^G c_d\frac{k}{V_2^G}$$

$$\kappa_2 = c_f \qquad \kappa_3 = c_d k \qquad \kappa_4 = c_q \qquad \kappa_5 = \lambda_o^G c_h$$

With these changes, the complete model can be written:

$$\min z = \int_{x^o\in\mathcal{A}}\Bigg\{\int_{x^i\in\mathcal{A}}\Bigg\{\sum_{b=i,o}\left(\alpha_1^b h_0^b(x) + \alpha_2(h_0^b(x))^{-1}\right)$$
$$+ \beta_1\Delta_C(x)^{-\frac{1}{2}} + \sum_{b=i,o}\sum_{e=A,G}\left(\beta_2\frac{\sqrt{\Delta_C(x)}}{h_1^{m,b}(x)} + \beta_3\frac{\Delta_C(x)}{h_1^{m,b}(x)} + \beta_4^b h_1^{m,b}(x)\right) + \beta_6\Delta_C(x)$$
$$+ \chi_1\Delta_P^{-\frac{1}{2}}(x) + \chi_2\Delta_P(x) + \chi_3\Delta_P^{\frac{1}{2}}(x)$$
$$+ \kappa_1\Delta_B^{-\frac{1}{2}}(x) + \kappa_2\Delta_B(x) + \kappa_3\frac{\Delta_B(x)^{\frac{3}{2}}}{h_2^G(x)} + \kappa_4\frac{\Delta_B(x)^2}{h_2^G(x)} + \kappa_5 h_2^G(x) + \Pi\Bigg\}dx^i\Bigg\}dx^o \tag{3.10a}$$

subject to:

$$\frac{\lambda_b(x)\delta(x)}{N_0 V_0} \leq h_0^b(x) \leq \frac{\lambda_b(x)\delta(x)}{V^0} \qquad \forall b \in \mathcal{B} \tag{3.10b}$$

$$\frac{\lambda_b^m(x)}{V_1} \leq \frac{\Delta_C(x)}{h_1^{m,b}(x)} \leq \frac{N_1^m \lambda_b^m(x)}{V_1} \qquad \forall b \in \mathcal{B}; m \in \mathcal{N} \tag{3.10c}$$

$$\frac{\lambda_b^A(x)}{V_2^A} \leq \Delta_P(x) \leq \frac{N_2^A \lambda_b^A(x)}{V_2^A} \qquad \forall b \in \mathcal{B} \tag{3.10d}$$

$$\frac{\bar{\lambda}^G}{V_2^G} \leq \frac{\Delta_B(x)^2}{h_2^G(x)} \leq \frac{N_2^G \bar{\lambda}^G}{V_2^G} \tag{3.10e}$$

$$0 < h_l^{m,b}(x) \leq H_l \qquad \forall b \in \mathcal{B}; m \in \mathcal{N}; l \in \mathcal{L} \tag{3.10f}$$

$$\Delta_P(x) \geq \frac{1}{\rho^2 \pi} \tag{3.10g}$$

The objective function now contains a significantly smaller set of decision variables, consisting only of terminal densities and headways. The constraints are a subset of those in the original model.

Writing the total cost as it appears in equation (3.10), it becomes clear that the problem decomposes by sets of decision functions. There are five distinct groups of decision functions that are not linked to the each other either in the objective function or the constraints. Therefore, the total cost model can be separated into five subproblems that determine the following variables:

1^o local outbound headways, $h_0^o(x)$

1^i local inbound headways, $h_0^i(x)$

2 consolidation terminal densities and access headways, $\Delta_C(x), h_1^{m,b}(x)$

3 airport densities, $\Delta_P(x)$

4 breakbulk terminal densities and longhaul ground headways, $\Delta_B(x), h_2^G(x)$

Subproblems 1^o and 1^i

We start with the first two subproblems to find inbound and outbound operating headways. These subproblems are the easiest to solve and help to introduce more complicated subproblems later. For $b = i, o$, the subproblems are:

$$\min z_{1^b} = \int_{x^o \in \mathcal{A}} \left(\int_{x^i \in \mathcal{A}} \left(\alpha_1^b h_0^b(x) + \frac{\alpha_2}{h_0^b(x)} \right) dx^i \right) dx^o \tag{3.11a}$$

subject to:

$$\frac{\lambda_b(x)\delta(x)}{N_0 V_0} \leq h_0^b(x) \leq \frac{\lambda_b(x)\delta(x)}{V_0} \tag{3.11b}$$

$$0 < h_0^b(x) \leq 1 \tag{3.11c}$$

Note that the subproblems can be further decomposed by area because the integrand and the constraints are local in nature. Hence, one can simply minimize the integrand for an area about x and then sum across all such subdivisions of the total area. This is easy to do because the integrand is a simple economic order quantity (EOQ) problem. The optimal headway is $h_0^b(x)^* = \sqrt{\frac{\alpha_2}{\alpha_1^b}}$, provided all constraints are met. Otherwise, the optimal solution will exist at one of the extreme points defined by the constraints. The solution should be intuitive. With higher transportation costs, headways should be increased, and with higher rent costs, headways should be lowered.

Subproblem 2

Subproblem 2 searches for the optimal consolidation terminal densities and access headways. It is defined as:

$$\min z_2 = \int_{x^o \in \mathcal{A}} \left(\int_{x^i \in \mathcal{A}} \left(\beta_1 \Delta_C(x)^{-1/2} + \sum_{b=i,o} \sum_{m=A,G} \left(\beta_2 \frac{\sqrt{\Delta_C(x)}}{h_1^{m,b}(x)} + \beta_3 \frac{\Delta_C(x)}{h_1^{m,b}(x)} + \beta_4^b h_1^{m,b}(x) \right) + \beta_6 \Delta_C(x) \right) dx^i \right) dx^o \tag{3.12a}$$

subject to:

$$\frac{\lambda_b^m(x)}{V_1} \leq \frac{\Delta_C(x)}{h_1^{m,b}(x)} \leq \frac{N_1 \lambda_b^m(x)}{V_1} \qquad b = i, o; m = A, G \qquad (3.12b)$$

$$0 < h_1^{m,b}(x) \leq H_1 \qquad b = i, o; m = A, G \qquad (3.12c)$$

As shown by (3.12), subproblem 2 is also a local problem, and therefore the concept of decomposition by area can be used. The decomposed minimization problem is more complicated than subproblem 1 because it contains a non-convex objective function and non-linear inequality constraints. However, it is possible to transform (3.12) into a convex problem with linear inequality constraints by introducing the following changes of variable:

$w_C = \ln(\Delta_C(x))$, $w_{m,b} = \ln(h_1^{m,b}(x)), b = i, o; m = A, G$

The transformed version of subproblem 2 is:

$$\min z_{2'} = \int_{x^o \in A} \left(\int_{x^i \in A} \left(\beta_1 e^{-\frac{w_C}{2}} + \sum_{b=i,o} \sum_{m=A,G} \left(\beta_2 e^{\frac{w_C}{2} - w_{m,b}} + \beta_3 e^{w_C - w_{m,b}} + \beta_4^b e^{w_{m,b}} \right) + \beta_6 e^{w_C} \right) dx^i \right) dx^o \qquad (3.13a)$$

subject to:

$$\ln\left(\frac{\lambda_b^m(x)}{V_1}\right) \leq w_C - w_{m,b} \leq \ln\left(\frac{N_1 \lambda_b^m(x)}{V_1}\right) \qquad b = i, o; m = A, G \qquad (3.13b)$$

$$w_{m,b} \leq ln(H_1) \qquad b = i, o; m = A, G \qquad (3.13c)$$

Since the transformed subproblem is convex, it can be solved using gradient search techniques for non-linear problems. For non-integrated networks, two versions of subproblem 2 are solved for each network. In integrated networks, the subproblems are joined. Although each network has separate headways, the consolidation terminals are shared across networks.

Subproblem 3

The same spatial decomposition and logarithmic transformation techniques reduce subproblems for 3 and 4 to simple convex programs. Recall that the sorting terms at break-bulk terminals and the main air hub are not local, yet when $\bar{\Delta}_P$ is estimated by $\Delta_P(x)$, and $\bar{\Delta}_B$ by $\Delta_B(x)$, these terms are local. Looking first at airport terminal density, subproblem 3 is defined as:

$$\min z_3 = \int_{x^o \in \mathcal{A}} \left(\int_{x^i \in \mathcal{A}} \left(\chi_1 \Delta_P^{-\frac{1}{2}}(x) + \chi_2 \Delta_P(x) + \chi_3 \Delta_P^{\frac{1}{2}}(x) \right) dx^i \right) dx^o \tag{3.14a}$$

subject to:

$$\frac{\lambda_b^A(x)}{V_2^A} \leq \Delta_P(x) \leq \frac{N_2^A \lambda_b^A(x)}{V_2^A} \qquad b = i, o \tag{3.14b}$$

$$\Delta_P(x) \geq \frac{1}{\rho^2 \pi} \tag{3.14c}$$

The following change of variable is introduced: $w_P = \ln(\Delta_P(x))$. The transformed version of subproblem 3 is:

$$\min z_{3'} = \int_{x^o \in \mathcal{A}} \left(\int_{x^i \in \mathcal{A}} \left(\chi_1 e^{-\frac{w_P}{2}} + \chi_2 e^{w_P} + \chi_3 e^{\frac{w_P}{2}} \right) dx^i \right) dx^o \tag{3.15a}$$

subject to:

$$\ln\left(\frac{\lambda_b^A(x)}{V_2^A} \right) \leq w_P \leq \ln\left(\frac{N_2^A \lambda_b^A(x)}{V_2^A} \right) \qquad b = i, o \tag{3.15b}$$

$$w_2 \geq ln(\frac{1}{\rho^2 \pi}) \tag{3.15c}$$

The optimal density of airports does not change from non-integrated to multimode multiservice networks. While deferred items may use these facilities, these items do not impact the airport location decisions. Therefore, subproblem 3 need only be solved once. Further, the decision of shifted demand is made exogenously, such that $\frac{n_2^{A,b}(x)}{\Delta_P(x)} \lambda_b^A \leq \nu V_2^A$.

The shift of deferred items to excess capacity in air is treated locally, although this too is not a local decision. The impact of this approximation on the optimal network cost is tested in the next chapter.

Subproblem 4

The density of breakbulk terminals and operating headways for longhaul ground transportation may change after integration. The same solution techniques are used, except $\bar{\lambda}^G$ is redefined by the average shift of deferred items to the air network. With the approximations of longhaul ground covered in Section 3.3.3 and vehicle repositioning between breakbulk terminals covered in Section 3.3.4, the entire longhaul ground network can be treated as a local problem. Subproblem 4 is:

$$\min z_4 = \int_{x^o \in \mathcal{A}} \left(\int_{x^i \in \mathcal{A}} \left(\kappa_1 \Delta_B^{-\frac{1}{2}}(x) + \kappa_2 \Delta_B(x) + \kappa_3 \frac{\Delta_B(x)^{\frac{3}{2}}}{h_2^G(x)} + \kappa_4 \frac{\Delta_B(x)^2}{h_2^G(x)} + \kappa_5 h_2^G(x) \right) dx^i \right) dx^o \quad (3.16a)$$

subject to:

$$\frac{\bar{\lambda}^G}{V_2^G} \leq \frac{\Delta_B(x)^2}{h_2^G(x)} \leq \frac{N_2 \bar{\lambda}^G}{V_2^G} \quad (3.16b)$$

$$0 < h_2^G(x) \leq H_2^G \quad (3.16c)$$

The terminal density and headway variables are transformed as follows: $w_B = \ln(\Delta_B(x))$, $w_2 = \ln(h_2^G(x))$. This results in the following transformed version of subproblem 4:

$$\min z_{4'} = \int_{x^o \in \mathcal{A}} \left(\int_{x^i \in \mathcal{A}} \left(\kappa_1 e^{-\frac{w_B}{2}} + \kappa_2 e^{w_B} + \kappa_3 e^{\frac{3w_B}{2} - w_2} + \kappa_4 e^{2w_B - w_2} + \kappa_5 e^{w_2} \right) dx^i \right) dx^o \quad (3.17a)$$

subject to:

$$\ln\left(\frac{\bar{\lambda}^G}{V_2^G}\right) \leq 2w_B - w_2 \leq \ln\left(\frac{N_2\bar{\lambda}^G}{V_2^G}\right) \tag{3.17b}$$

$$w_2 \leq ln(H_2^G) \tag{3.17c}$$

The entire cost model (3.9) can now be solved as a series of convex subproblems. This is done in the next chapter.

3.5 Model enhancements

Thus far, cost models have been formulated and optimization techniques developed for a distribution network with deterministic, time-independent demand. As mentioned earlier, there are three main benefits of integration: merged operation of facilities, local distribution savings, and more efficient utilization of longhaul ground vehicles and aircraft. These benefits have the potential to increase significantly when demand is uncertain or when seasonal fluctuations in demand exist. Therefore, in this section, the above analysis is generalized for the study of scenarios involving both time-independent but uncertain demand and time-dependent demand.

3.5.1 Uncertainty in demand

This section presents the case of stationary, random demand. Variations in demand are modeled as a stationary process with independent increments and a location-dependent index of dispersion (variance to mean ratio). More specifically, the demand in any time interval between any two regions of small area (e.g., about points x^o and x^i) are assumed to be independent of other demands if at least one of the following conditions is satisfied: (1) the two origin areas do not overlap; (2) the two destination areas do not overlap; (3) the two time intervals do not overlap. Inbound and outbound demands in a region have variance to mean ratios $\gamma_b^s(x), s \in \mathcal{S}, b \in \mathcal{B}$ (*items*). It is assumed that these values do not vary over time.

Controlling uncertainty in the non-integrated air network

Because of different characteristics, uncertainty is treated differently for each mode and service level. In the air network for express items, uncertainty is addressed by overdesigning the network to minimize the possibility of demand exceeding capacity. To accommodate this uncertainty in the design process, V_l^A is replaced with a smaller quantity $\theta_l^{A,b} V_l^A$ for some positive $\theta_l^{A,b} < 1$, such that:

$$\theta_l^{A,b} V_l^A + 3\sqrt{\theta_l^{A,b} \gamma^{E,b} V_l^A} \leq V_l^A \qquad \forall l \in \mathcal{L} \tag{3.18}$$

Recall that in local distribution, the capacity V_0^A may not represent a physical capacity, but rather the number of items that can be realistically delivered within time constraints depending on customer density. This quantity is reduced as in equation (3.18) when uncertainty is present. Across many days, this would leave an average excess capacity of $(1 - \theta_l^{A,b}) V_l^A$ in all air network vehicles, equivalent to three standard deviations of the expected vehicle load, but would ensure that overflows would be unlikely. Note from (3.18) that the capacity buffer increases with the level of demand uncertainty which can become extremely costly.

It may be possible to design systems that can be operated more efficiently in practice if one allows airplanes and air network ground vehicles to be rerouted dynamically to cover for other consolidation terminals and airports, as information becomes available. Express delivery carriers take advantage of this possibility, for example by designating some airplanes as "sweepers". Some such strategies have been discussed in the context of vehicle routing problems for ground vehicles in Daganzo and Erera (1999) and Erera (2000). Although the air transport problem is slightly different because airplanes are more limited in the number of stops that can be made en route to the hub, it is amenable to a similar treatment. Formulas can be developed to predict both the cost and the average amount of unused cargo space that would result from more realistic air strategies. In this case, the amount of wasted space will likely be lower than from (3.18), but still high enough to be a problem. This if left for future analysis.

Controlling uncertainty in the non-integrated ground network

In the ground network, additional strategies to handle uncertainty, such as rerouting vehicles, are easier to implement due to relaxed time constraints. Recall earlier discussions on deterministic empty vehicle repositioning in Section 3.3.4 and Appendix A. When uncertainty in demand is introduced in the ground networks, vehicles still travel full, but routing may change slightly each day. This does not change the full vehicle miles traveled at all levels, nor does it change the peddling costs. However, the need for empty vehicle repositioning increases. It is assumed that longhaul and access vehicles can be shared between terminals; empty vehicles may be rerouted to accommodate demand fluctuations. In particular, access vehicles may be repositioned at breakbulk terminals and airports and longhaul ground vehicles may need to be repositioned between breakbulk terminals. The goal is to maintain a constant supply of vehicles at terminals. No repositioning of delivery vans occurs between consolidation terminals because it is assumed that a sufficient supply of delivery vans exist at each consolidation terminal and fluctuations in demand can be absorbed by holding items across days.

Note that the number of empty vehicles systematically repositioned to (from) a terminal is equal to the difference between departures from and arrivals to that terminal. For a terminal serving customers in a surrounding area $\Delta_y^{-1}(x)$, the average inflow of vehicles is equal to the average outflow with the inclusion of deterministic repositioning introduced in Section 3.3.4. However, the standard deviation of the flows, $\sigma_y(x)$, is $\sqrt{\frac{\gamma_i^G(x)\lambda_i^G(x)+\gamma_o^G(x)\lambda_o^G(x)}{\Delta_y(x)V}}$, and this may require further (stochastic) adjustments.

The cost of repositioning vehicles due to stochastic effects alone can be approximated as a transportation problem of linear programming in a similar manner as the deterministic repositioning of vehicles between breakbulk terminals (see Section 3.3.4 and Appendix C). This cost is then added to the deterministic repositioning cost in the complete cost model to reflect these costs.[6] The average number of total empty vehicle miles across many days required to reposition trucks at the least cost each day is a function of the total area of the service region $\mathcal{A}$, the number of terminals, $N = \bar{\Delta}_y(x)|\mathcal{A}|$, and $\sigma_y(x)$. The stochastic repositioning cost per terminal per unit time is:

$$z_{reposition} = c_d^t \sigma_y(x) \Delta_y(x)^{-\frac{1}{2}} \left(1 + 0.078 \log_2(\bar{\Delta}_y(x)|\mathcal{A}|) \right)$$

In order to apply the area-decomposition solution technique, the expected terminal density must be replaced with the local terminal density, as is done with sorting costs.

Opportunities from integration

When networks are integrated, it is possible to fill excess capacity on air network vehicles with deferred items. The flexible deadlines of deferred items allow vehicles to travel full and hold items at terminals if sufficient capacity is not available. For express items, the reduced capacity is used in constraints, $n_l^{A,b}(x)\left(\frac{\lambda_b^E(x)}{\Delta_y(x)} h_l^{A,b}(x)\right) \leq \theta_l^{A,b} V_l^A$. However, for the total air network vehicle load, including less critical deferred items, the full physical capacity is used, $n_l^{A,b}(x)\left(\frac{\lambda_b^E + \omega_b(x)\lambda_b^D(x)}{\Delta_y(x)} h_l^{A,b}(x)\right) \leq V_l^A$.[7]

These changes for uncertain demand are implemented in the next chapter.

[6]This is conservative since this sum is the average cost obtained by a superposition of the deterministic solution and the TLP solution including only the stochastic deviation from the mean, which is a feasible (sub-optimal) solution of the real problem.

[7]For longhaul air routes, the right-hand side is νV_2^A.

3.5.2 Time-dependent demand

Other generalizations that incorporate time-dependent features of the problem more explicitly can be added to the logistic cost models as well. For systematic fluctuations (such as holiday demand surges), leased capacity is an attractive option. Barnhart and Schneur (1996) consider the use of commercial flights in the express delivery market. Typically, long term decisions such as aircraft acquisition and infrastructure investments are made based on peak demand. However, operating decisions for routing are made on a more continual basis to accommodate different demand levels. In the complete model (3.9), a third dimension for time can be added, τ, within a time horizon $\mathcal{H}$.

$$\min \int_{x^o \in \mathcal{A}} \int_{x^i \in \mathcal{A}} \int_{\tau \in \mathcal{H}} \{\text{ complete model, (3.9)}\} dx^i dx^o d\tau$$

The same solution techniques can be applied. Rather than simply decomposing by area, one can decompose by time as well by discretizing time into periods of near stationary demand. The terminal locations and aircraft fleet size should be determined by peak demand and then headways can be optimized within time regions. In these cases, integration should bring larger cost savings since large seasonal fluctuations will lead to more severely underutilized air vehicles and air terminals during off-peak periods. A detailed description of this work, however, is beyond scope of this thesis.

3.6 Concluding remarks

This chapter presents the functional form of logistic cost function with the solution techniques required to analyze complex networks. Extensions to the cost models can be added to incorporate a wider range of demand assumptions. In the next chapter, these models are used for network design and scenario analysis. The impact of demand uncertainty on integration is studied as well.

