2007

The hub covering flow problem and the stochastic p-hub center problem

Thaddeus Kim Teck Sim
University of Iowa

THE HUB COVERING FLOW PROBLEM AND

THE STOCHASTIC P-HUB CENTER PROBLEM

by

Thaddeus Kim Teck Sim

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Business Administration
in the Graduate College of
The University of Iowa

December 2007

Thesis Supervisors: Professor Timothy J. Lowe
Assistant Professor Barrett W. Thomas

ABSTRACT

The hub location problem seeks to find the best location for hubs and the assignment of non-hub nodes to hubs. The resulting structure is a hub-and-spoke network. In this dissertation, we study the hub location problem within the context of three different application areas.

We first consider the air travel industry, where many airline companies (especially the legacy airlines) operate hub-and-spoke networks. Important considerations in designing such a hub-and-spoke network are the cost of establishing the hubs, the cost of transporting passengers through the network, and the operating costs and maximum flying ranges of the different aircraft types in the fleet. We introduce the hub covering flow problem, which accounts for all these design considerations.

We next look at time-definite delivery services in the small package delivery industry, where a customer's package is guaranteed to be delivered within a certain time-window. With possible delays along the transportation route, a package may not be delivered on time. We introduce the stochastic p-hub center problem, which seeks to design a network where the longest origin-destination path in the network is minimized. The solution provides an upper bound on the delivery times, which can be used to design time-constrained service offerings, or to ascertain if current service guarantees are feasible. A service-level constraint is included into the model to ensure a high likelihood of on-time delivery.

In communication networks, two important service-level considerations are robustness (the ability of the network to perform when components become unavailable) and

response time (the time required to transfer information through the network). We propose another version of the stochastic p-hub center problem, which employs a two-stage stochastic programming formulation, that addresses the possibility of unavailable components or transmission delays in the network.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

_____
Thesis Supervisor

_____
Title and Department

_____
Date

THE HUB COVERING FLOW PROBLEM AND

THE STOCHASTIC P-HUB CENTER PROBLEM

by

Thaddeus Kim Teck Sim

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Business Administration
in the Graduate College of
The University of Iowa

December 2007

Thesis Supervisors: Professor Timothy J. Lowe
Assistant Professor Barrett W. Thomas

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Thaddeus Kim Teck Sim

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Business Administration at the December 2007 graduation.

Thesis Committee: _____
Timothy J. Lowe, Thesis Supervisor

_____
Barrett W. Thomas, Thesis Supervisor

_____
Samuel Burer

_____
Jeffery W. Ohlmann

_____
Pavlo Krokhmal

To my family

Praise God, from Whom all blessings flow;
Praise Him, all creatures here below;
Praise Him above, ye heavenly host:
Praise Father, Son, and Holy Ghost. Amen.

Thomas Ken, “Doxology” (1674)

ACKNOWLEDGEMENTS

With life as a college student coming to a close, it is with much gratitude that I recognize and thank the many people who have helped me along this journey.

I first thank my advisors Professors Timothy Lowe and Barrett Thomas, who guided me through the dissertation process and taught me the importance of being a careful researcher and writer. Thanks also are due to the members of my committee – Professors Samuel Burer, Jeffrey Ohlmann, and Pavlo Krokhmal – for their input throughout this process. A long-distance thank you to Professor Erhan Erkut for introducing me to the wonderful world of OR/MS and for nudging me towards a career in academia.

Echoing the lyrics of a Beatles' song: I got by with a little help from my friends. These include my office-mate Dengfeng Zhang, Hui Chen, the folks in the ISOR lab (Professor Nick Street, Brian Almquist, Chris Choi, Justin Goodson, Xin Ying Qiu, Ding Yuan, and Yi Zhang), and my good friend Kaan Ataman, with whom I have attended many Hawkeye football games, sneaked off to watch soccer matches on TV, and spent countless hours shooting the breeze.

My department secretary Barb Carr has been so helpful since my first day in the program. I am most thankful for her willingness to lend me her ear when I needed it. I would be remiss if I did not thank the college's Ph.D. coordinator Renea Jay, Carolyn Stoddard and Susan Collum at the copy center, and custodian Scott Zeman, who was a friendly face during the few late nights I spent at the office.

Sincere thanks to Fr. Ed Fitzpatrick, Fr. Jeff Belger, Troy Sheehan, and Joe Mat-

tingly for making the Newman Catholic Student Center a warm and welcoming place. Also, thanks to the many friends that I have come to know through the Newman Center. I will miss all the laughs, fellowships, and Village Inn pie nights.

Finally, I do not think that I can ever thank my family enough for their support during the many years that I have been away from home. Being away has meant many missed birthday celebrations, holidays, and life in general. I very much appreciate all the love and support that they have given me.

ABSTRACT

The hub location problem seeks to find the best location for hubs and the assignment of non-hub nodes to hubs. The resulting structure is a hub-and-spoke network. In this dissertation, we study the hub location problem within the context of three different application areas.

We first consider the air travel industry, where many airline companies (especially the legacy airlines) operate hub-and-spoke networks. Important considerations in designing such a hub-and-spoke network are the cost of establishing the hubs, the cost of transporting passengers through the network, and the operating costs and maximum flying ranges of the different aircraft types in the fleet. We introduce the hub covering flow problem, which accounts for all these design considerations.

We next look at time-definite delivery services in the small package delivery industry, where a customer's package is guaranteed to be delivered within a certain time-window. With possible delays along the transportation route, a package may not be delivered on time. We introduce the stochastic p-hub center problem, which seeks to design a network where the longest origin-destination path in the network is minimized. The solution provides an upper bound on the delivery times, which can be used to design time-constrained service offerings, or to ascertain if current service guarantees are feasible. A service-level constraint is included into the model to ensure a high likelihood of on-time delivery.

In communication networks, two important service-level considerations are robustness (the ability of the network to perform when components become unavailable) and

response time (the time required to transfer information through the network). We propose another version of the stochastic p-hub center problem, which employs a two-stage stochastic programming formulation, that addresses the possibility of unavailable components or transmission delays in the network.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

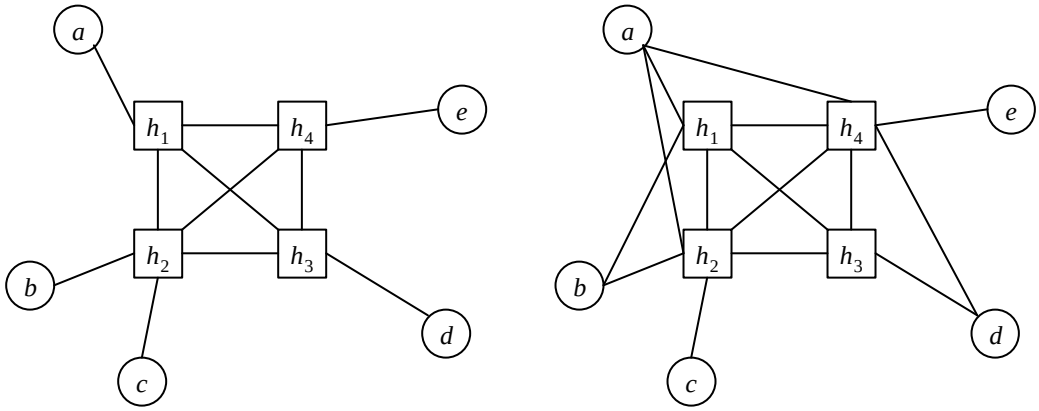
Figure

CHAPTER 1
INTRODUCTION

Hub-and-spoke systems are prevalent in many areas of everyday life from passenger travel through an airline's network of airports (Bryan and O'Kelly, 1999), to postal delivery (Ernst and Krishnamoorthy, 1996), communication (Klincewicz, 1998), and public transportation networks (Nickel et al., 2000). This has led to increased research on the hub location problem, which finds the best configuration of a hub-and-spoke network by locating hubs and assigning non-hub nodes to the hub nodes.

In the general form of the hub location problem on a network, a subset of the nodes is chosen to be hubs. The hub nodes have two functions: they collect demand from the non-hub nodes and transfer the demand to the other hubs in the network, or they receive demand from other hub nodes in the network and distribute this demand to the non-hub nodes that they serve. The hub nodes are assumed to be fully interconnected as shown in Figure 1.1, where the squares and circles represent hub and non-hub nodes, respectively. We will refer to this collection of hub nodes and the arcs connecting these nodes as the hub network. The non-hub nodes are connected to the hub nodes following a single-assignment or multiple-assignment rule. In a single-assignment network, each node is connected to *exactly* one hub node (Figure 1.1(a)), whereas in a multiple-assignment network, each node is connected to *at least* one hub node (Figure 1.1(b)). This network configuration creates at least one path between each origin-destination node pair and thus requires fewer links than a fully connected network.

Demand in a hub-and-spoke system is represented as a flow of demand units from an



(a) Single-assignment network. (b) Multiple-assignment network.

Figure 1.1: Hub-and-spoke network configurations. Squares and circles represent hub and non-hub nodes, respectively. The hub network is assumed to be fully connected.

origin node to a destination node. Demand flows from its origin node to the node's assigned hub (*collection phase*), through the hub network if necessary (*transfer phase*), and finally leaves the hub network arriving at the destination node (*distribution phase*). The unit cost factors (\$/distance) of the collection, transfer and distribution flow phases are χ, α and δ respectively. The transfer cost factor α is usually less than the collection and distribution cost factors ($\alpha < \chi$ and $\alpha < \delta$) due to economies of scale from the bundling of flows at the hubs and the transfer of these bundled flows through the hub network (O'Kelly and Bryan, 1998). These economies of scale are another benefit of the hub-and-spoke system.

1.1 Outline of Dissertation

In this dissertation, we study the hub location problem within the context of three different application areas. In Chapter 2, we consider the air travel industry where many

airline companies (especially the legacy airlines) operate hub-and-spoke networks. We introduce the hub covering flow problem, which fills a gap in the existing hub location literature. In brief, all aircraft have maximum flying ranges and the operating cost of one aircraft may be different from another. The uncapacitated hub location problem addresses the cost issue but not the maximum flying ranges. On the other hand, the hub covering location problem addresses the maximum flying ranges issue but it ignores operating costs. The hub covering flow problem addresses both issues.

The second application area is the small package delivery industry, which is dominated by well-known companies such as DHL, FedEx, UPS, and the United States Postal Service (USPS). A growing portion of this industry is the time-definite delivery service, where a company guarantees that a package will be delivered by a particular time of day (e.g., 8 a.m. next business day). In reality, the time to transport a package from its origin to its destination can vary because of delays along the transportation routes due to weather conditions (wind, inclement weather, and precipitation) and traffic congestion. Thus, it is important to know whether an offered time-definite delivery service is even feasible and if so, the reliability of the service. In Chapter 3, we model this problem as the chance-constrained stochastic p-hub center problem.

Finally, we study hub-and-spoke communication networks in Chapter 4. In communication networks, the transfer of information through a network may experience transmission delays along the links between points in the network, or processing delays at the hubs. Two important service level considerations for communication networks are robustness (the ability of the network to perform when certain components become unavailable)

and response time (the time required to transfer requested information through the network). We propose another version of the stochastic p-hub center problem, which employs a two-stage stochastic programming formulation.

In the remainder of this chapter, we provide a literature review of the hub location problem (Section 1.2) and of stochastic programming (Section 1.3).

1.2 Hub Location Problems

There are four core hub location problems: the uncapacitated hub location problem, the p-hub median problem, the hub covering location problem, and the p-hub center problem. We next review each of these four problems. The following is a list of the parameters used in the formulations:

N = set of nodes in the network,

d_{ij} = travel distance between nodes i and j (or t_{ij} = travel time),

W_{ij} = demand flow originating at node i that is destined for node j,

$O_i = \sum_{j \in N} W_{ij}$, the total demand flow originating from node i,

$D_i = \sum_{j \in N} W_{ji}$, the total demand flow destined for node i,

F_k = the annualized cost of establishing and operating a hub at node k,

α = unit cost factor for the transfer phase ($/distance unit),

χ = unit cost factor for the collection phase ($/distance unit), and

δ = unit cost factor for the distribution phase ($/distance unit).

1.2.1 The Uncapacitated Hub Location Problem

In the uncapacitated hub location problem (UHLP), the objective is to minimize the total cost of opening hubs and transporting demand flow through the network. Formulations for the UHLP can be found in O'Kelly (1992), Campbell (1994), and Klincewicz (1996), for example. Campbell et al. (2002) provide a multi-commodity flow formulation for the UHLP using the multi-commodity flow variable $Y_{kl}^i \geq 0, \forall i,k,l \in N$, to represent the amount of flow (or "commodity") originating at node i flowing through hub k first and then hub l. Also, let the variable $Z_{ik}, \forall i,k \in N$, to be 1 if node i is assigned to hub node k, and 0 otherwise. When $Z_{kk} = 1$, this indicates that node k is a hub. The UHLP formulation of Campbell et al. (2002) is

$$\text{UHLP}: \quad \min \quad \sum_{k\in N} F_k Z_{kk} + \alpha \sum_{i,k,l\in N} d_{kl} Y_{kl}^i + \sum_{i,k\in N} d_{ik}(\chi O_i + \delta D_i) Z_{ik} \tag{1.1a}$$

$$\text{s.t.} \quad Z_{ik} \leq Z_{kk}, \forall i,k \in N, \tag{1.1b}$$

$$\sum_{k\in N} Z_{ik} = 1, \forall i \in N, \tag{1.1c}$$

$$\sum_{j\in N} W_{ij} Z_{jk} + \sum_{l\in N} Y_{kl}^i = \sum_{l\in N} Y_{lk}^i + O_i Z_{ik}, \forall i,k \in N, \tag{1.1d}$$

$$Y_{kl}^i \geq 0, \forall i,k,l \in N, \tag{1.1e}$$

$$Z_{ik} \in \{0,1\}, \forall i,k \in N. \tag{1.1f}$$

The terms in objective function (1.1a) are the total annualized fixed cost of opening hubs, the cost of transporting demand through the hub network, and the cost of transporting demand from a non-hub node to a hub, respectively. Constraints (1.1b) require that a node be assigned to an open hub. Constraints (1.1c) and (1.1f) enforce the single-assignment rule. Constraints (1.1d) represent the flow conservation constraint at hub k for each node i.

If the multiple-assignment rule is needed, we replace the binary Z_{ik} variables with $Z_{ik} \in [0,1], \forall i,k \in N$. These continuous Z_{ik} variables are interpreted as the fraction of demand flowing from node i to k.

Several works have incorporated capacity constraints on the hubs (Aykin, 1994; Campbell, 1994; Bryan, 1998; Labbé et al., 2005; Rodíguez et al., 2007), capacity constraints on the links (Yaman, 2005; Yaman and Carello, 2005), or minimum flow thresholds on the links (Campbell, 1994). Marianov and Serra (2003) take a different approach in incorporating capacity-like constraints at the hub nodes. The authors model the hub airports as $M/D/c$ queues, and impose a constraint on the maximum number of airplanes that can be in a queue. The capacity constraints are modeled using chance constraints. Recent research on the hub location problem has focused on the polyhedral properties of the single-assignment problem. Work in this area includes those by Labbé and Yaman (2004), Labbé et al. (2005), Yaman (2005), and Yaman and Carello (2005).

Solution approaches for the hub location problem include dual-ascent methods (Klincewicz, 1996; Cánovas et al., 2007), a genetic algorithm based branch-and-bound method (Abdinnour-Helm and Venkataramanan, 1998), a branch-and-cut approach (Yaman and Carello, 2005; Labbé et al., 2005), tabu search (Yaman and Carello, 2005), simulated annealing (Ernst and Krishnamoorthy, 1999; Rodíguez et al., 2007), genetic algorithms (Topcuoglu et al., 2005; Cunha and Silva, 2007), and several hybrid heuristics (Abdinnour-Helm, 1998; Chen, 2007). In general, the exact methods have been shown to be able to solve small-sized problems (up to about 25 nodes) to optimality. The heuristic approaches have been applied to much larger problem instances (up to 200 nodes).

1.2.2 The p-Hub Median Problem

In the p-hub median problem (pHMP), the objective is to optimally locate p hubs in the network so that the total cost of transporting demand through the network is minimized. Unlike the UHLP, the number of hubs to be opened is given and is not a function of the the cost of opening the hubs. O'Kelly (1987) first formulated the pHMP as a quadratic programming problem. An equivalent mixed-integer linear programming formulation is obtained by replacing the quadratic variables with a four-index variable (Campbell, 1994). Ernst and Krishnamoorthy (1996) later formulate the pHMP using a similar multi-commodity formulation as with the UHLP presented in Section 1.2.1. The pHMP formulation of Ernst and Krishnamoorthy (1996) is

$$
\begin{aligned}
p\text{HMP}: \quad \min \quad & \alpha \sum_{i,k,l \in N} d_{kl} Y_{kl}^{i} + \sum_{i,k \in N} d_{ik}(\chi O_i + \delta D_i) Z_{ik} && \text{(1.2a)} \\
\text{s.t.} \quad & \sum_{k \in N} Z_{kk} = p, && \text{(1.2b)} \\
& Z_{ik} \leq Z_{kk}, \forall i, k \in N, && \text{(1.2c)} \\
& \sum_{k \in N} Z_{ik} = 1, \forall i \in N, && \text{(1.2d)} \\
& \sum_{j \in N} W_{ij} Z_{jk} + \sum_{l \in N} Y_{kl}^{i} = \sum_{l \in N} Y_{lk}^{i} + O_i Z_{ik}, \forall i, k \in N, && \text{(1.2e)} \\
& Y_{kl}^{i} \geq 0, \forall i, k, l \in N, && \text{(1.2f)} \\
& Z_{ik} \in \{0, 1\}, \forall i, k \in N. && \text{(1.2g)}
\end{aligned}
$$

The formulation of the pHMP is similar to that of the UHLP, with the only two differences being the omission of the hub opening cost in the objective function (1.2a), and the addition of constraint (1.2b), which requires that exactly p hubs be opened. As in the UHLP, if a multiple-assignment rule is required, the binary variables Z_{ik} should be replaced with their

continuous counterparts.

Solution approaches for the p-hub median problem include a LP-relaxation technique (Skorin-Kapov and Skorin-Kapov, 1994), branch-and-bound (Ernst and Krishnamoorthy, 1996), local neighborhood search heuristic (Klincewicz, 1991), GRASP (Klincewicz, 1992), tabu search (Klincewicz, 1992; Skorin-Kapov and Skorin-Kapov, 1994), simulated annealing (Ernst and Krishnamoorthy, 1996), genetic algorithms (Kratica et al., 2007), a path-relinking method (Pérez et al., 2004), and a neural network approach (Domínguez et al., 2003). Computational experiments for the exact methods were conducted on small data sets (up to 25 nodes), while the heuristics were tested on larger data sets (up to 200 nodes).

1.2.3 The Hub Covering Location Problem

In the hub covering location problem (HCLP), the objective is to find the best location of hubs in the network and assignment of nodes to hubs (subject to certain covering constraints) such that the total cost of opening the hubs is minimized. The HCLP includes cover constraints, which have the effect of restricting certain assignments of nodes to hubs. For example, in the facility location literature, a facility is said to be able to cover a node if the node is within a pre-specified distance from the facility. If the facility covers the node, then the node can be assigned to that facility.

In the HCLP, there are three definitions for coverage: an origin-destination pair (i, j) is covered if (1) the *total distance of the path* from node i to node j through hub k first then l does not exceed a specified value, (2) the *distance of each link in the path* from

node i to j through hub k first then l does not exceed a specified value, and (3) the *distance of each of the links between the hub and non-hub nodes in the path* (i.e., excluding the hub-to-hub link) from node i to j does not exceed a specified value (Campbell, 1994).

The first published formulation for the HCLP is a path-based formulation by Campbell (1994), where every possible path between each origin-destination node pair is enumerated. Campbell defines the path variable $Y_{ijkl}, \forall i, j, k, l \in N$, as the fraction of demand flow from node i to j through hub k first, then hub l (i.e., $i \rightarrow k \rightarrow l \rightarrow j$). In addition, let the variable $X_k, \forall k \in N$, be 1 if a hub is located at node k, and 0 otherwise. Also, set parameter $A_{ijkl}, \forall i, j, k, l \in N$, equal to 1 if hubs k and l cover the origin-destination pair (i, j), and 0 otherwise. This binary A_{ijkl} parameter can model any of the three coverage definitions described in the previous paragraph. The path-based formulation of Campbell (1994) is

$$\text{HCLP-Path}: \quad \min \quad \sum_{k \in N} F_k X_k \tag{1.3a}$$

$$\text{s.t.} \quad Y_{ijkl} \leq X_k, \forall i, j, k, l \in N, \tag{1.3b}$$

$$Y_{ijkl} \leq X_l, \forall i, j, k, l \in N, \tag{1.3c}$$

$$\sum_{k,l \in N} A_{ijkl} Y_{ijkl} \geq 1, \forall i, j \in N, \tag{1.3d}$$

$$X_k \in \{0, 1\}, \forall k \in N, \tag{1.3e}$$

$$Y_{ijkl} \in \{0, 1\}, \forall i, j, k, l \in N. \tag{1.3f}$$

Objective function (1.3a) minimizes the cost of opening hubs. Constraints (1.3b) and (1.3c) require that demand only flow through open hubs. Constraints (1.3d) enforce the coverage criterion.

Kara and Tansel (2003) propose several other mixed-integer linear programming formulations for the HCLP based on the first definition of coverage. Wagner (2007) builds upon the work of Kara and Tansel (2003) by proposing pre-processing methods to fix the values of some of the decision variables and to reduce the size of the problem by removing redundant constraints. The resulting model is shown to achieve better computational performance than those of Kara and Tansel (2003). Wagner (2007) also presents similar pre-processing approaches for the multiple-allocation version of the problem, and extends the single-assignment version to incorporate quantity-dependent discount factors, i.e., setting the discount factor α to be dependent on the amount of flow between pairs of hubs. Finally, Tan and Kara (2007) implement the covering version of the latest arrival hub problem of Kara and Tansel (2001).

1.2.4 The p-Hub Center Problem

In the p-hub center problem (pHCP), the objective is to simultaneously find the optimal location of p hubs and the assignment of non-hub nodes to the hubs so as to minimize the longest path in the network. O'Kelly and Miller (1991) first study the single hub problem (i.e., p = 1) in the plane, and propose several solution techniques. Campbell (1994) provides the first integer programming formulation for the pHCP using path variables. Kara and Tansel (2000) provide several alternative formulations that perform better computationally than that of Campbell (1994). The path-based formulation of Campbell (1994) is

$$p\text{HCP-Path}: \quad \min \quad \max_{i,j,k,l \in N} \{C_{ijkl} Y_{ijkl}\} \tag{1.4a}$$

$$\text{s.t.} \quad Y_{ijkl} \leq X_k, \forall i, j, k, l \in N, \tag{1.4b}$$

$$Y_{ijkl} \leq X_l, \forall i, j, k, l \in N, \tag{1.4c}$$

$$\sum_{k,l \in N} Y_{ijkl} = 1, \forall i, j \in N, \tag{1.4d}$$

$$\sum_{k \in N} X_k = p, \tag{1.4e}$$

$$X_k \in \{0, 1\}, \forall k \in N, \tag{1.4f}$$

$$Y_{ijkl} \in \{0, 1\}, \forall i, j, k, l \in N, \tag{1.4g}$$

where $C_{ijkl} = d_{ik} + \alpha d_{kl} + d_{jl}$ is the length of the path $i \rightarrow k \rightarrow l \rightarrow j$. Objective function (1.4a) minimizes the length of the longest path in the network. Other than constraints (1.4d), which together with constraints (1.4f) impose the single-assignment rule, all other constraints are similar to that of the HCLP in Section 1.2.3. If a multiple-assignment rule is required, we replace constraints (1.4d) with $\sum_{k,l \in N} Y_{ijkl} \geq 1, \forall i, j \in N$.

When introducing the pHCP at the beginning of this section, we stated that the objective is to minimize the longest path in the network. Referring to the three coverage definitions in Section 1.2.3, minimizing the longest path is based on the first coverage definition. If it is required that the characteristics of the paths follow one of the other two coverage definitions, the parameter C_{ijkl} can be appropriately defined to meet the requirements (e.g., setting $C_{ijkl} = +\infty$ if the path $i \rightarrow k \rightarrow l \rightarrow j$ does not meet the criteria of the desired path). See Campbell (1994) for a further discussion on this issue.

Kara and Tansel (2001) extend the pHCP to incorporate an operational-level issue where planes departing from a hub cannot leave until all planes arriving at the hub have arrived. Wagner (2004), however, writes that the min-max version of this latest arrival hub

location problem proposed by Kara and Tansel (2001) is no different than the basic p-hub center problem because the route that determines the longest path in the network is one where the transient or waiting times at the hubs are zero. Thus, the solution found by the basic p-hub center problem is the same as that of the min-max version of the latest arrival problem. A follow-up to the latest arrival hub network problem is that by Yaman et al. (2007), where stopovers are allowed between the non-hub and hub nodes (i.e., the route from a non-hub node to its hub may include a visit to another non-hub node).