Chapter 4

System design: evaluation and validation

4.1 Introduction

The logistic cost models and solution techniques presented in the previous chapter are applied here to design multimode, multiservice distribution networks and compare potential savings achieved from different levels of integration. The estimation of cost, demand and operating parameters for cost models introduced in Chapter 3 is covered in Section 4.2. Attempts are made to obtain realistic parameter estimates. However, as a result of modeling simplifications and imperfect information, there are differences between the design and operating plans developed here and those found in the industry today. In this research, we are asking general systematic questions with the goal of formalizing the process of design, operation, and evaluation of complex integrated logistics systems. With the methodology developed, package delivery companies can be more proactive and explore a wider range of "what if" scenarios. The results presented in Section 4.3 provide valuable insight into real world applications. In Section 4.4, the cost models are validated.

4.2 Description of test cases

The information required to estimate parameters is often difficult to obtain; only a portion of this information is public. One can obtain vehicle fleet details, location of larger network hubs, sorting time windows, and the total area of the service region from company literature such as annual reports and SEC Form 10K's. Detailed demand and cost data, however, can not be found from public information. The list of available information clearly expands when research is performed with the companies themselves, see Crainic and Rousseau (1986), Barnhart and Schneur (1996), and Kim *et al.* (1999). While package delivery carriers have been consulted as part of this research, there has been no direct sharing of private demand and cost information. Therefore, the cost data used in this research is based on earlier work on cost estimates by Han (1984), Han and Daganzo (1985), and Kiesling (1995). Public company literature from public package delivery companies complement these studies, Federal Express Corporation (1998a), Federal Express Corporation (1998b), and United Parcel Service (2000b). For specifics on cost data, see Appendix B.

In this chapter, a variety of geographic and demand/customer related assumptions are tested to identify the conditions most amenable to mode and service level integration. Recall from the introduction the two distinct views on integration held by UPS, a dominant deferred carrier, and Federal Express, a dominant express carrier. To assess these opposing views, networks of both types are studied. Further network/demand assumptions are included in the test cases to predict impacts of future trends in demand. As e-commerce grows, demand densities are changing. The extent to which these changes will continue is still unknown. Package delivery is switching from a primarily business-to-business context to a combination of business-to-business, business-to-consumer, consumer-to-business, and consumer-to-consumer. The continuous approximation models, used in conjunction with different demand density estimates, can be extremely valuable in evaluating network costs under different e-commerce scenarios. Further, this chapter examines the impact of demand uncertainty on the savings achieved through integration.

Two different service regions are studied in this research and a series of test cases are developed for each service region. Both service regions are comprised of subregions which are large enough to contain multiple terminals, yet small enough such that average network characteristics (demand levels, distances to main air hub, etc.) are representative of the entire subregion.

The first of these service regions, denoted SR1, is an idealized service area designed to reflect a wide array of network characteristics. As shown in Figure 4.1, the service region contains seventeen subregions, with different geographic features and customer densities, as well as demand rates and indices of dispersion (uncertainty level) in each direction.

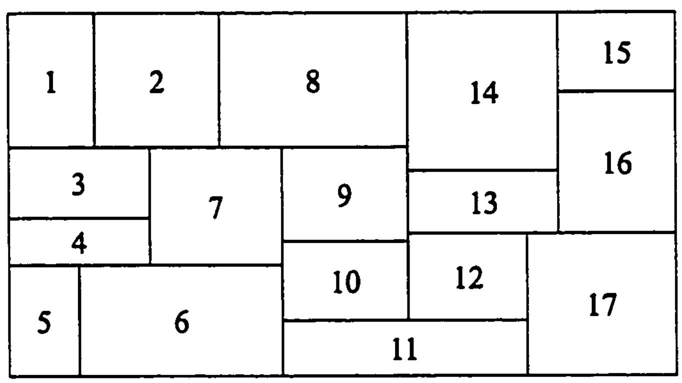


Figure 4.1: Service region 1: idealized network

The entire region is 125,000 mi^2, roughly the size of a medium-sized country such as Italy, Norway or Japan. The areas of the seventeen subregions range from 3,333 mi^2 to 12,000 mi^2, with an average of 7,353 mi^2. By studying the individual subregions, one can gain insights into the demand and geographic conditions that favor integration. See Appendix B for more details.

The second service region, SR2, is larger, roughly the size of the lower forty-eight United States, again with multiple subregions. Demand and geographic features are based on characteristics of the United States. While demand data is private, a common proxy for demand data used in the literature is population data (see, for example, Hall (1989a), Campbell (1993) and Kuby and Gray (1993)). The 1990 U.S. census includes population, housing counts, land area and geographic coordinates for the top Metropolitan Statistical Areas (MSA) in the United States; see U.S. Census Bureau (1990). Population density is used as a proxy for package demand level and housing density as a proxy for customer density. These MSA's are aggregated into groups of common demand and geographic features to form twenty subregions for SR2. The entire service region is 2,500,000 mi^2. The areas of the twenty subregions range from 16,383 mi^2 to 225,974 mi^2, with an average of 120,000 mi^2. More detailed information on the subregions can be found in Appendix B.

Twelve test cases, described in Table 4.1, are analyzed. The test cases are identified by service region ("SR1" or "SR2"), demand information available ("K" for known demand and "R" for random demand), and dominant service level ("D" for networks that primarily offer deferred services, "E" for networks specializing in express service levels, and "B" for networks with approximately balanced demand between the two service levels).

4.3 Network design and scenario evaluation

A key advantage of continuous approximation methods is in the simplicity of implementation. The entire cost model (3.9) can be solved for complex multimode, multiservice networks using standard spreadsheet technology. As a result, it is easy to test a variety of network assumptions, and determine how network configurations differ between test cases and integration levels.

The integration strategies described in Chapter 2 are tested against the non-integrated base case (BC). The strategies are as follows: facility-integrated networks with existing infrastructure (I1), facility-integrated networks with reduced infrastructure (I2), fully-

Scenario	Express Service		Deferred Service	
	Daily Demand	Number of Customers	Daily Demand	Number of Customers
SR1[K,B]	538,160	234,349	577,891	302,855
SR1[K,D]	72,944	37,220	722,363	361,506
SR1[K,E]	549,143	276,957	354,949	177,693
SR1[R,B]	538,160	234,349	577,891	302,855
SR1[R,D]	72,944	37,220	722,363	361,506
SR1[R,E]	549,143	276,957	354,949	177,693
SR2[K,B]	1,808,065	852,759	1,757,581	682,207
SR2[K,D]	878,791	341,103	8,824,772	1,705,517
SR2[K,E]	2,998,389	852,759	1,446,452	682,207
SR2[R,B]	1,808,065	852,759	1,757,581	682,207
SR2[R,D]	878,791	341,103	8,826,680	1,705,517
SR2[R,E]	2,998,389	852,759	1,446,452	682,207

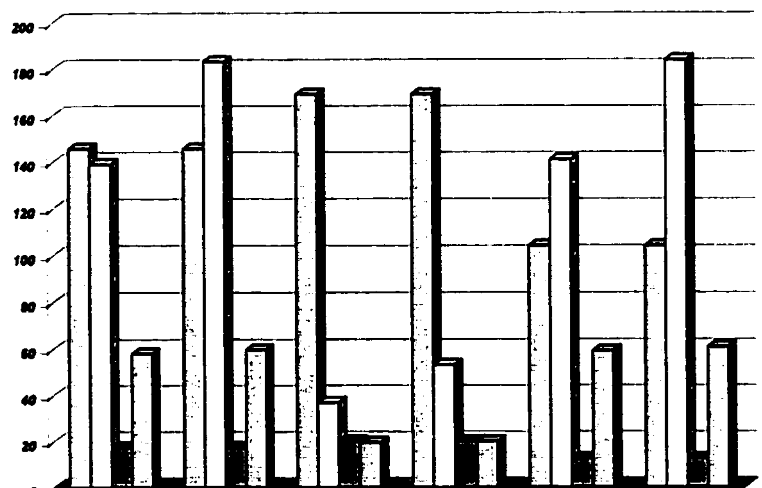
Table 4.1: Test case descriptions

integrated networks with reduced infrastructure (I3), and finally fully-integrated networks with reoptimized infrastructure (I4). In Section 4.3.1, the network configurations for integrated networks are compared against the base case to see how the infrastructure changes with integration scenarios. The costs of integration and the savings achieved are analyzed in Section 4.3.2. Chapter 5 explains how to translate these network configurations into a more detailed operating strategy.

4.3.1 Network configuration analysis

Terminal counts over all subregions in the non-integrated networks are presented for SR1 in Figure 4.2 and for SR2 in Figure 4.3. Because the two networks are operated independently, separate counts are presented for the air-network and ground-network consolidation terminals. The hierarchical structure of the distribution networks is clear from these figures. Each network consists of a large number of consolidation terminals that feed into a significantly smaller number of breakbulk terminals and airports. As the figures show, in the test cases with deterministic demand, consolidation terminal counts increase with service level demand. However, demand is not the only factor; this can be seen by comparing SR1[K,D] with SR1[K,E]. The first test case has twice the deferred demand level of the second, but only 1.6 times the number of ground consolidation terminals. The operational constraints, including those on headways and number of stops, account for this difference.

When demand is uncertain, it is necessary to overdesign the air portion of the network, and this is clear from the increase in air consolidation terminals. Local vehicles carry smaller average loads in uncertain air networks (see equation 3.14) and this increases the average transportation cost per item. With higher transportation costs, there is an incentive to reduce the linehaul distance between consolidation terminals and customers by increasing the density of consolidation terminals. The same effect also occurs with airport density and access costs, but it is less significant because the access and longhaul vehicles are larger and therefore less affected by uncertainty. More importantly, operational constraints on airport density (maximum service radius restrictions and the limit of two stops on longhaul trips) tend to overcapacitate the air network even in the deterministic cases. As a result, airport densities increase less significantly. The figures also show that the ground network is less sensitive to demand uncertainty. This is not surprising since it is assumed that ground vehicles can be repositioned between terminals within a time frame consistent with deferred deadlines and that items can be held over time to smooth stochastic fluctuations.

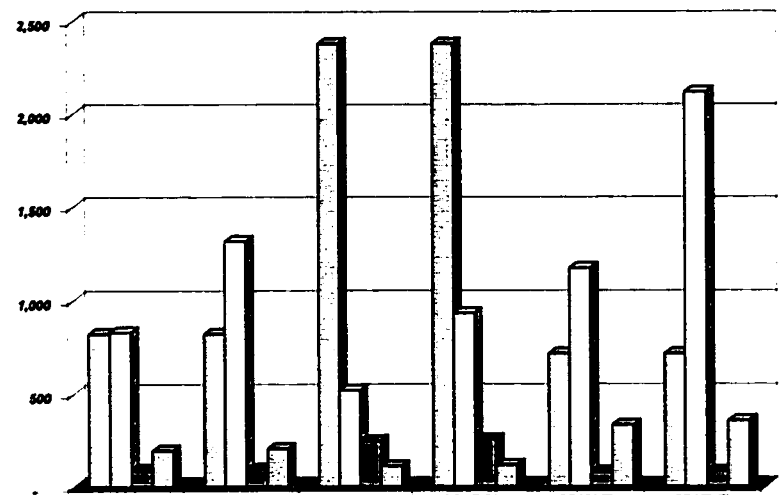


	SR1[K,B]	SR1[R,B]	SR1[K,D]	SR1[R,D]	SR1[K,E]	SR1[R,E]
# CT's (ground)	146	146	169	169	104	104
# CT's (air)	139	183	37	53	141	184
# BBT's	17	17	20	20	12	12
# airports	58	59	19	20	59	61

Figure 4.2: Base case: Terminals in idealized non-integrated networks

Test case results indicate that the largest changes in infrastructure resulting from integration occur with consolidation terminals. Figures 4.4 and 4.5 present the total (air plus ground) consolidation terminal counts for the four integration strategies (I1-I4)of SR1 and SR2, respectively.

For facility-integrated strategies (I1 and I2), the terminal counts for breakbulk ter-

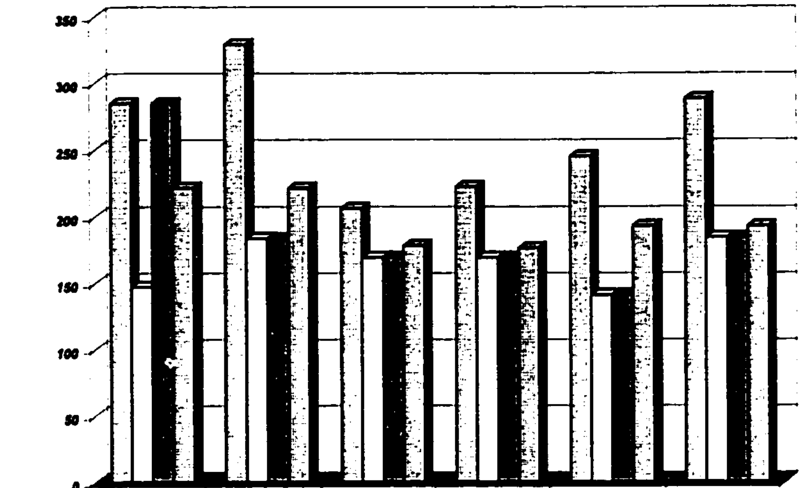


	SR2[K,B]	SR2[R,B]	SR2[K,D]	SR2[R,D]	SR2[K,E]	SR2[R,E]
# CT's (ground)	821	821	2,380	2,380	715	715
# CT's (air)	830	1,314	518	930	1,171	2,118
# BBT's	91	96	247	254	80	84
# airports	196	206	110	118	332	355

Figure 4.3: Base case: Terminals in US-based non-integrated networks

minals and airports are, by construction, the same as in the base case. With strategy I1, all existing facilities remain open and consolidation terminals may be shared between networks. The consolidation terminal count is simply the sum of the air consolidation terminal density and the ground consolidation terminal density from the base case. With strategy I2, the consolidation terminals of the smaller network are removed (i.e., ground consolidation terminals are removed in an express dominant network or air consolidation terminals in a deferred dominant network). When demand is deterministic and balanced between service levels, the densities of consolidation terminals for the two service levels are essentially the same and the total count is cut in half. The reduction in consolidation terminals is less dramatic when one service level significantly dominates the other. In the case of random demand when demand is balanced between service levels, the ground-network consolidation terminals are removed and this still leaves a fair number of (air) consolidation terminals.

When routing is integrated (I3), a choice is made between leaving all terminals open



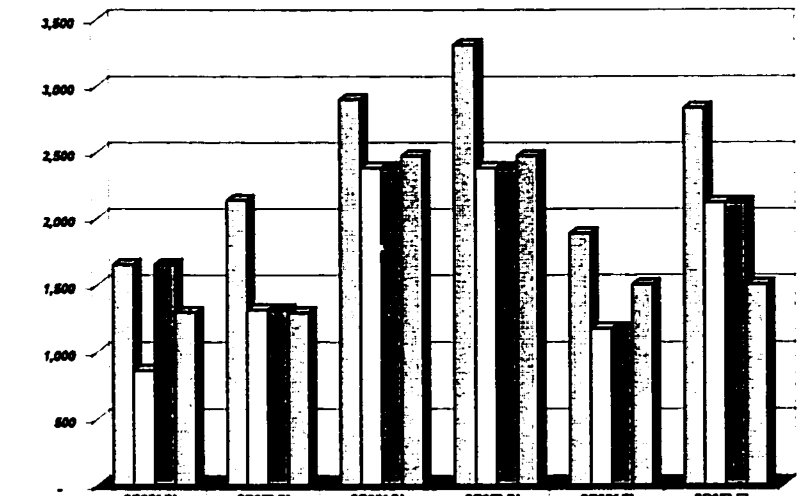
	SR1[K,B]	SR1[R,B]	SR1[K,D]	SR1[R,D]	SR1[K,E]	SR1[R,E]
# CT's (I1)	285	329	206	222	245	288
# CT's (I2)	147	183	169	169	141	184
# CT's (I3)	285	183	169	169	141	184
# CT's (I4)	221	221	178	176	193	192

Figure 4.4: Consolidation terminal counts in integrated, idealized networks

or closing the consolidation terminals of the smaller network.[1] In cases where one service level dominates the other, cost analysis suggests that it is best to close the consolidation terminals of the smaller network. In test cases with balanced, deterministic demand, removing consolidation terminals results in an underfacilitated network; therefore, cost savings are greater when all terminals are kept. However, when demand is random in these networks, there are sufficient terminals in the overdesigned air network to accommodate air and ground routing and it is cost-effective here to eliminate ground consolidation terminals.

In the final integration strategy (I4), the complete network, except for airports, is reoptimized for integrated operations. As the figures show, in some cases the terminal configurations do not change much. In others, the reduction in consolidation terminals is quite significant. A small change in the density of breakbulk terminals is observed as well,

[1]These decisions can be made at a finer scale by looking at each subregion and deciding whether or not to keep consolidation terminals open in that subregion.



	SR2[K,B]	SR2[R,B]	SR2[K,D]	SR2[R,D]	SR2[K,E]	SR2[R,E]
# CT's (I1)	1,651	2,136	2,897	3,310	1,886	2,833
# CT's (I2)	860	1,314	2,380	2,380	1,171	2,118
# CT's (I3)	1,651	1,314	2,380	2,380	1,171	2,118
# CT's (I4)	1293	1292	2471	2472	1503	1502

Figure 4.5: Consolidation terminal counts in integrated, US-based networks

but not included in the figures. One must look closer at costs and see how sensitive costs are to these changes. Since it is very costly to relocate terminals, only a large difference in cost savings between strategies I3 and I4 would justify such action.[2] In the next section, the cost analysis to facilitate these decisions is performed.

4.3.2 Cost analysis

The development of cost components in Chapter 3 include discussion on how these components change under each integration strategy. Here these changes are quantified. The analysis focuses on two specific test cases from the US-based service region, (SR2[K,B] and SR2[R,D]). Results for the remaining test cases are qualitatively similar and are also discussed.

The cost components of the two cases are presented in Figures 4.6 and 4.7. The figures

[2]It is also possible to change the number of terminals without relocating terminals, by selectively closing a fraction of existing terminals.

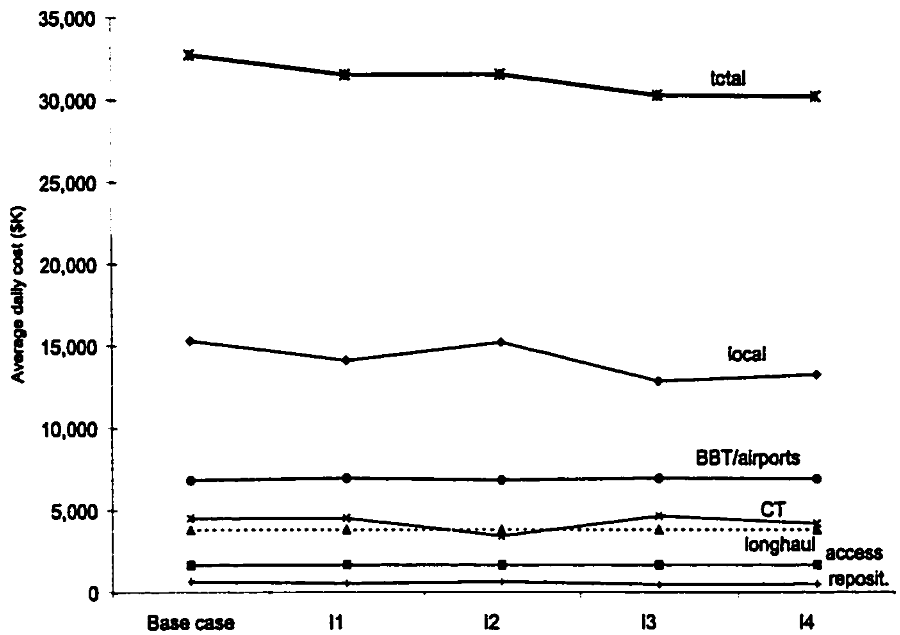


Figure 4.6: Integration cost analysis: deterministic balanced demand in SR2

plot the average daily costs for SR2[K,B] and SR2[R,D], respectively, under the base case and the four integration strategies. The total cost of operation is represented by the top line in each figure, along with various cost components as defined in Chapter 3. The two networks consist of very different demand assumptions, yet the same patterns in cost savings are observed in both cases, although the magnitude of these savings differ. The same phenomena is observed across all test cases.

One striking feature of the cost figures presented (and of all cases) is the dominance of local transportation costs. This should not be surprising since local transportation consists of many trips made in small vehicles operating on short headways. In turn, changes in local costs have a large impact on total cost. With facility integration I1, the local transportation costs decrease due to shorter linehaul distances from increased terminal densities in both SR2[K,B] and SR2[R,D]. Recall Figure 3.2 predicts this change. In addition, when looking at the cost savings in individual subregions, the total savings are greatest in regions where local transportation costs account for over 45% of total costs.