Ernst et al. (2006b) propose a new mixed-integer linear programming formulation for the single- and multiple-allocation p-hub center problem based on a concept they refer to as the radius of hubs. The authors provide bounds on the objective function based on different heuristics, and propose a shortest-path branch-and-bound approach for the multiple-allocation version of the problem. This branch-and-bound approach is similar to the one proposed in Ernst and Krishnamoorthy (1998). The authors conducted computational tests on problem sets with up to 200 nodes.

A stream of work related to the pHCP is the p-hub center assignment problem, which is the pHCP with the hub locations given. Campbell et al. (2007) explore the single- and multiple-allocation version of the problem, and show that several special cases of these two problems can be solved in polynomial time. Ernst et al. (2006a) propose a new formulation for the single-allocation problem based on the idea of the radius of a hub. The authors show that the problem is NP-hard, provide bounds on the objective function value found using different heuristics, and propose a population-based meta-heuristic solution approach.

1.3 Stochastic Programming

The mathematical programming models formulated in hub location studies are deterministic optimization models (with the lone exception being Marianov and Serra (2003)). In the three application areas considered in this dissertation, there clearly exists some uncertainty in travel times or processing times, for example. Thus, it is natural to model these problems as stochastic programs.

In stochastic programming, the parameters of the mathematical programming problem are not fixed. These parameters are typically known to take on certain values with a known probability function. There are two common approaches to modeling a stochastic programming problem: the two-stage formulation and the chance constrained formulation. We next briefly discuss these two stochastic programming formulation approaches, and direct the reader to the texts by Kall and Wallace (1994) and Birge and Louveaux (1997), and to the Stochastic Programming Community Home Page (http://www.stoprog.org) for further discussion on stochastic programming.

1.3.1 Two-Stage Formulation

In the two-stage formulation, a decision is made in the first-stage. Following this decision, some realization of a random event is observed and a second decision then is made. This recourse decision made in the second-stage is typically based on finding the optimal decision to mitigate against an undesirable outcome, or to improve upon a current situation following the effect of the random events on the decisions made in the first-stage.

Birge and Louveaux (1997) give the following general formulation for the *two-stage stochastic programming problem*:

$$\begin{aligned} \text{SP-2S}: \quad \min \quad & c^T x + E_{\boldsymbol{\xi}} Q(x, \xi(\omega)) \\ \text{s.t.} \quad & Ax = b, \\ & x \in X, \end{aligned}$$

where

$$\begin{aligned} Q(x, \xi(\omega)) = \quad \min \quad & q(\omega)^T y(\omega) \\ \text{s.t.} \quad & T(\omega)x + W(\omega)y(\omega) = h(\omega), \\ & y(\omega) \in Y. \end{aligned}$$

The $n_1 \times 1$ vector x represents the first-stage variables. Vector c is $n_1 \times 1$, while matrix A and vector b are $m_1 \times n_1$ and $m_1 \times 1$, respectively. Following a decision on the first-stage problem, some random event $\omega \in \Omega$ (where Ω is the set of all random events) occurs. For a particular realization ω, the values of the second-stage parameters $q(\omega), T(\omega), W(\omega)$, and $h(\omega)$ are known, where $q(\omega)$ is an $n_2 \times 1$ vector, $T(\omega)$ is an $m_2 \times n_1$ matrix, $W(\omega)$ is an $m_2 \times n_2$ matrix, and $h(\omega)$ is a $m_2 \times 1$ vector. Once the $q(\omega), T(\omega), W(\omega)$, and $h(\omega)$ values corresponding to a realization ω are known, the second-stage decision $y(\omega)$ is then made. The second-stage vector $y(\omega)$ has dimension $n_2 \times 1$. Letting $T_{i\cdot}$ and $W_{i\cdot}$ represent the i-th row of $T(\omega)$ and $W(\omega)$, respectively, we define the vector $\xi^T(\omega) = (q(\omega)^T, T_{1\cdot}(\omega), \ldots, T_{m_2\cdot}(\omega), W_{1\cdot}(\omega), \ldots, W_{m_2\cdot}(\omega), h(\omega)^T)$. The random vector $\xi(\omega)$ is an $N \times 1$ vector, where $N = n_2 + (m_2 \times n_1) + (m_2 \times n_2) + m_2$. Also, let $\Xi \in \Re^N$ be the support of $\boldsymbol{\xi}$; that is, the smallest closed subset in $\Re$ such that $P(\boldsymbol{\xi} \in \Xi) = 1$.

Matrices W and T are called the *recourse* and *technology* matrices, respectively. If the recourse matrix is not random (i.e., $W(\omega) = W$), the problem is said to have *fixed recourse*. The problem is said to have *relatively complete recourse* if for every solution x satisfying the first-stage constraints $Ax = b$, and for any $\xi \in \Xi$, there is a feasible solution to the second-stage problem. If there is a feasible solution to the second-stage problem for all $x \in X$ and $\xi \in \Xi$, then the problem has *complete recourse*. Finally, if $W = [I, -I]$, we have *simple recourse*.

Problem SP-2S can also be written in the following *deterministic equivalent program* (DEP):

$$\begin{aligned} \text{SP-DEP}: \quad \min \quad & c^T x + \mathcal{Q}(x) \\ \text{s.t.} \quad & Ax = b, \\ & x \in X, \end{aligned}$$

where $\mathcal{Q}(x) = E_{\boldsymbol{\xi}} Q(x, \xi)$ is called the *value function* or *recourse function*. We note that this two-stage approach to modeling a stochastic programming problem can be extended to modeling multi-stage stochastic programming problems in the obvious way if necessary.

A common approach for solving two-stage stochastic programming problems is via a branch-and-cut technique similar to Benders decomposition (Benders, 1962), where a master problem solves for the first-stage variables and a subproblem then addresses the recourse function given the values of the first-stage variables. Van Slyke and Wets (1969) introduce the L-shaped method for the two-stage stochastic linear programming problem (continuous first- and second-stage variables). The L-shaped method generates cuts at each

iteration to approximate the recourse function. Laporte et al. (1994) extend the L-shaped method to solve stochastic integer programming problems with binary first- and second-stage variables. Examples of the application of the integer L-shaped method include the probabilistic traveling salesman problem (Laporte et al., 1994) and the stochastic vehicle routing problem (Gendreau et al., 1995; Laporte et al., 2002).

Other solution approaches include variations on branch-and-bound by Norkin et al. (1996, 1998), cutting plane algorithms (Carøe and Tind, 1997, 1998), and branch-and-cut (Ahmed and Shapiro, 2002; Ahmed et al., 2004). For further discussion on solution approaches for stochastic programming problems, see Klein Haneveld and van der Vlerk (1999).

1.3.2 Chance-Constrained Formulation

In deterministic linear programming, the optimal solution to the problem has to satisfy all the constraints; that is, none of the constraints can be violated. Due to the presence of uncertainty in stochastic programming problems, it may not be possible to find a feasible solution where none of the constraints are violated. Instead, one approach is to allow violation of some or all of constraints up to a certain probability threshold. We refer to these constraints as *probabilistic* or *chance constraints*.

The chance-constrained formulation was first proposed by Charnes and Cooper (1959, 1963). The stochastic programming problem with chance constraints can be written as

$$\text{SP-CC}: \quad \min \quad c^T x \tag{1.5a}$$

$$\text{s.t.} \quad P\{A(\omega)x \leq b(\omega)\} \geq \gamma, \tag{1.5b}$$

$$x \in X, \tag{1.5c}$$

where constraints (1.5b) are the probabilistic or chance constraints. When the vector $\gamma = e$ (i.e., the vector of ones), then the chance constraints are said to hold *almost surely* (i.e., with probability 1). In SP-CC, some constraints may be required to hold almost surely, while others may only be required to be satisfied at a minimum probability value.

CHAPTER 2
THE HUB COVERING FLOW PROBLEM

2.1 Introduction

The health of the North American airline industry has declined sharply in the past decade beginning with a downturn in the domestic economy. This economic downturn, together with advances in communications technology (for example, video-conferencing and the Internet), has led to fewer business travelers, who constitute a profitable revenue stream for airline companies. Other issues have exacerbated the problems of the airline industry: the terrorist events of September 11, 2001; the installation of airport security service fees following September 11, 2001; sky-rocketing fuel costs; high labor costs as a result of long-standing union contracts for the legacy airlines; increased competition from low-cost airline companies such as Southwest Airlines and JetBlue; and high gate prices at the larger airports.

It is not surprising then that by the end of the third quarter of 2005, four of the top seven US airline companies – United Airlines, Northwest Airlines, US Airways, and Delta Air Lines – had filed for Chapter 11 bankruptcy protection, while American Airlines, the largest airline company, had narrowly avoided filing for bankruptcy in 2003. Even Air Canada, the national airline of Canada, had spent 18 months under bankruptcy protection from April 2003 to September 2004 (CBC News Online, 2005).

One of the many ways that these legacy airlines have sought to address these problems under Chapter 11 bankruptcy protection is to re-organize the operations of their hub-

and-spoke networks to improve efficiencies. In a hub-and-spoke network, low capacity feeder aircraft transport passengers from the typically low demand, smaller-sized non-hub airports to the larger hub airports where passengers are transported toward their final destinations on larger capacity aircrafts. This hub-and-spoke network configuration allows for the connection of smaller cities into the national air transportation system and thus, more effective use of a company's high-cost assets such as airports, aircraft, and equipment (Ott, 1993; CNN.com, 2004). Since the smaller cities do not generate high demand, airline companies typically use smaller-sized aircraft to serve these cities. These aircraft are flown as frequently as needed to meet demand while maximizing aircraft utilization rates (Toh and Higgins, 1985; Butler and Huston, 1990).

Several airline companies have simplified their fleet by reducing the number of different aircraft types they operate. American Airlines, for example, has reduced the number of aircraft types by half so as to lower maintenance, training, and inventory costs (Fiorino, 2002). Airline companies also have retired older, less efficient aircrafts and purchased more fuel-efficient models in response to the recent record high fuel prices (CNNMoney.com, 2005). With lower demand for air travel, airline companies have reduced operating capacity by mothballing planes in the desert regions of Southern California, New Mexico, and Arizona (St. Louis Business Journal, 2002; The New York Times, 2003). To put all these changes in fleet sizes and composition into perspective, CNNMoney.com (2005) reports that the number of planes taken out of service and the number of people who have lost jobs due to the reduction in flights is equivalent in size to a company such as American Airlines going out of business.

Airline companies have not focused solely on changes to their fleet, but also on their hub infrastructure. For instance, in 2004, Spain's Iberia Airlines shifted hubbing services for their Central America market from their hub in Miami, Florida, to new hubs in Guatemala and Panama. This permits Iberia to provide direct flights from its main hub in Spain to their two priority Central American markets using the larger, more efficient Airbus 340 aircrafts (Iberia Airlines, 2004).

The decision as to where hub airports should be located is also dependent on the company's aircraft fleet. The short-haul feeder routes between the non-hub airports and hub airports are typically non-stop routes and since aircrafts have a specified maximum flying distance, this operational constraint must be considered when designing a hub-and-spoke network. The limited flight ranges are similar to the coverage constraints in the location theory literature (e.g., the set covering problem (Toregas et al., 1971)); that is, a facility is said to cover a demand point if the distance between the facility and the demand point is within a specified distance. With regards to the air travel industry, we say that a hub airport covers a non-hub airport if the maximum flying range of the aircraft assigned to this route is at least as long as the length of the route.

These decisions regarding mothballing, retiring and/or purchasing aircraft, as well as the establishment or elimination of hubs, are all very important strategic choices. Once made, a company must operate within this new environment for many years to come. Since these decisions may dictate whether a company will emerge from bankruptcy protection as an operationally efficient and financially sound business, these choices are not to be made haphazardly.

This network configuration problem is a form of the hub location problem. The two hub location problems that closely resemble this network configuration problem are the hub covering location problem (HCLP) and the uncapacitated hub location problem (UHLP). First, the maximum flying ranges of the aircraft is the third definition of coverage described in Section 1.2.3. We could then formulate this problem as a HCLP. Although the HCLP ensures that all nodes are covered while minimizing the annual cost of maintaining hubs, it does not account for the cost of transferring demand flows through the network. Air travel providers are in the business of transporting passengers cost-effectively (and hopefully, profitably too). Hence, transportation costs should not be ignored. The UHLP, on the other hand, minimizes the total cost of opening hubs and transporting demand flow through the network but it does not account for the coverage constraints on the assignment of nodes to hubs.

It is clear then that neither the HCLP nor the UHLP simultaneously address the two issues of (1) the coverage constraint (i.e., the maximum flying distance between hub and non-hub airports), and (2) the cost of transporting passengers through the network. The *hub covering flow problem* (HCFP), which we propose in this chapter, fills this gap in the literature. Two mixed-integer linear programming formulations for the HCFP are provided in Section 2.2. The HCFP essentially combines the HCP and UHLP, and thus, it possesses some of the characteristics and assumptions of these two models. One assumption, which will be discussed in greater detail in Section 2.2, is that all aircraft types have the same per unit operating cost and maximum flying range. This assumption is limiting, and so we extend one of the two HCFP formulations to account for aircraft-types with different

operating costs and maximum flying distances.

The HCFP can be used as a tool by airline companies to help decide on the location of hubs and the composition of an aircraft fleet to service geographically dispersed demand. In Section 2.3, we give an example of how the HCFP can be used for doing so. Section 2.4 provides another numerical example showing the impact of not solving the hub location and the network configuration problem simultaneously. Intuitively, one would expect the hubs to be situated at locations with high concentration of demand flows. Thus, the omission of demand flow in the design problem may lead to poor hub locations. We compare the results from solving the HCFP against a sequential optimization procedure where the hub locations are first determined by solving the HCLP, then solving the UHLP to incorporate the cost of transporting demand flow through these set of hub locations.

2.2 The Hub Covering Flow Problem

The hub covering flow problem (HCFP) finds the best configuration of a hub-and-spoke network that minimizes the total cost of opening hubs and transporting demand flow, while satisfying a maximum flying distance constraint between the hub and non-hub airports. We now provide two mixed-integer linear programming formulations for the HCFP based on the UHLP formulation in Section 1.2.1.

For our first HCFP formulation, we replace constraint (1.1b) in UHLP with the following coverage constraints

$$Z_{ik} \leq A_{ik} Z_{kk}, \forall i, k \in N,$$

where the parameter $A_{ik}, \forall i, k \in N,$ indicates if a hub at node k can cover node i; that is,

$A_{ik} = 1$ if the distance between hub k and node i is within the maximum flight range R of the aircraft, and 0 otherwise. To show that the above coverage constraints capture the coverage constraint and provide a feasible solution to HCFP, recall that the third definition of coverage states that the distance of each of the links between the hub and non-hub nodes in the path from node i to j should not exceed a specified value. By this coverage definition, there is thus a feasible path between all origin-destination pairs (i, j) if and only if the distance between each node and its assigned hub node is within the coverage distance. This implies that a feasible solution to the HCFP is obtained when all the assignments of nodes to hubs satisfy the coverage requirement. Thus, the constraints above ensure that every node is only assigned to a hub that can cover it. The first HCFP formulation is

$$\text{HCFP1}: \quad \min \quad \sum_{k\in N} F_k Z_{kk} + \alpha \sum_{i,k,l\in N} c_{kl} Y_{kl}^i + \sum_{i,k\in N} c_{ik}(\chi O_i + \delta D_i) Z_{ik} \tag{2.1a}$$

$$\text{s.t.} \quad Z_{ik} \leq A_{ik} Z_{kk}, \forall i, k \in N, \tag{2.1b}$$

$$\sum_{k\in N} Z_{ik} = 1, \forall i \in N, \tag{2.1c}$$

$$\sum_{j\in N} W_{ij} Z_{jk} + \sum_{l\in N} Y_{kl}^i = \sum_{l\in N} Y_{lk}^i + O_i Z_{ik}, \forall i, k \in N, \tag{2.1d}$$

$$Y_{kl}^i \geq 0, \forall i, k, l \in N, \tag{2.1e}$$

$$Z_{ik} \in \{0, 1\}, \forall i, k \in N. \tag{2.1f}$$

Formulation HCFP1 has $|N|^3$ continuous variables, $|N|^2$ binary variables, and $2|N|^2 + |N|$ linear constraints.

Unlike objective function (1.1a) in UHLP, we have opted for the use of c_{kl} and c_{ik} parameters in objective function (2.1a) instead of the d_{kl} and d_{ik} distance parameters. This is to emphasize that the cost of transferring demand flow through the network is not

necessarily a function of distance alone. A standard cost metric in the airline industry is the cost per available seat-mile (CASM), which is calculated as total operating costs divided by available seat miles. In other words, CASM considers aircraft capacity, fixed and variable operating costs, and travel distance. Since we are working with a network of nodes where the distances between nodes are known, we thus define the *cost per seat* c_{ik} as the product of CASM and d_{ik}, i.e., $c_{ik} = \text{CASM} \times d_{ik}$.

For our second HCFP formulation, we use a penalty-function approach to implement the coverage requirement. Similar to the Big-M approach in linear programming, we penalize the objective function value by a large value M each time a non-hub node is assigned to a hub that cannot cover it. We redefine the cost per seat (CPS) parameter c_{ik} as

$$\hat{c}_{ik} = \begin{cases} c_{ik}, & \text{if } 0 \leq d_{ik} \leq R \\ M, & \text{if } d_{ik} > R \end{cases} . \tag{2.2}$$

If node i is assigned to hub k and the distance between the two is within the coverage radius R, then the CPS $\hat{c}_{ik} = c_{ik}$ as before. If the distance between the two nodes exceeds R, then the CPS $\hat{c}_{ik} = M$. By setting a large enough value for M, all infeasible node to hub assignments will be discouraged. Our second formulation for the HCFP is

$$\begin{aligned} \text{HCFP2}: \quad \min \quad & \sum_{k \in N} F_k Z_{kk} + \alpha \sum_{i,k,l \in N} c_{kl} Y_{kl}^i + \sum_{i,k \in N} \hat{c}_{ik}(\chi O_i + \delta D_i) Z_{ik} \\ \text{s.t.} \quad & (1.1\text{b}), (2.1\text{c}) - (2.1\text{f}) \end{aligned} \tag{2.3}$$

Similar with HCFP1, HCFP2 has $|N|^3$ continuous variables, $|N|^2$ binary variables, and $2|N|^2 + |N|$ linear constraints.

The CPS coefficients in the second and third terms in the objective function of both HCFP1 and HCFP2 imply that regardless of which aircraft type is used to provide

service, the operating costs are identical (due to the constant CASM value). Moreover, the constant coverage radius R, which is used to define the coverage indicator parameter A_{ik} in constraints (2.1b) and $\hat{c}_{ik}$ in equation (2.2) implies that the maximum flying ranges are the same for all aircraft types. These implications are unrealistic as, for example, the CPS and maximum flying distance of an aircraft used for regional service are different from those of an aircraft used for long-haul service (Swan and Adler, 2006). Also, it is quite common for airline companies to use different aircraft types to operate the feeder routes from a single hub: an aircraft of a particular size for some routes, while another type of aircraft for another set of routes, and so on. Since both HCFP1 and HCFP2 assume that all aircraft have the same CPS values and maximum flying distances, we shall refer to these two formulations as the *one aircraft-type* case.

We can account for multiple aircraft types in formulation HCFP2 via a redefinition of the penalty-cost function (2.2). Suppose there are two types of aircraft at a hub. The first aircraft has a maximum flying range of R_1 with a CPS of $c^1_{ik}(d_{ik})$, which is defined as a function of the distance between nodes i and k. The ik subscripts in c^1_{ik} allows the use of different cost functions depending on which hub serves node i. The second aircraft has a longer maximum flying range ($R_2 > R_1$) but a larger CPS (i.e., $c^2_{ik} > c^1_{ik}$). (See Swan and Adler (2006) where they show that certain aircraft types have higher CPS values than others over the same trip distance.) Any airport that is farther than R_2 distance units from the hub cannot be served by that hub. We re-define $\hat{c}_{ik}$ to allow for this *two aircraft-type*

case as follows:

$$\hat{c}_{ik} = \begin{cases} c^1_{ik}(d_{ik}), & \text{if } 0 \leq d_{ik} \leq R_1 \\ c^2_{ik}(d_{ik}), & \text{if } R_1 < d_{ik} \leq R_2 \\ M, & \text{if } d_{ik} > R_2 \end{cases} . \tag{2.4}$$

Equation (2.4) can be extended in the obvious way if there are more than two aircraft-types operating out of a hub.

This modified $\hat{c}_{ik}$ parameter (2.4) accounts for the different CPS and coverage radii of the two aircraft. Furthermore, the optimal allocation of airports to hubs from HCFP2 indicates which type of aircraft to use to serve each airport; hence endogenously solving the fleet assignment problem (see, for example, Subramanian et al. (1994) for a description of the fleet assignment problem). Moreover, the objective function allows the model to weigh the trade-off between opening hubs and using different combinations of aircraft types each with a different coverage radius.

It is important to note that HCFP2 remains a mixed-integer linear program regardless of the nature of the cost functions c^1_{ik} and c^2_{ik} in (2.4). These cost functions – be they discrete or continuous, linear or non-linear – simply modify the $\hat{c}_{ik}$ coefficients, which then are used as input parameters when solving HCFP2. This added flexibility in parameterizing the optimization model is useful in situations where, for example, a concave cost function is required to further emphasize decreasing marginal transportation costs.

Finally, if a maximum distance constraint between hub nodes is required, this is easily done by redefining the c_{kl} parameter in the second term of objective function (2.3) with a $\hat{c}_{kl}$ parameter similar to (2.2). Imposing coverage constraints between hub nodes, in addition to those between hub and non-hub nodes, results in the second definition of

coverage that the *distance of each link in the path* from node i to j through hub k first then l does not exceed a specified value (see Section 1.2.3). This coverage definition is applicable in situations where transported items "require some preserving or rejuvenating process, such as heating or cooling, which is available only at the hub locations," or to limit continuous driving by truck drivers, for example (Campbell, 1994). As with the $\hat{c}_{ik}$ parameter in (2.2), $\hat{c}_{kl}$ neither alters the structure of HCFP2, nor does it introduce additional complexities to the model.

2.3 Numerical Examples

We now provide some examples to illustrate how the HCFP can be used in decision making. All the models were solved using the XPress-MP solver (Dash Optimization, 2004) via the NEOS Server for Optimization (Czyzyk et al., 1998; Dolan, 2001; Gropp and Moré, 1997).

For our experiments, we use the 50-node Australia Post (AP) data set, which is available from the OR-Library (Beasley, 1990, 2004). The AP data set contains information on node coordinates; demand flow; collection, transfer and distribution cost factors; as well as the fixed costs of opening hubs. (We note that another commonly used data set for hub location model testing is the Civil Aeronautics Board (CAB) data set, but it only has demand flow between nodes.) In the AP data set, distances between pairs of nodes are the Euclidean distances calculated from the given node coordinates, and demand flow between nodes is not symmetrical ($W_{ij} \neq W_{ji}$). The collection cost factor χ, transfer cost factor α, and distribution cost factor δ are 3, 0.75 and 2, respectively. These values are commonly

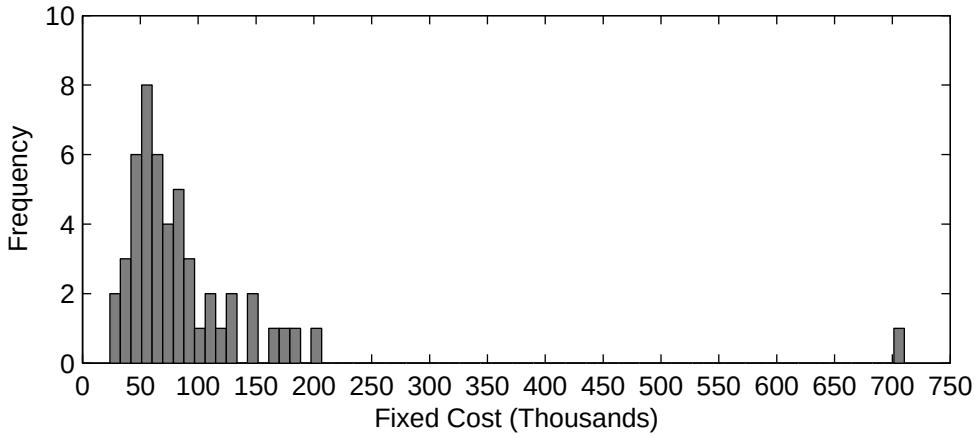


Figure 2.1: Distribution of the fixed cost values in the "tight" data set. The average fixed cost value is 94343.

used for the AP data set in computational experiments reported in the hub location literature (e.g., Ernst and Krishnamoorthy, 1996).

There are two sets of hub opening (fixed) costs: "tight" and "loose". Figure 2.1 shows the tight fixed cost values for all 50 nodes. In the tight fixed cost set, it is generally more expensive to open a hub at a node with large total demand flow, which is defined as the sum of demand originating and terminating at a node (i.e., $DF_i = O_i + D_i$). (A regression analysis of the fixed cost of opening a hub with total demand flow shows a coefficient of correlation of 0.998.) Even though nodes with large demand flow are intuitively good locations for hubs, the high fixed cost values in the tight data set has a tendency to deter the model from doing so. This correlation between total demand and fixed costs is not present in the loose fixed cost set (Figure 2.2). The tight data set has fixed cost values ranging from about 20,000 to 200,000 with an outlier at about 710,000, while the loose data set has values ranging from 20,000 to 35,000.

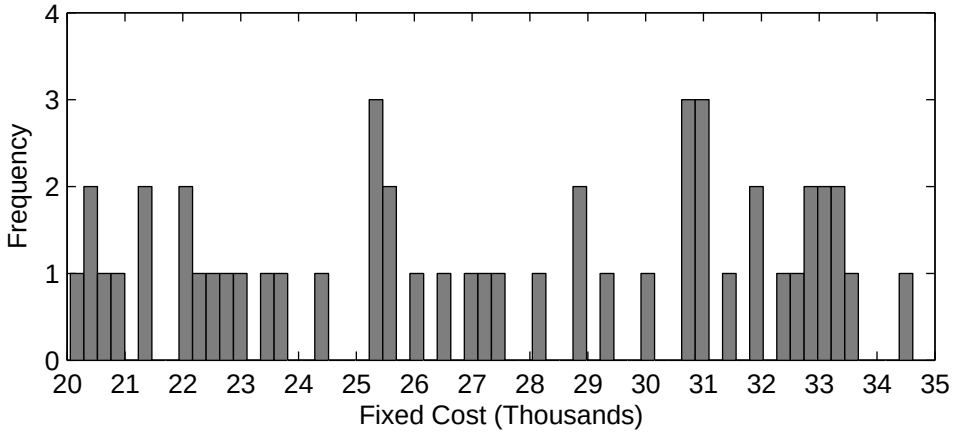


Figure 2.2: Distribution of the fixed cost values in the "loose" data set. The average fixed cost value is 27404.

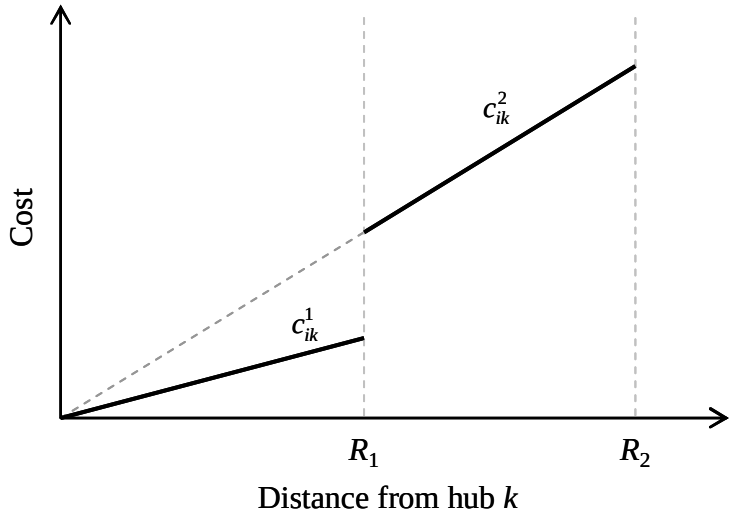


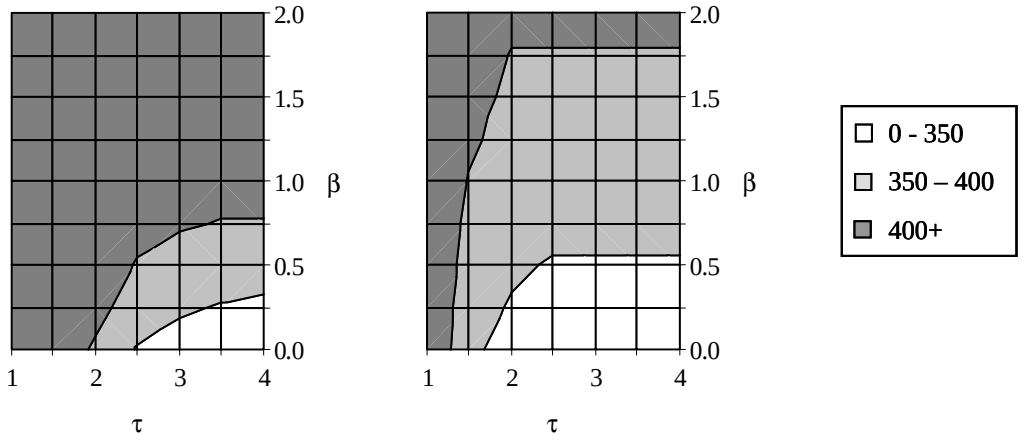
Figure 2.3: The cost per seat (CPS) function $\hat{c}_{ik}$ for equation (2.5).

In our experiments, we consider the case with two aircraft-types (see equation (2.4)). Suppose the cost per seat (CPS) of the first aircraft is $c^1_{ik} = d_{ik}$, with this aircraft having a maximum flight distance of R_1. The CPS of the second aircraft is $c^2_{ik} = d_{ik}(1+\beta)$, where $\beta > 0$ is the cost premium of using the longer ranging aircraft. The maximum flight distance of the second aircraft is R_2 and is set to a multiple of R_1 (i.e., $R_2 = \tau R_1$). The cost per seat (CPS) function is

$$\hat{c}_{ik} = \begin{cases} c^1_{ik} = d^1_{ik}, & \text{if } 0 \le d_{ik} \le R_1 \\ c^2_{ik} = d^1_{ik}(1+\beta), & \text{if } R_1 < d_{ik} \le R_2 \\ M, & \text{if } d_{ik} > R_2 \end{cases} , \tag{2.5}$$

which is shown in Figure 2.3. We solve problems with $R_1 = 10$ and 15, and experiment with values of $\tau = 1, 1.5, 2, \ldots, 4$. Each (R_1, τ, β) combination can be interpreted as operating parameters for a particular two aircraft-type fleet.

For comparison purposes, we consider options for developing a hub-and-spoke infrastructure that incurs a yearly cost of approximately 400,000 units. One such design involves only one aircraft-type and where every hub serves demand points that are within 20 distance units (i.e., $R = 20$). We use the tight fixed cost data set and find from HCFP that four hubs are opened with a resulting cost close to 400,000. To investigate alternative designs, the light grey regions in Figure 2.4 indicate different combinations of (R_1, τ, β) that have total cost values between 350,000 and 400,000. Table 2.1 shows the number of open hubs for the corresponding (R_1, τ, β) combinations. Since the (R_1, τ, β) combinations represent fleets of different types of aircraft, a decision maker could use this information to determine the best mix of aircraft types at the hubs depending on what aircraft types are



(a) $R_1 = 10$. (b) $R_1 = 15$.

Figure 2.4: Total cost values (in thousands) for different (R_1, τ, β) combinations and the tight fixed cost data set, where $R_2 = \tau R_1$.

readily available to the firm, while maintaining a total cost close to 400,000.

If a decision maker prefers having three instead of four hubs, she can use Figure 2.4 and Table 2.1 to find the many fleet combinations and network designs with three hubs at approximately the same cost. For example, the optimal number of hubs for the instance $(R_1, \tau, \beta) = (10, 2.5, 0.5)$ is four hubs at a cost about 400,000. For an optimal network configuration with only three hubs at a similar cost, one option is to use a fleet corresponding to $(R_1, \tau, \beta) = (15, 2, 1)$. Thus, the first aircraft-type has a maximum flight range of 15 distance units, and the second flying up to 30 distance units at twice the per unit distance operating cost of the first. Another option is $(R_1, \tau, \beta) = (10, 2.5, 0.25)$. For a two hub configuration, $(R_1, \tau, \beta) = (10, 4, 0.5)$ might be considered.

In summary, Figure 2.4 and Table 2.1 can be used to evaluate the interaction be-

Table 2.1: Number of open hubs for different (R_1, τ, β) combinations and the tight fixed cost data set, where $R_2 = \tau R_1$.

		β								
R_1	τ	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
10	1.0	11	11	11	11	11	11	11	11	11
	1.5	7	7	7	7	7	7	8	8	8
	2.0	4	5	5	5	6	6	6	6	6
	2.5	3	3	4	4	5	5	5	5	5
	3.0	3	3	3	3	4	4	5	5	5
	3.5	2	2	3	3	4	4	4	5	5
	4.0	1	2	2	3	4	4	4	5	5
15	1.0	7	7	7	7	7	7	7	7	7
	1.5	4	4	4	4	4	4	5	5	5
	2.0	3	3	3	3	3	4	4	4	4
	2.5	2	2	2	2	2	4	4	4	4
	3.0	1	2	2	2	2	4	4	4	4
	3.5	1	2	2	2	2	4	4	4	4
	4.0	1	2	2	2	2	4	4	4	4

tween total cost, number of open hubs, and fleet configuration. The decision maker can fix one (or two) of the variables and obtain optimal values for the remaining variable(s).

Next, consider the area of the zone between the maximum flight ranges of the two aircraft-types, i.e., $\pi((\tau R_1)^2 - (R_1)^2)$. It is not surprising that the number of open hubs is non-increasing as the size of this zone increases (increasing τ while holding β constant in Table 2.1), since increasing τ creates a larger coverage radius for a hub. Also from Table 2.1, the number of open hubs is non-decreasing as the cost premium β increases (with τ held constant). This is to be expected because as the longer ranging aircraft becomes more costly to use, the shorter ranging aircraft becomes more appealing, which then results in more open hubs. When the cost premium β decreases (while holding τ constant), the total number of the longer ranging aircrafts employed at a hub may increase because these aircrafts cost less to operate. This lower transportation cost may result in fewer open hubs (see Table 2.1) while the number of airports served by the longer ranging aircraft is (monotonically) increasing (see Table 2.2).

Although we had anticipated the number of airports served by the longer ranging aircraft to increase as the size of the zone between the maximum flight ranges increases, the results in Table 2.2 do not fully support our expectations. For example, consider $R_1 = 10$ and $\beta = 1$. As τ increases from 2 to 3, the corresponding number of airports served are 12, 10, and 14. The lack of empirical support for our hypothesis is partially attributed to conducting the numerical experiments on a network instead of a continuous plane. Nevertheless, the results in Table 2.2 do exhibit an increasing trend, although not the anticipated monotonic increase.

Table 2.2: Number of airports served by the longer ranging aircraft for different (R_1, τ, β) combinations and the tight fixed cost data set, where $R_2 = \tau R_1$.

		β								
R_1	τ	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
10	1.5	12	12	12	12	12	12	9	9	9
	2.0	29	17	17	17	12	12	12	12	12
	2.5	26	26	19	19	10	10	10	10	10
	3.0	22	22	22	22	14	14	10	10	10
	3.5	27	26	22	20	14	14	14	10	10
	4.0	33	26	26	20	14	14	14	10	10
15	1.5	10	9	8	8	7	7	4	4	4
	2.0	13	10	10	10	10	5	5	5	5
	2.5	16	16	12	12	12	5	5	5	5
	3.0	18	16	12	12	12	5	5	5	5
	3.5	18	16	12	12	12	5	5	5	5
	4.0	18	16	12	12	12	5	5	5	5

2.4 Impact of Demand Flow and Coverage Radius on the Hub Network

At the beginning of this chapter, we mentioned that demand flow will likely dictate the location of hub nodes in the network. In this section, we compare the optimal network configurations via the hub covering location problem (HCLP) and the hub covering flow problem (HCFP), and show that demand flow does influence the location and number of hubs. The results using the 50-node AP data set with the "loose" fixed costs indicate a preference for hub locations at nodes that are closer spatially to the center of the service region, and at nodes that have high total demand flows.

We use the one aircraft-type HCFP formulation to demonstrate the impact of demand flow, transportation costs and coverage radius on the design of the network. To provide for a meaningful comparison between HCFP and HCLP, we first solve the HCLP for the optimal hub locations and then, fixing the optimal Z_{kk} variables from HCLP, use HCFP to find the optimal flows using this set of open hubs. This latter problem is basically an assignment problem (see, for example, Campbell et al., 2007) since the hub nodes are fixed. We refer to this two-stage optimization problem as HCLP+FLOW. For each problem solved, the A_{ik} parameters were set using the same R value.

For the HCLP, we use our arc-based formulation:

$$\begin{aligned} \text{HCLP-Arc}: \quad \min \quad & \sum_{k \in N} F_k Z_{kk} \\ \text{s.t.} \quad & Z_{ik} \leq A_{ik} Z_{kk}, \forall i, k \in N, \\ & \sum_{k \in N} Z_{ik} = 1, \forall i \in N, \end{aligned}$$

Table 2.3: Number of open hubs for HCLP+FLOW and HCFP using the loose fixed cost data set.

Radius, R	HCLP+FLOW	HCFP
5	22	22
10	10	10
15	5	6
20	4	5

$$Z_{ik} \in \{0, 1\}, \forall i, k \in N.$$

Notice that this formulation is similar to that of the set covering problem (Toregas et al., 1971) and is more parsimonious than the path-based formulation HCLP-Path: formulation HCLP-Arc has $|N|^2$ variables and $|N|^2 + |N|$ constraints, while HCP-Path has $|N|^4 + |N|^2$ variables and $2|N|^4 + |N|^2$ constraints.

Table 2.3 shows the optimal number of open hubs for both HCLP+FLOW and HCFP for different coverage radii and the loose fixed cost data set. In three instances, HCFP opens one more hub than HCLP+FLOW because the total benefits of lower per unit transportation cost (due to α) within the hub network and the reduction in the number of costly links between hub and non-hub nodes more than offset the fixed cost of the additional hub. From Table 2.4, the cost savings from this trade-off may be as high as 15% when an additional hub is used. With the tight fixed cost set, the cost savings between the HCFP and HCLP+FLOW were nil except when $R = 15$, where the results showed a cost

Table 2.4: Total costs of HCFP and HCLP+FLOW for different coverage radii using the loose fixed cost data set.

Radius, R	HCFP	HCLP+FLOW	% Difference
5	0.658	0.659	0.06%
10	0.379	0.379	0.17%
15	0.305	0.338	9.63%
20	0.282	0.334	15.41%

reduction of 0.13%. The lack of cost savings for the tight fixed cost set is attributable to the wide ranging and higher fixed cost values (see Figure 2.1), which makes the interchanging of hub locations rather prohibitive.

Table 2.5 lists the indices of the hubs opened by HCLP+FLOW but not by HCFP, and vice versa. Figure 2.5 shows the locations of these hubs in relation to the 50 nodes (black dots) in the network. The numbers next to the circles correspond to the hub indices in Table 2.5. The shaded circles in Figure 2.5(a) are those hubs opened by HCLP+FLOW but not by HCFP, while the unshaded circles in Figure 2.5(b) are those opened by HCFP but not HCLP+FLOW. Also, the size of these circles are roughly proportional to the total demand flow $DF_i = O_i + D_i$ for the nodes. Figure 2.5 shows that with the loose fixed cost set, HCFP favors locating hubs toward the center of the service region and at nodes with high total demand values to reduce the costly transportation costs between the hubs and non-hub nodes. This preference for interior hub locations emphasizes an important practical consideration when locating hubs in that the ideal hub should have 360 degrees of

Table 2.5: List of hubs opened in HCLP+FLOW but not in HCFP, and vice versa, using the loose fixed cost data set.

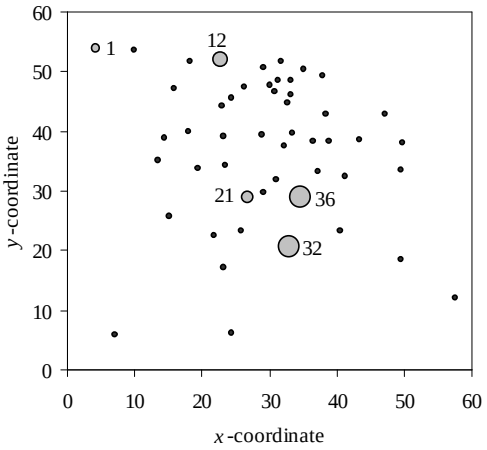
	Coverage radius, R			
	5	10	15	20
Hubs in HCLP+FLOW not in HCFP	12, 32	36	-	1, 21
Hubs in HCFP not in HCLP+FLOW	13,33	35	35	4, 22, 35

flow traffic (Ott, 1993).

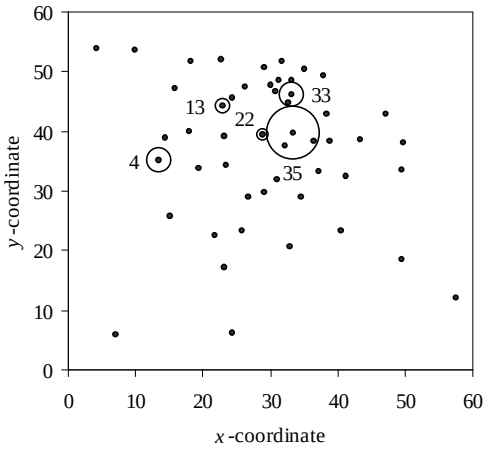
Of course, HCP+FLOW is at an obvious disadvantage in terms of total costs because it solves the location problem separately from the allocation/flow problem. HCFP, on the other hand, solves both problems concurrently and thus, it can evaluate the trade-off between the cost of opening hubs and transportation costs. These results are meant to emphasize the importance of including demand flows when designing a hub-and-spoke network especially when demand flow and the cost of transferring it across the network are important considerations in the design.

2.5 Conclusion

The hub covering flow problem (HCFP) finds the optimal design for a hub-and-spoke network accounting for the costs of opening hubs, demand flow, and coverage constraints. With demand represented by flows from an origin to a destination, coverage is defined on the path of these flows: an origin-destination pair (i, j) is considered covered if the distance of each of the links between the hub and non-hub nodes in the path does not exceed a specified value.



(a) Hubs opened in HCLP+FLOW but not in HCFP.



(b) Hubs opened in HCFP but not in HCLP+FLOW

Figure 2.5: Hubs opened by HCLP+FLOW but not HCFP, and vice versa, for the loose fixed cost set.

The hub covering location problem (Campbell, 1994) solves for the optimal location of hubs and allocation of nodes to hubs but it omits the cost of demand flow in the optimization problem. For air travel service providers who are in the business of cost-effectively transporting travelers, transportation cost is an important consideration and should not be ignored. This omission may result in much higher operating costs and poor choices for hub locations.

We provide two formulations for the HCFP. The first (HCFP1) includes the coverage requirements in the one of the constraints – an approach that borrows from the traditional set covering problem of location theory. We also discuss a second formulation (HCFP2) that models the coverage requirements in the objective function instead of in the constraints. This is done via a modification of one of the transportation cost parameters.

An assumption in these formulations of the coverage requirements is that only one aircraft-type operates out of a hub. In reality, there are many different aircraft types with maximum flying distances and operating costs. We show how formulation HCFP2 can be altered to include this multiple aircraft-type issue, and further modifications to HCFP2 allows us to include a second definition of coverage into the model.

HCFP2 has many benefits: it more closely reflects reality without requiring a more complex model (it is of the same size as the HCFP1), it has an additional benefit of providing a solution to the fleet assignment problem via the optimal assignments of non-hub nodes to hubs, and it aids in decision-making by quantifying the costs and listing the characteristics of optimal network designs for different fleet combinations.

Real-life hub-and-spoke networks may contain thousands of nodes. For example, The World Factbook states that there are over 5000 paved airports in the U.S. alone (Central Intelligence Agency, 2006). We believe that solving the HCFP for such a large data set using commercial solvers may be impractical, and alternative solution approaches for the HCFP is needed.

CHAPTER 3
STOCHASTIC P-HUB CENTER PROBLEM: CHANCE-CONSTRAINED FORMULATION

3.1 Introduction

The small package delivery service market is a lucrative industry, generating billions of dollars in revenues annually. For instance, UPS, the largest package delivery company in the world reported total revenues of $47.5 billion in 2006 (United Parcel Services, 2006). Other primary players in the package delivery industry in the United States include FedEx, DHL, and the United States Postal Service (USPS), with UPS and FedEx representing approximately 80% of this market (Hannon, 2005). A significant portion of the $47.5 billion revenue generated by UPS is through their Next Day Air service, which guarantees the delivery of packages within a specified time period. This expedited delivery service, which generated $6.8 billion in revenue in 2006 (a 6.2% increase from 2005) from delivery of 1.27 million packages a day, accounts for 14% of UPS' total revenues or 22% of its domestic revenues.

With such large volumes of packages transported between many different origin-destination points during a single year, it is paramount that the delivery networks operate efficiently so as to be able to meet the service guarantees. The configuration of the network has an important role in this regard. Package delivery companies operate complex hub-and-spoke networks, which have many sorting hub centers and many more source and destination local service centers. With a particular service network, a company may wish to understand the ability of the network to support the service guarantees provided to its

customers. The company may also be interested in what kind of service guarantee it could provide its customers given design constraints on its service network.

In the hub location literature, the p-hub center problem addresses this latter question. The p-hub center problem (pHCP) looks to simultaneously find the optimal location of p hubs and the assignment of non-hub nodes to the p hubs so as to minimize the longest origin-destination path in the network. The solution to the pHCP in essence provides an upper bound on the delivery times in the hub-and-spoke network. This information can then be used to design time-constrained service offerings to customers, or to ascertain if certain current service guarantees are feasible.

The pHCP has received limited research attention in the hub location literature. The first mathematical programming formulation for the pHCP is due to Campbell (1994), and Kara and Tansel (2000) provide several alternative formulations that have better computational performance. Kara and Tansel (2001) extend the pHCP to incorporate an operational-level issue where planes departing from a hub cannot leave until all planes arriving at the hub have arrived. Pamuk and Sepil (2001) describe a single-reallocation heuristic with tabu search to solve the weighted pHCP. Ernst et al. (2006b) propose a new mixed-integer linear programming formulation for the single- and multiple-allocation p-hub center problem based on a concept they refer to as the radius of hubs, and also discuss several solution approaches. A related stream of work is that by Campbell et al. (2007) and Ernst et al. (2006a) who study the p-hub center assignment problem, which is the pHCP with the hub locations given.

An assumption in all the works mentioned above is that travel time is deterministic.

This is not the case in many real-life applications as weather or traffic issues could hasten or prolong travel. With variability in the time required to transport a package to its destination, there is the possibility that a package may not be delivered on time. A failure in on-time delivery of a package may result in a company needing to appease an unhappy customer. For example, FedEx and UPS refunds the delivery service charges in the event of a service failure (FedEx, 2007; United Parcel Services, 2007). While this refund cost is known, it is more difficult to quantify the cost of lost goodwill, and thus a company may insist on a minimum service level instead.

In this chapter, we introduce the stochastic p-hub center problem (SpHCP), which employs a chance-constrained formulation to model the minimum service level requirement. This model accounts for the uncertain travel times in designing the hub network that minimizes the maximum travel time through the network. We will show that the presence of stochasticity in the problem affects the optimal configuration of the hub-and-spoke network.