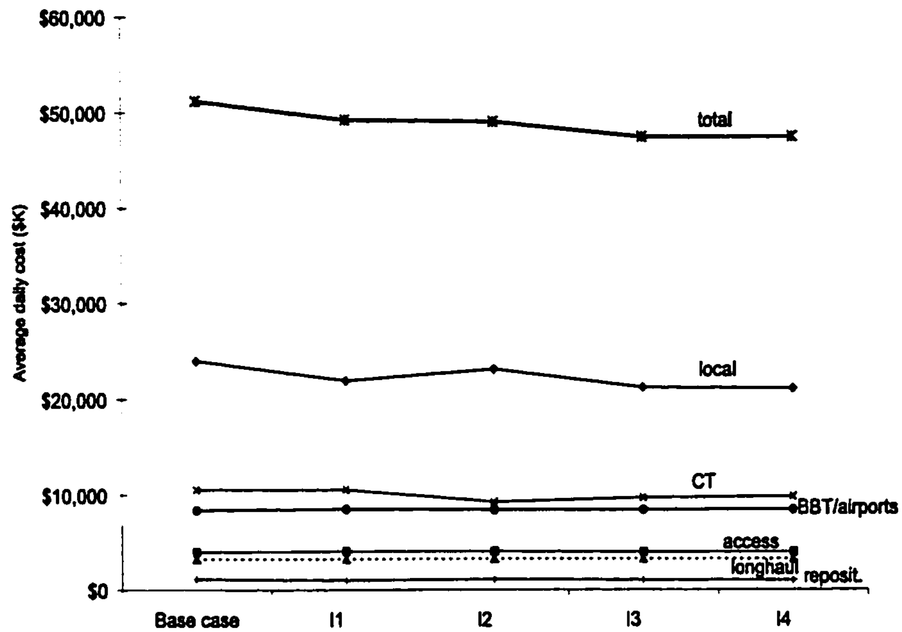


Figure 4.7: Integration cost analysis: random, deferred-dominant demand in SR2

These regions typically have low demand levels and low customer densities. With the rise of e-commerce, the importance of local distribution to individual customers should increase and the incentive for integrating local distribution between service levels should increase too.

Since each network maintains separate routing with strategy I1, items of different networks do not interact at consolidation terminals. Therefore, no additional sort is required and the costs of these terminals remain unchanged. The outbound sorting costs to consolidation terminals increase with the logarithm of the new increased consolidation terminal densities. As a result, slight increases in breakbulk terminal and airport costs occur. These changes, however, are too small to see on the figures. Furthermore, access costs increase slightly with strategy I1, and this change too is undetectable on the graphs. On access trips, vehicles may be required to visit more consolidation terminals to collect or distribute items; however, the number of detours is still fairly small compared with the number of stops on a local delivery or pickup route. Longhaul transportation costs do not change. For reasons related to the reduction in local transportation costs, repositioning costs are lower since the service radius of each consolidation terminal decreases. Repositioning costs account for a small fraction of total costs, however.

For facility integration with fewer consolidation terminals (I2), there is an obvious decrease in consolidation terminal costs. At the same time, this decrease in consolidation terminal density increases the local transportation costs, and, to a lesser extent, repositioning costs. The decrease in consolidation terminal costs is comparable to the increase in local transportation costs, and no large change in total cost is detectable between strategies I1 and I2.

When facility integration is accompanied by route integration, larger changes in total costs are observed. Recall that in balanced deterministic cases such as SR2[K,B], the total number of consolidation terminals from the base case are maintained. In random, deferred cases such as SR2[R,D], the air network consolidation terminals are removed. This difference accounts for part of the cost savings differential between Figures 4.6 and 4.7. For SR2[K,B], the drop in local transportation costs is quite sharp, and a less dramatic drop is observed for SR2[R,D]. In both cases, the density of customers on local transportation routes increases (see Figure 3.2) when air and ground networks are integrated. This lowers the detour portion of local transportation costs. In SR2[K,B] local transportation costs are further reduced as a result of the shorter linehaul distances associated with higher consolidation terminal densities. This is not the case for SR2[R,D]; however, consolidation terminal costs are considerably lower. Longhaul costs are reduced in both cases, but since the cost of longhaul transportation is small compared with local transportation costs, this change is difficult to see in the figures. In SR2[K,B], 120,000 deferred items (roughly 8% of the total deferred demand) are shifted to the air network. This has a small impact on the density of breakbulk terminals, and also allows the entire ground network to be run more efficiently. In SR2[R,D], 200,000 deferred items are shifted to the air network. Since the network is so large (the shifted items represent only 2.5% of the total deferred demand), the savings, measured as a percent of total cost, is quite small.

Interestingly, in both cases, there is not a large change in total costs between scenarios I3 and I4. In SR2[K,B], despite the 50% decrease in consolidation terminal density from I3 to I4, total costs change only 0.2%. The rise in local transportation costs almost offsets the decrease in consolidation terminal and other costs. This cost savings is not likely to justify the cost of relocating terminals. In SR2[R,D], a smaller change in consolidation terminal count is observed between the two strategies, and an even smaller change in total cost. Across all test cases, the largest change between integration strategies with existing infrastructure and relocating terminals is only 0.5%.

Impact of network and demand assumptions

The original question of what conditions make integration attractive brings with it the challenge of comparing vastly different networks in terms of geographic scale, service level mix and daily package volumes. Looking at the twelve test cases, it is difficult to say conclusively what conditions are best for integration because the base cases are so different. Comparing absolute cost savings or percentage of total cost savings can give misleading results since the magnitude of savings is not transferable between networks of different sizes. A five million dollar savings per day in a small air network might appear inconsequential to a larger ground network. Using percentage of total costs saved as a metric is again dependent on the base network. Even a financially desirable savings from integration when compared to a large ground network would be dwarfed by the size of the ground network. Fortunately the question can be answered without comparing multiple base cases if one poses the question as follows: given an existing single mode, single service network (be it ground or air), when does it make sense to integrate operations with another single mode, single service network? A natural metric to use here would be the total savings achieved (ground and air) through integration as a percent of the total pre-integration cost of the original single-mode network.

This approach is used to evaluate the merits of integrating ground service with an existing air network for SR1 and SR2, both with deterministic and random demand. A series of new test cases are considered using the air network of the SR1[*,B] and SR2[*,B] and varying the levels of deferred service demand with the goal of determining the level of deferred demand required to justify integration. Based on analysis presented in the previous section, integration strategy I3 (fully-integrated, existing infrastructure) is the integration strategy of choice and costs savings are based on the adoption of this strategy.

In Figures 4.8 and 4.9, the results are presented for SR1 and SR2, respectively, with a fixed average level of express demand of 538,160 packages per day for SR1 and 1,808,065 packages per day for SR2. Values on the x-axis represent the total average deferred demand and the figures indicate the point at which deferred demand exceeds express demand. Values on the y-axis plot the total (air and ground) network savings divided by the total pre-integration air network costs. On each graph, deterministic and random demand scenarios are tested.

One can see from both figures that, for low levels of deferred demand, savings are small. Express carriers may be reluctant to integrate operations with deferred carriers when the level of deferred demand is less than that of express. However, savings grow quickly as deferred demand increases. At a point, the growth rate of savings decreases and savings reach an asymptote since excess air capacity is filled and the maximum benefits of local transportation integration are realized.

Of course, there are other costs and benefits to integration not considered here that could impact decisions. Integration gives carriers the ability to move deferred items quickly in response to routing problems in the ground network (weather, surge in demand,

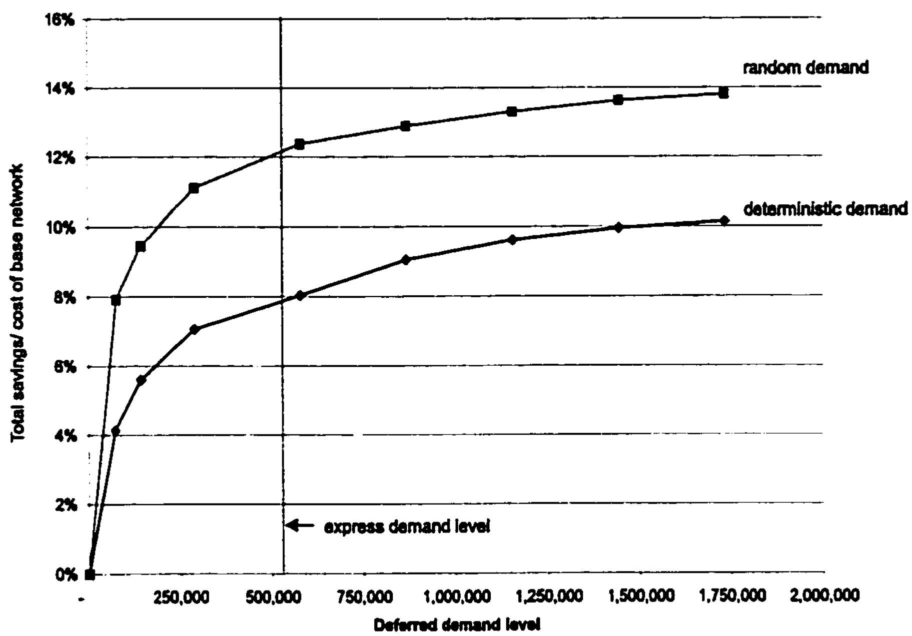


Figure 4.8: Savings comparison: idealized network

etc.). Further, overhead costs including administrative costs, sales costs, etc. can be reduced through integration. Additional savings can be achieved when it is necessary to overcapacitate for seasonal demand fluctuations.

4.4 Model validation

In this section, the approximations required as part of the solution techniques covered in Section 3.4 of Chapter 3 are validated. The cost approximations for vehicle routing models used here have been successfully compared with costs obtained with simulated annealing techniques in Robuste *et al.* (1990). Recent work by Erera (2000) has shown that the costs of even advanced local distribution strategies can be approximated within 5% of cost results from simulation. Costs from hub location decisions obtained with continuous approximations have also been validated against discrete cost models in the literature; see Campbell (1993). The validation of cost functions to estimate the repositioning costs

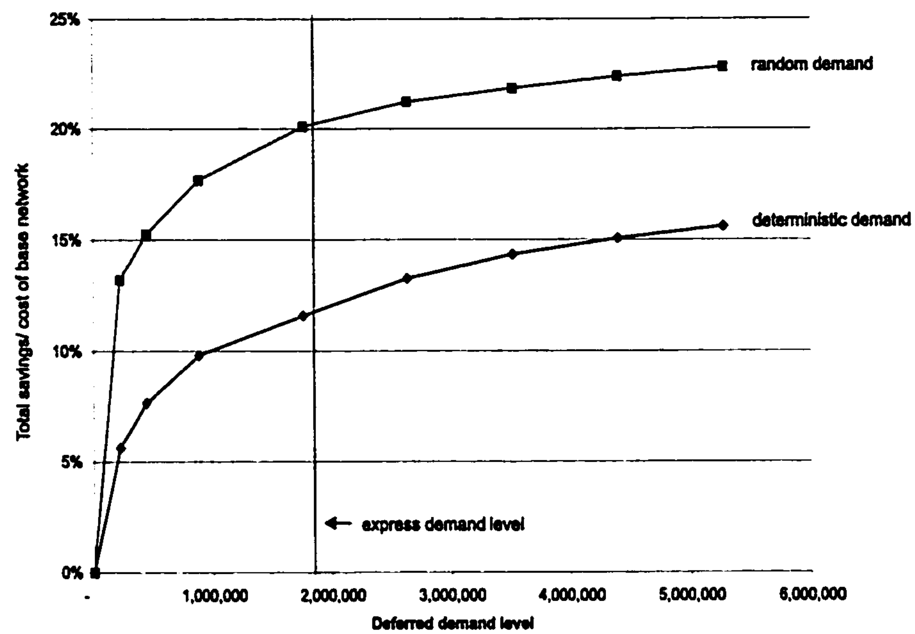


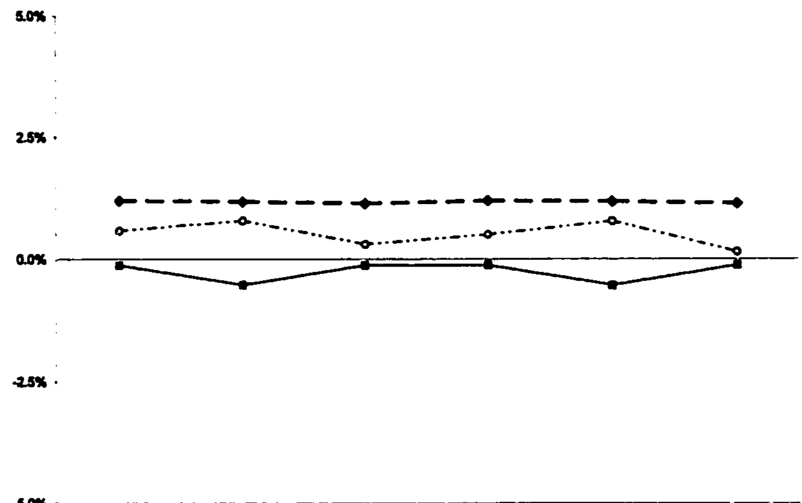
Figure 4.9: Savings comparison: US-based network

is included in Appendix C. The remaining cost approximations are validated here.

4.4.1 Terminal cost approximations

To decompose terminal costs by area, two key assumptions are made. First, the total number of breakbulk terminals and airports are approximated by local conditions (i.e., $\bar{\Delta}_B$ approximated by $\Delta_B(x)$ and $\bar{\Delta}_P$ by $\Delta_P(x)$). Second, the impact of regional imbalance on sorting is neglected for breakbulk terminal costs(i.e., $\lambda_i^G(x)c_k \log\left(\frac{1}{\Delta_B(x)}\right) + \lambda_o^G(x)c_k \log(\Delta_B(x)) \approx 0$). This imbalance is not a factor for airport/main hub sorting costs.

After optimizing operations for all levels of integration, approximated breakbulk terminal and airport costs are compared with their true values without making the two approximations. In Figures 4.10 and 4.11, the relative errors in terminal costs caused by these approximations are presented for SR1 and SR2, respectively. Terminal cost errors



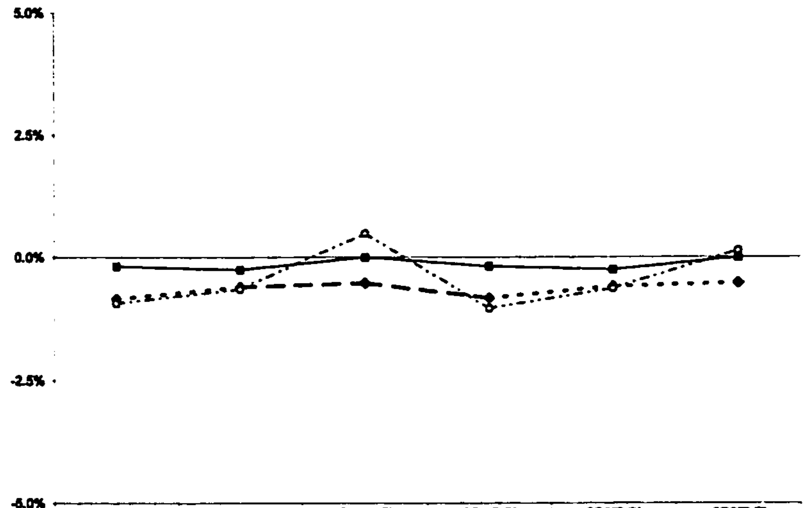
	SR1[K,B]	SR1[K,D]	SR1[K,E]	SR1[R,B]	SR1[R,D]	SR1[R,E]
BBT (BC)	1.2%	1.2%	1.1%	1.2%	1.2%	1.1%
airport (BC)	-0.1%	-0.5%	-0.1%	-0.1%	-0.5%	-0.1%
BBT (I4)	0.6%	0.8%	0.3%	0.5%	0.8%	0.1%

Figure 4.10: Terminal cost comparison: idealized network

between approximated costs z_{app} and true costs z_{true} are measured as follows:

$$error = \frac{z_{true} - z_{app}}{z_{true}}$$

The figures show errors in breakbulk terminal costs, BBT(BC), and airport/air hub costs, Airport(BC) in the base case, and breakbulk terminal costs under integration strategy I4, BBT(I4), when breakbulk terminal densities are reoptimized. As the figures show, the maximum error in cost is under 2% for both networks. One can see as well that the errors are smaller for SR2 than SR1, suggesting that the accuracy of approximations improve as the problem size increases. Additional tests compare the isolated effects of the two approximations individually; these results show relative errors within 1%-2%. Further, results confirm the cost overestimation of approximating $\bar{\Delta}_B$ by $\Delta_B(x)$ and $\bar{\Delta}_P$ by $\Delta_P(x)$ and the reduction of this overestimation with problem size.



	SR2[K,B]	SR2[K,D]	SR2[K,E]	SR2[R,B]	SR2[R,D]	SR2[R,E]
BBT (BC)	-0.8%	-0.6%	-0.5%	-0.8%	-0.6%	-0.5%
airport (BC)	-0.2%	-0.2%	0.0%	-0.2%	-0.3%	0.0%
BBT (I4)	-0.9%	-0.6%	0.5%	-1.0%	-0.6%	0.1%

Figure 4.11: Terminal cost comparison: US-based network

4.4.2 Longhaul transportation approximations

Decisions regarding longhaul transportation are also approximated locally. Chapter 5 uses numerical optimization techniques to obtain more detailed estimates of longhaul operating costs, including the allocation of the deferred items to excess air capacity and vehicle repositioning. Table 4.2 shows a comparison of costs obtained with the two approaches. The continuous approximation results include operating rules on longhaul transportation (multiple stops on collection only) that are relaxed in the numerical optimization. Therefore, costs obtained with numerical optimization could be lower if it is found that these operating rules are too restrictive. However, the size of the problem requires crude time discretization which may lead to higher results with numerical optimization.

Since allocation decisions are made at consolidation terminals, the following costs are compared: ground vehicle access and longhaul costs, access and longhaul vehicle repositioning for the ground network, storage costs for deferred items at terminals, and

Scenario	Continuous Approximation	Discrete Lower Bound	Discrete Upper bound
SR1[K,D]	1,693,429	1,477,262 [3]	2,575,995
SR1[K,E]	800,662	685,704	1,513,360

Table 4.2: Comparison of access and longhaul transportation costs

marginal costs of sending deferred items by air. Details on the numerical optimization results are presented in the next chapter. Additional validation for larger problems requires advancements in the numerical optimization techniques. This is left for future work.

[3]Results for SR1[K,D] are for six iterations. Better bounds are expected with further iterations and added cuts.

Chapter 5

System operation: distribution of deferred items

5.1 Introduction

In previous chapters the decisions involving the allocation of deferred items to excess air capacity and the routing of ground vehicles to serve the items not sent by air are approximated as local decisions. However, once long term decisions on network design are made using the continuous approximation models, shorter term routing decisions can be considered in greater detail. In particular, the longhaul routing of deferred items can be studied in depth, including an exploration of the benefits of sending a fraction of these items via excess air capacity. By intelligently choosing which deferred items to shift to the air network, the ground network can be operated more efficiently.

This chapter demonstrates how to translate continuous approximation results into more detailed tactical plans. Methodologies are developed to evaluate the extent to which sharing aircraft capacity can allow all longhaul vehicles (air and ground) to be utilized better. Since aircraft schedules are determined quarterly or yearly for express package delivery, it is assumed here that aircraft schedules are fixed for shorter term problems. Local distribution routes feeding into consolidation terminals are unaffected as well. Therefore, this chapter focuses exclusively on the simultaneous routing of deferred items and ground vehicles on the access/longhaul network. The deferred item and vehicle routing problem is formulated as a multicommodity network flow model with side constraints.

As mentioned in Chapter 1, multicommodity network flow problems have been studied extensively in the literature, both because of the many applications that exist (especially in transportation, see for example Powell and Sheffi (1983), Dejax and Crainic (1987), Barnhart and Schneur (1996), Powell (1996), Newman and Yano (2000), and Crainic (2000)) and because of the challenges of solving these problems, see Farvolden *et al.* (1993), Jones *et al.* (1993), Ball *et al.* (1995), and Barnhart *et al.* (1998). Problem size and the existence of integer variables makes the simultaneous routing of items and vehicles difficult to solve. Solution methods that decompose the item and vehicle routing have been successful in solving large multicommodity network flow problems; see Crainic and Rousseau (1986) and Armacost (2000). Both the issues of problem size and integrality are covered in this chapter.

Section 5.2 describes the deferred item and vehicle routing network in further detail, and Section 5.3 presents the model formulation with limited computational results for small test cases. Section 5.4 presents several heuristics to allow for the solution of large-scale problems. Implementation issues, including the development of discrete data sets from the continuous approximation results of Chapter 4, and computational results are discussed in Section 5.5. Finally, Section 5.6 presents conclusions and extensions.