3.2 Model Formulation

In the stochastic p-hub center problem, the problem is to locate p hubs in the network and assign each non-hub node to a hub node (i.e., the single-assignment rule) so that the longest path duration in the network is minimized for a given service level γ. It is reasonable to assume that that γ will be close to 1 (e.g., $\gamma = 0.95$). Let $t_{ij}(\omega)$ represent the travel time between nodes i and j for the random event ω, and $t_{ij}(\omega) = t_{ji}(\omega)$. With $Z_{ik} = 1$ indicating that node i is assigned to hub k (0 otherwise), and p being the number

of hubs to be opened in the network, the chance-constrained formulation for the SpHCP is

$$\text{S}p\text{HCP-CC}: \quad \min \quad \beta \tag{3.1a}$$

$$\text{s.t.} \quad P\left\{\sum_{k\in N}(t_{ik}(\omega)+\alpha t_{kl}(\omega))Z_{ik}+t_{jl}(\omega)Z_{jl}\leq\beta\right\}\geq\gamma, \quad \forall i,j,l\in N, \tag{3.1b}$$

$$\sum_{k\in N}Z_{kk}=p, \tag{3.1c}$$

$$Z_{ik}\leq Z_{kk}, \forall i,k\in N, \tag{3.1d}$$

$$\sum_{k\in N}Z_{ik}=1, \forall i\in N, \tag{3.1e}$$

$$Z_{ik}\in\{0,1\}, \forall i,k\in N. \tag{3.1f}$$

3.2.1 SpHCP-CC: Certainty Equivalent Problem

Constraints (3.1b) can be expressed as a set of linear constraints if we assume that we can enumerate all possible scenarios (or outcomes), i.e., $\Omega=\{\omega_1,\ldots,\omega_{|S|}\}$, where S is the set of all scenarios. Each scenario $\omega_s, s\in S$, has a corresponding positive probability p_s. For notational convenience, let

$$T_{ijls}(Z)=\sum_{k\in N}(t^s_{ik}+\alpha t^s_{kl})Z_{ik}+t^s_{jl}Z_{jl}, \tag{3.2}$$

which is the total travel time from node i to node j through a hub k first then hub l under scenario s. The chance constraints (3.1b) can be written as

$$\sum_{s\in S}p_s 1_{[0,\beta]}(T_{ijls}(Z))\geq\gamma, \forall i,j,l\in N, \tag{3.3}$$

with the indicator function

$$1_{[0,\beta]}(T_{ijls}(Z))=\begin{cases}1, & \text{if } 0\leq T_{ijls}(Z)\leq\beta\\ 0, & \text{otherwise}\end{cases}.$$

Subtracting 1 from both sides of (3.3) gives

$$\sum_{s\in S} p_s 1_{(\beta,\infty)}(T_{ijls}(Z)) \leq 1-\gamma,$$

where

$$1_{(\beta,\infty)}(T_{ijls}(Z)) = \begin{cases} 1, & \text{if } T_{ijls}(Z) > \beta \\ 0, & \text{otherwise} \end{cases}.$$

The above indicator function can be expressed as the following set of linear-integer constraints:

$$T_{ijls}(Z) - \beta \leq Y_{ijls} M, \tag{3.4}$$

$$Y_{ijls} \in \{0,1\}, \tag{3.5}$$

where M is a large number. Constraint (3.4) sets $Y_{ijls} = 1$ if the coverage constraint is not satisfied. We can re-formulate the chance-constrained stochastic p-hub center problem as the following *certainty equivalent problem,* which is an integer programming problem:

$$\begin{aligned}
\text{S}p\text{HCP-CEP}: \quad & \min \quad \beta \\
& \text{s.t.} \quad \sum_{k\in N} (t^s_{ik} + \alpha t^s_{kl}) Z_{ik} + t^s_{jl} Z_{jl} - \beta \leq Y_{ijls} M, \\
& \qquad\qquad \forall i,j,l \in N, s \in S, && \text{(3.6a)} \\
& \sum_{s\in S} p_s Y_{ijls} \leq 1-\gamma, \forall i,j,l \in N, && \text{(3.6b)} \\
& \sum_{k\in N} Z_{kk} = p, && \text{(3.6c)} \\
& Z_{ik} \leq Z_{kk}, \forall i,k \in N, && \text{(3.6d)} \\
& \sum_{k\in N} Z_{ik} = 1, \forall i \in N, && \text{(3.6e)} \\
& Y_{ijls} \in \{0,1\}, \forall i,j,l \in N, s \in S, && \text{(3.6f)} \\
& Z_{ik} \in \{0,1\}, \forall i,k \in N. && \text{(3.6g)}
\end{aligned}$$

3.2.2 SpHCP-CC with Normally Distributed Travel Times

In problem SpHCP-CEP, complete enumeration of all possible scenarios leads to a combinatorial explosion in the size of the integer programming problem. In this section, we present a formulation for the SpHCP where the travel time on a link is normally distributed. We will assume normally distributed travel times for the remainder of this chapter.

Consider a unit of demand originating at node i destined for node j traveling through hub k first then hub l, i.e., $i \to k \to l \to j$. Suppose that the travel time on link ij, represented by the random variable T_{ij}, is normally distributed with mean t_{ij} and variance σ_{ij}^2. We make the assumption that the travel time on a link is independent of the travel time on all other links in the network. The total travel time of the path $i \to k \to l \to j$ is the random variable $T_{ijkl} = T_{ik} + \alpha T_{kl} + T_{jl}$, with $T_{ijkl} \sim N\left(t_{ik} + \alpha t_{kl} + t_{jl}, \sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{jl}^2\right)$.

For path $i \to k \to l \to j$, its chance constraint is $P(T_{ijkl} \leq \beta) \geq \gamma$, or equivalently,

$$\beta \geq t_{ik} + \alpha t_{kl} + t_{jl} + z_\gamma \sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{jl}^2}, \tag{3.7}$$

where z_γ is the z-value corresponding to the 100γ-th percentile from the standard normal distribution. (Recall that, with the service level parameter γ assumed to be at least 0.5, z_γ is non-negative.) With normally distributed travel times T_{ij}, constraints (3.1b) can be written as:

$$\begin{aligned} \beta &\geq \sum_{k \in N} \left(z_\gamma \sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{jl}^2} + (t_{ik} + \alpha t_{kl}) \right) Z_{ik} + t_{jl} Z_{jl} - (1 - Z_{jl})\sigma_{jl} \\ &= \sum_{k \in N} \left(z_\gamma \sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{jl}^2} + (t_{ik} + \alpha t_{kl}) \right) Z_{ik} \\ &\qquad + (t_{jl} + \sigma_{jl}) Z_{jl} - \sigma_{jl}, \forall i, j, l \in N. \end{aligned} \tag{3.8}$$

To show the correctness of the above chance constraint, we need to show that for any given

node-pair (i, j), the resulting chance constraint is non-binding when Z_{ik} or Z_{jl} is zero. Consider the following three scenarios:

- Suppose $Z_{ik} = Z_{jl} = 0$. Then, constraints (3.8) become $\beta \geq -\sigma_{jl}$. Since β is non-negative, this constraint is non-binding.

- Suppose $Z_{ik} = 1$ and $Z_{jl} = 0$. From (3.8), we have

$$\beta \geq z_\gamma \sqrt{\sigma_{ik}^2 + \alpha^2 \sigma_{kl}^2 + \sigma_{jl}^2} + (t_{ik} + \alpha t_{kl}) - \sigma_{jl}. \tag{3.9}$$

  This constraint, however, is not necessarily binding, or more accurately, there exists another path in the network that dominates this path. Note that there exists at least one node j' such that $Z_{j'l} = 1$. This follows since each possible location for a hub is the location of a demand point, and a demand point at a hub location will be assigned to that hub. Then, with $Z_{ik} = 1$, we have from (3.8):

$$\begin{aligned} \beta &\geq \left(z_\gamma \sqrt{\sigma_{ik}^2 + \alpha^2 \sigma_{kl}^2 + \sigma_{j'l}^2} + (t_{ik} + \alpha t_{kl}) \right) + (t_{j'l} + \sigma_{j'l}) - \sigma_{j'l} \\ &= z_\gamma \sqrt{\sigma_{ik}^2 + \alpha^2 \sigma_{kl}^2 + \sigma_{j'l}^2} + (t_{ik} + \alpha t_{kl} + t_{j'l}) \end{aligned} \tag{3.10}$$

  We will show that constraint (3.10) is more binding than constraint (3.9); that is, the right-hand side of (3.9) is less than or equal to the RHS of (3.10), or

$$\sqrt{\sigma_{ik}^2 + \alpha^2 \sigma_{kl}^2 + \sigma_{jl}^2} - \sigma_{jl} \leq \sqrt{\sigma_{ik}^2 + \alpha^2 \sigma_{kl}^2 + \sigma_{j'l}^2} + t_{j'l}.$$

  Beginning with the left-hand side of the above inequality, first note that

$$\sqrt{\sigma_{ik}^2 + \alpha^2 \sigma_{kl}^2 + \sigma_{jl}^2} - \sigma_{jl} \leq \sqrt{\sigma_{ik}^2 + \alpha^2 \sigma_{kl}^2}.$$

The above inequality holds because if we add σ_{jl} to both sides, then taking the square of both sides, we have

$$\begin{aligned}\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{jl}^2 &\leq (\sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2} + \sigma_{jl})^2 \\ &= (\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2) + \sigma_{jl}^2 + 2\sigma_{jl}\sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2},\end{aligned}$$

which is always true since $\sigma_{jl} \geq 0$. Finally, we have that

$$\begin{aligned}\sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2} &\leq \sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{j'l}^2} \\ &\leq \sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{j'l}^2} + t_{j'l}\end{aligned}$$

because $t_{j'l} \geq 0$. This then gives $\sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{jl}^2} - \sigma_{jl} \leq \sqrt{\sigma_{ik}^2 + \alpha^2\sigma_{kl}^2 + \sigma_{j'l}^2} + t_{j'l}$ as desired.

- Suppose $Z_{ik} = 0$ and $Z_{jl} = 1$. From (3.8), we have $\beta \geq t_{jl}$. This constraint, however, is not necessarily binding. For node i, there exists some hub k' such that $Z_{ik'} = 1$. With $Z_{jl} = 1$, we have from (3.8):

$$\begin{aligned}\beta &\geq \left(z_\gamma\sqrt{\sigma_{ik'}^2 + \alpha^2\sigma_{k'l}^2 + \sigma_{jl}^2} + (t_{ik'} + \alpha t_{k'l})\right) + (t_{jl} + \sigma_{jl}) - \sigma_{jl} \\ &= z_\gamma\sqrt{\sigma_{ik'}^2 + \alpha^2\sigma_{k'l}^2 + \sigma_{jl}^2} + (t_{ik'} + \alpha t_{k'l} + t_{jl}),\end{aligned}$$

which dominates constraint $\beta \geq t_{jl}$.

Thus, with the assumption that travel times on the links are independent normal random variables, we have the following MILP formulation for the stochastic p-hub center problem

with chance constraints:

$$
\begin{aligned}
\min \quad & \beta \\
\text{s.t.} \quad & \beta \geq \sum_{k \in N} \left(z_\gamma \sqrt{\sigma_{ik}^2 + \alpha^2 \sigma_{kl}^2 + \sigma_{jl}^2} + (t_{ik} + \alpha t_{kl}) \right) Z_{ik} \\
& \qquad + (t_{jl} + \sigma_{jl}) Z_{jl} - \sigma_{jl}, \forall i, j, l \in N, \\
& \sum_{k \in N} Z_{kk} = p, \\
& Z_{ik} \leq Z_{kk}, \forall i, k \in N, \\
& \sum_{k \in N} Z_{ik} = 1, \forall i \in N, \\
& Z_{ik} \in \{0, 1\}, \forall i, k \in N.
\end{aligned}
$$

The above formulation is similar in structure and size to that proposed by Kara and Tansel (2000) for the pHCP (i.e., the deterministic version).

3.3 The SpHCP-CC on the Line

In this section, we present results for the special case of the chance-constrained SpHCP with normally distributed travel times on the line. The motivation for considering this special case is to extract insight from the model, and to characterize the optimal solutions of the model. The characteristics of the optimal solution may provide suggestions for solution approaches for the chance-constrained SpHCP.

Let the length of the line be N units with nodes located at the points $0, 1, \ldots, N$, with a link connecting every pair of nodes. The mean travel time t_{ij} on the link connecting nodes i and j is the straight line distance between the two nodes with variance σ_{ij}^2. We assume that $\sigma_{ij} = \nu t_{ij}$, where ν is a common coefficient of variation. We further assume that the travel time on any link between a pair of adjacent nodes is independent of that on

any other link in the network.

We define the β-*path* as the path whose duration equals or exceeds that of all other paths in the network. We use the notation $\mathcal{P}(ABC)$ to represent the path from node A to node B then to node C, and $\mathcal{LP}_\gamma(ABC)$ to represent the 100γ-th percentile duration of $\mathcal{P}(ABC)$, i.e., the "length" of path $\mathcal{P}(ABC)$. For notational convenience, we will drop the subscript γ in $\mathcal{LP}_\gamma(ABC)$ as γ is usually a given fixed value.

3.3.1 $p = 1$

Consider the 1-hub problem on the line. Let nodes A and B be located at points 0 and N, respectively, as shown in Figure 3.1. For any given location of the single hub $K \in \{0, 1, \ldots, N\}$, the only three candidates for the β-path are:

- $\mathcal{P}(AKA)$ with $\mathcal{LP}(AKA) = 2K + z_\gamma\sqrt{2\nu^2K^2} = 2K + z_\gamma\nu\sqrt{2K^2}$,
- $\mathcal{P}(AKB)$ with $\mathcal{LP}(AKB) = K + (N-K) + z_\gamma\sqrt{\nu^2K^2 + \nu^2(N-K)^2} = N + z_\gamma\nu\sqrt{K^2 + (N-K)^2}$, and
- $\mathcal{P}(BKB)$ with $\mathcal{LP}(BKB) = 2(N-K) + z_\gamma\sqrt{2\nu^2(N-K)^2} = 2(N-K) + z_\gamma\nu\sqrt{2(N-K)^2}$.

All paths linking any other origin-destination node-pairs are dominated by at least one of these three paths.

Theorem 1 *If N is even, the optimal location for the single hub is at $K^\star = \frac{1}{2}N$. If N is odd, $K^\star = \lfloor\frac{1}{2}N\rfloor$ or $\lceil\frac{1}{2}N\rceil$.*

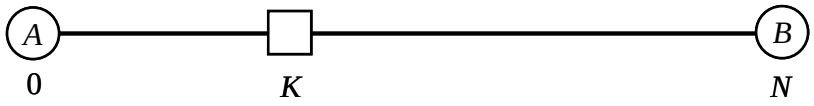


Figure 3.1: Illustration of the 1-hub SpHCP-CC on the line.

Proof: We will prove this theorem for the case where N is even. The proof for when N is odd follows in the same manner. Our approach for this proof is to show that for a given configuration where the hub is not located at the center of the line, the length of the β-path does not increase as we shift the hub towards the center of the line.

Without loss of generality, suppose the hub is located at the point K, with $K < \frac{1}{2}N$. Thus, we have $\mathcal{LP}(AK) < \mathcal{LP}(KN)$. Of the three β-path candidates listed above, $\mathcal{P}(BKB)$ is the β-path. Now, consider shifting the hub to the right towards the center of the line by one unit. Let $K' = K + 1$ represent the new hub location, and assume that $\mathcal{LP}(AK') \leq \mathcal{LP}(K'N)$. For this new hub location, the path length of the three β-path candidates are:

- $\mathcal{LP}(AK'A) = 2(K+1) + z_\gamma \nu \sqrt{2(K+1)^2}$,
- $\mathcal{LP}(AK'B) = (K+1) + (N-K-1) + z_\gamma \nu \sqrt{(K+1)^2 + (N-K-1)^2}$, and
- $\mathcal{LP}(BK'B) = 2(N-K-1) + z_\gamma \nu \sqrt{2(N-K-1)^2}$.

$\mathcal{P}(BK'B)$ dominates the other two candidate paths. However,

$$\begin{aligned} \mathcal{LP}(BK'B) &= 2(N-K-1) + z_\gamma \nu \sqrt{2(N-K-1)^2} \\ &< 2(N-K) + z_\gamma \nu \sqrt{2(N-K)^2} \\ &= \mathcal{LP}(BKB). \end{aligned}$$

Thus, the length of the β-path is non-increasing as we shift the hub towards the center of the line. One, therefore, can simply shift the hub to the center of the line (or to $\lfloor \frac{1}{2}N \rfloor$ or $\lceil \frac{1}{2}N \rceil$ in the case where N is odd) without increasing the length of the β-path. ■

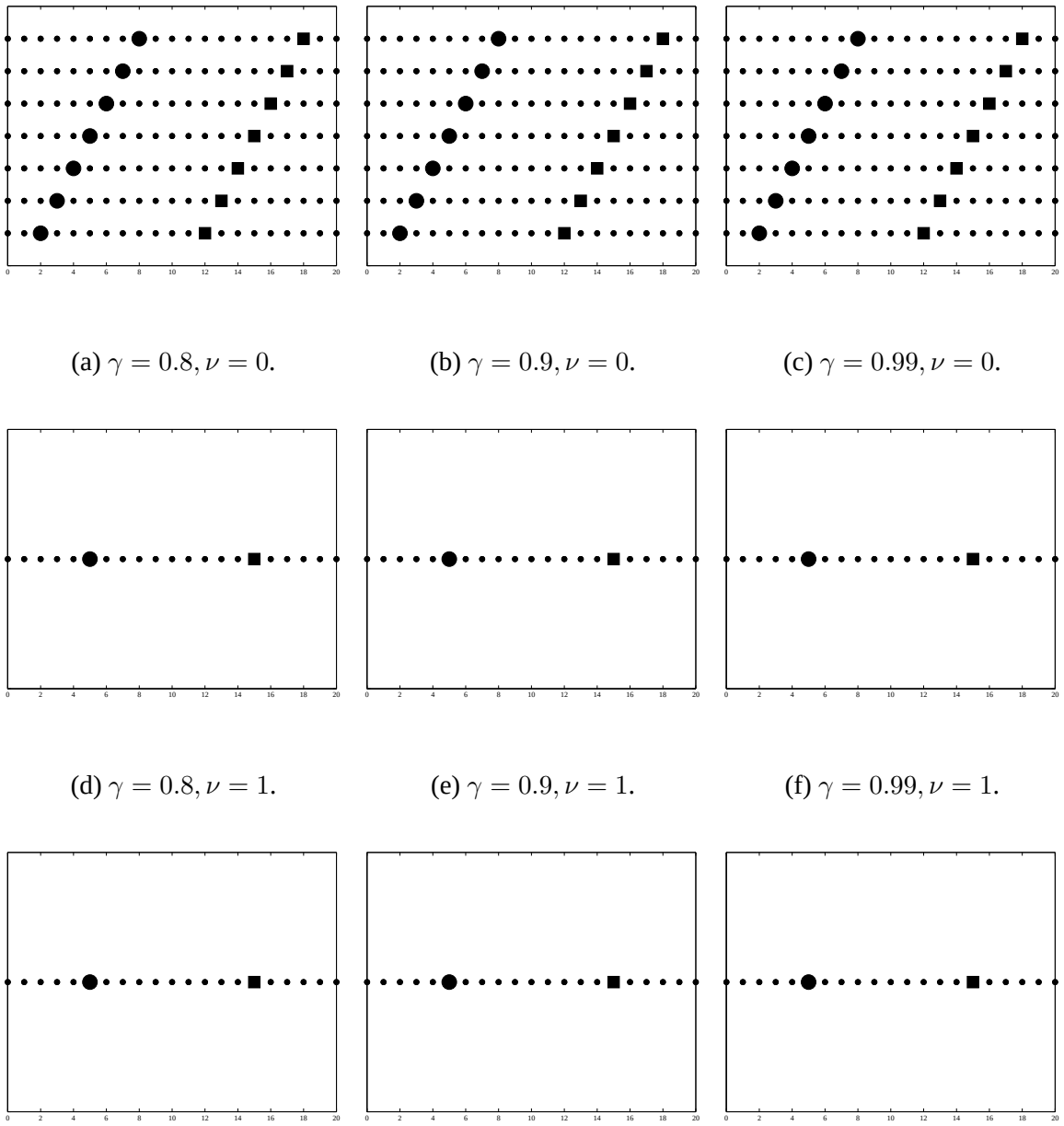
We note that the optimal location of the hub at the center of the line is influenced by neither the service level parameter γ (with the assumption that $\gamma \geq 0.5$), nor the value of ν.

3.3.2 $p = 2$

In this section, we first present observations on the results from our numerical experiments on the 2-hub problem, and then provide an explanation for these observations. In our experiments, we set $N = 20$ and $\alpha = 0.75$, and solved the following hub center single-allocation problem (HCSAP) over all $\binom{20}{2}$ 2-hub configurations H:

$$
\begin{aligned}
\text{HCSAP}: \quad \min \quad & \beta \\
\text{s.t.} \quad & \beta \geq \sum_{k \in H} \left(z_\gamma \nu \sqrt{t_{ik}^2 + \alpha^2 t_{kl}^2 + t_{jl}^2} + (t_{ik} + \alpha t_{kl}) \right) Z_{ik} \\
& \qquad + (t_{jl} + \nu t_{jl}) Z_{jl} - \nu t_{jl}, \forall i, j \in N, l \in H, \\
& \sum_{k \in H} Z_{ik} = 1, \forall i \in N, \\
& Z_{ik} \in \{0, 1\}, \forall i \in N, k \in H.
\end{aligned}
$$

Figure 3.2 shows the results for different values of the service level parameter γ and the coefficient of variation ν. The optimal hub locations are identified by the larger icons, and if there are multiple lines in a box (for example, the problem instances with $\nu = 0$), this indicates multiple optima. The results show that there exist optimal solutions that have



(a) $\gamma = 0.8, \nu = 0$. (b) $\gamma = 0.9, \nu = 0$. (c) $\gamma = 0.99, \nu = 0$.

(d) $\gamma = 0.8, \nu = 1$. (e) $\gamma = 0.9, \nu = 1$. (f) $\gamma = 0.99, \nu = 1$.

(g) $\gamma = 0.8, \nu = 2$. (h) $\gamma = 0.9, \nu = 2$. (i) $\gamma = 0.99, \nu = 2$.

Figure 3.2: Optimal hub locations for the 20-node, 2-hub SpHCP-CC on the line.

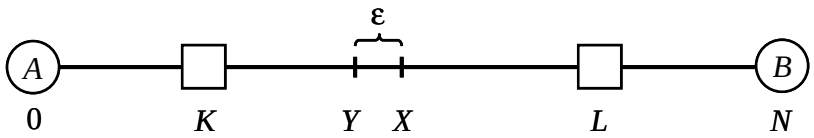


Figure 3.3: Illustration of the 2-hub SpHCP-CC on the line.

a symmetric configuration; that is, the two hubs are located such that the distance along the line from the left-most node (i.e., the node at point 0) to its closest hub is equal to the distance from the right-most node (i.e., the node at point 20) to its closest hub. (This is true for one of the multiple optimal solutions for the problem instances with $\nu = 0$.) Also, notice that in the solution for the problem instances with $\nu > 0$, the optimal location of the two hubs are the same regardless of the value of γ and/or ν. We next show why the optimal location of the two hubs do not change as γ or ν varies.

We will assume that the symmetric configuration is an optimal solution for the $p = 2$ problem, and present our analysis for the problem on the continuous line. Let A and B represent the left-most and right-most nodes on the line, respectively, and K and L be the two hubs as shown in Figure 3.3. With our assumption of a symmetric network, we have $t_{AK} = t_{BL}$. Suppose we define the split point X to be the point on the line between hubs K and L where all points to the left of X are assigned to hub K, and all points to the right of X are assigned to hub L.

Theorem 2 *The optimal position of X is equidistant to both hubs K and L, i.e., $t_{XK} = t_{XL}$.*

Proof: Let X be the midpoint between hubs K and L, and Y be the point $\varepsilon > 0$ units to the left of X as shown in Figure 3.3. As we are considering a continuous problem, we are

Table 3.1: List of paths with splits at points X and Y.

Paths with split at Y		Paths with split at X
$\mathcal{LP}(AKA)$	$=$	$\mathcal{LP}(AKA)$
$\mathcal{LP}(AKY)$	$<$	$\mathcal{LP}(AKX)$
$\mathcal{LP}(AKLY)$	$>$	$\mathcal{LP}(AKLX)$
$\mathcal{LP}(AKLB)$	$=$	$\mathcal{LP}(AKLB)$
$\mathcal{LP}(YKY)$	$<$	$\mathcal{LP}(XKX)$
$\mathcal{LP}(YKLY)$	$>$	$\mathcal{LP}(XKLX)$
$\mathcal{LP}(YKLB)$	$<$	$\mathcal{LP}(XKLB)$
$\mathcal{LP}(YLY)$	$>$	$\mathcal{LP}(XLX)$
$\mathcal{LP}(YLB)$	$>$	$\mathcal{LP}(XLB)$
$\mathcal{LP}(BLB)$	$=$	$\mathcal{LP}(BLB)$

indifferent to which of the two hubs K and L the split points X and Y are assigned. For each of the two split points, we consider both cases where the point is assigned to hubs K and L. For the midpoint X to be optimal, it suffices to show that the length of the β-path with the split at Y is greater than or equal to the length of the β-path with the split at X.

Consider the lengths of the β-path candidates with splits at points Y and X shown in Table 3.1. There are three instances where the length of the path with the split at X is greater than that at Y. Notice, however, that

- $\mathcal{LP}(AKY) < \mathcal{LP}(AKX) < \mathcal{LP}(AKLY)$

- $\mathcal{LP}(YKY) < \mathcal{LP}(XKX) < \mathcal{LP}(YLY)$
- $\mathcal{LP}(YKLB) < \mathcal{LP}(XKLB) < \mathcal{LP}(AKLY)$.

Thus, the β-path defined at Y is greater than or equal to the β-path defined at X. With Y to the right of the point X, a similar analysis follows. ■

With $t_{AK} = t_{BL}$ and $t_{XK} = t_{XL}$, we now determine which of the 10 candidate paths we only need to consider for the β-path. Through a process of elimination due to equivalence or dominance, we have

- $\mathcal{LP}(AKA) = \mathcal{LP}(BLB)$
- $\mathcal{LP}(AKLB) > \mathcal{LP}(AKA)$
- $\mathcal{LP}(XKX) = \mathcal{LP}(XLX)$
- $\mathcal{LP}(XKLX) > \mathcal{LP}(XKX)$
- $\mathcal{LP}(XKLB) > \mathcal{LP}(AKX)$
- $\mathcal{LP}(XKLB) > \mathcal{LP}(XLB)$
- $\mathcal{LP}(XKLB) = \mathcal{LP}(AKLX)$
- $\mathcal{LP}(AKLX) \leq \mathcal{LP}(AKLB)$ when $t_{AK} \geq t_{XK}$, or $\mathcal{LP}(AKLX) < \mathcal{LP}(XKLX)$ when $t_{AK} < t_{XK}$.