The notation used in the models presented in this chapter is defined in the following sections. Note that some symbols from previous chapters have new meanings.

5.2 Network description

Because this chapter focuses on longhaul and access routing, demand is aggregated to consolidation terminals and they are now treated as the origins and destinations of the model. To model the routing, a time-space network, G = (N,A) is created, where each node in the set N represents a physical terminal location (either a consolidation terminal, a breakbulk terminal, or an air hub) and an instant of discrete time within the planning horizon; see Figure 5.1. Note that airports do not appear in the network. Instead, items shifted to the air network are assumed to flow directly between consolidation terminals and the air hubs. This is done because it is assumed that sufficient capacity exists on air network access routes to handle aircraft loads. Therefore, the routing of access vehicles from consolidation terminals to airports does not have to be considered.

We present a more general formulation here that allows for more than one air hub. However, within the context of the distribution networks studied in this research, only one main air hub is assumed.

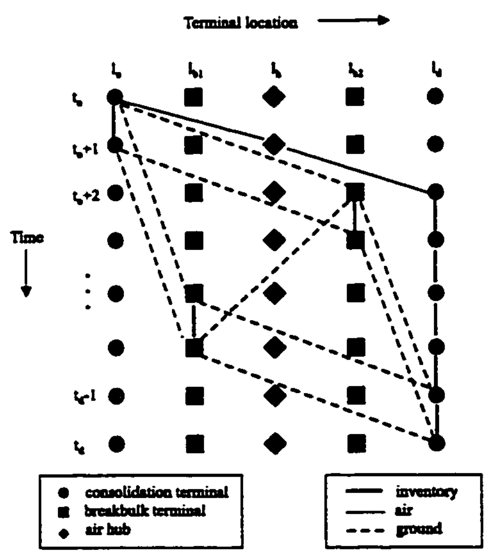


Figure 5.1: Network representation

The set of nodes, N, is divided into three subsets: consolidation terminals (C), breakbulk terminals (B), and air hubs (H). The set of arcs, A, that link the nodes of the network is partitioned into three subsets: inventory arcs (IA) for holding items at a node until the next time period, ground arcs (GA) for transporting items by ground vehicle or repositioning vehicles, and express arcs (EA) for transporting items over the air network.

Inventory arcs exist between adjacent time periods (i.e., from (l, t) to $(l, t+1)$ for $l \in N$). A ground arc (denoting the possibility of travel) from (l_1, t_1) to (l_2, t_2) for $l_1, l_2 \in C \cup B$ will be included in the network if the ground travel time between l_1 and l_2 is $t_2 - t_1$. Similarly, there is an air arc (representing both an access route to/from an airport and a flight to/from an air hub) from (l_1, t_1) to (l_2, t_2) for $l_1 \in C, l_2 \in H$ or $l_1 \in H, l_2 \in C$, for all *scheduled* departure times t_1 if the air network travel time between l_1 and l_2 is $t_2 - t_1$. Perfect information and time-dependent demand are assumed.

As with many dynamic network problems, the choice of planning horizon is challenging. In order to negate beginning and ending effects, the planning horizon should be long compared with a typical delivery window. However, increasing the time horizon can lead to computational problems as the number of nodes and arcs increases. Rolling horizon methods do not work well for systems with long memories (see Daganzo and Erera (1999) and Erera (2000)) because the solution depends on the recourse which is unknown.

Fortunately, when the data are periodic in nature (e.g., weekly or daily cycles), as in the case of package delivery, an infinite horizon can be simulated if one restricts oneself to the exploration of periodic solutions. This is done by introducing periodic boundary conditions as constraints, which can be easily modeled by treating the time dimension as a closed loop; i.e., by wrapping the network on a cylinder and linking the last time period with the first, see Figure 5.2.

Item demand can now specified by origin, arrival date, destination, and number of cylinder rotations. It is assumed in Section 5.3 that the number of cylinder rotations is 1. Arcs between nodes are placed on the cylinder, with some arcs connecting nodes at

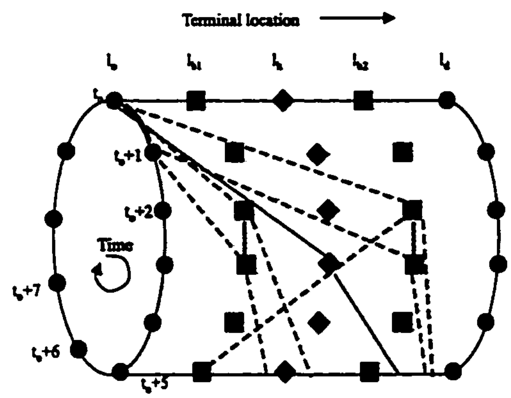


Figure 5.2: Cyclic network with periodic boundary conditions

the end of one rotation to nodes at the beginning of another.

5.3 Model formulation

The objective of the deferred item and vehicle routing problem is to minimize the cost of distributing deferred items. Costs include marginal item transportation costs by ground and air, vehicle costs for ground transportation, and inventory costs at terminals. Figures 5.2 and 5.1 illustrate possible routes from an origin (l_o, t_o) to a destination (l_d, t_d). Items can be routed between consolidation terminals either by air or by ground through a series of intermediate breakbulk terminals, shown in the figure as l_{b2} and l_{b1}. Since aircraft schedules and express package flow are given, excess air capacity can be determined for all flight legs, and this is input data for the model. The routing of ground vehicles (and the resulting capacity on ground arcs); however, is determined within the deferred item and vehicle routing problem, consistent with vehicle balancing constraints.

The following parameters and decision variables are included in the model:

Parameters

c_a^l marginal cost per item over arc $a \in A$ ($/item)

c_a marginal cost per vehicle over arc $a \in GA$ (\$/vehicle-mile)

V^i ground vehicle capacity for vehicle type $i \in \mathcal{V}$ (items)

e_a excess air capacity (items) for arc $a \in EA$

$d_{(l_o,t_o),(l_d,t_d)}$ origin-destination demands with origin times and delivery due dates (items)

Variables

f_a^k amount of commodity k sent over arc a, (items, continuous)

x_a^i number of loaded and empty ground vehicles of type $i \in \mathcal{V}$ sent over arc $a \in GA$, (vehicles, integer)

In this section, two equivalent arc-based formulations of the problem are presented, derived from two commodity definitions for the set K of all commodities. In the first, we define a commodity k by an origin node (l_o, t_o) and a destination node (l_d, t_d), and origin-destination demand, $d_{(l_o,t_o),(l_d,t_d)} = d^k > 0$. The origin and destination nodes are indexed by commodity k: $(l_o, t_o)^k, (l_d, t_d)^k$.

In the second formulation, demand is aggregated over destinations and due dates and commodities are defined by origin node only. A commodity, $k \in K$, is now defined by $(l_o, t_o)^k$. The origin-destination demand now must be indexed by commodity and destination, $d_{(l_o,t_o)^k,(l_d,t_d)} = d^k_{(l_d,t_d)} > 0$.

5.3.1 Formulation I: disaggregated origin-destination formulation

The first formulation of the deferred item and vehicle routing problem (DIVRP-D) considers commodities at the disaggregate level, $k = (l_o, t_o)^k, (l_d, t_d)^k$. The formulation is as follows:

DIVRP-D

$$min \sum_{k\in K, a\in A} c'_a f_a^k + \sum_{a\in GA, i\in \mathcal{V}} c_a x_a^i \tag{5.1a}$$

subject to

$$\sum_{(m,n)\in A} f_{m,n}^k - \sum_{(n,m)\in A} f_{n,m}^k = \begin{cases} -d^k, & \text{if } n = (l_o, t_o)^k \\ d^k, & \text{if } n = (l_d, t_d)^k \\ 0, & \text{otherwise} \end{cases} \quad \forall n \in N, k \in K \tag{5.1b}$$

$$\sum_{k\in K} f_a^k \le e_a \qquad \forall a \in EA \tag{5.1c}$$

$$\sum_{k\in K} f_a^k \le \sum_{i\in \mathcal{V}} V^i x_a^i \qquad \forall a \in GA \tag{5.1d}$$

$$\sum_{(m,n)\in GA} x_{m,n}^i - \sum_{(n,m)\in GA} x_{n,m}^i = 0 \qquad \forall n \in C \cup B, \forall i \in \mathcal{V} \tag{5.1e}$$

$$x_a^i \ge 0, \text{ integer} \qquad \forall a \in GA, \forall i \in \mathcal{V} \tag{5.1f}$$

$$f_a^k \ge 0 \qquad \forall a \in A, k \in K \tag{5.1g}$$

We minimize the objective function, which is the total fixed and variable transportation costs and variable holding costs. The first set of flow balance constraints (5.1b) ensures item flow conservation at each node for all commodities. Built into these equations are delivery time windows implicit in the time-space network: items cannot depart from an origin before a set time and must reach the destination by the due date. Vehicle capacity constraints limit allocation to air by excess capacity available (5.1c) and require sufficient ground vehicles to cover item flow (5.1d). Flow balance constraints (5.1e) for ground vehicles create feasible routings between nodes. All decision variables must be non-negative and ground vehicle variables must satisfy integrality constraints.

This formulation includes several vehicle types: smaller trucks for access trips and larger tractor-trailers for longhaul routes. Arcs can be designated for vehicle repositioning exclusively. This is discussed in Section 5.5.1.

5.3.2 Formulation II: demand aggregated by destination

As problem sizes increase, the numbers of decision variables and constraints in formulation DIVRP-D become quite large as the number of commodities is quadratic in the number of consolidation terminals. The number of commodity-specific decision variables and constraints can be reduced by a factor when demand is aggregated by destination and due date. A commodity, $k \in K$, is now defined by $(l_o, t_o)^k$ together with a set of destinations and due dates, D^k. The second formulation, DIVRP-A, is defined as follows.

DIVRP-A

$$min \sum_{k\in K, a\in A} c'_a f^k_a + \sum_{a\in GA, i\in \mathcal{V}} c_a x^i_a \tag{5.2a}$$

subject to

$$\sum_{(m,n)\in A} f^k_{m,n} - \sum_{(n,m)\in A} f^k_{n,m} = \begin{cases} -\sum_{j\in D^k} d^k_j, & \text{if } n = (l_o, t_o)^k \\ d^k_n, & \text{if } n \in D^k \\ 0, & \text{otherwise} \end{cases} \qquad \forall n \in N, k \in K \tag{5.2b}$$

$$\sum_{k\in K} f^k_a \le e_a \qquad \forall a \in EA \tag{5.2c}$$

$$\sum_{k\in K} f^k_a \le \sum_{i\in \mathcal{V}} V^i x^i_a \qquad \forall a \in GA \tag{5.2d}$$

$$\sum_{(m,n)\in GA} x^i_{m,n} - \sum_{(n,m)\in GA} x^i_{n,m} = 0 \qquad \forall n \in C \cup B, \forall i \in \mathcal{V} \tag{5.2e}$$

$$x^i_a \ge 0, \text{ integer} \qquad \forall a \in GA, \forall i \in \mathcal{V} \tag{5.2f}$$

$$f^k_a \ge 0 \qquad \forall a \in A, k \in K \tag{5.2g}$$

Formulation DIVRP-A is more compact than the original formulation, with fewer flow variables $\{f_a^k\}$ and fewer item flow balance constraints (5.2b). The remainder of the formulation is the same as the disaggregated formulation. Here the number of commodities grows linearly in the number of consolidation terminals.

Table 5.1 shows this reduction in problem size for a relatively small problem with fifty-five consolidation terminals, five breakbulk terminals and eight time periods. The number of constraints, total variables, and integer variables for both DIVRP-D and DIVRP-A are shown in the table. While DVIRP-A has fewer total variables and constraints, the number of integer variables is the same in both formulations because the physical network has not changed, only the definition of (continuous) commodities.

The reduction in problem size in DVIRP-A translates into a reduction in solution time for the linear relaxation of DVIRP-A as well. Results indicate that solving the aggregated problem requires fewer simplex iterations and less processing time ("CPU time" in the Table 5.1). Integer solutions to the test instance for both formulations are obtained with the CPLEX 6.5.1 solver, using a branch and bound algorithm. As the results in the table show, the branch and bound algorithm could not find a feasible solution for formulation DIVRP-D within the twenty-five hour time limit. For DIVRP-A, a feasible integer solution within 2.3% of the upper and lower bounds is obtained within the limit. This solution requires the exploration of 2,160 nodes in the branch and bound tree.

While these results suggest the aggregated formulation outperforms the disaggregated formulation for both linear and integer solutions, the results also indicate that solving a realistic size problem (hundreds or thousands of nodes and arcs) with either DIVRP-A or DIVRP-D involves an excessive number of variables and constraints and the models presented here cannot be solved without additional refinements. These refinements are presented in the following sections.

Instance size	**DIVRP-D**	**DIVRP-A**
Number of constraints	216,519	32,874
Number of variables	781,597	152,487
Number of integer variables	10,477	10,477
Linear relaxation		
Simplex iterations	134,644	36,964
CPU time (seconds)	694	141
Objective	180,613	180,613
Integer solution		
CPU time	25 hours(limit)	25 hours(limit)
Branch and Bound nodes	NA	2,160
Best integer solution (upper bound)	no solution	184,964
Best linear solution (lower bound)	NA	180,618
Optimality gap	NA	2.3%

Table 5.1: Problem size and solution statistic comparison by formulation: 55 consolidation terminals, 5 breakbulk terminals and 8 time periods

5.4 Solution approach

In this section, a two-stage solution heuristic is proposed to find feasible integer solutions from optimal linear relaxations of the problem. In the following subsections, systematic improvements to the solution approach are introduced to solve larger problems and obtain better upper and lower bounds. By including these methods, the two-stage procedure can be employed iteratively until the gap between bounds is acceptable.

Solution approach:

Stage 1: lower bound Solve the linear relaxation of DIVRP-A or D, using approaches described in Sections 5.4.1.

Stage 2: upper bound Obtain a feasible integer solution from the linear relaxation, using approaches described in Section 5.4.2.

5.4.1 Lower bounding techniques

5.4.1.1 Dantzig-Wolfe reformulation

As the size of the node and arc sets increases, the arc-based formulations (DIVRP-D and DIVRP-A) become impractical/impossible to implement due to the excessive memory requirements. Therefore, a path-based formulation of the deferred item and vehicle routing problem is introduced. In the earlier arc-based formulations, optimal commodity flows for each network arc are chosen. In the path-based formulation, a series of arcs that a commodity traverses from origin to destination is aggregated into a single path variable. The path-based reformulations of the problem are equivalent to the arc-based formulations; however, the decision variables are the sets of commodity flows over each path, and (as with arc-based formulations) the set of loaded and empty ground vehicles over the ground arcs. This path-based reformulation is used to solve linear relaxations of the deferred item and vehicle routing problem. In the following, both disaggregated and aggregated formulations are discussed.

Disaggregated reformulation

The fraction of commodity k flowing over path P, comprised of all arcs $a \in P$, is defined as λ_P. Let $\mathcal{P}^k$ be the set of available origin-destination paths for each commodity. Then the total flow of commodity k on a specific arc a is:

$$f_a^k = \sum_{P \in \mathcal{P}^k : a \in P} \lambda_P d^k$$

The linear relaxation of the deferred item and vehicle routing problem can now be written with the new path flow variables as

$$min \sum_{k \in K} \sum_{P \in \mathcal{P}^k} c_P \lambda_P + \sum_{a \in A, i \in \mathcal{V}} c_a x_a^i \tag{5.3a}$$

subject to

$$\sum_{P \in \mathcal{P}^k} \lambda_P = 1 \qquad \forall k \in K \tag{5.3b}$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}^k : a \in P} \lambda_P d^k \leq e_a \qquad \forall a \in EA \tag{5.3c}$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}^k : a \in P} \lambda_P d^k \leq \sum_{i \in \mathcal{V}} x_a^i V^i \qquad \forall a \in GA \tag{5.3d}$$

$$\sum_{(m,n) \in GA} x_{m,n}^i - \sum_{(n,m) \in GA} x_{n,m}^i = 0 \qquad \forall n \in N, \forall i \in \mathcal{V} \tag{5.3e}$$

$$x_a^i \geq 0 \qquad \forall a \in GA, \forall i \in \mathcal{V} \tag{5.3f}$$

$$\lambda_P \geq 0 \qquad \forall k \in K, P \in \mathcal{P}^k \tag{5.3g}$$

where

c_P is the cost of path P for commodity $k, \forall k \in K, \forall P \in \mathcal{P}^k, i.e., c_P = \sum_{a \in P} c'_a d^k$.

In the path formulation, the original item flow balancing constraints (5.1b) from the arc-based formulation are replaced with "convexity constraints" (5.3b) that ensure the total demand for each commodity is satisfied. The two sets of arc capacity constraints (5.3c and 5.3d) maintain feasible flows over each arc. Again, the flow of ground vehicles must be balanced (5.3e).

Periodic boundary conditions can be implemented easily with the path-based formulation; recall Figure 5.2. Shortest paths between origins and destinations are created as described above. However, when paths that traverse multiple rotations are added to the master problem, they "wrap around" the cylinder.

Column generation

The number of paths is exponential in the number of arcs. Therefore, for large problems, rather than enumerating all origin-destination paths, a subset of these paths is generated as needed. Using column generation, new paths can be generated iteratively if the inclusion of such paths in $\mathcal{P}^k$ could reduce the total cost.

The path-based problem is solved with a feasible subset of paths first. Then, additional paths, not in the current subset, are considered for inclusion in $\mathcal{P}^k$. A candidate path P can be obtained with the following single-commodity shortest path problem:

Pricing Problem

$$min \sum_{a \in A} (c'_a - \pi^a) f^k_a - \sigma^k \tag{5.4a}$$

subject to

$$\sum_{(m,n) \in A} f^k_{m,n} - \sum_{(n,m) \in A} f^k_{n,m} = \begin{cases} -d^k, & \text{if } n = (l_o, t_o)^k \\ d_k, & \text{if } n = (l_d, t_d)^k \\ 0, & \text{otherwise} \end{cases} \qquad \forall n \in N \tag{5.4b}$$

$$f^k_a \geq 0 \qquad \forall a \in A \tag{5.4c}$$

The objective function of the shortest path problem is the reduced cost of a candidate path P:

$$\bar{c}_P = \sum_{a \in A} (c'_a - \pi^a) f_a^k - \sigma^k$$

where π^a is the dual variable of the arc capacity constraint for arc a (5.3c and 5.3d) and σ^k is the dual variable for the commodity-specific convexity constraint for commodity k (5.3b). Since a shift of flow to the new path P does not violate the ground vehicle balance constraints (5.3e) directly, the dual variables of these constraints do not appear. If the reduced cost of candidate path P, $\bar{c}_P$, is negative, then the current set of paths, $\mathcal{P}^k$, is not optimal and P should be added.

The paths generated in the pricing problems are added to the master problem (equations 5.3). After each iteration of the master problem, dual variables are updated and the pricing problem is run to check for new columns to add. In addition, columns with highly positive reduced costs are removed from the master problem to maintain a manageable problem size. This process is repeated until optimality conditions of the linear relaxation are met for all commodities (all paths have non-negative reduced costs), or a less-restrictive stopping criteria is satisfied.

Aggregated reformulation

When commodities are aggregated, the pricing problems generate solutions for multiple origin-destination pairs by creating shortest path trees from origin nodes (defined by origin location and time) to all destinations (defined by locations and due dates).

The master problem for the aggregated path formulation is the same with slight modifications to the capacity constraints. Again the master problem finds the optimal flows on each tree among existing trees for each commodity while satisfying vehicle capacity and repositioning constraints. Depending of the shortest path algorithm used, trees may be generated in the disaggregated version as well.

Pricing Problem

$$min \sum_{a \in A} (c'_a - \pi^a) f_a^k - \sigma^k \tag{5.5a}$$

subject to

$$\sum_{(m,n) \in A} f_{m,n}^k - \sum_{(n,m) \in A} f_{n,m}^k = \begin{cases} -\sum_{j \in D^k} d_j^k, & \text{if } n = (l_o, t_o)^k \\ d_n^k, & \text{if } n \in D^k \\ 0, & \text{otherwise} \end{cases} \quad \forall n \in N \tag{5.5b}$$

$$f_a^k \geq 0 \quad \forall a \in A \tag{5.5c}$$

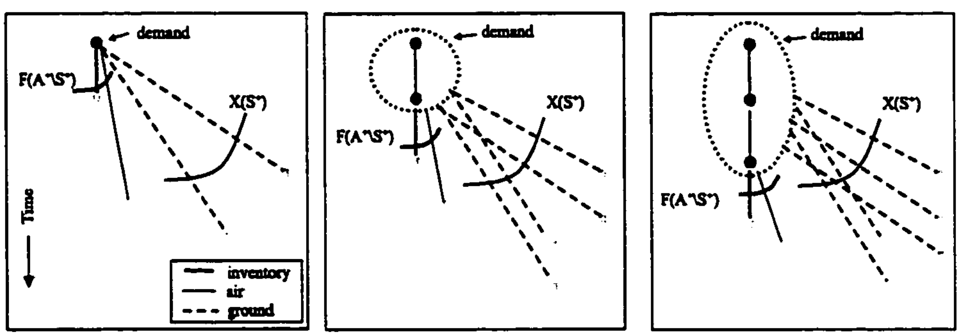
The number of subproblems is linear in the number of consolidation terminals for the aggregated version, and quadratic in the number of consolidation terminals for the disaggregated version. However, results from the solution of a series of test cases indicate that while the aggregated formulation requires fewer pricing problems at each master problem iteration, more master problem iterations are required for convergence with the aggregated formulation. A tree may contain several optimal origin-destination paths; however, poor paths within the tree may keep that tree out of consideration. Further iterations are required until the sub-optimal paths are replaced with better paths. These results are consistent with earlier results on multicommodity network flow problems by Jones *et al.* (1993), even with ground vehicle balancing constraints added.

Therefore, a hybrid decomposition algorithm is used to generate lower bounds in the solution heuristic. Shortest path trees from an origin to all destinations are obtained with aggregated pricing problems. These trees are then disaggregated by destination and a column for each origin-destination path with a negative reduced cost is added to the master problem. This increases the number of columns in the master problem; however, the convergence rate is faster and the number of master problem iterations decreases.

5.4.1.2 Cutting planes

In order to improve the lower bound, we look for constraints to add to the linear relaxation of the deferred item and vehicle routing problem that eliminate, or "cut", fractional solutions without eliminating feasible integer solutions. These constraints are referred to as cutting planes, or each constraint referred to simply as a cut. For a detailed discussion on this subject, see Wolsey (1998) and Nemhauser and Wolsey (1999). This section focuses on the addition of a set of constraints for each consolidation terminal to reduce fractional ground vehicle solutions for arcs outbound from that terminal.

As shown in Figure 5.3, flow out of a node (including air and inventory arcs) must be at least the total demand at that node. In Figure 5.3(a), the node under consideration is a consolidation terminal in the first time period. Therefore, inbound flow is equal to the amount of items arriving at the consolidation terminal from local tours for distribution, plus any demand from previous days that have wrapped around the cylinder.



a b c

Figure 5.3: Outbound cuts: (a) first time period (b) second (c) third

For ease of explanation, cuts are described for the arc-based formulation. Since the arc-based and path-based formulations are equivalent, the cuts are valid for both. In formulation DIVRP-D, the flow balance constraints at an origin consolidation terminal n for commodity k ($n = (l_o, t_o)^k$) are given by (5.1b). Let A_n^+ be the set of all arcs outbound from that node and let A_n^- be the set of all inbound arcs. With this notation, 5.1b can be written:

$$\sum_{a \in A_n^-} f_a^k - \sum_{a \in A_n^+} f_a^k = -d^k \qquad \text{where } n = (l_o, t_o)^k \tag{5.6a}$$

Let S_n^+ be the set of outbound ground vehicle arcs for node n. Outbound flow can then be decomposed by ground vehicle flow and other flow (air and inventory arcs). Rearranging the terms of (5.6a), we obtain

$$\sum_{a \in S_n^+} f_a^k + \sum_{a \in A_n^+ \setminus S_n^+} f_a^k = d^k + \sum_{a \in A_n^-} f_a^k \qquad \text{where } n = (l_o, t_o)^k \tag{5.6b}$$

Let K_n be the set of all commodities with origin n. Then the total demand arriving at n is $D_n = \sum_{k \in K_n} d^k$. Constraint (5.6b) can be summed over all commodities in K_n. Recall that ground vehicle arcs must satisfy capacity constraints, and that ground vehicle types are restricted to certain arcs. All ground vehicles traveling to/from a consolidation terminal have the same capacity, V. Let x_a be the flow of this vehicle type over an arc. For each arc $a \in S_n^+$, $\sum_{k \in K_n} f_a^k \leq x_a V$. Therefore, the following constraint derived from (5.6b) is valid

$$\sum_{a \in S_n^+} x_a V + \sum_{k \in K_n} \sum_{a \in A_n^+ \setminus S_n^+} f_a^k \geq D_n \qquad \text{where } n = (l_o, t_o)^k \tag{5.6c}$$

Note the flow over inbound arcs $a \in A_n^-$ does not appear in (5.6c). This flow can be ignored and (5.6c) is still valid, since their inclusion would only raise the right-hand side of (5.6c). Sufficient outbound flow to serve this inbound flow is still enforced through balance constraints in DVIRP-D.

Let $X(S_n^+)$ be the vehicle flow on arcs in the set S_n^+ and $F(A_n^+ \setminus S_n^+)$ be the total item flow over all commodities on outbound arcs not in S_n^+. Then constraint (5.6c) can be rewritten

$$VX(S_n^+) + F(A_n^+ \setminus S_n^+) \geq D_n \qquad \text{where } n = (l_o, t_o)^k \tag{5.6d}$$

Problem	Lower bound	Gap	Lower bound with cuts	Improvement in gap	Number of cuts	Time (min)
ID2	1,395,328	36%	-	-	1215	395[1]
ID3	763,444	19%	853,579	10%	360	1
ID4	426,434	44%	426,581	0.02%	558	11
ID5	341,268	50%	367,482	4%	342	2
ID6	327,752	51%	334,255	1%	450	5
SR1[K,E]	680,506	55%	685,704	0.4%	1269	642

Table 5.2: Improving lower bounds with cutting planes

The mixed integer cut-set inequality as defined in Wolsey (1998), Nemhauser and Wolsey (1999), and Atamtürk (2000) can be defined for flow emanating from a node given constraint (5.6d) as follows:

$$rX(S_n^+) + F(A_n^+ \backslash S_n^+) \geq r\eta \qquad \text{where } n = (l_o, t_o)^k \tag{5.7}$$

where

$\eta = \lceil \frac{D_n}{V} \rceil$, minimum number of vehicles needed if all flow is sent by ground arcs

$r = D_n - (\eta - 1)V$

In later time periods (Figure 5.3 (b) and (c)), nodes for the same physical location of the origin and multiple time periods are aggregated to create "super nodes". These super nodes contain both the nodes and the inventory arcs between the nodes. Flow is then defined as flow inbound to and outbound from the super node, and additional cuts are generated for these super nodes.

After solving the linear relaxation by column generation, the cuts defined by (5.7) are added for all n. Table 5.2 presents results from the addition of these cuts to the test

[1]Computational limits reached. For ID2, no solution found before this time. For SR1[K,E], time limit reached before optimal solution is found.

cases described in Section 5.5.1. The first columns present lower bounds obtained with column generation and the gap between these lower bounds and the best upper bounds, as described in Section 5.4.2. The improved lower bounds obtained with the discussed cuts are shown next, along with the improvement to the gap. The number of cuts added and the solution time required to resolve the linear relaxation with these cuts are listed in the final two columns. As the table shows, the solution time of the linear relaxation with cuts grows quickly with problem size.

Additional cuts can be added to further improve lower bounds. The set S_n^+ can be reduced by checking for arcs such that $\sum_{k \in K_n} f_a^k < r x_a$ for $a \in S_n^+$. These arcs should be moved from S_n^+ to $A_n^+ \backslash S_n^+$. This would provide a richer class of cuts. Similar cuts for inbound consolidation terminal flows, as well as all breakbulk terminals can be added. To add cuts for breakbulk terminals with multiple vehicle types visiting a node, the cuts must be redefined; see Atamtürk (2000). It has also been shown that residual capacity inequalities can be added for each ground arc to improve lower bounds; see Atamtürk and Rajan (2000).

5.4.2 Upper bounding techniques

In this section, several rounding techniques are explored to obtain upper bounds.

5.4.2.1 Rounding approach 1

The first rounding approach obtains feasible integer solutions by running two additional linear programs, after the initial linear relaxation of the DIVRP-D. First, all fractional ground vehicle variables are rounded up to the nearest integer, and these integers are then used as lower bounds on arcs of a network flow model for ground vehicles only (the vehicle flow problem). The vehicle flow problem minimizes the transportation costs of ground vehicles, without considering item flows. However, sufficient capacity to transport items flows is guaranteed by the lower bounds on ground vehicles established in the first step. Vehicle balance constraints are the same as (5.1e) in the original formulation since merely rounding fractional values does not guarantee balanced vehicle flows. This formulation of the vehicle flow problem decomposes by vehicle type, and can be solved separately for all vehicle types. Due to the total unimodularity of the network flow matrix and integral right-hand sides, the solution to the flow routing problem is integer.

These integer values are maintained in the item flow problem performed next to determine item flows over the fixed ground network from the vehicle flow problem. The item flow problem allows path flows to be reallocated based on the fixed ground network, provided convexity constraints and capacity constraints are satisfied. This approach is described formally below.

(a) Round up fractional ground vehicle values: $\ell_a^i = \lceil \bar{x}_a^i \rceil, \quad \forall a \in GA, \forall i \in \mathcal{V}$ where $\{\bar{x}_a^i\}$ is the solution to DIVRP-D.

(b) Solve the vehicle flow problem to balance ground vehicle movements

Vehicle flow problem

$$min \sum_{a \in GA, i \in \mathcal{V}} c_a x_a^i \tag{5.8a}$$

subject to

$$\sum_{(m,n) \in GA} x_{m,n}^i - \sum_{(n,m) \in GA} x_{n,m}^i = 0 \qquad \forall n \in N, i \in \mathcal{V} \tag{5.8b}$$

$$x_a \geq \ell_a^i \qquad \forall a \in GA, i \in \mathcal{V} \tag{5.8c}$$

(c) Reflow items over fixed ground network with integer values $\{\hat{x}_a^i\}$ where $\{\hat{x}_a^i\}$ is the solution from step (b).

Item flow problem

$$min \sum_{k \in K} \sum_{P \in \mathcal{P}^k} c_P \lambda_P + \sum_{a \in A, i \in \mathcal{V}} c_a \hat{x}_a^i \tag{5.9a}$$

subject to

$$\sum_{P \in \mathcal{P}^k} \lambda_P = 1 \qquad \forall k \in K \qquad (5.9b)$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}^k : a \in P} \lambda_P d^k \leq e_a \qquad \forall a \in EA \qquad (5.9c)$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}^k : a \in P} \lambda_P d^k \leq \sum_{i \in \mathcal{V}} \hat{x}_a^i V^i \qquad \forall a \in GA \qquad (5.9d)$$

$$\lambda_P \geq 0 \qquad \forall k \in K, P \in \mathcal{P}^k \qquad (5.9e)$$

While this approach guarantees feasibility of both capacity constraints and ground vehicle balancing constraints, it could lead to a heavily over-capacitate ground network and a large optimality gap.

5.4.2.2 Rounding approach 2

In this alternative approach, near-integer variables are rounded to the nearest integer, either up or down in step (a). The vehicle flow problem is run to ensure vehicle flow balancing constraints are met; however, in this approach, step (c) may be infeasible if there is insufficient capacity for the transportation of items since some variables are rounded down in step (a).

5.4.2.3 Rounding approach 3

A variation of the first approach ignores the vehicle flow problem in step (b) and re-runs the entire deferred item and vehicle routing problem with a fraction of ground vehicle variables fixed. This approach is run iteratively, fixing more variables at each stage and then running DIVRP-D to check for feasible, integer solutions. This approach is implemented as follows:

(a) Define a set of candidate fractional ground vehicle variables to fix, $a \in GA'$

(b) Fix *some* variables in GA' at integer values

(c) Rerun DIVRP-D with selected ground variables fixed

Repeat steps (a) - (c) until no fractional ground vehicle variables remain or an infeasible solution is reached

In step (a), the candidate set of ground vehicle variables can be defined in several ways. Two examples are listed here based on the near-integrality of variables $\bar{x}_a^i$. The parameter $\alpha \in [0,1]$ can be adjusted to control the number of variables considered.

i. $a \in GA'$ for vehicle type i if $\min\{\lceil \bar{x}_a^i \rceil - \bar{x}_a^i, \bar{x}_a^i - \lfloor \bar{x}_a^i \rfloor\} < \alpha$

ii. $a \in GA'$ for vehicle type i if $\lceil \bar{x}_a^i \rceil - \bar{x}_a^i < \alpha$

In step (b), a fraction β of these candidate variables are selected by generating a random number $\rho \in [0,1]$ for each candidate variable and selecting that variable if $\rho < \beta$. Those variables chosen may be rounded to the nearest integer or rounded up to the nearest integer. Several variations of this rounding approach based on the above options can be defined.

The trick here is to choose the appropriate number of fractional variables to fix at integer values. Fixing too many (high values of α and β) may lead to an infeasible solution, both in terms of vehicle balance and item capacity (when arcs can be rounded down). Fixing too few (low values of α and β) may lead to a highly fractional solutions, much like the linear relaxation itself, which require a large number of iterations to reach an integer solution.

5.4.2.4 Rounding approach 4

To solve the problem of infeasibility in approach 3, near-integer variables can be bounded rather than fixed, as a hybrid of approaches 1 and 3. This approach provides a feedback loop where item and vehicle flows are adjusted in DIVRP-D to accommodate the new set of bounds.

(a) Define a set of candidate fractional ground vehicle variables to bound, $a \in GA'$

(b) Set bounds for some variables in GA'.

(c) Rerun DIVRP with new lower bounds on these ground variables

Repeat steps (a) - (c) until no fractional arcs remain

The same options for selecting candidate variables and rounding these variables from approach 3 are used here. Iterative solutions to approach 4 may not produce integer solutions; however, approach 1 can be employed at various iterations of approach 4 to obtain integer upper bounds from the current fractional solution obtained with approach 4.

Table 5.3 presents upper bounds for test problem ID5 obtained with various rounding approaches. Rather than simply employing these approaches after all columns have been added to the master problem in the column generation phase, the four approaches are applied at different stages during the column generation to obtain potentially better bounds. For each approach, the best integer solution is listed, along with the optimality gap and the time to run the rounding heuristic. Times are given in minutes. Results for several variations of approaches 3 and 4 are presented. For these runs, the values of α and β are shown, along with the average number of iterations of the rounding heuristic for that run and the maximum number of iterations allowed. After the maximum number of iterations, if the solution is still fractional, but feasible, approach 1 is run.

The variations of approaches 3 and 4 are defined by the method to select candidate variables and the rounding performed. In the runs of approaches 3a and 4a, variables are considered if $\lceil \bar{x}_a^i \rceil - \bar{x}_a^i < \alpha$. These variables, if selected, are rounded up. For approaches 3b and 4b, variables are considered if $\min\{\lceil \bar{x}_a^i \rceil - \bar{x}_a^i, \bar{x}_a^i - \lfloor \bar{x}_a^i \rfloor\} < \alpha$, and then the selected variables are rounded up. Finally, in approaches 3c and 4c, variables are selected if $\min\{\lceil \bar{x}_a^i \rceil - \bar{x}_a^i, \bar{x}_a^i - \lfloor \bar{x}_a^i \rfloor\} < \alpha$. Selected variables are rounded to the nearest integer.

Approach	Upper bound	Gap	Time	α	β	Average Iterations	Maximum Iterations
1	721,351	49%	0.5	-	-	-	-
2	infeasible	NA	0.5	-	-	-	-
3a	706,000	48%	21	0.15	0.25	38	500
3a	684,290	46%	32	0.1	0.1	76	500
3a	698,249	47%	48	0.2	0.025	376	500
3b	728,891	50%	5	0.25	0.15	500	500
3b	731,988	50%	26	0.5	0.25	250	250
3c	infeasible	NA	20	0.25	0.15	150	150
4a	721,085	49%	14	0.15	0.25	51	500
4a	705,444	48%	30	0.1	0.1	129	500
4a	695,255	47%	51	0.25	0.025	340	500
4b	730,581	50%	5	0.25	0.15	500	500
4b	733,604	50%	18	0.5	0.25	250	250
4c	733,490	50%	15	0.25	0.15	150	150

Table 5.3: Comparison of rounding heuristics for ID5

As expected with approaches 3 and 4, lower values of α and β produce better bounds, for the most part. However, raising these values often does not yield significantly higher upper bounds, and there is a large reduction in the number of approach iterations performed and overall solution time. By fixing variables are their nearest integer variable (approaches 2 and 3c) feasible solutions could not be found.

Rounding approach 3 appears to produce the best solutions; yet solutions times for this heuristic are quite high. For larger problems, it may not be practical to run this approach, and approach 1, with a significantly shorter solution time, may be preferred. The best upper bound, 684,290, is used in Tables 5.2 and 5.5.

Another possible rounding heuristic would be to consider fractional values in cycles in step (a), rather than as individual arcs. Therefore, ground vehicle variables can be adjusted in cycles in such a manner that would ensure that items could flow feasibly over the network and ground vehicle balancing constraints could be maintained. This is left for future research.

5.5 Results and discussion

5.5.1 Implementation issues

This section describes some key implementation issues when applying the deferred item and vehicle routing problem to integrated networks designed with continuous approximation techniques. In particular, the test cases of SR1 are considered. The location of breakbulk terminals and consolidation terminals are determined based on the terminal densities for each of the seventeen subregions. A typical business day is divided into smaller time units, consistent with pickup and delivery times. Distances between terminals are converted into integer multiples of time units based on vehicle speeds for the various vehicle types. The distances and travel times between terminals include slack for expected stochastic travel delays and systematic delays at hubs and terminals. It is assumed that detailed demand information (location of origins and destinations, as well as arrival times at origin and due dates for both express and deferred services) is known before allocation decisions are made.

Continuous cost parameters described in Chapter 4 and Appendix B have been translated into equivalent discrete cost parameters. Figures 5.4 and 5.5 show both continuous and discrete vehicle costs for access and longhaul trips, respectively. The figures capture the range of distances on access and longhaul trips, obtained from continuous approximation results. Other network costs are translated from continuous models as well, including terminal storage costs and item handling costs.

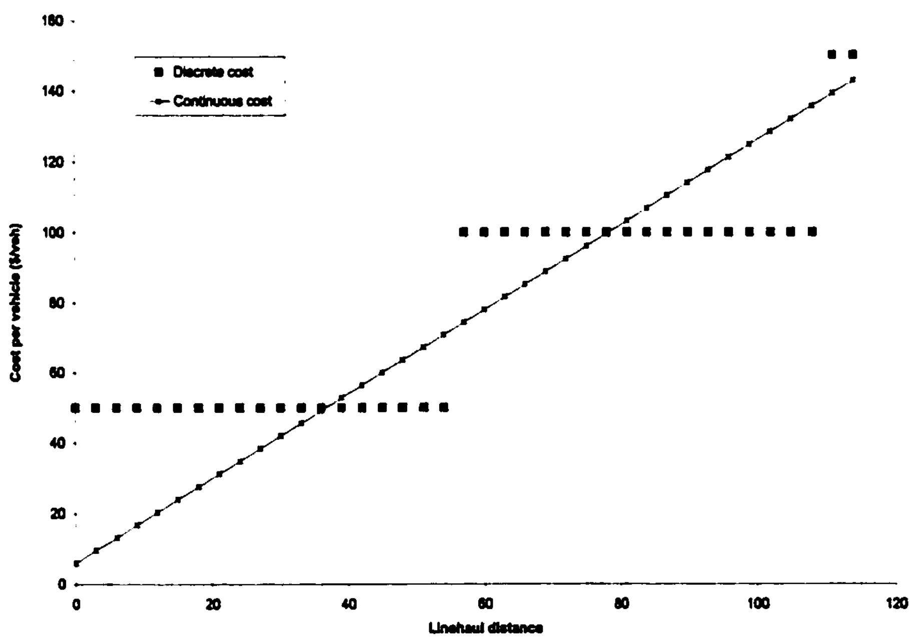


Figure 5.4: Access cost per vehicle: continuous and discrete

Insights from continuous approximation results can be used to reduce the set of arcs. For example, results from continuous approximation models show that access routes typically do not stop at multiple consolidation terminals en route to breakbulk terminals. Therefore, arcs between consolidation terminals exist for vehicle repositioning only (i.e., arcs where all vehicles have an item capacity, $V = 0$). In addition, the distance an access

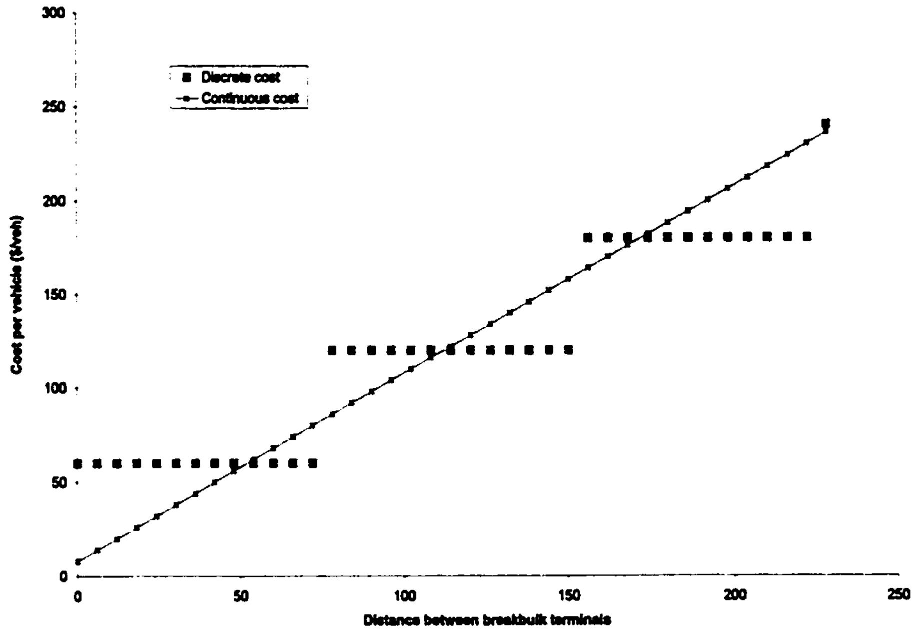


Figure 5.5: Longhaul cost per vehicle: continuous and discrete

vehicle can travel from a consolidation terminal to a breakbulk terminal is limited based on continuous approximation results. These insights are shown to significantly reduce the number of integer variables, by almost 50% in some cases.