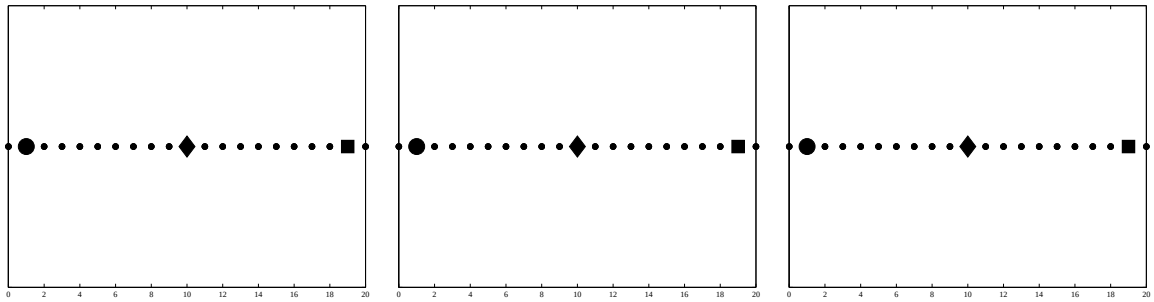
In the end, we are left with two candidates for the β-path: $\mathcal{P}(AKLB)$ and $\mathcal{P}(XKLX)$. The 2-hub SpHCP with normally distributed travel times reduces to

$$\begin{aligned} \min \quad & \beta \text{ over all symmetric configurations} \\ \text{s.t.} \quad & \beta \geq 2AK + \alpha KL + z_\gamma \nu \sqrt{2AK^2 + \alpha^2 KL^2} \\ & \beta \geq 2XK + \alpha KL + z_\gamma \nu \sqrt{2XK^2 + \alpha^2 KL^2} \\ & \text{by locating hubs } K \text{ and } L. \end{aligned}$$

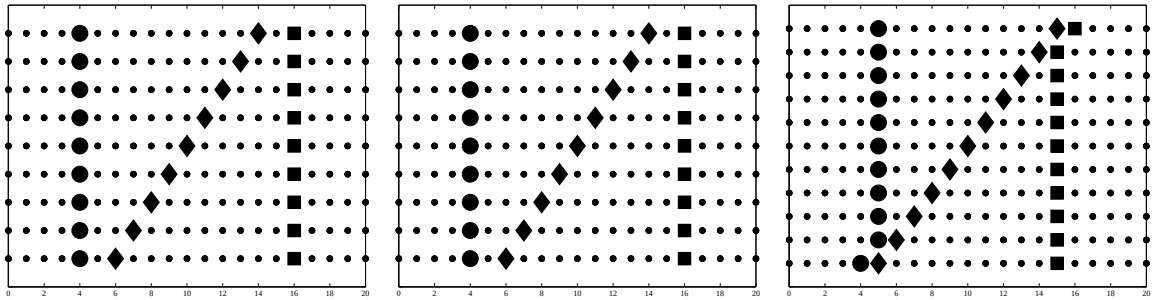
The optimal solution to the above problem is to set $t_{AK} = t_{XK}$; that is, locate hubs K and L at points $\frac{1}{4}N$ and $\frac{3}{4}N$, respectively. It is clear from the mathematical programming formulation above that the optimal location of hubs K and L is not influenced by γ or ν. Thus, the optimal locations of the hubs K and L do not change as γ or ν varies.

3.3.3 $p = 3$

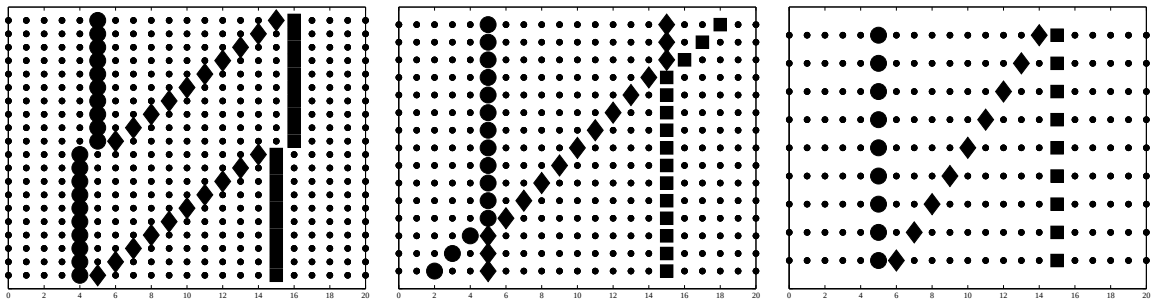
Figure 3.4 shows the numerical results for the 3-hub case with $N = 20$ and $\alpha = 0.75$. In each of the nine problem instances, an optimal solution to the problem instance exhibits a symmetric configuration: a hub located in the middle between the two "boundary" (or outer) hubs, and the distance between the left-most hub and node 0 is equal to the distance between the right-most hub and node N (except for a slight difference with the $\gamma = 0.8$ and $\nu = 2$ problem instance). We notice that the two boundary hubs tend to shift towards the middle of the line as the service level parameter γ and/or the coefficient of variation ν increases. We will show why this is likely to occur using the problem on the continuous line. We assume that the symmetric configuration is an optimal solution for this set of problem instances. Considering symmetric configurations only, in our computational



(a) $\gamma = 0.8, \nu = 0$. (b) $\gamma = 0.9, \nu = 0$. (c) $\gamma = 0.99, \nu = 0$.



(d) $\gamma = 0.8, \nu = 1$. (e) $\gamma = 0.9, \nu = 1$. (f) $\gamma = 0.99, \nu = 1$.



(g) $\gamma = 0.8, \nu = 2$. (h) $\gamma = 0.9, \nu = 2$. (i) $\gamma = 0.99, \nu = 2$.

Figure 3.4: Optimal hub locations for the 20-node, 3-hub SpHCP-CC on the line.

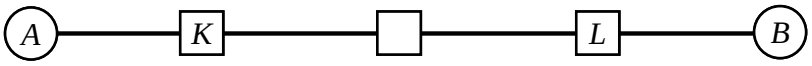


Figure 3.5: Illustration of the 3-hub SpHCP-CC on the line.

experiments with $\alpha = 0.75$, the β-path has always been the path that begins at node 0 through the left-most hub, then the right-most hub, and finally to node N. Using the notation shown in Figure 3.5, path $\mathcal{P}(AKLB)$ has been the β-path. Since an increase in γ has a similar effect as an increase in ν, we will assume that γ is constant and will explain why the boundary hubs tend to shift towards the center of the line as the coefficient of variation ν varies as long as $\mathcal{P}(AKLB)$ remains the β-path.

The length of path $\mathcal{P}(AKLB)$ is

$$\mathcal{LP}(AKLB) = t_{AK} + \alpha t_{KL} + t_{BL} + z_\gamma \nu \sqrt{t_{AK}^2 + \alpha^2 t_{KL}^2 + t_{BL}^2}.$$

Since we restrict this discussion to symmetric solutions only, it suffices to show that both boundary hubs shift inward by some value $\Delta > 0$ as ν increases. The length of $\mathcal{P}(AKLB)$ following the shift of the hubs is

$$\begin{aligned} \mathcal{LP}_\Delta(AKLB) &= (t_{AK} + \Delta) + \alpha(t_{KL} - 2\Delta) + (t_{BL} + \Delta) \\ &\quad + z_\gamma \nu \sqrt{(t_{AK} + \Delta)^2 + \alpha^2 (t_{KL} - 2\Delta)^2 + (t_{BL} + \Delta)^2}. \end{aligned}$$

We can express $\mathcal{LP}_\Delta(AKLB)$ as $f(\Delta) + \nu g(\Delta)$, where

$$\begin{aligned} f(\Delta) &= (t_{AK} + \Delta) + \alpha(t_{KL} - 2\Delta) + (t_{BL} + \Delta), \quad \text{and} \\ g(\Delta) &= z_\gamma \nu \sqrt{(t_{AK} + \Delta)^2 + \alpha^2 (t_{KL} - 2\Delta)^2 + (t_{BL} + \Delta)^2}. \end{aligned}$$

Lemma 1 *The first-order derivative of $f(\Delta)$ is a positive constant value.*

Proof: The first-order derivative of $f(\Delta)$ is $f'(\Delta) = 2(1-\alpha)$, which is a positive constant value since $0 < \alpha < 1$. ■

Lemma 2 *$g(\Delta)$ is a strictly convex function.*

Proof: To show that $g(\Delta)$ is a convex function, it suffices to show that the second-order derivative of the square root term in $g(\Delta)$ is non-negative since z_γ is positive for the service level parameter γ (assuming $\gamma \geq 0.5$). Let $h(\Delta)$ represent the square root term in $g(\Delta)$.

$$\begin{aligned}
h(\Delta) &= \sqrt{(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2} \\
h'(\Delta) &= \frac{1}{2}\left[(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2\right]^{-\frac{1}{2}} \\
&\qquad \times \left(2t_{AK} + 2\Delta - 4\alpha^2 t_{KL} + 8\alpha^2\Delta + 2t_{BL} + 2\Delta\right) \\
&= \frac{\Delta(4\alpha^2+2) + (t_{AK} + t_{BL} - 2\alpha^2 t_{KL})}{\sqrt{(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2}}. \\
h''(\Delta) &= \Big[(4\alpha^2+2)\sqrt{(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2} \\
&\qquad - \left[\Delta(4\alpha^2+2) + (t_{AK} + t_{BL} - 2\alpha^2 t_{KL})\right]^2 \\
&\qquad\quad \times \left[(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2\right]^{-\frac{1}{2}}\Big] \\
&\qquad \times \left[(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2\right]^{-1}.
\end{aligned}$$

Since the denominator of $h''(\Delta)$ (i.e., $(t_{AK}+\Delta)^2+\alpha^2(t_{KL}-2\Delta)^2+(t_{BL}+\Delta)^2$) is positive, it remains to show that the numerator is non-negative, i.e.,

$$\begin{aligned}
&(4\alpha^2+2)\sqrt{(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2} \\
&\quad - \frac{[\Delta(4\alpha^2+2) + (t_{AK} + t_{BL} - 2\alpha^2 t_{KL})]^2}{\sqrt{(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2}} \\
&= \left[(t_{AK}+\Delta)^2 + \alpha^2(t_{KL}-2\Delta)^2 + (t_{BL}+\Delta)^2\right]^{-\frac{1}{2}}
\end{aligned}$$

$$\times \left[(4\alpha^2+2)((t_{AK}+\Delta)^2+\alpha^2(t_{KL}-2\Delta)^2+(t_{BL}+\Delta)^2)\right.$$
$$\left.-[\Delta(4\alpha^2+2)+(t_{AK}+t_{BL}-\alpha^2 t_{KL})]^2\right].$$

The denominator of the above equation is again positive, so we focus on the numerator. Expanding the expression in the numerator gives:

$$\begin{aligned}
&(4\alpha^2+2)t_{AK}^2+2(4\alpha^2+2)\Delta t_{AK}+(4\alpha^2+2)\Delta^2+\alpha^2(4\alpha^2+2)t_{KL}^2\\
&-4\alpha^2(4\alpha^2+2)\Delta t_{KL}+4\alpha^2(4\alpha^2+2)\Delta^2+(4\alpha^2+2)t_{BL}^2+2(4\alpha^2+2)\Delta t_{BL}\\
&+(4\alpha^2+2)\Delta^2-(4\alpha^2+2)^2\Delta^2-t_{AK}^2-t_{AK}t_{BL}+2\alpha^2 t_{AK}t_{KL}-t_{AK}t_{BL}\\
&-t_{BL}^2+2\alpha^2 t_{BL}t_{KL}+2\alpha^2 t_{AK}t_{KL}+2\alpha^2 t_{BL}t_{KL}-4\alpha^4 t_{KL}^2-2(4\alpha^2+2)\Delta t_{AK}\\
&-2(4\alpha^2+2)\Delta t_{BL}+2\alpha^2(\alpha^2+1)\Delta t_{KL}\\
=&\,(4\alpha^2+1)t_{AK}^2+2\alpha^2 t_{KL}^2+(4\alpha^2+1)t_{BL}^2-2t_{AK}t_{BL}+4\alpha^2 t_{AK}t_{KL}+4\alpha^2 t_{BL}t_{KL}\\
=&\,4\alpha^2 t_{AK}^2+4\alpha^2 t_{BL}^2+(t_{AK}-t_{BL})^2+2\alpha^2 t_{KL}^2+4\alpha^2 t_{AK}t_{KL}+4\alpha^2 t_{BL}t_{KL},
\end{aligned}$$

which is a positive value. Therefore, $g(\Delta)$ is a strictly convex function. ■

Figure 3.6 shows a plot of functions $f(\Delta)$ and $g(\Delta)$ for different values of Δ with $\gamma = 0.8$ and hubs K and L located at points 1 and 19 on the line. Let Δ_ν represent the optimal Δ value for a given ν value, and $\Delta_{\nu'}$ correspond to that for $\nu' = \nu + \varepsilon, \varepsilon > 0$. If our conjecture that the boundary hubs shift inward as ν increases, then we must have $\Delta_{\nu'} > \Delta_\nu$.

Theorem 3 *Let Δ_ν be the minimizer of $\mathcal{LP}_\Delta(AKLB)$ for coefficient of variation ν. Let $\Delta_{\nu'}$ be the minimizer of $\mathcal{LP}_\Delta(AKLB)$ for $\nu' = \nu + \varepsilon$, $\varepsilon > 0$. Then $\Delta_{\nu'} > \Delta_\nu$.*

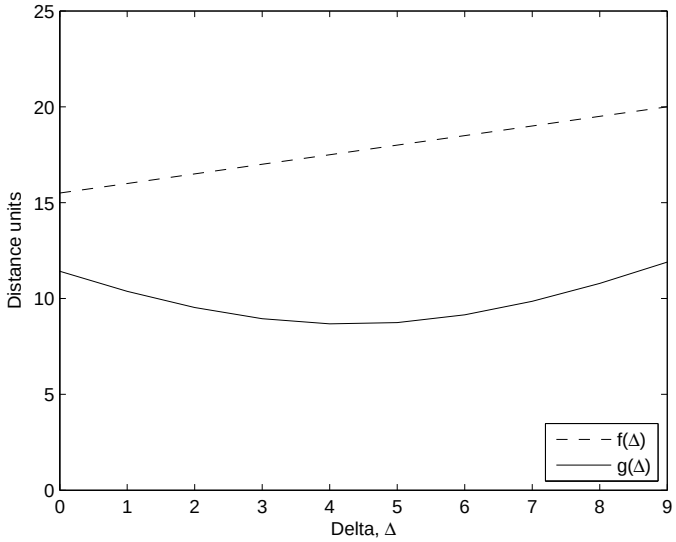


Figure 3.6: Plot of functions $f(\Delta)$ and $g(\Delta)$ for different values of Δ with $\gamma = 0.8$.

Proof: Since Δ_ν and $\Delta_{\nu'}$ are minimizers of $\mathcal{LP}_\Delta(AKLB)$ for ν and ν', respectively, the following two equations must hold:

$$f'(\Delta_\nu) + \nu g'(\Delta_\nu) = 0 \tag{3.13}$$

and

$$f'(\Delta_{\nu'}) + \nu' g'(\Delta_{\nu'}) = 0. \tag{3.14}$$

(Equations (3.13) and (3.14) have interior minimizers so long as $\nu > 0$ since $\mathcal{LP}_\Delta(AKLB)$ is a convex function.) From equations (3.13) and (3.14), we have

$$f'(\Delta_\nu) + \nu g'(\Delta_\nu) = f'(\Delta_{\nu'}) + \nu' g'(\Delta_{\nu'}). \tag{3.15}$$

Since $f'(\Delta)$ is a constant, we can eliminate $f'(\Delta_\nu)$ and $f'(\Delta_{\nu'})$ from equation (3.15),

giving:

$$\begin{aligned} \nu g'(\Delta_\nu) &= (\nu+\varepsilon) g'(\Delta_{\nu'}) \\ g'(\Delta_{\nu'}) &= \frac{\nu}{\nu+\varepsilon} g'(\Delta_\nu). \end{aligned}$$

From equation (3.13), we have $g'(\Delta_\nu) = -f'(\Delta_\nu)/\nu$, which is negative-valued since $f'(\Delta_\nu) > 0$ (from Lemma 1), and $\nu > 0$. Therefore,

$$g'(\Delta_{\nu'}) > g'(\Delta_\nu). \tag{3.16}$$

Finally, since $g(\Delta)$ is a convex function, the function $g'(\Delta)$ is a monotonically increasing function. Thus, $g'(\Delta_{\nu'}) > g'(\Delta_\nu)$ implies that $\Delta_{\nu'} > \Delta_\nu$. ■

Theorem 3 shows that the primary driver for the shifting of the boundary hubs towards the center of the line is due to the function $g(\Delta)$, which is the square root term in the calculation of the path length $\mathcal{LP}_\Delta(AKLB)$. An increase in the coefficient of variation ν (or the service level parameter γ) increases the contribution of the square root term, and thus $g(\Delta)$, to the length of the path. Since $g(\Delta)$ is a convex function, its impact is more visible when γ and/or ν is large.

3.4 Computational Experiments

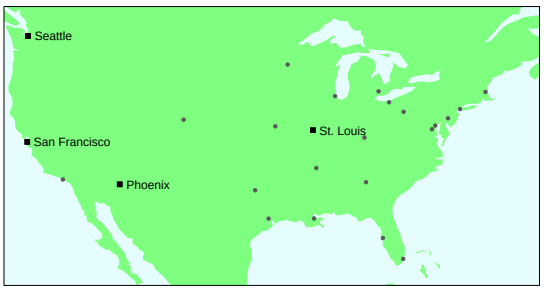
The results in the previous section indicate that there is often a centrally located hub ($p = 1$ and 3 cases), and that the boundary hubs tend to shift in towards the central hub as γ or ν increases ($p = 3$ case). In this section, we demonstrate via computational experiments that similar results are observed with more a general data set. We experiment with the SpHCP-CC with normally distributed travel times using the standard CAB data

set (Beasley, 2004), which consists of the 25 largest cities in the United States. The first set of experiments assume that the coefficient of variation is the same for every link in the network (i.e., $\nu_{ij} = \nu$). In the second set of experiments, we set different coefficient of variation values for each link. We will continue to assume independence in the travel times among the links in the network.

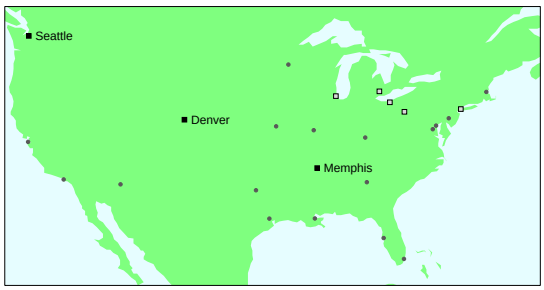
3.4.1 Homogeneous CV

For this set of experiments, we use the 10, 15, 20, and 25 node CAB data set. We solve SpHCP-CC with $p = 2, 3, 4, 5$, coefficient of variation $\nu = 0, 1, 2$ (where $\nu = 0$ represents the deterministic case), and service level parameter $\gamma = 0.8, 0.9, 0.99$. As seen in Figures 3.2 and 3.4, a problem instance may have multiple optima. Thus, to ensure a complete analysis of this model, we enumerated all possible hub configurations for each problem instance (N, p, ν, γ). We solve the corresponding assignment problem HCSAP using CPLEX, which solved the HCSAP within a few seconds.

Figure 3.7 shows the graphical results for the 25-node, 4-hub problems with $\gamma = 0.9$. Similar to the results observed for the SpHCP-CC problem on the line, there is typically a central hub in the service region and the remaining hubs scattered around this central hub. In Figure 3.7(a), three of the four hubs are located on the west coast of the country with a hub in the center of the country at St. Louis. For problem instance $\nu = 1$, there were multiple optimal solutions. Each of these five optimal solutions share three common hubs: Denver, Memphis, and Seattle. These common hubs are represented by the dark square icons in Figure 3.7(b). The optimal hub locations in each of these five optimal solutions



(a) $\nu = 0$. The optimal hub locations are Phoenix, San Francisco, Seattle, and St. Louis.



(b) $\nu = 1$. There are multiple optimal solutions for this problem instance. Denver, Memphis and Seattle (indicated by the shaded square icons) appear as hubs in all optimal solutions, while the fourth hub in an optimal solution is one of the many unshaded square icons.

(c) $\nu = 2$. Denver, Detroit and New Orleans appear as hubs in all optimal solutions, while the fourth hub in an optimal solution is one of the many unshaded square icons.

Figure 3.7: Optimal hub locations for the 25-node, 4-hub SpHCP-CC.

is a combination of the three common hubs and one of the "uncommon" hubs, which are represented by the light square icons. Notice that there is still a central hub, which is now located in Memphis, and that the hubs in San Francisco and Phoenix have now shifted to one in Denver and the other around the Great Lakes region. This relocation of the hubs show a weak shift of the boundary hubs towards the center of the service region with an increase in the coefficient of variation ν. Problem instance $\nu = 2$ also had multiple optimal solutions. Figure 3.7(c) shows that there are three common hubs: Denver, Detroit, and New Orleans. These three common hubs alone show a more pronounced shift in the hub locations towards the center of the service region with a further increase in ν.

The clustering of the hubs towards the center of the service region as γ or ν increases can be shown by calculating the diameter of the hub network. The hub network diameter is the length of the longest link (based on the mean length of the link) connecting any two pairs of hubs in the network. The diameter acts as a proxy for how clustered the hub nodes are in the network. Tables 3.2 and 3.3 provide the hub network diameter for all problem instances. For those problem instances where there are multiple optima, we report the smallest diameter value over all the optimal solutions. In general, the diameter of the hub network decreases as γ or ν increases. The only exceptions are for $(N, p, \nu) = (25, 4, 0.80)$ and $(25, 4, 0.9)$. These exceptions are likely due to the spread of the cities across the country.

Table 3.2: Minimum diameter of the hub network using the CAB data set with $p = 2$ and 3, and the number of multiple optimal solutions for each problem instance.

			Diameter				# of optimal solutions			
p	γ	ν	$N = 10$	15	20	25	$N = 10$	15	20	25
2	0.80	0	1080	1608	1266	781	1	1	1	1
		1	1080	1608	1266	781	1	1	1	1
		2	1080	1080	1266	781	1	1	1	1
	0.90	0	1080	1608	1266	781	1	1	1	1
		1	1080	1608	1266	781	1	1	1	1
		2	1010	1080	1266	781	1	1	1	1
	0.99	0	1080	1608	1266	781	1	1	1	1
		1	1010	1080	1080	781	1	1	1	1
		2	1010	1080	1080	781	1	1	1	1
3	0.80	0	1217	2600	2144	1217	1	1	1	2
		1	1217	1979	1217	1217	1	1	1	1
		2	1144	1144	1144	1144	1	2	1	1
	0.90	0	1217	2600	2144	1217	1	1	1	2
		1	1217	1144	1217	1217	1	2	1	1
		2	1144	1144	1144	1144	1	2	1	1
	0.99	0	1217	2600	2144	1217	1	1	1	2
		1	1144	1144	1144	1144	2	2	1	1
		2	1010	1144	1144	1144	3	1	1	1

Table 3.3: Minimum diameter of the hub network using the CAB data set with $p = 4$ and 5, and the number of multiple optimal solutions for each problem instance.

			Diameter				# of optimal solutions			
p	γ	ν	$N = 10$	15	20	25	$N = 10$	15	20	25
4	0.80	0	1765	2600	2397	1736	1	1	2	1
		1	1217	1715	1217	2090	3	4	19	3
		2	1144	1208	1144	1144	6	8	14	24
	0.90	0	1765	2600	2397	1736	1	1	2	1
		1	1144	1715	1217	1873	1	8	14	5
		2	1144	1208	1144	1144	6	8	13	17
	0.99	0	1765	2600	2397	1736	1	1	2	1
		1	1144	1208	1144	1144	11	8	9	11
		2	1010	1208	1144	1144	12	4	8	8
5	0.80	0	1765	2600	2453	2087	2	6	4	8
		1	1217	1715	1217	1873	13	30	151	3
		2	1144	1208	1144	1144	15	48	91	234
	0.90	0	1765	2600	2453	2087	2	6	4	8
		1	1144	1715	1217	1873	5	56	91	91
		2	1144	1208	1144	1144	15	48	78	136
	0.99	0	1765	2600	2453	2087	2	6	4	8
		1	1144	1208	1144	1144	25	40	36	55
		2	1010	1208	1144	1144	19	18	28	28

3.4.2 Non-Homogeneous Coefficient of Variation

In the previous section, we observed a clustering of the hub nodes towards the center of the service region as the common coefficient of variation ν increased. A follow-up question, then, is whether this phenomenon is also seen when the coefficient of variation among all links is different. We use the 25-node CAB data set with $p = 4$ hubs and the service level parameter $\gamma = 0.95$. We set the mean coefficient of variation $\overline{\nu}$ to 1 and 2, and generated 10 random scenarios for each of $\overline{\nu} = 1$ and $\overline{\nu} = 2$ with the coefficient of variation ν_{ij} for link ij sampled from the uniform distribution $U(0.8\overline{\nu}, 1.2\overline{\nu})$. We again enumerated all possible hub configurations and solve the assignment problem HCSAP using CPLEX.