Table 5.4 describes three test cases based on SR1. In addition, several other test cases are considered. A three-day delivery window is assumed with new items arriving each morning for distribution. For each problem, the number of consolidation and breakbulk terminals are listed first. Since a hybrid aggregation formulation is used (see Section 5.4.1), the number of commodities is a quadratic function of the number of consolidation terminals. Demands between consolidation terminals are determined by demand densities from the continuous approximation. The availability of excess capacity on aircraft is obtained from continuous approximation results as well. The next two columns of Table 5.4 represent the total number of nodes and arcs in the time-space graph. The total number of arcs includes ground vehicle and air arcs only since the number of inventory arcs does not significantly impact computation. The final two columns show the number

Problem	CT's	BBT's	Commodities	Total nodes	Total arcs[2]	Rows	Initial columns
SR1[K,B]	217	17	140,616	2,820	68,787	NA	NA
SR1[K,D]	169	19	85,176	2,268	48,648	110,364	135,564
SR1[K,E]	141	17	59,220	1,908	35,223	77,751	94,368
ID2	135	18	54,270	1,848	102,681	94,875	158,274
ID3	40	12	4,680	636	5,688	9,744	10,920
ID4	62	15	11,346	936	12,327	21,201	24,474
ID5	38	15	4,218	648	7,425	11,187	12,270
ID6	50	15	7,350	792	12,630	16,644	20,742

Table 5.4: Test problems

of rows and initial columns in the Dantzig-Wolfe reformulation of DIVRP-D.

While the smaller problems tested in Section 5.3 are obtained using AMPL and the CPLEX 6.5.1 solver, the solution algorithms for large-scale problems in Section 5.4 are implemented with the CPLEX Callable Library directly. All programs are run on SunOS 4.1.1 machines.

5.5.2 Computational results

Computational results for applications of the solution approach to the test problems described in Table 5.4 are discussed in this section. In Table 5.5, the quality of the best solutions obtained for each test problem are compared. The lower bound obtained with column generation for the linear relaxation of the deferred item and vehicle routing problem are presented first. In the next column, the improved lower bound found with cutting planes are displayed. The fourth and fifth columns show the best upper bound and the gap between the upper and lower bounds. The final column lists the rounding

[2]not including inventory arcs

Problem	Lower bound	Lower bound with cuts	Upper bound	Gap	Approach used
SR1[K,D][3]	1,477,262	NA	2,575,995	43%	1 (iteration 0)
SR1[K,E]	680,393	685,704	1,513,360	55%	1 (iteration 0)
ID2	1,395,328	NA	2,184,844	35%	1 (iteration 24)
ID3	763,444	853,579	941,555	9%	3a (iteration 12)
ID4	426,434	426,581	757,526	44%	3a (iteration 18)
ID5	341,268	367,482	684,290	46%	3a (iteration 12)
ID6	327,752	333,839	663,976	50%	1 (iteration 0)

Table 5.5: Test case results

approach that produced the best upper bound and the rounding iteration at which that best bound is obtained.

The problems have different cost structures, vehicle capacities and demand levels, so it should not be surprising that the gaps do vary. It is difficult to say without knowing the true optimal solution if the larger gaps are caused by poor upper bounds, poor lower bounds, or both. The test problem with the smallest gap between bounds is problem ID3 which is also the problem for which the added cuts improve the lower bound most. This suggests that, perhaps, the 35% gap for problem ID2 could be reduced appreciably if a lower bound could be obtained with outbound consolidation terminal flow cuts. Due to computational limits, a solution for these cuts has not been found. Several additional cuts mentioned in Section 5.4.1.2 may prove useful in further improving lower bounds for all test problems, and consequently reducing the listed gaps.

We now take a closer look at the upper bounds. Results suggest that while rounding approaches 1 and 3 outperform approaches 2 and 4, upper bounds obtained with ap-

[3]Runs for SR1[K,D] could not be completed due to computational limits. Cuts not finished for ID2 due to computational limits.

proach 4 are often close to those found with 1 and 3. Approach 2, on the other hand, often fails to produce feasible integer bounds. Like the results for problem ID5 in Table 5.3, the gap obtained for problem ID4 using approach 1 is close to the listed gap in Table 5.5 for approach 4 (46% rather than 44%), but the time required to run approach 1 is considerably shorter (6 minutes as opposed to 358 minutes). The same is not true for problem ID3 where approach 3 yields a 9% gap and approach 1 a 17% gap. Interestingly, with approach 1, it is sometimes the case that the solution does not improve after the first iteration. Note that for SR1[K,D] the number of iterations is constrained by computational limits.

In Table 5.6, solution times for the complete solution heuristic are presented in detail. All times are given in minutes. The time required to run the column generation is divided into time spent generating new columns with pricing problems and time spent solving the master problem at each iteration. The total number of master problem iterations (i.e., the number of times the master problem is solved with a new set of columns) are listed along with times for column generation. In addition, solution times for the rounding approaches and cutting plane methods to improve bounds are listed. The total solution time is provided in the last column. Solution time for linear relaxations within column generation and cut generation appear to be the largest bottleneck in terms of processing time.

5.6 Conclusions and future work

Using the methodologies presented in this chapter, deferred item and vehicle routing problems of realistic size can be solved. These results can be used to validate continuous approximation results from Chapter 4 and show how operating guidelines developed

[4]Runs for SR1[K,D] could not be completed due to computational limits. Cuts not finished for ID2 due to computational limits.

	Column Generation			Bounding		
Test case	Pricing problem	Master problem	Iterations	Rounding	Cuts	Total
SR1[K,D][4]	1	1,404	6	73	NA	1,479
SR1[K,E]	2	472	25	99	642	1,215
ID2	5	1,890	24	145	395	2,435
ID3	0.03	4	9	50	1	55
ID4	0.2	41	19	313	4	358
ID5	0.05	6	12	32	2	40
ID6	0.1	17	13	2	3	22

Table 5.6: Time comparisons for test cases (minutes)

with continuous approximation methods can be translated into more detailed solutions specifying exact item and vehicle routing. Future work on the deferred item and vehicle routing problem includes improvements to solution techniques, including more intelligent rounding heuristics and additional cutting planes. Since the time required to solve the master problem within column generation is prohibitively large, this should be a major focus of future research. Other relaxations of the problem are possible and could be explored in future work. In addition, new solution techniques to solve larger problems (including networks of the size of SR2) are also left for future work.

Chapter 6

Conclusions

6.1 Summary of results

In this research, a complete design and planning framework is developed for complex multimode, multiservice logistics networks. This is the first application of a hybrid continuous approximation/numerical optimization methodology to the design and operation of large-scale integrated distribution networks with shipment choice. As such, advancements in both continuous approximation and numerical optimization techniques, and in the integration of the two approaches, are required. A key component of this research is the integration of the two techniques. The two-stage research approach employs continuous approximation models for the design of integrated networks, and numerical optimization techniques for more detailed service planning.

As mentioned in the introduction, this research addresses significant gaps identified in the continuous approximation literature. The systems modeled in this research include multiple time windows and multiple transportation modes for distribution. The continuous approximation cost functions used are capable of realistically modeling complex distribution systems with uncertain demands. Distribution activities include multiple transshipments, peddling tours, and shipment choice. All key distribution costs are included in these functions, including sorting, facility charges, and vehicle repositioning, as well as transportation and inventory. This research demonstrates that creative solution techniques can be applied to reduce these complex cost models to a series of subproblems that can be solved with standard spreadsheet technology.

Cost components are shown to model costs accurately using independent cost validation. In some cases, numerical optimization techniques are used to validate continuous cost models. For example, simulations of the transportation problem of linear programming are used in Appendix C to validate new cost models to estimate empty vehicle repositioning costs, and the approximation of longhaul ground transportation costs are validated in Chapter 4 with mathematical programming techniques.

This research demonstrates the use of continuous approximation models to better understand and better plan operating strategies for package delivery companies. While complete demand and cost data are not available for this research, proxies for this information are developed and an application to package distribution systems is provided. A variety of integration scenarios are analyzed both for a small distribution network, and a larger network roughly the size of the United States. The advantages of integrated operations are quantified and compared across different demand and network assumptions. Qualitative conclusions suggest that benefits of integration are greater when deferred demand exceeds express demand. Further, benefits increase significantly when demands are uncertain. This insight helps to explain the different business strategies of United Parcel Services and Federal Express. A large deferred carrier such as UPS should realize greater cost savings from integration.

Chapter 5 presents a variety of formulations and solution techniques for the deferred item and vehicle routing problem. Results from continuous approximation models proved useful here in reducing the problem size. With the iterative solution heuristic detailed in Chapter 5, large problems can be studied. The quality of solutions from these heuristics are tested with upper and lower bounds. This chapter introduces many ideas for improvements to the bounds presented and are discussed in the next section.

6.2 Areas of future research

Throughout this thesis, several areas of future research are identified. This work is a first step in exploring complex, multimode, multiservice logistics systems; there are many other scenarios to explore. As mentioned at the end of Chapter 3, seasonal demand fluctuations can be incorporated into the current cost modeling. The hybrid methodology can be used to assess how these fluctuations impact network design. The option of leasing equipment can be incorporated and new approaches to accommodating seasonal fluctuations can be developed and analyzed.

One key advantage of continuous approximation models is their ability to model large systems. Ideas developed here can be extended to the design of global package delivery networks to maintain acceptable service levels while operating in a cost efficient way. The modeling approach can also be extended to examine the performance of multiple hub air networks, perhaps even including multimodal hubs. Multimodal hubs would enable inbound and outbound transportation of deferred items to be modally decoupled at the hubs. Items traveling between a given origin/destination pair could then be served by a combination of modes with the mode transfer occuring at the main hub. This variation would provide greater flexibility to balance loads on aircraft into and out of the hub. These extensions can be easily incorporated into numerical optimization models as well. Computational problems may occur, however, as the problem sizes increase.

Furthermore, other supply-chain models for e-commerce applications can be studied. Models can be used to consider how the performance of multimode, multiservice networks affects the optimal configuration of on-line commerce firms such as web-based grocers and other types of retailers that sell and purchase multiple products with service delivery windows.

Results presented in Chapter 4 highlight the importance of local distribution and the potential opportunities for savings from integration. The study of local distribution has been, and should continue to be, a rich area for research. Specific local distribution research for integrated networks could focus on more intelligent strategies that exploit the advantages of integration. For example, delivery vehicles could be loaded with both express and deferred items at consolidation terminals each morning for delivery. Express items with tight deadlines could be delivered first before deferred items. Afternoon item pickup could be performed in the reverse order. This would allow vehicles to make better use of capacity and more evenly distribute driver work shifts throughout the day.

In addition, several areas of future research are identified in relation to the deferred item and vehicle routing problem in Chapter 5. One of the major road blocks to solving large instances of the deferred item and vehicle routing problem is the solution time for linear relaxations. Improving solution time is a critical next step in this area. Results suggest that improvements in the gaps between upper and lower bounds can be improved with refinements to both bounds. Chapter 5 introduces several types of cutting planes that can be added to the solution approach to improve lower bounds obtained from the linear relaxation, including additional cut-set inequalities, as well as residual arc capacity inequalities. Upper bounds can be improved with more intelligent rounding heuristics. Again, solution speed is critical here.

Another important area of future research would be to incorporate demand uncertainty into the deferred item and vehicle routing models to make the models more useful for day-to-day planning.

To explore larger and more complex problems, further integration of continuous and discrete approaches should be explored. This research has demonstrated how to incorporate different modeling techniques to analyze a class of complex logistics systems.

Appendix A

Routing in air networks

In this appendix, some unique savings opportunities in the air network from integration are discussed. Hub location models are reviewed in Magnanti and Wong (1984), Campbell (1993), and Ball *et al.* (1995). Time windows and multiple time zones have been included in analytic and discrete models (see Han (1984), Daganzo (1987a), Daganzo (1987b), Kiesling (1995), and Barnhart and Schneur (1996)). Studies have shown that pure hub and spoke networks may not be optimal for express package delivery, yet it is often assumed, for simplicity of analysis, that routing between main hubs and regional airports are symmetric inbound and outbound (see Hall (1989a) and Kuby and Gray (1993)). However, further work (Barnhart and Schneur (1996)) has shown that asymmetric routing can lower costs. In this appendix, the use of asymmetric routing to reduce fleet size is studied.

Recall the representation of a typical air distribution network with one main hub in Figure 2.1. As shown in that figure, peddling tours between airports and the hub may be introduced (as opposed to a pure hub-and-spoke structure) to operate the network with a smaller fleet and still maintain daily frequencies to meet time deadlines. Due to the high cost of purchasing additional aircraft, the goal here is to design an overnight network that operates with the fewest number of aircraft in the fleet. This is often a difficult goal, because one must be careful not to design the network too tightly. There must be slack in operations to ensure on-time delivery. This appendix focuses specifically on the impact of hub location and routing strategies on fleet size given a known density of airports.

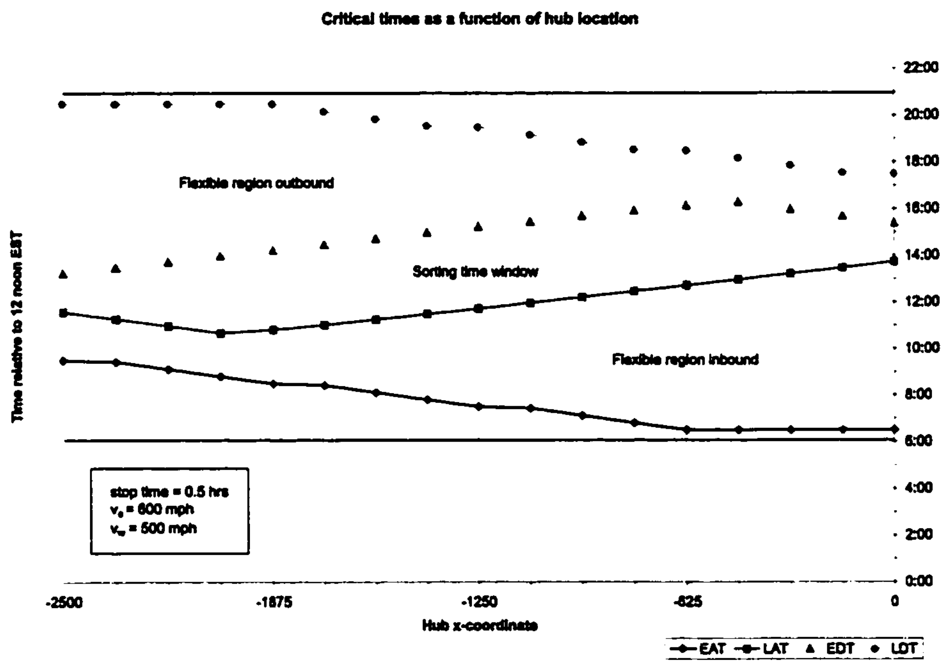


Figure A.1: Arrival and departure patterns as a function of hub location

Earlier work on air network design by Hall (1989a) studied the impact of time zones and sorting time windows on hub location. Figure A.1 extends observations made in that paper. The arrival and departure times of aircraft at the main hub are plotted for a pure hub and spoke network as a function of hub location. It is assumed that the hub is located in the center of the vertical axis of the region, and only the horizontal coordinate changes in Figure A.1. It is further assumed that aircraft fly directly from regional airports to the main hub in the evening and return directly to regional airports the next morning for delivery. The predefined evening departure time (in local time) is the same at all airports and the cut-off time for arrival to each airport the next morning is also the same across airports. As a result, the following times can be determined:

EAT the earliest arrival time at the main hub. This is the time at which the first aircraft arrives at the main hub.

LAT the time of the latest aircraft arrival at the hub. This is the time at which the last aircraft arrives at the main hub for sorting. No other aircraft can depart from the hub until items on this last aircraft have been offloaded, sorted, and loaded onto aircraft for morning delivery.

EDT the earliest departure from the hub to meet the cut-off arrival time. This is the time at which aircraft for the regional airport with the most severe delivery time constraints must leave the main hub in order to arrive the next morning in time for delivery.

LDT the latest departure time to meet the cut-off arrival time. This is the latest time at which an aircraft can leave the main hub and still perform morning delivery on time in the most favorably located regions.

These time values are plotted in Figure A.1 as a function of the hub location. Times are measured relative to noon EST. The time between the latest arrival time and the earliest departure time is the sorting window. This is the time when all items are at the main hub and can be sorted and loaded onto aircraft. The figure shows flexible zones where aircraft arrive hours before the last arrival to the airport (the beginning of the sorting window) or can depart hours after the first aircraft must depart (the end of the sorting window). If time flexibility exists, the size of the aircraft fleet can be lowered by allowing planes to make multiple stops at airports. However, as the figure shows, because of time zone differences, aircraft that have time flexibility inbound may not have flexibility outbound (and vice versa). Here we explore ways to exploit these differences and make more productive use of aircraft by designing asymmetric routing strategies between airports and the main hub, allowing some aircraft to make a second stop inbound but not outbound and others to stop twice outbound, but not inbound. Asymmetric strategies could introduce empty redistribution miles; however, these miles may be offset by savings in fleet size, especially if empty miles can be filled with deferred items when networks are integrated.

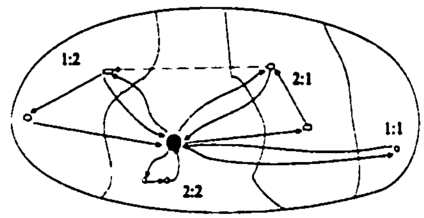


Figure A.2: Asymmetric routing strategies

Figure A.2 illustrates the concept of asymmetric routing strategies. The service region has been divided into four routing zones. Close to the hub, where flight times are shorter and fewer time zone changes exist, it is possible to stop twice inbound and twice outbound, thereby serving two airports with just one aircraft. These regions are referred to as 2:2 regions. Far from the hub, where time constraints make multiple stops on both inbound and outbound routes impossible, all airports must be served directly. These regions are referred to as 1:1 regions. As Figure A.1 shows, there is some flexibility in the other two regions, where some locations have flexibility inbound (2:1 regions) and others outbound (1:2 regions) depending on their distance and direction from the hub. Thus, airports in the 1:2 region can be paired with airports in the 2:1 region and empty planes can be repositioned during the day. As a results, four airports can be served with only three aircraft whereas four aircraft would be needed with symmetric routing.

In Figure A.3, an operating plan for an asymmetric routing strategy is illustrated for an idealized network where all airports have the same demand level equal to one-half of an aircraft load. One aircraft is dedicated to 2:1 region. This aircraft leaves the hub in the morning for the furthest airport (half-full with airport's delivery volume). Later that afternoon, the aircraft departs with the outbound volume from that airport. However,

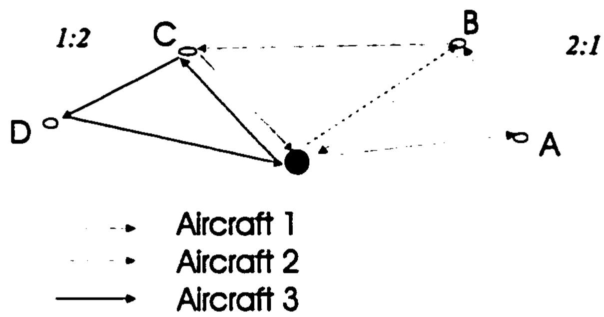


Figure A.3: Asymmetric regions: operating plan

before returning to the hub, the aircraft stops at a second airport in the 2:1 region to collect items from that region and then returns to the hub for overnight sorting. A second aircraft is dedicated to 1:2 region. This aircraft leaves the main hub each morning and makes two stops in the 1:2 region to deliver items. However, when the aircraft returns to the hub that evening, time constraints due to the distance from the hub and time zone effects permit only one stop. A third aircraft is shared between the regions. This aircraft serves one airport in the 2:1 region in the morning. During the day, the aircraft is repositioned to serve an airport in the 1:2 region inbound to the hub that evening.

Time is the biggest determinant of the number of stops permissible in a region. To determine the area of each routing region, the multiple-stop symmetric region (2:2) is maximized first and possible asymmetric regions are introduced; however, the regions must be balanced, so the areas of the two asymmetric regions are determined by the minimum of the two and the remainder becomes a 1:1 region.

A series of time constraints for the duration of one-stop tours (T_1) and two-stop tours (T_2) are defined to find the area of each service routing zone. It is important to consider the impact of time zones. The calculations for trip durations are presented below. Included in the equation for T_2 is the travel time between airports. The distance between neighboring airports is approximated by $k\Delta_P(x)^{-\frac{1}{2}}$.

$$T_1(x) = \tau_s + \Gamma(x, x_h)/v \tag{A.1a}$$

$$T_2(x) = 2\tau_s + \Gamma(x, x_h))/v + k\Delta_P(x)^{-\frac{1}{2}}/v \tag{A.1b}$$

where

τ_s : aircraft stopping time (loading, unloading, etc.)(*time/stop*)

v : velocity of an aircraft, may be direction dependent(*distance/time*)

x_h : location of main hub(*planar coordinates*)

$\Gamma(x, x_h)$: distance function between an airport at x and the hub(*distance*)

The number of stops $n_2^{A,b}(x)$ for $b = i, o$ can then be determined by:

$$n_2^{A,i}(x) = \begin{cases} 2, & \text{if } TZ(x) + T_2(x) < LAT \\ 1, & \text{otherwise} \end{cases} \tag{A.2a}$$

$$n_2^{A,o}(x) = \begin{cases} 2, & \text{if } -TZ(x) - T_2(x) < EDT \\ 1, & \text{otherwise} \end{cases} \tag{A.2b}$$

where

$TZ(x)$: Time zone at x (relative to PST)

The above equations are used to draw service regions of area, $A^{i:o}$, with i stops inbound and o stops outbound. Due to the discrete nature of time zones, continuous models alone cannot be used with this formulation. Therefore, a hybrid formulation is needed. For each time zone, continuous approximation expressions (A.1a and A.1b) determine routing regions in each time zone. Then regions are combined and balanced across time zones. Then fleet size and aircraft miles can be calculated for each region.

As an approximation for fleet size, the number of planes per airport are integrated over the routing region.

Fleet size

1 : 1 region:	$\int_{A^{1:1}} \Delta_P(x)dx$
2 : 2 region:	$\frac{1}{2}\int_{A^{2:2}} \Delta_P(x)dx$
1 : 2/2 : 1 region:	$\frac{3}{4}\int_{A^{1:2/2:1}} \Delta_P(x)dx$

To calculate loaded miles, the same approach is used and the miles are integrated over the region. Routes in 2:2 and 2:1 and 1:2 regions include a detour for additional stops. Note that loaded miles may not be fully loaded, as shown in Figure A.2.

Loaded miles

1 : 1 region:	$\int_{A^{1:1}} \Delta_P(x)2r(x)dx$
2 : 2 region:	$\frac{1}{2}\int_{A^{2:2}} \Delta_P(x)\left(2r(x) + \Delta_P(x)^{-\frac{1}{2}}\right)dx$
1 : 2/2 : 1 region:	$\frac{3}{4}\int_{A^{1:2/2:1}} \Delta_P(x)\left(2r(x) + \Delta_P(x)^{-\frac{1}{2}}\right)dx$

Finally, we need to consider the repositioning miles required in asymmetric regions.

Repositioning miles

1 : 2/2 : 1 region:	$\frac{1}{2}\int_{A^{1:2/2:1}} (\bar{d})\Delta_P(x)dx$

where $\bar{d}$ = average distance between regions

In Figure A.4, the above methodology is used to determine the routing strategies for given hub locations in an idealized service region. Demand is assumed to be uniform and balanced. Hub locations in the center of the region yield the largest values for $A^{2:2}$ and $A^{1:2/2:1}$. These results are slightly different than results obtained when minimizing total vehicle miles rather than aircraft fleet size (see Hall (1989a)). There is clearly a trade-

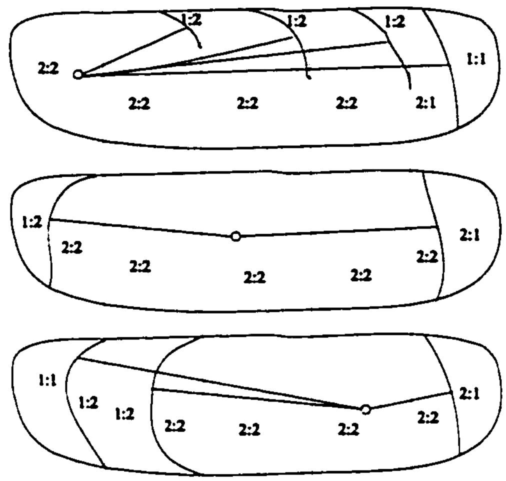


Figure A.4: Impact of hub location on routing regions

off between fleet size and empty miles, since asymmetric strategies required additional repositioning miles. Further exploration of this trade-off is left for future work.

This methodology is then applied to data from the United States. As a proxy for the largest package demand centers, the hundred largest Metropolitan Statistical Areas are used. The optimal hub location, along with the routing subregions are shown in Figure A.5. Operating asymmetric routes would require only seventy-six aircraft. A pure hub-and-spoke network would require one hundred aircraft and a symmetric network (only 1:1 and 2:2 regions) would require eighty-five aircraft.

Additional tests are performed to analyze the sensitivity of these results to hub location. Tests are conducted with hub locations for pure hub and spoke network where goal is to minimize travel distance (i.e. results in Hall (1989a) which suggest hubs should located slightly east of the center of the region) and with current hub locations for package delivery firms.

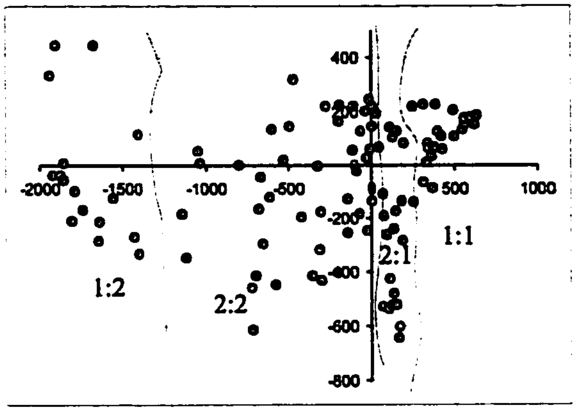


Figure A.5: Hub location and routing with 100 Metropolitan Statistical Areas

Hub location	Fleet size	Cities in $A^{2:1}$	Cities in $A^{2:2}$
Louisville, KY	95	2	9
Memphis, TN	78	13	31
Indianapolis, IN	94	3	9

Table 1: Impact of hub location on fleet size

Of course, there are other considerations in hub location, including cost of land, weather, etc. This appendix has shown that there are benefits to focusing on reducing fleet size rather than miles. Such an objective does, however, create empty repositioning mileage. There are opportunities in multimode, multiservice distribution networks to allocate this empty aircraft space to other service levels and transform empty miles into loaded miles while keeping the aircraft fleet size down.

Appendix B

Parameter estimates

As mentioned in Chapter 4, much of the information needed to estimate cost function parameters is private. In this appendix, parameter values are defined, and where appropriate the derivation of these values is included. The appendix covers cost and operational parameters, as well as demand parameters.

B.1 Operational and cost parameters

Cost estimates are derived from work by Kiesling (1995). Vehicle costs and operating statistics are derived from earlier work by Kiesling (1995) and company literature. Vehicle capacity is given in units of items. In each test case, a single aircraft size is assumed. However, it is possible to reoptimize aircraft size by region. To do so, some consideration must be given to repositioning aircraft between regions with the same aircraft size. A shift factor, ν of 85% is used.

The average distance between breakbulk terminals in SR1 is 192 miles and the max-

[1]Using continuous approximation models, it became clear that operating the idealized deferred dominant scenarios with the same large aircraft as in the other scenarios did not make sense economically and a smaller fleet was used.

	c_d	c_d'	c_q	V	N	H
Level 0	0.2	1.5	1.75	50	50	1
Level 1	0.1	1.2	6	500	10	1
Level 2^A	0.05	10	20	10,000[1]	2	1
Level 2^G	0.075	1	8	1,000	5	3
	c_f	c_f'	c_k	c_h		
CT	1370	0.25	0.045	0.005		
BBT	2557	0.25	0.045	0.005		
Airport	2557	0.5	0.045	0.005		

Table B.1: Cost and operating parameters

imum service radius of an airport is 50 miles. For SR2, the average distance between breakbulk terminals is 1,118 miles and the maximum service radius of an airport is 100 miles.

B.2 Demand parameters

Total demand for each test case is described in Chapter 4. For SR1, demand and customer densities, as well as uncertainty levels for each subregion are generated randomly. Values for SR2 are based on the 1990 United States census with some randomization to allow for inbound and outbound demand imbalances and differences between express and deferred demands within regions. Customer density is estimated with housing counts as a proxy for customer locations and population for demand. All randomization is performed while ensuring total network balance of item flows. The following tables present demand and geographic data for the test cases. The percent of the total service region area is listed, as well as the distance to the main hub for each region. Demand and customer densities are listed for deferred and express service levels. Random and deterministic cases have the same parameters with the exception that $\gamma_b^s(x) = 0$ for all directions and service levels in the deterministic cases. Therefore, only random cases are shown.

	area	r_2^A	λ_o^D	λ_i^D	γ_o^D	γ_i^D	δ^D	λ_o^E	λ_i^E	γ_o^E	γ_i^E	δ^E
1	3%	238	10.5	9.5	3.5	0.6	5.4	7.7	8.9	0.5	0.3	3.4
2	5%	202	4.1	4.9	3.8	0.6	2.3	4.5	4.7	0.5	0.0	2.0
3	3%	202	10.5	12.2	1.7	0.0	6.1	8.0	9.3	0.3	1.3	3.5
4	4%	235	6.8	6.7	1.7	0.6	3.5	6.7	6.7	1.3	2.5	3.1
5	4%	176	7.4	6.9	0.1	0.4	3.6	6.0	4.6	0.9	4.7	2.2
6	8%	100	3.4	2.7	3.7	0.4	1.6	3.1	3.1	4.1	0.9	1.5
7	9%	146	2.4	3.6	0.1	4.6	1.6	3.3	2.9	1.3	1.7	1.4
8	7%	100	3.0	4.1	2.8	3.8	1.8	3.2	2.7	0.9	0.0	1.3
9	4%	35	6.6	6.3	4.1	2.6	3.3	5.7	4.3	0.2	0.6	2.1
10	4%	43	7.4	4.8	0.3	1.6	3.1	6.1	7.7	0.5	0.2	2.9
11	7%	108	2.9	4.0	0.0	1.2	2.0	3.4	3.8	3.2	1.7	1.6
12	10%	117	2.7	2.7	1.3	2.0	1.4	2.6	2.3	0.0	0.2	1.2
13	5%	90	6.2	5.8	0.1	0.4	3.1	4.6	5.9	1.7	1.6	2.2
14	4%	94	6.7	6.4	2.4	2.4	3.5	6.4	5.3	0.3	1.6	2.6
15	4%	224	6.2	5.4	4.0	1.9	3.0	5.2	5.9	0.5	1.0	2.4
16	8%	200	3.8	3.1	0.9	1.7	1.8	3.4	4.0	0.0	3.3	1.6
17	10%	204	3.1	3.0	0.2	0.7	1.6	3.3	2.9	2.6	0.8	1.4

Table B.2: Demand parameters for SR1[R,B]

	area	r_2^A	λ_o^D	λ_i^D	γ_o^D	γ_i^D	δ^D	λ_o^E	λ_i^E	γ_o^E	γ_i^E	δ^E
1	3%	238	13.1	11.9	3.5	1.1	6.6	1.0	1.3	0.8	0.4	0.6
2	5%	202	5.1	6.2	4.3	4.9	2.7	0.6	0.7	1.5	1.3	0.3
3	3%	202	13.2	15.2	3.8	0.0	7.4	1.6	1.0	0.3	1.4	0.7
4	4%	235	8.5	8.4	1.8	0.8	4.2	0.7	1.0	1.4	3.0	0.4
5	4%	176	9.3	8.7	0.6	4.5	4.3	0.7	1.0	4.6	4.9	0.4
6	8%	100	4.2	3.4	4.1	4.6	2.0	0.4	0.4	4.1	1.2	0.2
7	9%	146	3.1	4.6	4.4	4.9	1.8	0.3	0.4	4.3	3.3	0.2
8	7%	100	3.8	5.1	3.0	5.0	2.2	0.4	0.4	2.4	0.0	0.2
9	4%	35	8.3	7.9	4.1	3.4	4.0	0.9	0.8	0.8	2.2	0.5
10	4%	43	9.2	6.0	1.2	3.6	3.6	0.8	0.8	0.7	0.3	0.4
11	7%	108	3.7	5.1	0.1	1.6	2.4	0.5	0.5	3.7	2.5	0.2
12	10%	117	3.3	3.4	3.4	2.6	1.7	0.4	0.3	0.4	2.7	0.2
13	5%	90	7.7	7.2	3.5	4.9	3.6	0.9	0.7	3.0	4.0	0.4
14	4%	94	8.4	7.9	4.3	4.8	4.2	0.9	0.9	0.8	2.1	0.5
15	4%	224	7.8	6.8	4.4	3.5	3.6	1.0	0.9	0.9	1.1	0.5
16	8%	200	4.7	3.8	1.7	3.2	2.1	0.3	0.4	0.0	4.4	0.2
17	10%	204	3.8	3.7	0.4	4.2	1.9	0.3	0.4	3.9	0.8	0.2

Table B.3: Demand parameters for SR1[R,D]

	area	r_2^A	λ_o^D	λ_i^D	γ_o^D	γ_i^D	δ^D	λ_o^E	λ_i^E	γ_o^E	γ_i^E	δ^E
1	3%	238	5.0	4.3	0.5	0.4	2.2	7.8	9.1	0.2	0.0	4.2
2	5%	202	2.9	2.5	1.4	0.9	1.4	4.6	4.8	0.7	1.2	2.3
3	3%	202	6.6	8.2	0.1	1.0	3.6	8.2	9.5	0.2	1.3	4.2
4	4%	235	4.2	5.3	1.3	2.9	2.4	6.9	6.8	0.7	1.4	3.7
5	4%	176	4.3	4.6	3.4	0.9	2.3	6.2	4.7	0.9	3.2	2.7
6	8%	100	2.0	1.7	0.9	0.3	0.9	3.1	3.2	2.5	0.4	1.7
7	9%	146	2.1	1.5	3.4	0.9	0.9	3.4	3.0	4.2	3.1	1.6
8	7%	100	2.0	2.2	1.5	0.0	1.1	3.3	2.8	1.2	0.0	1.5
9	4%	35	3.8	3.6	0.2	0.4	1.8	5.8	4.4	0.2	1.0	2.5
10	4%	43	4.1	4.0	0.1	0.2	2.1	6.2	7.8	0.1	0.1	3.5
11	7%	108	1.8	2.2	2.6	0.1	1.0	3.5	3.9	1.5	0.4	1.9
12	10%	117	1.8	1.9	0.0	2.7	0.9	2.7	2.4	0.2	2.7	1.3
13	5%	90	3.6	3.8	2.8	2.7	1.9	4.7	6.1	1.0	0.9	2.7
14	4%	94	4.1	3.7	0.0	0.1	1.9	6.5	5.4	0.3	2.0	3.0
15	4%	224	4.6	4.1	0.5	1.0	2.1	5.3	6.0	0.4	0.4	2.9
16	8%	200	1.9	1.7	0.0	3.2	0.9	3.5	4.1	0.0	3.5	1.9
17	10%	204	2.0	2.2	3.7	0.0	1.1	3.4	2.9	1.4	0.1	1.6

Table B.4: Demand parameters for SR1[R,E]

	area	r_2^A	λ_o^D	λ_i^D	γ_o^D	γ_i^D	δ^D	λ_o^E	λ_i^E	γ_o^E	γ_i^E	δ^E
1	6%	1236	0.3	0.4	0.5	6.0	0.2	0.4	0.4	2.8	1.3	0.2
2	5%	427	0.5	0.4	0.3	2.3	0.2	0.5	0.6	1.4	4.2	0.2
3	4%	800	0.9	1.1	3.1	1.4	0.4	0.9	1.1	4.2	3.1	0.5
4	6%	1262	0.8	0.7	4.5	4.2	0.3	0.7	0.7	2.5	3.5	0.3
5	7%	1172	0.6	0.4	3.2	6.3	0.2	0.6	0.6	0.5	2.2	0.3
6	9%	914	0.8	0.8	0.1	0.2	0.3	0.8	0.9	0.5	3.7	0.3
7	8%	307	0.8	0.7	2.9	2.9	0.2	0.8	0.7	1.5	0.8	0.3
8	7%	721	0.8	1.0	0.0	1.5	0.3	0.7	0.8	6.1	5.8	0.4
9	1%	1256	5.0	3.4	0.1	6.0	1.9	3.0	2.5	4.7	4.8	2.3
10	2%	574	1.5	1.7	1.3	2.8	0.6	1.9	1.5	0.1	2.1	0.7
11	4%	1082	1.6	1.8	0.2	1.3	0.6	1.8	1.5	1.0	4.8	0.7
12	3%	303	0.5	0.4	1.5	2.0	0.2	0.6	0.4	1.9	4.2	0.2
13	8%	732	0.2	0.2	0.2	1.4	0.1	0.2	0.3	4.5	5.4	0.1
14	4%	597	0.2	0.1	1.0	1.9	0.0	0.2	0.2	1.3	3.2	0.0
15	5%	220	0.4	0.6	0.6	2.1	0.2	0.6	0.6	0.3	3.7	0.3
16	9%	535	0.4	0.4	0.7	4.2	0.2	0.4	0.5	3.6	5.3	0.2
17	6%	766	0.5	0.7	3.1	4.5	0.3	0.7	0.7	4.1	6.3	0.3
18	5%	956	0.9	0.6	2.1	2.2	0.4	1.0	0.8	3.5	2.8	0.5
19	1%	1397	1.1	1.9	2.1	4.4	0.7	1.2	1.4	2.7	3.2	0.9
20	1%	1320	1.8	1.4	1.9	0.2	0.7	1.9	2.2	1.8	3.4	0.9

Table B.5: Demand parameters for SR2[R,B]

	area	r_2^A	λ_o^D	λ_i^D	γ_o^D	γ_i^D	δ^D	λ_o^E	λ_i^E	γ_o^E	γ_i^E	δ^E
1	6%	1236	2.0	2.2	4.9	8.8	0.4	0.1	0.2	8.8	2.8	0.1
2	5%	427	2.3	2.5	2.9	3.9	0.5	0.2	0.2	1.4	7.9	0.1
3	4%	800	4.2	4.9	6.6	5.3	0.9	0.5	0.6	7.2	7.1	0.2
4	6%	1262	3.3	3.6	4.9	6.8	0.6	0.4	0.3	7.0	5.8	0.1
5	7%	1172	2.9	3.2	5.0	7.7	0.5	0.3	0.2	5.3	5.0	0.1
6	9%	914	3.7	4.3	2.5	4.4	0.7	0.4	0.4	0.8	3.7	0.1
7	8%	307	4.0	3.3	5.3	4.2	0.6	0.4	0.3	4.4	0.8	0.1
8	7%	721	3.7	4.1	1.6	5.4	0.7	0.4	0.5	6.4	5.8	0.1
9	1%	1256	14.7	12.2	8.3	7.1	4.6	2.5	1.7	6.0	4.8	0.9
10	2%	574	9.4	7.7	8.7	9.5	1.4	0.7	0.8	4.1	9.2	0.3
11	4%	1082	9.0	7.5	2.3	3.5	1.4	0.8	0.9	5.7	4.8	0.3
12	3%	303	2.6	2.0	6.5	3.3	0.4	0.2	0.2	6.8	4.2	0.1
13	8%	732	1.1	1.3	5.9	2.7	0.2	0.1	0.1	9.7	5.4	0.0
14	4%	597	0.2	0.5	2.3	3.0	0.1	0.1	0.0	8.3	9.6	0.0
15	5%	220	2.5	2.6	0.9	2.9	0.6	0.2	0.3	4.8	3.7	0.1
16	9%	535	2.0	2.4	7.4	4.4	0.4	0.2	0.2	5.1	6.9	0.1
17	6%	766	3.3	3.5	4.8	5.5	0.6	0.2	0.3	5.4	7.1	0.1
18	5%	956	5.0	4.2	3.7	3.1	1.0	0.4	0.3	6.0	2.8	0.2
19	1%	1397	8.8	8.4	5.8	7.5	1.8	0.5	0.9	7.2	5.7	0.4
20	1%	1320	9.6	10.8	4.0	3.2	1.8	0.9	0.7	4.2	6.3	0.4