Table 3.4 shows the (smallest) diameter of the hub network for each of the 10 different scenarios for $\overline{\nu} = 1, 2$. We also provide the hub network diameter for the deterministic case ($\nu = 0$) for comparison. In general, we see that the diameter of the hub network is non-increasing as $\overline{\nu}$ increases – indicating a clustering of the hub nodes. Thus, empirical evidence shows that as the mean coefficient of variation increases (and similarly, as the service level parameter γ increases), the hub nodes tend to cluster around the center of the service region, regardless of whether these two parameters are the same or different across all the links in the network.

3.5 Solution Methodology

Our observations of the optimal locations for the p hubs in the chance-constrained SpHCP for the problem on the line (Section 3.3) and for the CAB dataset (Section 3.4) lead us to believe that we could exploit the structure of the optimal hub configuration for

Table 3.4: Minimum diameter of the hub network for the 25-node, 4-hub problem using the CAB data set with $\gamma = 0.95$ for each of the 10 different non-homogeneous coefficient of variation scenarios.

	Diameter			# of opt. solutions	
Scenario	$\nu = 0$	$\overline{\nu} = 1$	$\overline{\nu} = 2$	$\overline{\nu} = 1$	$\overline{\nu} = 2$
1	1736	1715	1715	1	1
2	1736	1967	1967	4	5
3	1736	1765	1765	5	5
4	1736	2090	2090	1	1
5	1736	1873	1873	3	1
6	1736	2090	1626	3	9
7	1736	1217	1217	15	13
8	1736	2027	1208	1	43
9	1736	1715	1302	3	14
10	1736	1217	1217	17	15
Average	1736	1217	1208		

use in a solution heuristic. In particular, an optimal solution to the SpHCP typically has a hub located in the center of the service region with the remaining $p-1$ hubs surrounding the central hub. We next describe our radial heuristic, which finds the best hub network configuration with this particular structure.

3.5.1 Radial Heuristic

Our radial heuristic is similar to the heuristic proposed by Dyer and Frieze (1985) for the p-center problem except in how the the first center is chosen. In Dyer and Frieze (1985), the heuristic arbitrarily chooses the first center, while we solve the vertex 1-center problem on the network (using the mean link lengths) for the first center. The procedure to locate the remaining $p-1$ hubs is similar to Dyer and Frieze (1985). Recall that depending on the value of γ and ν, the outer $p-1$ hubs may be closer to the central hub than to the boundary of the service region. Thus, to find the best hub configuration, we replicate this process of generating a hub configuration several times. At each iteration, we restrict the set of candidate locations for the $p-1$ hubs.

Figure 3.8 shows the steps in the radial heuristic. In step 1, we allow the p hubs to be located at any of the nodes in the network; thus, we initialize the set of candidate locations C with the set of nodes N. In step 2, we generate a hub configuration $H(C)$ based on the set C (see Figure 3.10). The last operation in step 3 removes the hub node farthest away from the 1-center hub H_1 from the set C; thus, eliminating that node as a candidate hub location in the next iteration. This has the effect of reducing the diameter of the hub network in the subsequent iteration of the heuristic. The radial heuristic stops when

Step 1: Set $C \leftarrow N$, and set $\beta^\star$ to be some large number.

Step 2: If $|C| \geq p$, generate a hub configuration $H(C)$ based on the set C (see Figure 3.10). Else, go to step 4.

Step 3: Solve HCSAP for $\hat{\beta}$ using configuration $H(C)$.

If $\hat{\beta} < \beta^\star$, set $H^\star \leftarrow H(C)$ and $\beta^\star \leftarrow \hat{\beta}$.

Set $C \leftarrow C \setminus \arg\max_{i \in C}\{t_{iH_1}\}$, and go to step 2. ($H_1$ is the 1-center hub.)

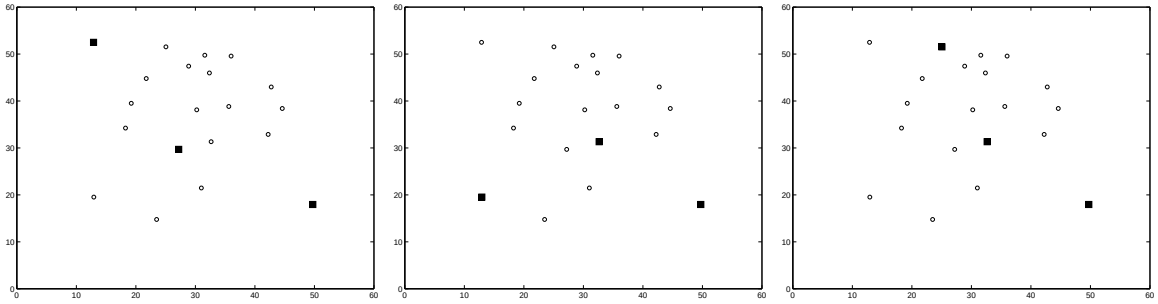
Step 4: Return $H^\star$ and $\beta^\star$.

Figure 3.8: Radial heuristic.

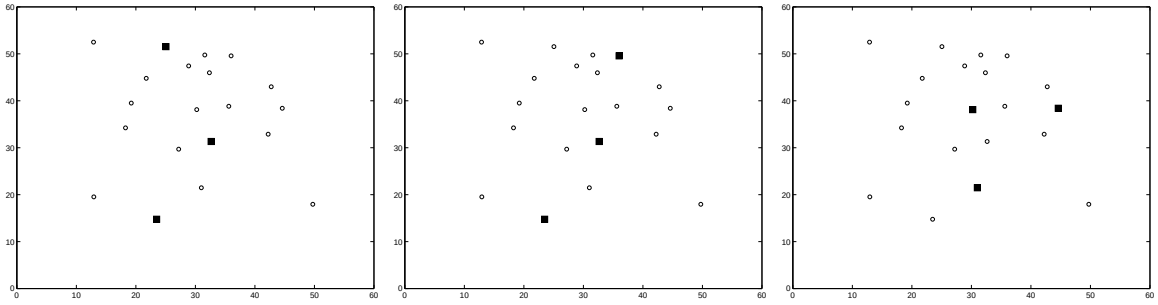
the number of candidate nodes in the set C is less than p. Figure 3.9 shows the locations of the hubs for the first 9 iterations of the radial heuristic using the 20-node AP data set (Ernst and Krishnamoorthy, 1996) with $p = 3$ hubs.

The process to generate a hub configuration in step 2 is to first find the 1-center in the network. This can be done in $\mathcal{O}(|C|)$ time using the algorithm shown in Figure 3.10 (Daskin, 1995). Next, we use a construction heuristic based on the heuristic of Dyer and Frieze (1985) to locate the remaining $p - 1$ hubs. The construction heuristic to generate a hub configuration is given in Figure 3.10. The radial heuristic is coded in MATLAB (The Mathworks Inc., 1997), and the assignment problem HCSAP is solved using CPLEX via CPLEX-MEX (Musicant, 2000).

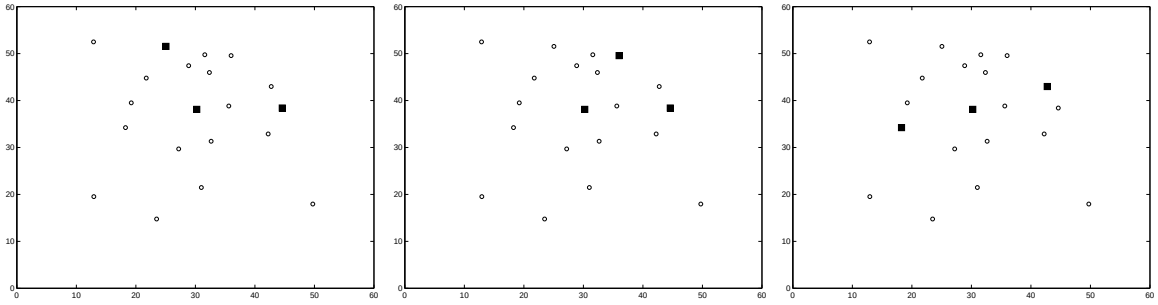
We test the radial heuristic using the 10-, 15-, 20-, and 25-node CAB data set, as well as the 10-, 20-, and 25-node AP data set. We set $p = 2, 3, 4, 5$, $\nu = 0, 1, 2$, and $\gamma = 0.8, 0.9, 0.99$. For each problem instance, we calculate the percentage gap, which



(a) Iteration 1. (b) Iteration 2. (c) Iteration 3.



(d) Iteration 4. (e) Iteration 5. (f) Iteration 6.



(g) Iteration 7. (h) Iteration 8. (i) Iteration 9.

Figure 3.9: Locations of the hub nodes generated in the first 9 iterations of the radial heuristic based on the 20-node AP data set with $p = 3$.

% Find central hub

Set the incumbent travel time value $T^\star$ to some large number.

Let H_1 represent the index of the optimal central hub.

For each node $h \in C$

Let $T = \max_{i \in C}\{t_{ih}\}$.

If $T < T^\star$, set $T^\star \leftarrow T$ and $H_1 \leftarrow h$.

% Locate remaining $p - 1$ hubs

Set $H_S \leftarrow H_1$.

While $|H_S| \leq p$

For each node $i \in C \setminus \{H_S\}$, set $T_i \leftarrow \min_{j \in H_S}\{t_{ij}\}$.

Set $i^\star \leftarrow \arg\max_{i \in C \setminus \{H_S\}}\{T_i\}$.

Set $H_S \leftarrow H_S \cup \{i^\star\}$.

Return H_S as the hub configuration.

Figure 3.10: Construction heuristic to generate a hub configuration.

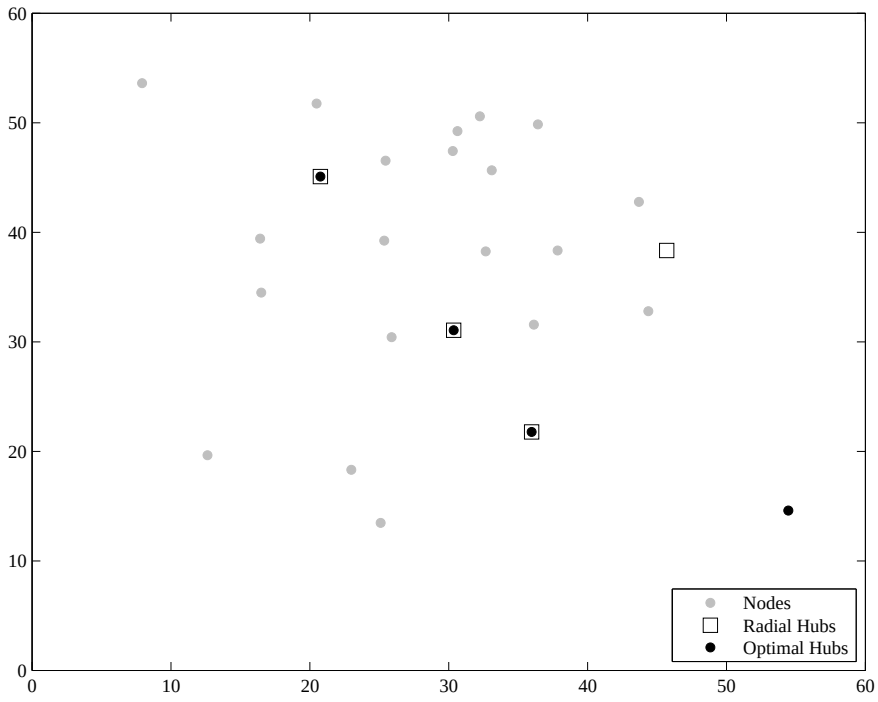


Figure 3.11: Plot of the optimal hub locations versus the hub locations found by the radial heuristic for the 25-node AP data set with $p = 4$, $\nu = 1$ and $\gamma = 0.9$.

measures how far the objective function value of the best solution found from the radial heuristic is from that of the optimal solution. Despite the identified solution structure, the radial heuristic had limited success. The average percentage gaps were 6.4% and 4.8% for the CAB and AP data sets, respectively. Some problem instances had percentage gaps of over 20%. Visual inspection of the solutions show that the radial heuristic finds all but one or two of the optimal hubs (see, for example, Figure 3.11). We see from this plot that a local search heuristic beginning with the hub configuration found by the radial heuristic will likely find the optimal solution. In the next section, we propose one such local search heuristic.

Step 1: Generate an initial (feasible) hub configuration H.

Let H be the incumbent hub configuration.

Solve HCSAP for β using configuration H.

Step 2: Let $N_C \leftarrow N \setminus H$ be the set of non-hub nodes to be candidate hub nodes.

Randomly choose a non-hub node $i \in N_C$.

Step 3: For each hub $h \in H$, swap node i with hub h.

Let H_h represent this new hub configuration.

Solve HCSAP for β_h using configuration H_h.

Step 4: Let $\hat{\beta} = \min_{h \in H}\{\beta_h\}$ and $\hat{h} = \arg\min_{h \in H}\{\beta_h\}$.

If $\hat{\beta} < \beta$, set $H \leftarrow H_{\hat{h}}$, $\beta \leftarrow \beta_{\hat{h}}$, and go to step 2; else, go to step 5.

Step 5: Choose another node $i \in N_C$ not tried previously. Go to step 3.

If all nodes in N_C have been evaluated without any improvement to the incumbent solution, then stop.

Figure 3.12: Teitz-Bart heuristic.

3.5.2 The Teitz-Bart Heuristic

For the local search heuristic, we use a well-known heuristic from the facility location literature proposed by Teitz and Bart (1968) for the p-median problem in the network (Hakimi, 1964, 1965). The Teitz-Bart heuristic is essentially a best-improving, single-swap heuristic. A general outline of the heuristic to solve the chance-constrained SpHCP is shown in Figure 3.12. The heuristic is coded in MATLAB (The Mathworks Inc., 1997).

Step 1 of the Teitz-Bart heuristic requires an initial feasible solution. We use the

Table 3.5: Computational results for the small CAB and AP data sets (up to 25 nodes).

	CAB data set		AP data set	
	TB+Radial	TB+Random	TB+Radial	TB+Random
Average % optimality gap	0.636%	0.894%	0.675%	0.823%
Average $\sigma_\beta/\bar{\beta}$	0.0004	0.0102	0.0001	0.0119

best hub configuration found from our radial heuristic for the initial feasible solution. An approach that is commonly used to initialize the heuristic is to generate a random feasible solution. We compare both initialization methods, and show that there is a benefit to initializing the heuristic with the best radial heuristic solution. As there is a random component to the Teitz-Bart heuristic in step 2, we run the heuristic 10 times for each initial hub configuration. When initializing with a random configuration, we generated 10 random hub configurations for each problem instance of (N, p, γ, ν) and report the average β value $\bar{\beta}$ found.

Table 3.5 summarizes the results of the computational experiments using the CAB and AP data sets. For each problem instance (N, p, γ, ν), we calculated the average $\bar{\beta}$ of the 10 computational runs. The "Average % optimality gap" is the average of $100(\bar{\beta}/\beta^\star)\%$ over all problem instances, where $\beta^\star$ is the optimal objective function value for a problem instance. We first note that that the Teitz-Bart heuristic improves upon the best hub configuration found from the radial heuristic – reducing the average percentage gap from 6.4% to 0.636% for the CAB data set, and from 4.8% to 0.675% for the AP data set. Additionally, initializing the Teitz-Bart heuristic with the best hub configuration found from the radial

heuristic results in a lower optimality gap than if we randomly initialized the heuristic. Even though there is a random component to the Teitz-Bart heuristic in step 2, the heuristic consistently found the same final objective function value β when initialized with the radial heuristic, as shown by the average standard-deviation-to-mean ratio $\sigma_\beta/\bar{\beta}$. In comparison, initializing with random hub configurations resulted in some variation in the final β value found. While the run-time for the Teitz-Bart heuristic is similar for the two different initialization schemes (see Table 3.6), these results show that there is a benefit to using the radial heuristic as an initialization procedure because it leads to more consistent results and lower β values on average.

In this set of computational experiments, the Teitz-Bart heuristic initialized by the radial heuristic was not able to find the optimal solution in some of the problem instances. If such a suboptimal hub network is used, the service level provided by this network will be lower than the optimal network. The implied service level is calculated by

$$\Phi^{-1}(\beta^\star, \mu_{\beta\text{-path}}, \sigma_{\beta\text{-path}}),$$

where Φ^{-1} is the inverse normal cumulative distribution function, $\beta^\star$ is the optimal objective function value, and $\mu_{\beta\text{-path}}$ and $\sigma_{\beta\text{-path}}$ are the mean and standard deviation of the β-path, respectively, of the suboptimal network. (Recall, the β-path is the "longest" path in the network.) Tables 3.7 and 3.8 provide the difference between the desired service level parameter γ and the implied service level value for the different problem instances. (We note that the implied service level value can be calculated only for those problem instances where the coefficient of variation is positive; i.e., $\nu = 1, 2$.) The average reduction in service level is 0.10% and 0.15% for the CAB and AP data sets, respectively, with the largest

Table 3.6: Average run-times in seconds for the small CAB and AP data sets (up to 25 nodes).

		CAB data set		AP data set	
$\|N\|$	p	TB+Radial	TB+Random	TB+Radial	TB+Random
10	2	2.3	2.7	2.2	2.7
	3	4.9	6.9	5.4	6.9
	4	6.9	9.9	8.9	9.9
	5	8.2	12.5	9.3	12.5
15	2	5.3	17.8		
	3	16.9	84		
	4	29.3	184.5		
	5	42.1	314.7		
20	2	13.2	17.8	13	17.8
	3	54.3	84	53.7	84.0
	4	97.3	184.5	111.6	184.5
	5	143.8	314.7	202	314.7
25	2	25.7	36.7	25.9	36.7
	3	121.9	191	117.1	191.0
	4	235.7	395.3	248.6	395.3
	5	354.4	555	418.7	555.0

service level reduction being 2.6%. Thus, even when the heuristic does not find the optimal solution, the final hub configuration returned by the heuristic provides a service level that is close to the desired service level.

We next applied the Teitz-Bart heuristic using the two different initialization procedures on the larger 40-node AP data set. Tables 3.9 and 3.10 show that initializing with the radial heuristic also resulted in the Teitz-Bart heuristic consistently finding the same final solution for this large data set. In 10 of the 36 problem instances, both initialization procedures resulted in the same final solution (indicated by equal $\bar{\beta}$ values and zero $\sigma_\beta/\bar{\beta}$ values). In the remaining 26 problem instances, the radial initialization heuristic found a better average β value than the random initialization procedure in 19 of these problem instances. This leads us to suggest that when faced with solving problems with a large number of nodes, by using the radial heuristic to find an initializing set of hubs, one only needs to run the Teitz-Bart heuristic a handful of times (so as to account for the random component in the heuristic) to obtain a relatively good final solution. Moreover, if we consider the magnitude increase in run times when increasing the number of hubs p or the number of nodes N (e.g., from the 20-node to 40-node data set), we can further appreciate having a consistent solution approach that provides good solutions without having to perform a lot of computational runs. Sensitivity analysis of the problem, if needed, can then also be performed as it is not prohibitive to do so.

We do acknowledge, however, that the random initialization resulted in better $\bar{\beta}$ values in 7 of the 36 problem instances. Recall that when running the Teitz-Bart heuristic with the random initialization method, we generated 10 random initial hub configurations

Table 3.7: Difference between the desired service level γ and the implied service level from the best hub configuration found using the Teitz-Bart heuristic initialized with the radial heuristic for the CAB and AP data sets with $p = 2$ and 3.

	CAB data set				AP data set		
(p, ν, γ)	$N = 10$	$N = 15$	$N = 20$	$N = 25$	$N = 10$	$N = 20$	$N = 25$
(2, 1, 0.80)	2.0%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%
(2, 1, 0.90)	1.5%	0.7%	0.0%	0.0%	0.0%	0.0%	0.0%
(2, 1, 0.99)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
(2, 2, 0.80)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
(2, 2, 0.90)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%
(2, 2, 0.99)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
(3, 1, 0.80)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
(3, 1, 0.90)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%
(3, 1, 0.99)	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%
(3, 2, 0.80)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
(3, 2, 0.90)	0.0%	0.0%	0.0%	0.0%	0.0%	1.0%	0.0%
(3, 2, 0.99)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%

Table 3.8: Difference between the desired service level γ and the implied service level from the best hub configuration found using the Teitz-Bart heuristic initialized with the radial heuristic for the CAB and AP data sets with $p = 4$ and 5.

	CAB data set				AP data set		
(p, ν, γ)	$N = 10$	$N = 15$	$N = 20$	$N = 25$	$N = 10$	$N = 20$	$N = 25$
(4, 1, 0.80)	0.0%	0.0%	1.2%	1.1%	2.6%	0.0%	0.0%
(4, 1, 0.90)	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
(4, 1, 0.99)	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	0.0%
(4, 2, 0.80)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
(4, 2, 0.90)	0.0%	0.0%	0.0%	0.0%	0.0%	2.0%	0.0%
(4, 2, 0.99)	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%
(5, 1, 0.80)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
(5, 1, 0.90)	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
(5, 1, 0.99)	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	0.0%
(5, 2, 0.80)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
(5, 2, 0.90)	0.0%	0.0%	0.0%	0.0%	0.0%	2.0%	0.0%
(5, 2, 0.99)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table 3.9: Results for the 40-node AP data set with $p = 2$ and 3.

	$\bar{\beta}$		$\sigma_\beta/\bar{\beta}$		Run-time (sec.)	
(p, ν, γ)	Radial	Random	Radial	Random	Radial	Random
(2, 0, 0.80)	65435	65060	0.0000	0.0182	173	189
(2, 0, 0.90)	65435	63559	0.0000	0.0311	184	168
(2, 0, 0.99)	65435	64309	0.0000	0.0282	179	179
(2, 1, 0.80)	98877	98488	0.0000	0.0021	176	182
(2, 1, 0.90)	117020	117230	0.0000	0.0103	174	171
(2, 1, 0.99)	158440	161380	0.0000	0.0064	171	147
(2, 2, 0.80)	132940	133620	0.0000	0.0035	177	189
(2, 2, 0.90)	168400	168400	0.0000	0.0000	157	158
(2, 2, 0.99)	256860	256860	0.0000	0.0000	134	138
(3, 0, 0.80)	59558	60337	0.0000	0.0190	1032	1338
(3, 0, 0.90)	59558	59591	0.0000	0.0228	1094	1098
(3, 0, 0.99)	59558	58850	0.0000	0.0110	2035	1075
(3, 1, 0.80)	95001	93223	0.0000	0.0187	1035	1957
(3, 1, 0.90)	113530	109150	0.0000	0.0212	1507	1213
(3, 1, 0.99)	145550	145990	0.0000	0.0040	1497	987
(3, 2, 0.80)	122400	122690	0.0000	0.0025	1072	1664
(3, 2, 0.90)	154080	159600	0.0000	0.0399	931	1186
(3, 2, 0.99)	229300	229890	0.0000	0.0042	959	830

Table 3.10: Results for the 40-node AP data set with $p = 4$ and 5.

	$\bar{\beta}$		$\sigma_\beta/\bar{\beta}$		Run-time (sec.)	
(p, ν, γ)	Radial	Random	Radial	Random	Radial	Random
(4, 0, 0.80)	52266	52880	0.0000	0.0262	2137	2380
(4, 0, 0.90)	52266	53899	0.0000	0.0300	4124	3772
(4, 0, 0.99)	52266	52445	0.0000	0.0108	3344	2992
(4, 1, 0.80)	86545	87271	0.0000	0.0175	2134	4977
(4, 1, 0.90)	101800	103190	0.0000	0.0313	5404	3832
(4, 1, 0.99)	138020	138530	0.0000	0.0116	4051	3001
(4, 2, 0.80)	115720	115720	0.0000	0.0000	2796	4807
(4, 2, 0.90)	146230	146870	0.0000	0.0114	3196	4409
(4, 2, 0.99)	219210	219210	0.0000	0.0000	2998	2445
(5, 0, 0.80)	49741	50349	0.0000	0.0120	8092	5325
(5, 0, 0.90)	49741	50280	0.0000	0.0126	8067	4181
(5, 0, 0.99)	49741	50774	0.0000	0.0276	7201	9097
(5, 1, 0.80)	86545	86545	0.0000	0.0000	4074	9321
(5, 1, 0.90)	101800	101800	0.0000	0.0000	7180	5809
(5, 1, 0.99)	138020	138020	0.0000	0.0000	7147	5111
(5, 2, 0.80)	115720	115720	0.0000	0.0000	4114	6530
(5, 2, 0.90)	146230	146230	0.0000	0.0000	4655	5022
(5, 2, 0.99)	219210	219210	0.0000	0.0000	3660	3962

for each problem instance. For each problem instance, we take the best solution (i.e., the one with the lowest $\bar{\beta}$ value) from the 10 final solutions and compared that to the solution found using the radial initialization. Over the 36 problem instances, the $\bar{\beta}$ value from the best solution is better than that of the radial solution in 10 problem instances. As seen in Table 3.11, the relative difference (i.e., radial $\bar{\beta}$ divided by best random $\bar{\beta}$) ranged from 0.4% to 6.1%. Notice that the radial initialization approach was outperformed only for the problem instances where p is small. Although the best solution found by using the random initialization betters that of the radial in 10 of the 36 problem instances, we emphasize again the difference in the $\sigma_{\beta}/\bar{\beta}$ values of the two initialization procedures.