Table B.6: Demand parameters for SR2[R,D]

	area	r_2^A	λ_o^D	λ_i^D	γ_o^D	γ_i^D	δ^D	λ_o^E	λ_i^E	γ_o^E	γ_i^E	δ^E
1	6%	1236	0.3	0.3	4.9	8.8	0.2	0.5	0.7	8.8	2.8	0.2
2	5%	427	0.4	0.5	2.9	3.9	0.2	0.8	0.7	1.4	7.9	0.2
3	4%	800	0.8	0.9	6.6	5.3	0.4	1.6	1.9	7.2	7.1	0.5
4	6%	1262	0.5	0.6	4.9	6.8	0.3	1.4	1.2	7.0	5.8	0.3
5	7%	1172	0.5	0.5	5.0	7.7	0.2	1.0	0.8	5.3	5.0	0.3
6	9%	914	0.6	0.7	2.5	4.4	0.3	1.3	1.4	0.8	3.7	0.3
7	8%	307	0.6	0.5	5.3	4.2	0.2	1.4	1.2	4.4	0.8	0.3
8	7%	721	0.6	0.7	1.6	5.4	0.3	1.4	1.7	6.4	5.8	0.4
9	1%	1256	2.4	2.0	8.3	7.1	1.9	8.0	6.4	6.0	4.8	2.3
10	2%	574	1.5	1.2	8.7	9.5	0.6	2.4	2.8	4.1	9.2	0.7
11	4%	1082	1.4	1.2	2.3	3.5	0.6	2.6	3.0	5.7	4.8	0.7
12	3%	303	0.5	0.4	6.5	3.3	0.2	0.8	0.7	6.8	4.2	0.2
13	8%	732	0.2	0.2	5.9	2.7	0.1	0.4	0.3	9.7	5.4	0.1
14	4%	597	0.2	0.2	2.3	3.0	0.0	0.2	0.2	8.3	9.6	0.0
15	5%	220	0.5	0.5	0.9	2.9	0.2	0.6	1.0	4.8	3.7	0.3
16	9%	535	0.3	0.4	7.4	4.4	0.2	0.7	0.7	5.1	6.9	0.2
17	6%	766	0.5	0.6	4.8	5.5	0.3	0.9	1.1	5.4	7.1	0.3
18	5%	956	0.8	0.7	3.7	3.1	0.4	1.5	1.1	6.0	2.8	0.5
19	1%	1397	1.0	1.1	5.8	7.5	0.7	1.8	3.0	7.2	5.7	0.9
20	1%	1320	1.5	1.7	4.0	3.2	0.7	3.0	2.4	4.2	6.3	0.9

Table B.7: Demand parameters for SR2[R,E]

Appendix C

Transportation problem approximation

To obtain a simple model for expected empty vehicle repositioning miles traveled between terminals, an approximation for the transportation problem of linear programming is developed. Similar formulas have already been developed for the traveling salesman problem, or "TSP" (Eilon *et al.* (1971), Karp (1977), Daganzo (1984b)), and for the vehicle routing problem, or "VRP" (Eilon *et al.* (1971), Daganzo (1984a), Haimovich *et al.* (1985), Newell and Daganzo (1986a), Newell and Daganzo (1986b), Newell (1986)). The results apply to problems where N points are randomly and homogeneously distributed on a region of a metric plane with area $|\mathcal{A}|$, and density $\Delta = N/|\mathcal{A}|$. In all cases the distance traveled per point for the TSP, or the "detour" distance per point for the VRP, tends to a fixed multiple of $\Delta^{-1/2}$ as N and $|\mathcal{A}|$ are increased in a fixed ratio; i.e., the average detour distance per point is bounded. Systematic design methodologies for complex "many-to-one" and "one-to-many" logistics problems based on these formulae have been developed; see for example Bramel and Simchi-Levi (1997), Daganzo (1999), and the references in these books.

Similar formulae are developed for TLP's where points lie on a region of a linear normed space, but the results are different. More detailed derivations and general results can be found in Daganzo and Smilowitz (2000).

Dimensional analysis yields exact formulae for any version of the TLP that can be completely specified in terms of just three constants. Specifically, for the vehicle repositioning problem, information about the area and shape of the service region, the number and location of terminals, the distances between terminals, and the vehicle demand/supply at each terminal can be summarized with just three parameters. It is assumed that the service region has a fixed shape; i.e., all regions are equal except for a scale parameter. Terminal locations are drawn from a uniform distribution over the service area. The supply/demand of vehicles (vehicles sent - vehicles received) at the end of each day is normally distributed with mean 0 and standard deviation σ. The problem is assumed to be balanced, i.e., total supply equals total demand over the service area. The three relevant parameters are therefore service area ($|\mathcal{A}|$), density of terminals, (Δ) and standard deviation in vehicle demand(σ). Area is measured in units of distance^2, standard deviation in vehicle demand in units of vehicles and average vehicle miles traveled per terminal ($\langle p \rangle$) in units of vehicle-miles. The number of terminals is dimensionless. Using dimensional analysis, we can eliminate two parameters and define the problem in terms of two independent dimensionless quantities, Λ_1 and Λ_2

$$\Lambda_1 = \langle p \rangle \frac{\sqrt{\Delta}}{\sigma} \qquad\qquad \Lambda_2 = \Delta|\mathcal{A}|$$

Writing Λ_1 as a function of Λ_2, the equation for vehicle miles traveled as a function of the three parameters becomes

$$\langle p \rangle = \frac{\sigma}{\sqrt{\Delta}} f_s(\Delta|\mathcal{A}|)$$

where f_s is the only unknown left to be determined. This function will generally depend on the type of problem; e.g., regions shape s and the norm. Since f_s is dimensionless, it will be called the "dimensionless distance per point". It was found in Daganzo and Smilowitz (2000) that if $N = \Delta|\mathcal{A}|$ is large, an approximation for the dimensionless distance per point in 2-D in a square region, is $f(N) \approx k_1 + k_2 \log_2(N)$ where k_1 and k_2 are constants dependent on the distance metric. The constants were estimated using a Monte Carlo simulation of the transportation problem of redistributing empty vehicles. Figure 2 displays the results on a diagram of $f(N)$ vs. $\log_2(N)$ where $N = \Delta|\mathcal{A}|$. The results speak for themselves. A function for $f(N) = 0.42 + 0.031 \log_2(N)$ was obtained for $N \in [25, 5000]$ with a Euclidean metric. The deviations from the line are consistent with the standard errors estimated from the simulation. Since the average number of items supplied or demanded per point is $\frac{\sigma}{\sqrt{2\pi}}$ in the case of normal demands, we see that the average distance traveled per item in the Euclidean case is $f(N)\sqrt{2\pi}\Delta^{-1/2}$. Further, we see that

$$< \text{distance per terminal} >= \sigma\Delta^{-1/2}(1 + 0.078 \log_2(\Delta|\mathcal{A}|)) \tag{C.1}$$

As a point of reference, this distance is about twice as long as for the Euclidean TSP, for the values of N one is likely to encounter in actual logistics problems ($N \approx 2^5 \rightarrow 2^{10}$). It is shown in Daganzo and Smilowitz (2000) that the coefficient 0.078 is independent of zone shape.

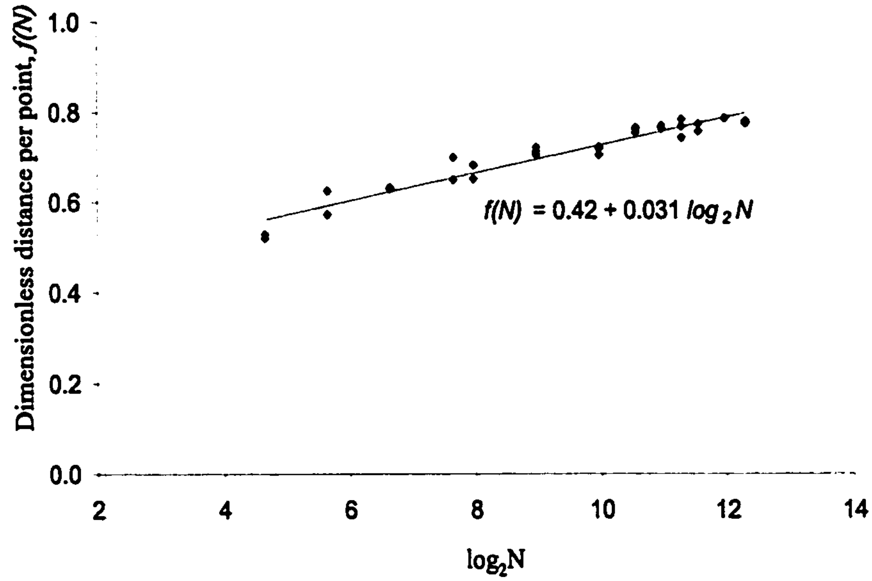


Figure C.1: Dimensionless distance per point v. $\log_2(N)$

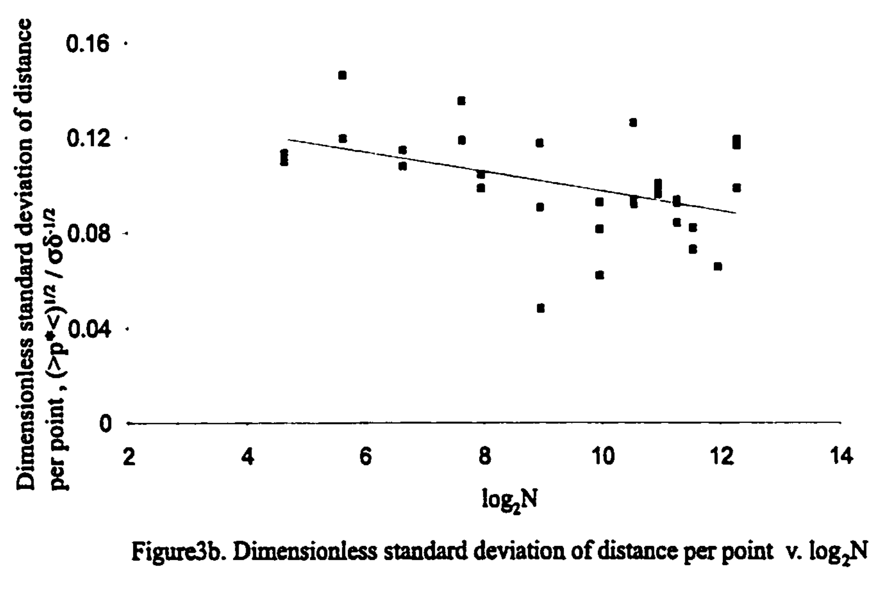


Figure C.2: Dimensionless standard deviation of distance per point v. $\log_2(N)$

Bibliography

Ahuja, R., Magnanti, T., and Orlin, J. (1993). *Network Flows: Theory, Algorithms and Applications.* Prentice-Hall, Inc., Englewood Cliffs, N.J.

Armacost, A. (2000). *Composite Variable Formulations for Express Shipment Service Network Design.* PhD dissertation, Massachusetts Institute of Technology.

Atamtürk, A. (2000). On capacitated network design cut–set polyhedra. Technical report, IEOR, University of California at Berkeley.

Atamtürk, A. and Rajan, D. (2000). On splittable and unsplittable capacitated network design arc-set polyhedra. Technical report, IEOR, University of California at Berkeley.

Ball, M., Magnanti, T., Monma, C., and Nemhauser, G., editors (1995). *Network Models,* volume 7 of *Handbooks in Operations Research and Management Science.* Elsevier Science Publishing, New York.

Barnhart, C. and Schneur, R. R. (1996). Air network design for express shipment service. *Operations Research,* **44**(6), 852–863.

Barnhart, C., Johnson, E., Nemhauser, G., Savelsbergh, M., and Vance, P. (1998). Branch and price: Column generation for solving huge integer programs. *Operations Research,* **46**(3), 316–329.

Blumenfeld, Dennis, E., Burns, L., Frick, M., Daganzo, C., and Hall, R. (1987). Reducing logistics costs at General Motors. *Interfaces,* **17**, 26–47.

Bramel, J. and Simchi-Levi, D., editors (1997). *The Logic of Logistics*. Springer, Heidelberg, Germany.

Campbell, J. F. (1993). Continuous and discrete demand hub location problems. *Transportation Research B*, **27B**(6), 473–482.

Crainic, T. (2000). Service network design in freight transportation. *European Journal of Operational Research*, **122**(2), 272–288.

Crainic, T. and Rousseau, J.-M. (1986). Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem. *Transportation Research B*, **20B**(3), 225–242.

Daganzo, C. F. (1984a). The distance traveled to visit N points with a maximum of C stops per vehicle: An analytic model and an application. *Transportation Science*, **18**(4), 331–350.

Daganzo, C. F. (1984b). The length of tours in zones of different shapes. *Transportation Research B*, **18B**, 135–146.

Daganzo, C. F. (1987a). Modeling distribution problems with time windows: Part i. *Transportation Science*, **21**(3), 171–179.

Daganzo, C. F. (1987b). Modeling distribution problems with time windows: Part ii. *Transportation Science*, **21**(3), 180–187.

Daganzo, C. F. (1999). *Logistics Systems Analysis*. Springer, New York.

Daganzo, C. F. and Erera, A. L. (1999). On planning and design of logistics systems for uncertain environments. In M. G. Speranza and P. Stahly, editors, *New Trends in Distribution Logistics*, pages 100–105. Springer, Berlin.

Daganzo, C. F. and Newell, G. F. (1986). Configuration of physical distribution networks. *Networks*, **16**, 113–132.

Daganzo, C. F. and Smilowitz, K. R. (2000). Asymptotic approximations for the transportation problem of linear programming and other scalable networks. Working paper, UCB-ITS-WP-2000-2, University of California, Berkeley.

Daskin, M. S. (1985). Logistics: An overview of the state of the art and perspectives on future research. *Transportation Research*, **19A**(5/6), 383–398.

Dejax, P. J. and Crainic, T. G. (1987). A review of empty flows and fleet management models in freight transportation. *Transportation Science*, **21**(4), 227–247.

Eilon, S., Watson-Gandy, C. T., and Christofides, N. (1971). *Distribution Management*. Griffin, London.

Erera, A. L. (2000). *Design of Logistics Systems for Uncertain Environments*. PhD dissertation, University of California, Berkeley, Institute of Transportation Studies.

Farvolden, J. M., Lustig, I. J., and Powell, W. B. (1993). A primal partitioning solution for the arc-chain formulation of a multicommodity network flow problem. *Operations Research*, **41**(4), 669–693.

Federal Express Corporation (1998a). 1998 SEC Form 10K.

Federal Express Corporation (1998b). Investor relations information packet.

FedEx Corporation (2000). Unleashing the power of a global brand. http://www.fedex.com/us/about/ download/fedex_brochure.pdf.

Geoffrion, A. (1976). The purpose of mathematical programming is insight, not numbers. *Interfaces*, **7**(1), 81–92.

Haimovich, M., Rinnooy-Kan, A., and Stougie, L. (1985). Bounds and heuristics for capacitated routing problems. *Math. Opns. Res.*, **10**, 527–542.

Hall, R. (1986). Discrete models / continuous models. *Omega, International Journal of Management Science*, **14**, 213–220.

Hall, R. (1989a). Configuration of an overnight package air network. *Transportation Research A*, **23A**(2), 139–149.

Hall, R. (1989b). Dispatching regular and express shipments between a supplier and manufacturer. *Transportation Research B*, **23B**(3), 195–211.

Hall, R. (1991). Characteristics of multi-stop / multi-terminal delivery routes with backhauls and unique items. *Transportation Research B*, **25B**(6), 391–403.

Han, A. F.-W. (1984). *One-to-Many Distribution of Nonstorable Items: Approximate Analytical Models.* PhD dissertation, University of California, Berkeley, ITS.

Han, A. F.-W. and Daganzo, C. F. (1985). Distribution of perishable items: transshipments allowed. ITS research report, University of California, Berkeley.

Jones, K. L., Lustig, I. J., Farvolden, J. M., and Powell, W. B. (1993). Multicommodity network flows: the impact of formulation on decomposition. *Mathematical Programming*, **62**, 95–117.

Jordan, W. and Burns, L. (1984). Truck backhauling on two terminal networks. *Transportation Research B*, **18B**(6), 487–503.

Karp, R. (1977). Probabilistic analysis of partitioning algorithms for the traveling salesman problem. *Math. Opns. Res.*, **2**, 209–224.

Kiesling, M. K. (1995). *A comparison of freight distribution costs for combination and* *dedicated carriers in the air express industry*. PhD dissertation, University of California, Berkeley, Institute of Transportation Studies.

Kim, D., Barnhart, C., Ware, K., and Reinhardt, G. (1999). Multimodal express package delivery: A service network design application. *Transportation Science*, **33**(4), 391–407.

Kuby, M. J. and Gray, R. G. (1993). The hub network design problem with stopovers and feeders: The case of Federal Express. *Transportation Research A*, **27A**(1), 1–12.

Langevin, A., Mbaraga, P., and Campbell, J. (1996). Continuous approximation models in freight distribution: An overview. *Transportation Research B*, **30B**(3), 163–188.

Magnanti, T. and Wong, R. (1984). Network design and transportation planning: Models and algorithms. *Transportation Science*, **18**(1), 1–55.

Nemhauser, G. and Wolsey, L. (1999). *Integer and Combinatorial Optimization*. Wiley, New York.

Newell, G. (1973). Scheduling, location, transportation, and continuum mechanics: some simple approximations to optimization problems. *SIAM, Journal of Applied Mathematics*, **25**, 346–360.

Newell, G. (1986). Design of multiple vehicle delivery tours–iii: Valuable goods. *Transportation Research B*, **20B**(5), 377–390.

Newell, G. and Daganzo, C. (1986a). Design of multiple vehicle delivery tours–i: A ring-radial network. *Transportation Research B*, **20B**(5), 345–364.

Newell, G. and Daganzo, C. (1986b). Design of multiple vehicle delivery tours–ii: Other metrics. *Transportation Research B*, **20B**(5), 365–376.

Newman, A. M. and Yano, C. A. (2000). Scheduling direct and indirect trains and containers in an intermodal setting. *Transportation Science*, **34**(3), 256–270.

O'Reilly, B. (2000). They've got mail. *Fortune*, **141**(3).

Popken, D. (1994). An algorithm for the multiattribute, multicommodity flow problem with freight consolidation and inventory costs. *Operations Research*, **42**(2), 274–286.

Powell, W. (1996). A stochastic formulation of the dynamic assignment problem, with an application to truckload motor carriers. *Transportation Science*, **30**, 195–219.

Powell, W. and Sheffi, Y. (1983). The load planning problem of motor carriers: Problem description and a proposed solution approach. *Transportation Research A*, **17A**(6), 471–480.

Robuste, F., Daganzo, C. F., and Souleyrette II, R. (1990). Implementing vehicle routing models. *Transportation Research B*, **24**(4), 263–286.

Rocks, D. (2000). Going nowhere fast in cyberspace. *Business Week*.

Rosenfield, D., Engelstein, I., and Feigenbaum, D. (1992). An application of sizing service territories. *European Journal of Operations Research*, **63**, 164–172.

United Parcel Service (2000a). About UPS. http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=UPS & script=2100.

United Parcel Service (2000b). Annual report.

United Parcel Service (2000c). Domestic services. http://www.ups.com/using/services/domestic/dom-guide.html.

United Parcel Service (2000d). The UPS story. http://www.ups.com/about/story.htm.

U.S. Census Bureau (1990). Tiger/geographic identification code scheme.

Wolsey, L. (1998). *Integer Programming*. Wiley, New York.