3.6 Conclusion

In this chapter, we introduce the chance-constrained stochastic p-hub center problem, which can be used as a strategic planning tool to design a hub-and-spoke network for a small package delivery company. Chance constraints are included in the model to specify a minimum service level required of the network's ability to transport packages from their origins to destinations within a period of time.

We observed that the optimal network configuration typically has a hub located in the center of the service region with the remaining hubs surrounding this central hub. The surrounding hubs tend to be located closer to the center of the service region when the level of uncertainty is high. We exploit this structure as an initial starting configuration in a single-swap, best-improvement heuristic proposed by Teitz and Bart (1968). Our computational results show that this solution approach consistently finds the same final solution,

Table 3.11: Percentage gap between the radial $\bar{\beta}$ value and the best $\bar{\beta}$ value found by the random heuristic for the problem instances where the best random $\bar{\beta}$ value is lower than the radial $\bar{\beta}$ value.

(p, ν, γ)	Radial $\bar{\beta}$	Best Random $\bar{\beta}$	% gap
(2,0,0.80)	65435	61682	6.1%
(2,0,0.90)	65435	61682	6.1%
(2,0,0.99)	65435	61682	6.1%
(2,1,0.80)	98877	98391	0.5%
(2,1,0.90)	117020	116550	0.4%
(3,0,0.80)	59558	58748	1.4%
(3,0,0.90)	59558	58192	2.3%
(3,0,0.99)	59558	58192	2.3%
(3,1,0.80)	95001	92101	3.1%
(3,1,0.90)	113530	107940	5.2%

and in many cases, the final objective function value found by this approach is better than that from using a randomized initialization procedure for the Teitz-Bart heuristic.

CHAPTER 4
STOCHASTIC P-HUB CENTER PROBLEM: TWO-STAGE FORMULATION

4.1 Introduction

Many real-life communication networks are two-level hierarchical networks that have a hub-and-spoke configuration (Kim et al., 1995). The first level, called the backbone network, is the network consisting of the hub nodes (switches, concentrators, gates, control/transfer points, etc.). The backbone network can take on many configurations, including ring, tree, star, and fully-meshed (or fully-connected) networks. The second level – the access (or tributary or local) network – consists of the network formed from the connection of the user (or terminal) nodes to the hub nodes. As with backbone networks, access networks can be ring, tree, star, or fully-connected networks. For a survey of the literature on these different types of two-level networks, the reader is directed to the paper by Klincewicz (1998). In this chapter, we shall only focus on networks with fully-connected backbone networks (where all the hub nodes are connected to each other) and star access networks (where a terminal or non-hub node is connected to only one hub node).

Due to the dependence on communication services for work and/or leisure activities, there is a constant concern about the dependability or robustness of the communication networks. Borrowing from the definition of survivability by Brush and Marlow (1990), Kim et al. (1995) define robustness as the ability of the network to perform required functions after a specified set of components become unavailable. A common approach to ensure the robustness of the network is to configure the network so that there are multiple paths be-

tween every pair of nodes in the network. Since the hub network is fully-connected, there are indeed several paths from one hub node to another. As for the access network, one could require multiple homing for every access node; that is, each access node is connected to more than one hub node. In Kim et al. (1995), the authors discuss a hub-and-spoke network where each user node is connected to exactly two hub nodes. This is the multiple-assignment version of the hub location problem. The solution to their optimization problem indicates which of the two hub nodes are the primary or secondary homing nodes for a user node.

In addition to the dependability of the network, another important performance measure is response time. The response time is the time required to transfer requested information through the network. There may often be congestion at a hub node due to processing times by switches at the hubs or repeaters and/or signal amplifiers along the links. These delays increase the latencies in the transfer of data through the network. Thus, it is desirable to minimize these as much as possible.

Research on the problem of minimizing the maximum eccentricity (i.e., the longest transmission time between any two pairs of nodes in a network) with fixed transmission times between nodes include Farley et al. (2000), McMahan and Proskurowski (2004), Wu (2004), and Chen et al. (2007). These works, however, assume that the transmission times are fixed and do not consider the possibility of delays in transmission due to congestion. When incorporating this uncertainty in transmission times into the mathematical programming model, the result is a stochastic programming problem.

In this chapter, we present the two-stage stochastic p-hub center problem (SpHCP),

which looks at constructing a fully-connected/star communication network with the objective of minimizing the expected maximum eccentricity (i.e., the longest path) in the network. The first-stage problem of this two-stage stochastic programming problem is to find the best location for p hubs in the network. Given the location of the hubs, and a realization of congestion in the network, the second-stage problem then is to assign each node to a single hub (i.e., single-assignment) so as to minimize the longest connection path in the network.

Given a realization of delay within the network, traffic originating from a user node destined for another will choose the path that minimizes the total transfer time. Thus, a node may be connected to a particular hub for some realizations of delay, while the same node may be connected to another hub for a different set of delay realizations, and so on. Unlike Kim et al. (1995), who explicitly require dual homing local connections, we do not specify the total number of hubs to which a node can be assigned. The solution to the SpHCP will indicate which hub will serve as the primary hub for the user node, which will be the secondary hub, and so on.

4.2 The Stochastic p-Hub Center Problem (SpHCP)

We formulate the SpHCP as a two-stage stochastic integer programming (SIP) problem. Let N be the set of all nodes in the network. The first-stage problem is to choose p nodes from the set of nodes N to be hubs. Once a realization of the uncertainties has been observed, the second-stage problem then is to assign nodes to hubs so as to minimize the longest path in the network.

Suppose that a unit of information (e.g., a packet) being transmitted along a link may or may not experience a transmission delay. Using the notation described in Section 1.3.1, we let $\omega \in \Omega$ represent a realization of a random event from the set of random events Ω. For our two-stage SpHCP, ω is an $|N| \times |N|$ matrix, where the ij-th entry in the matrix ω_{ij} indicates whether there is indeed a delay or not. For example, $\omega_{ij} = \{\text{Delay}, \text{No Delay}\}$, or $\omega_{ij} = \{\text{Available}, \text{Unavailable}\}$). Let $t_{ij}(\omega_{ij})$ be the transmission time between nodes i and j, which is dependent on ω_{ij}. (For notational simplicity, we will simply write $t_{ij}(\omega_{ij})$ as $t_{ij}(\omega)$.) The two-stage stochastic p-hub center problem with time delays on the links is

$$\text{S}p\text{HCP}: \quad \min \quad E_{\boldsymbol{\xi}} Q(\boldsymbol{X}, \xi(\omega)) \tag{4.1a}$$

$$\text{s.t.} \quad \sum_{k \in N} X_k = p, \tag{4.1b}$$

$$X_k \in \{0,1\}, \forall k \in N, \tag{4.1c}$$

where $\boldsymbol{X}$ is the column vector $[X_1 \dots X_N]^T$ and the second-stage problem $Q(\boldsymbol{X}, \xi)$ is

$$Q(\boldsymbol{X}, \xi(\omega)) = \quad \min \quad \beta \tag{4.2a}$$

$$\text{s.t.} \quad \beta \geq \sum_{k \in N} (t_{ik}(\omega) + \alpha t_{kl}(\omega)) Z_{ik} + t_{jl}(\omega) Z_{jl}, \forall i, j, l \in N, \tag{4.2b}$$

$$Z_{ik} \leq X_k, \forall i, k \in N, \tag{4.2c}$$

$$\sum_{k \in N} Z_{ik} = 1, \forall i \in N, \tag{4.2d}$$

$$Z_{ik} \in \{0,1\}, \forall i, k \in N. \tag{4.2e}$$

Objective function (4.1a) minimizes the expected longest path (in terms of transmission time) in the network, with constraint (4.1b) requiring that p hubs be located. The right-hand-side of constraints (4.2b) calculates the length of the path from node i to node j,

while constraints (4.2c) and (4.2d) ensure that each node is connected to exactly one open hub. Problem SpHCP is a stochastic integer programming problem with binary first- and second-stage decision variables.

Now, suppose instead that the transmission delays are observed only at the hubs. Let $d_k(\omega_k)$ represent the delay incurred at hub k due to event ω_k (for example, $\omega_k = \{\text{Low congestion}, \text{High congestion}\}$). Again, for simplicity, we simply write $d_k(\omega_k)$ as $d_k(\omega)$. Notice that the random event ω here is an $|N| \times 1$ vector. We then replace constraints (4.2b) with

$$\beta \geq \sum_{k \in N} (t_{ik} + \alpha t_{kl} + d_k(\omega)) Z_{ik} + (t_{jl} + d_l(\omega)) Z_{jl} - d_l(\omega) Z_{il}, \forall i, j, l \in N. \tag{4.3}$$

To show the correctness of constraints (4.3), we first note that any path in the network can be stated as a path from an origin node i to a destination node j that first visits hub k, then followed by hub l, i.e., $i \to k \to l \to j$. Next, consider the following five possible paths in the hub-and-spoke network:

- **Case 1:** All the nodes in the path are different, i.e., $i \neq k \neq l \neq j$. With nodes i and j assigned to hubs k and l, respectively (i.e., $Z_{ik} = Z_{jl} = 1$ and thus $Z_{il} = 0$), we have

$$t_{ik} + \alpha t_{kl} + t_{jl} + d_k(\omega) + d_l(\omega).$$

- **Case 2:** The path contains only a single hub node with $i \neq j$, i.e., $k = l$ and $i \to k \to j$. With both nodes i and j assigned to hub k, and assuming that travel time from a node to itself is zero, we have

$$(t_{ik} + \alpha t_{kk} + d_k(\omega)) + (t_{jk} + d_k(\omega)) - d_k(\omega) = t_{ik} + t_{jk} + d_k(\omega).$$

- **Case 3:** Similar to Case 2 but with the origin and destination nodes being the same node, i.e., $i = j$ and $i \rightarrow k \rightarrow i$. This scenario occurs when a customer sends a package to a recipient who is situated in the same service area (or zip code), but the parcel first has to be processed at the sorting hub before delivery to the recipient. We have

$$(t_{ik} + \alpha t_{kk} + d_k(\omega)) + (t_{ik} + d_k(\omega)) - d_k(\omega) = t_{ik} + t_{ik} + d_k(\omega).$$

- **Case 4:** The originating node is a non-hub node but the destination node is a hub node, i.e., $k = l = j$ and $i \rightarrow k$. We have

$$(t_{ik} + \alpha t_{kk} + d_k(\omega)) + (t_{kk} + d_k(\omega)) - d_k(\omega) = t_{ik} + d_k(\omega).$$

- **Case 5:** The originating node is a hub node but the destination node is a non-hub node, i.e., $i = k = l$ and $k \rightarrow j$. We have

$$(t_{kk} + \alpha t_{kk} + d_k(\omega)) + (t_{jk} + d_k(\omega)) - d_k(\omega) = t_{jk} + d_k(\omega).$$

Finally, if delay is observed both at the hubs and on the links, we would use

$$\beta \geq \sum_{k \in H} (t_{ik}(\omega) + \alpha t_{kl}(\omega) + d_k(\omega)) Z_{ik} + (t_{jl}(\omega) + d_l(\omega)) Z_{jl} - d_l(\omega) Z_{il}, \forall i, j, l \in N, \quad (4.4)$$

in place of constraints (4.2b).

4.3 Solution Methodology

Any two-stage stochastic programming problem can be written in the following general form:

$$\text{SP:} \quad \min_{x \in X} c^T x + \mathcal{Q}(x).$$

There are several approaches for solving a two-stage stochastic programming problem, with the most common being the (continuous) L-shaped method first introduced by Van Slyke and Wets (1969), or the integer L-shaped method of Laporte et al. (1994) for stochastic integer programming (SIP) problems. In the integer L-shaped method and other existing solution methods for SIP problems, the cost vector c of the first-stage variables x in problem SP is used to direct the optimization procedure in the search for the optimal values of the first-stage variables. In objective function (4.1a) of the SpHCP, however, the cost vector c is a zero vector. One could set $c = e$, where e is the vector of 1's, so that c is non-zero but this does little to the original problem SpHCP since $c^T x$ will always have a value of p in any feasible solution due to constraint (4.1b). Moreover, setting $c = e$ still does not provide any guidance to the solution procedure on how or where to search for the optimal solution $x^\star$. Thus, existing SIP solution methods are not particularly useful for the SpHCP, and a novel approach is required.

The remainder of this section is organized as follows: in Section 4.3.1, we take a further look at the second-stage problem. An oft-encountered difficulty in stochastic programming problems is the calculation of the recourse function due to the potentially numerous scenarios to be evaluated. Instead of calculating the recourse function exactly, we describe a Monte Carlo sampling method to approximate the recourse function in Section 4.3.2. Finally, the role of the master problem in the SpHCP can be thought of as a process of generating feasible hub configurations $\boldsymbol{X}$ with which the recourse function $\mathcal{Q}(\boldsymbol{X})$ is calculated. In Section 4.3.3, we discuss our observations of the optimal hub locations for several problem instances. We then propose a heuristic to generate good hub

configurations based on our observations.

4.3.1 The Second-Stage Problem: The p-Hub Center Single-Allocation Problem

Consider formulation (4.2) for the second-stage problem $Q(\boldsymbol{X}, \xi(\omega))$ of SpHCP with uncertain travel times. In this problem, it is assumed that the hub locations are known. Let H represent the set of hubs determined from the first-stage problem, i.e., $H = \{i \in N : X_i = 1\}$. We eliminate constraints (4.2c) as they are now redundant. For a given realization of ξ, the second-stage problem

$$
\begin{aligned}
Q(\boldsymbol{X}, \xi(\omega)) = \quad & \min \quad \beta \\
& \text{s.t.} \quad \beta \geq \sum_{k \in H} (t_{ik}(\omega) + \alpha t_{kl}(\omega)) Z_{ik} + t_{jl}(\omega) Z_{jl}, \forall i, j \in N, l \in H, \\
& \qquad \sum_{k \in H} Z_{ik} = 1, \forall i \in N, \\
& \qquad Z_{ik} \in \{0, 1\}, \forall i \in N, k \in H,
\end{aligned}
$$

is the *p-hub center single-allocation problem* (HCSAP). The second-stage problem incorporating delays at the hubs also reduces to the HCSAP for fixed first-stage variables and a given realization of ξ, as does the SpHCP with delays at the hubs and on the links. Ernst et al. (2006a) show that the HCSAP is NP-hard, while Campbell et al. (2007) show that several special cases of HCSAP can be solved in polynomial time.

For any $H \neq \varnothing$, there are finitely many ways to assign nodes to hubs. The longest path can then be determined from these assignments. With positive bounded transmission times and a fully-connected hub network, the path lengths are necessarily finite. Thus, problem SpHCP has complete recourse.

4.3.2 Evaluating the Recourse Function

Recall that the recourse function $\mathcal{Q}(\boldsymbol{X})$ is the expectation function of $Q(\boldsymbol{X}, \xi)$ for a given hub configuration $\boldsymbol{X}$. In statistical terms, $\mathcal{Q}(\boldsymbol{X})$ is the population mean with the population consisting of many $Q(\boldsymbol{X}, \xi)$ values.

Consider the recourse function (4.1a). Suppose that the number of scenarios (or realizations) for the SpHCP is finite, i.e., support $\Xi = \{\xi_1, \ldots, \xi_S\}$. Each realization $\xi_s, s = 1, \ldots, S$, has positive probability p_s, i.e., $P(\boldsymbol{\xi} = \xi_s) = p_s > 0$. The recourse function for a particular hub configuration $\boldsymbol{X}$ is calculated as

$$\mathcal{Q}(\boldsymbol{X}) = E_{\boldsymbol{\xi}} Q(\boldsymbol{X}, \boldsymbol{\xi}) = \sum_{s=1}^{S} p_s Q(\boldsymbol{X}, \xi_s). \tag{4.5}$$

Notice that for every feasible hub configuration $\boldsymbol{X}$, we have to solve S instances of the HCSAP to obtain $\mathcal{Q}(\boldsymbol{X})$. If the number of scenarios S is large, solving for $\mathcal{Q}(\boldsymbol{X})$ may be intractable even for just a single hub configuration $\boldsymbol{X}$. Instead of evaluating $Q(\boldsymbol{X}, \xi)$ for all possible values of ξ, we evaluate $Q(\boldsymbol{X}, \xi)$ on a sample of M ξ values; that is, we approximate the population mean $\mathcal{Q}(\boldsymbol{X})$ with its sample mean $\hat{\mathcal{Q}}_M(\boldsymbol{X})$. This is called the *sample average approximation* (SAA) approach. Shapiro (2003) discusses several theoretical properties of SAA, and how one could use the SAA to make statistical inferences. An overview of the use of Monte Carlo sampling methods for solving stochastic programming problems within the L-shaped method is covered in Birge and Louveaux (1997, Chapter 10), and a branch-and-bound scheme utilizing sample average approximations is discussed in Ahmed and Shapiro (2002). Further details of the SAA can be found in Kleywegt et al. (2001), with computational results provided in Verweij et al. (2003) and Linderoth et al. (2006).

For any realization ξ of the random variable $\boldsymbol{\xi}$, and for a given hub configuration $\boldsymbol{X}$, there is a corresponding $Q(\boldsymbol{X}, \xi)$. Thus, it can be seen that $Q(\boldsymbol{X}, \boldsymbol{\xi})$ is also a random variable because of $\boldsymbol{\xi}$. Next, suppose we generate a random sample $Q(\boldsymbol{X}, \xi_1), \ldots, Q(\boldsymbol{X}, \xi_M)$ from the distribution of $Q(\boldsymbol{X}, \boldsymbol{\xi})$. The mean of this sample is

$$\hat{\mathcal{Q}}_M(\boldsymbol{X}) = \frac{1}{M} \sum_{m=1}^{M} Q(\boldsymbol{X}, \xi_m).$$

In the statistics literature, it is commonly assumed that the statistic $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ is a function of a random sample $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M$ of random variables. So, we express the sample average $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ as a function of the independent and identically distributed random variables $\boldsymbol{\xi}_m$:

$$\hat{\mathcal{Q}}_M(\boldsymbol{X}) = \frac{1}{M} \sum_{m=1}^{M} Q(\boldsymbol{X}, \boldsymbol{\xi}_m). \tag{4.6}$$

Each random variable $\boldsymbol{\xi}_m, m = 1, \ldots, M$, has the same distribution as $\boldsymbol{\xi}$.

The sample average $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ is an unbiased and consistent estimator of $\mathcal{Q}(\boldsymbol{X})$ (Shapiro, 2003). Thus, we have a means of approximating $\mathcal{Q}(\boldsymbol{X})$ without resorting to evaluating $Q(\boldsymbol{X}, \xi_s)$ for all possible scenarios $\xi_s \in \Xi, s = 1, \ldots, S$. The estimator $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ is the *sample average approximation* for $\mathcal{Q}(\boldsymbol{X})$, and the recourse (or expectation) function is calculated based on an empirical distribution instead of the actual distribution of $Q(\boldsymbol{X}, \boldsymbol{\xi})$.

Finally, there is the question about the size of the sample that should be used to calculate the approximate recourse function $\hat{\mathcal{Q}}_M(\boldsymbol{X})$. Ideally, the sample size M should be large enough such that the sample average estimator $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ and the sample error are "stable" (Kleywegt et al., 2001). As an example, we randomly generated a hub configuration using the 25-node CAB data set with $p = 5$. Figure 4.1 plots the sample average estimator $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ against the number of samples M used to calculate $\hat{\mathcal{Q}}_M(\boldsymbol{X})$, while Figure 4.2

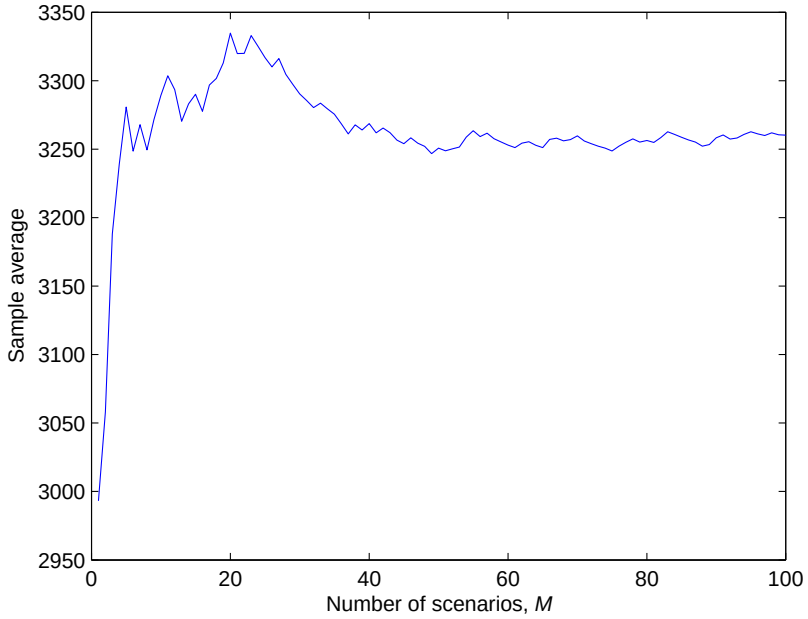


Figure 4.1: Plot of the sample average $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ as a function of the number of scenarios $M = 1, \ldots, 100$.

is a similar plot with the vertical axis representing the standard deciation of the sample for different sample sizes M. One notices that the sample average $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ stabilizes at approximately $M = 50$, while the standard error stabilizes when 30 to 40 scenarios are used. Therefore, a reasonable sample size for M is 50 scenarios for this hub configuration. Further tests with different problem instances support our choice of $M = 50$.

4.3.3 The First-Stage Problem

The role of the first-stage problem in the SpHCP can be considered as generating feasible hub configurations $\boldsymbol{X}$ from which the recourse function $\mathcal{Q}(\boldsymbol{X})$ is calculated. For a $|N|$ node problem with p hubs, there are $\binom{|N|}{p}$ possible feasible hub configurations.

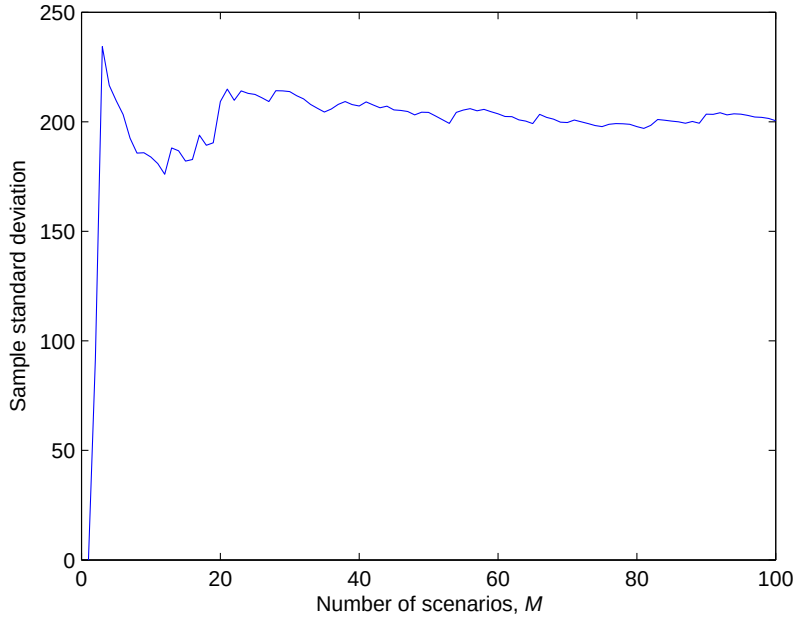


Figure 4.2: Plot of the sample standard deviation for sample sizes $M = 1, \ldots, 100$.

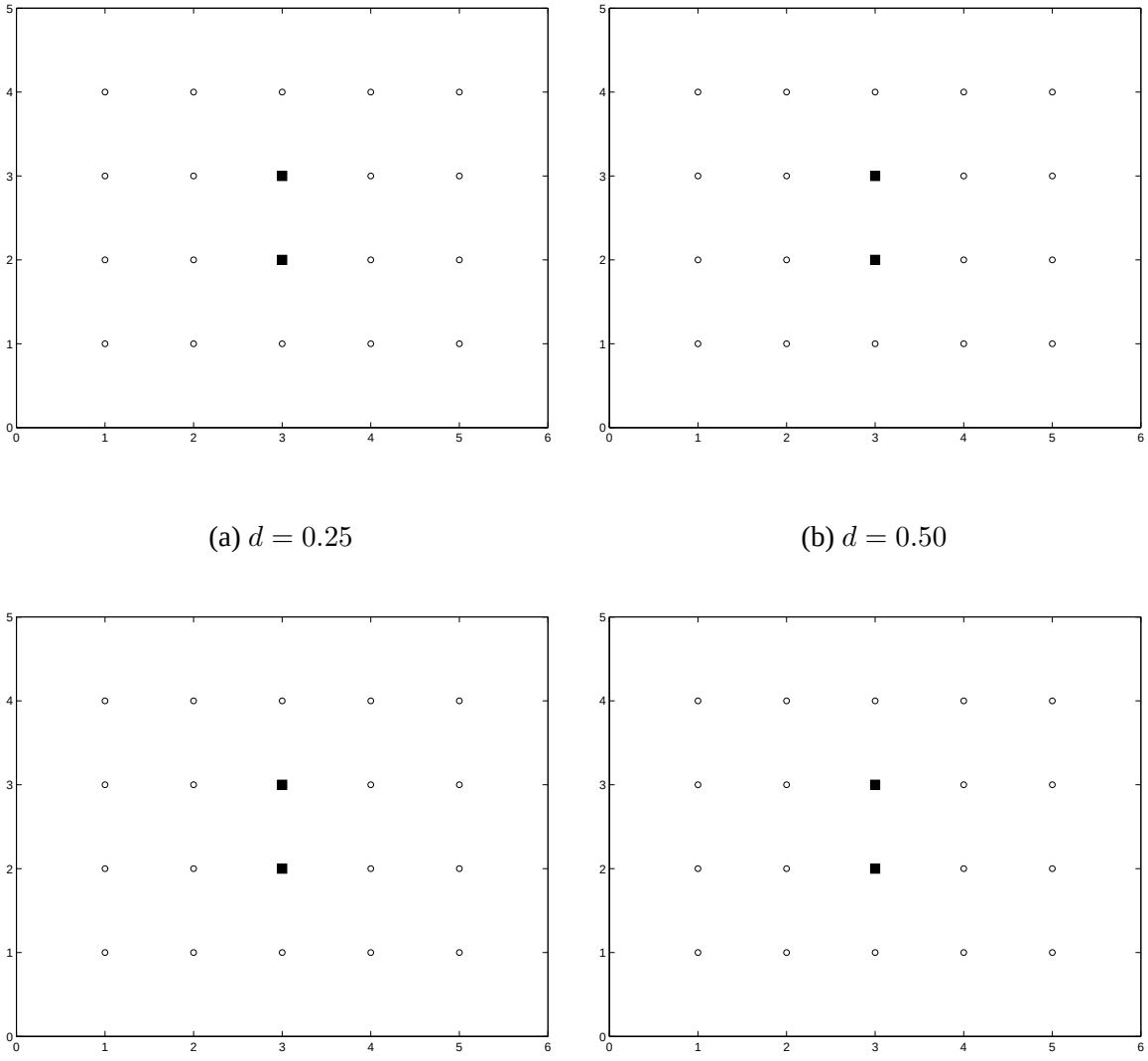
The task of finding the best hub configuration from among all these possible options is intractable, especially when one has to solve the NP-hard second-stage problem M times just to obtain a good approximation of the quality for a particular hub configuration. While there are many solution approaches that can be used for the first-stage problem (local search methods, tabu search, simulated annealing, genetic algorithms, etc.), we instead study the optimal solutions for small problems to see if there is a particular structure in the locations of the p hubs that we can use to design a heuristic to solve the problem.

We first consider the problem where the nodes (and thus, the candidate hub locations) are on a rectangular lattice. The problem has 20 nodes with 5 nodes along the horizontal axis and 4 nodes along the vertical axis. Using the same setup as described in Sec-

tion 4.2, the transmission time on a link $t_{ij}(\omega)$ is dependent on the random event ω_{ij}, where $\omega_{ij} = \{\text{Delay}, \text{No Delay}\}$. Delay on a link occurs with probability $0 < P(\text{Delay}) \leq 1$, and thus $P(\text{No Delay}) = 1 - P(\text{Delay})$. When there is no delay on a link, the transmission time $t_{ij}(\text{No Delay}) = t_{ij}$, where t_{ij} is the straight-line distance between nodes i and j. However, if a delay is experienced on a link, the transmission time is $t_{ij}(\text{Delay}) = (1 + d)t_{ij}$, where $d > 0$ is the relative increase in transmission time due to the delay. We thus call d the delay factor. We assume that the probability of experiencing a delay on a link is independent of that on all other links.

We generated problem instances using different combinations of the number of hubs p, the probability of delay $P(\text{Delay})$, and the delay factor d. We set $p = 2, 3, 4$, $P(\text{Delay}) = 0.25, 0.50, 0.75$, and $d = 0.25, 0.50, 0.75, 1.00$. For each problem instance $(p, P(\text{Delay}), d)$, we enumerated all possible hub configurations and for each hub configuration $\boldsymbol{X}$, solve the second-stage problem over $M = 50$ scenarios using CPLEX. The best hub configuration for a problem instance is the one that returns the lowest $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ value. Figures 4.3 to 4.5 are representative of the results from our computational experiments for the problem on the lattice. The hubs are generally spread out evenly around the lattice. In some problem instances, there is also a hub located in the center of the lattice.

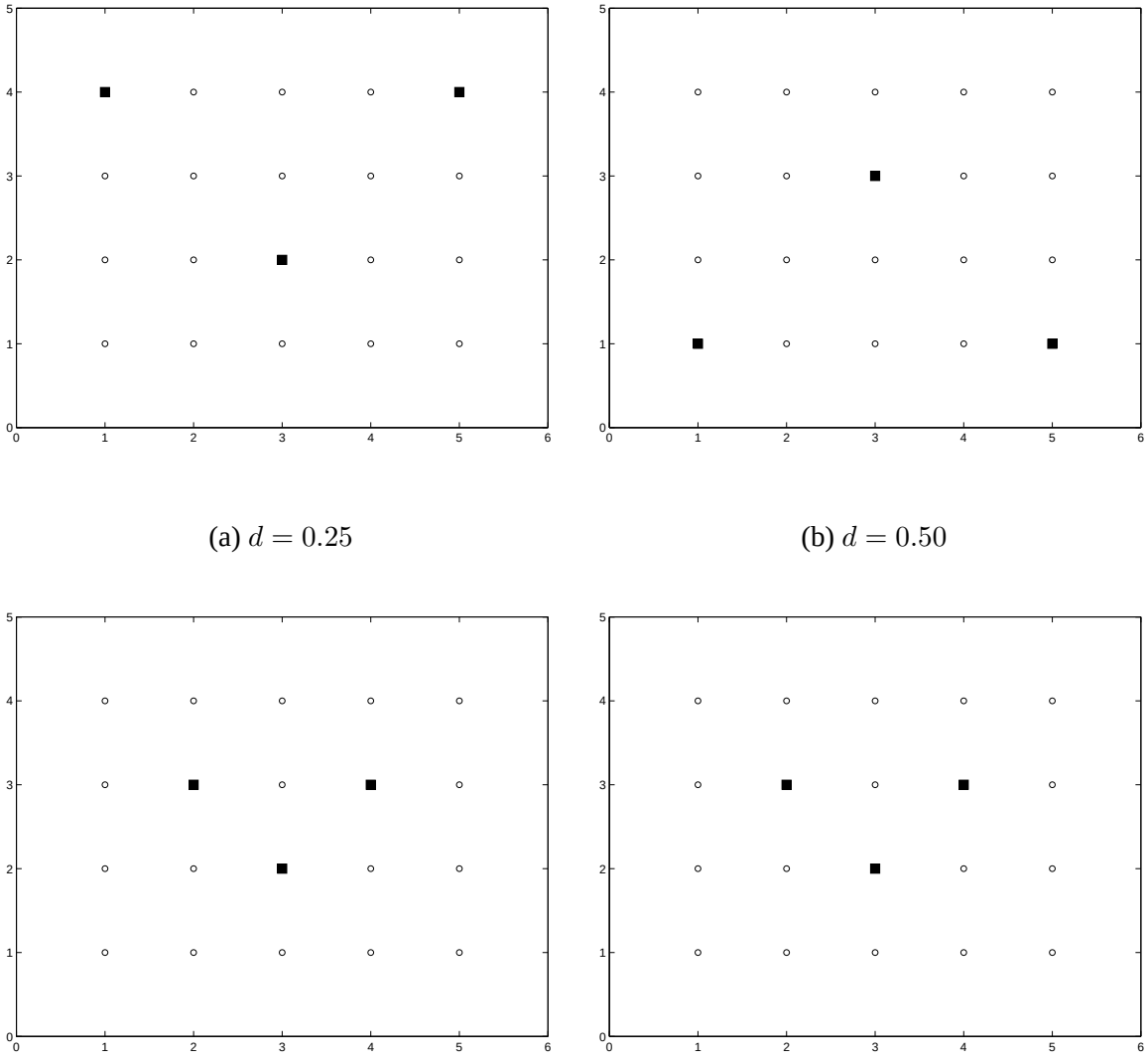
We also carried out the same test using the CAB and AP data sets, which are available from the OR-Library (Beasley, 2004). The results from these experiments also show the hubs spread evenly around the service region, and in some problem instances, there is a hub located close to the center of the service region. These observations of the optimal hub locations for the lattice, CAB, and AP data sets suggest that the radial heuristic solution



(a) $d = 0.25$ (b) $d = 0.50$

(c) $d = 0.75$ (d) $d = 1.00$

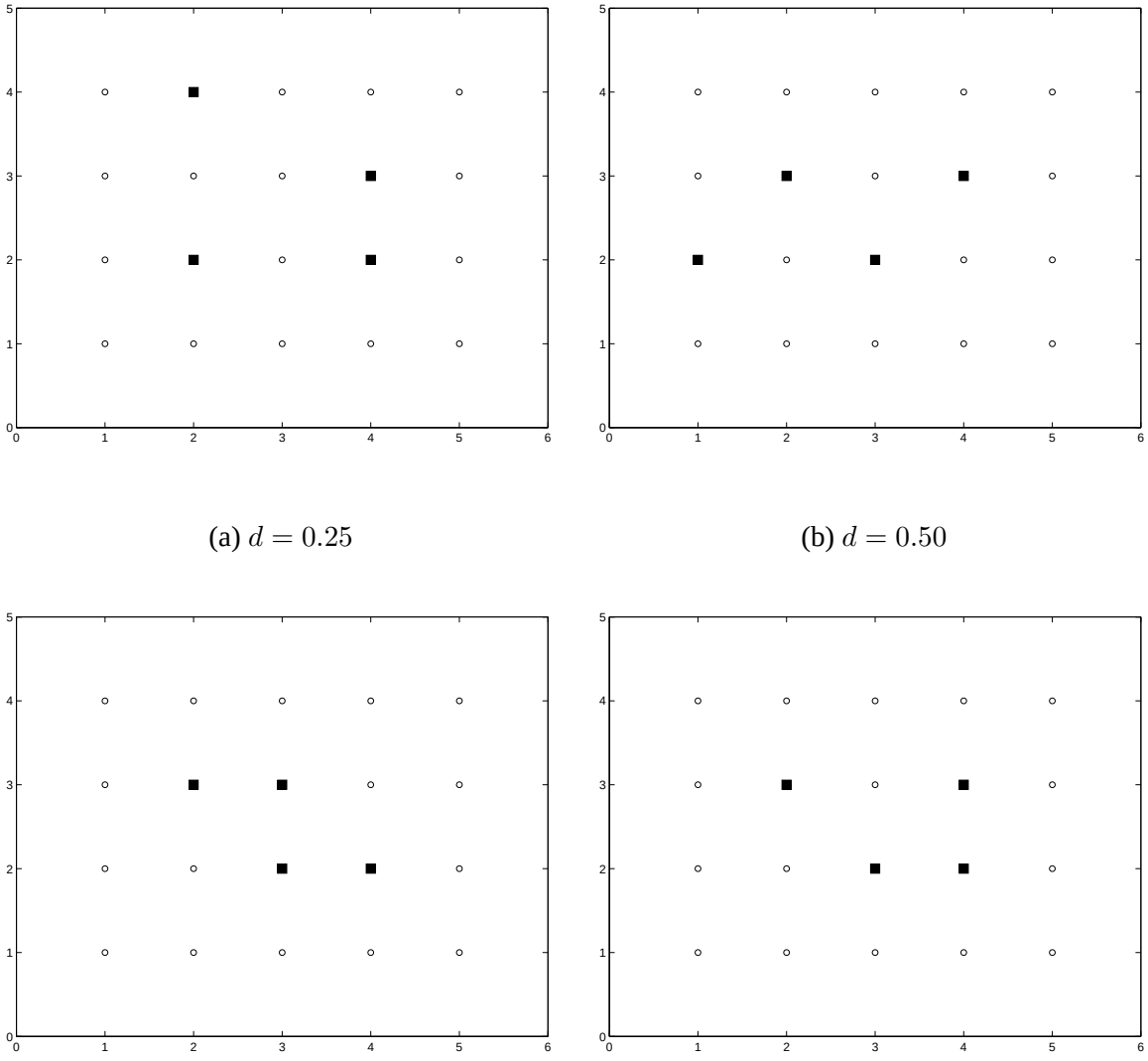
Figure 4.3: Best hub configurations for the 20-node two-stage SpHCP on the lattice with $p = 2$ hubs, probability of delay $P(\text{Delay}) = 0.25$, and varying values of the delay factor d.



(a) $d = 0.25$ (b) $d = 0.50$

(c) $d = 0.75$ (d) $d = 1.00$

Figure 4.4: Best hub configurations for the 20-node two-stage SpHCP on the lattice with $p = 3$ hubs, probability of delay $P(\text{Delay}) = 0.75$, and varying values of the delay factor d.



(a) $d = 0.25$ (b) $d = 0.50$

(c) $d = 0.75$ (d) $d = 1.00$

Figure 4.5: Best hub configurations for the 20-node two-stage SpHCP on the lattice with $p = 4$ hubs, probability of delay $P(\text{Delay}) = 0.25$, and varying values of the delay factor d.

approach first introduced in Chapter 3 will find good solutions for the two-stage SpHCP as it generates hub configurations with similar structures.

4.4 Computational Experiments

We use the CAB and AP data sets for the computational experiments of the two-stage SpHCP with random transmission times in this section. The random event $\omega_{ij} = \{\text{Delay}, \text{No Delay}\}$ is independent of other links in the network, and delay on a link occurs with probability $P(\text{Delay})$. A delay on a link increases the transmission time on that link by the delay factor $d > 0$, i.e., $t_{ij}(\text{Delay}) = (1 + d)t_{ij}$. For each node-hub $(|N|, p)$ combination, we set $P(\text{Delay}) = 0.2, 0.4, 0.6$, and $d = 0.25, 0.50, 0.75, 1.00$.

In Section 4.4.1, we present our radial heuristic, and discuss its performance. The heuristic is coded in MATLAB (The Mathworks Inc., 1997) using the CPLEX-MEX interface (Musicant, 2000) to call CPLEX to solve the second-stage problem. Section 4.4.2 discusses the benefits of solving the stochastic programming problem, which accounts for the stochastic nature of the problem, instead of simply solving the deterministic version of the problem using the expected values of the stochastic parameters in the optimization problem.

4.4.1 Radial Heuristic

Our radial heuristic is an iterative heuristic. At each iteration of the heuristic, we first locate a central hub, and then locate the remaining $p - 1$ hubs around the central hub. The outline of the radial heuristic is provided in Figure 4.6, and Figure 4.7 shows the steps to construct a hub configuration for the radial heuristic.

Step 1: Set $C \leftarrow N$, and set $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ to be some large number.

Step 2: If $|C| \geq p$, generate a hub configuration $\boldsymbol{X}(C)$ based on the set C (see Figure 4.7). Else, go to step 4.

Step 3: Calculate $\hat{\mathcal{Q}}_M(\boldsymbol{X}(C))$.

If $\hat{\mathcal{Q}}_M(\boldsymbol{X}(C)) < \hat{\mathcal{Q}}_M(\boldsymbol{X})$, set $\boldsymbol{X}^\star \leftarrow \boldsymbol{X}(C)$ and $\hat{\mathcal{Q}}_M(\boldsymbol{X}) \leftarrow \hat{\mathcal{Q}}_M(\boldsymbol{X}(C))$.

Set $C \leftarrow C \setminus \arg\max_{i \in C}\{t_{iH_1}\}$, and go to step 2. ($H_1$ is the 1-center hub.)

Step 4: Return $\boldsymbol{X}^\star$ and $\hat{\mathcal{Q}}_M(\boldsymbol{X})$.

Figure 4.6: Radial heuristic.

% *Find the central hub*

Set the incumbent travel time value $T^\star$ to some large number.

Let H_1 represent the index of the optimal central hub.

For each node $h \in C$

Let $T = \max_{i \in C}\{t_{ih}\}$.

If $T < T^\star$, set $T^\star \leftarrow T$ and $H_1 \leftarrow h$.

% *Locate remaining* $p - 1$ *hubs*

Set $\boldsymbol{X} \leftarrow H_1$.

While $|\boldsymbol{X}| \leq p$

For each node $i \in C \setminus \{\boldsymbol{X}\}$, set $T_i \leftarrow \min_{j \in \boldsymbol{X}}\{t_{ij}\}$.

Set $i^\star \leftarrow \arg\max_{i \in C \setminus \{\boldsymbol{X}\}}\{T_i\}$, and $\boldsymbol{X} \leftarrow \boldsymbol{X} \cup \{i^\star\}$.

Return $\boldsymbol{X}$ as the initial hub configuration.

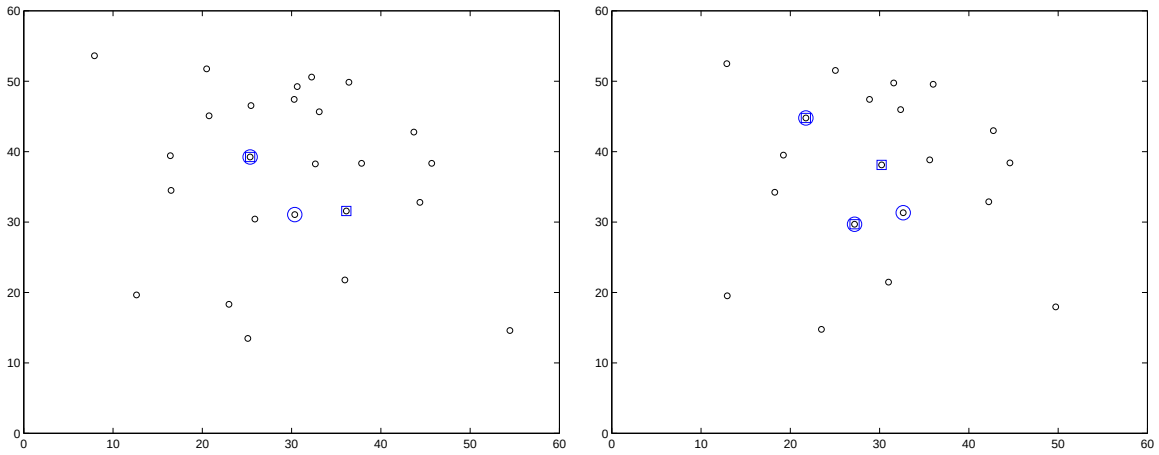
Figure 4.7: Construction heuristic to generate a hub configuration.

We compare the best $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ value found using the radial heuristic against the best $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ value found by enumerating all possible hub configurations. (We will refer to the solution obtained from the complete enumeration as simply the optimal solution.) A t-test on the null hypothesis of whether the sample means $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ of the two samples are equal will be used. Table 4.1 shows the number of problem instances for each $(|N|, p)$ combination where the t-test results in the rejection of the null hypothesis at a 5% significance level. (Note that there are 12 problem instances for each $(|N|, p)$ combination because we set $P(\text{Delay}) = 0.2, 0.4, 0.6$, and $d = 0.25, 0.50, 0.75, 1.00$.) The radial heuristic performs well for the CAB data set, but its performance for the AP data set leaves much to be desired. When we compare the hub locations between the optimal solution and that found by the radial heuristic, we noticed that in many problem instances, if we were to swap a hub with its closest non-hub node, we would arrive at the same hub locations as the optimal solution. An example of this result is shown in Figure 4.8. In Figure 4.8(a), the radial heuristic locates one of the two hubs at the optimal hub location. If we swapped the second hub located by the radial heuristic with its closest non-hub node, we end up with the optimal solution for this problem instance. In Figure 4.8(b), a hub may need to be swapped with its second closest non-hub node to obtain the optimal hub configuration. These two examples indicate that incorporating a simple improvement heuristic into the radial heuristic is a reasonable approach towards finding the optimal solution.

We implement the improvement heuristic as follows: at each iteration of the radial heuristic, the improvement heuristic swaps the hub nodes with their closest non-hub nodes one at a time to generate new hub configurations. For example, suppose the hub configu-

Table 4.1: Total number of problem instances where the null hypothesis (the sample mean of the optimal solution is equal to the best sample mean found using the radial heuristic) is rejected (5% significance level).

$\|N\|$	p	CAB data set	AP data set
10	2	1 out of 12	0 out of 12
	3	8 out of 12	3 out of 12
	4	10 out of 12	3 out of 12
	5	5 out of 12	4 out of 12
15	2	4 out of 12	
	3	7 out of 12	
20	2	4 out of 12	3 out of 12
	3	6 out of 12	9 out of 12
25	2	0 out of 12	2 out of 12
	3	12 out of 12	3 out of 12
Total		57 out of 120	29 out of 96



(a) $|N| = 25, p = 2, P(\text{Delay}) = 0.6, d = 0.75$.

(b) $|N| = 20, p = 3, P(\text{Delay}) = 0.2, d = 0.75$.

Figure 4.8: Plot of the hub locations of the optimal solution (indicated by the larger circles) and that found by the radial heuristic (indicated by the square icons) using the AP data set.

ration generated at the current iteration for a 3-hub problem has hubs located at nodes 2, 8, and 12. Let non-hub nodes 3, 9, and 13 be the closest non-hub nodes for the three hubs, respectively. The improvement heuristic will then create three new hub configurations: (3,8,12), (2,9,12), and (2,8,13). We also extended this improvement heuristic to swap a hub node with its two closest non-hub nodes, thus generating six new hub configurations at each iteration of the radial heuristic. We refer to the first improvement heuristic as Improve1, and the second as Improve2.

We see from Table 4.2 that using the Improve1 heuristic together with the radial heuristic resulted in a significant reduction in the total number of rejections of the null hypothesis for the AP data set, while it only had a minimal affect on the CAB data set. However, when the Improve2 heuristic is used, we see a significant reduction in the total

Table 4.2: Total number of problem instances where the null hypothesis (the sample mean of the optimal solution is equal to the best sample mean found using a combination of the radial and improvement heuristics) is rejected (5% significance level).

		CAB data set		AP data set	
$\|N\|$	p	Improve1	Improve2	Improve1	Improve2
10	2	0 out of 12	0 out of 12	0 out of 12	0 out of 12
	3	6 out of 12	0 out of 12	3 out of 12	1 out of 12
	4	4 out of 12	0 out of 12	3 out of 12	3 out of 12
	5	2 out of 12	2 out of 12	0 out of 12	0 out of 12
15	2	0 out of 12	0 out of 12		
	3	3 out of 12	1 out of 12		
20	2	0 out of 12	0 out of 12	0 out of 12	0 out of 12
	3	0 out of 12	0 out of 12	2 out of 12	1 out of 12
25	2	0 out of 12	0 out of 12	2 out of 12	0 out of 12
	3	12 out of 12	5 out of 12	1 out of 12	1 out of 12
Total		27 out of 120	8 out of 120	11 out of 96	6 out of 96

number of rejects for both data sets.

4.4.2 Value of the Stochastic Solution

A stochastic programming problem, compared to a deterministic linear programming problem, is usually more difficult to solve. Thus, it is common to question whether it is worth the considerable effort to model and solve a stochastic programming problem. An answer to this question is to consider the *value of the stochastic solution*. For our two-stage SpHCP, this involves comparing the optimal SpHCP solution against the *expected result of using the expected value solution* (EEV). The EEV is obtained by first solving the *expected (or mean) value problem*, which is the problem obtained by setting the random variables to their expected values. From the solution to this deterministic optimization problem, we fix the values of the first-stage variables and solve the second-stage problem over all possible scenarios. However, we will solve the second-stage problem using a sample of $M = 50$ scenarios to keep the process tractable. The objective function values from the 50 expected value solutions are then averaged to obtain the EEV. The difference between the EEV and the solution to the stochastic programming problem SpHCP is called the *value of the stochastic solution* (VSS = EEV - SpHCP). The VSS measures how good (or bad) a decision based on the solution to the expected value problem is when compared to the decision based on solving the stochastic programming problem (Birge and Louveaux, 1997).

For each $(|N|, p)$ combination, we still have 12 different problem instances as described in Section 4.4.1. Table 4.3 shows the average, minimum, and maximum VSS/EEV values over the 12 problem instances for each $(|N|, p)$ combination. The VSS/EEV value

Table 4.3: Average, minimum, and maximum VSS/EEV percentage values, where the VSS/EEV indicates the relative improvement in the sample average $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ if the stochastic programming approach is used instead of the expected value approach.

		CAB data set			AP data set		
$\lvert N\rvert$	p	Average	Minimum	Maximum	Average	Minimum	Maximum
10	2	4.2%	0.0%	9.7%	0.7%	0.0%	5.4%
	3	2.3%	0.0%	8.7%	6.7%	0.0%	16.9%
	4	1.1%	0.0%	4.3%	3.9%	0.3%	9.8%
	5	3.1%	0.0%	8.5%	0.1%	0.0%	0.5%
15	2	5.2%	0.0%	17.6%	0.0%	0.0%	0.0%
	3	6.6%	0.0%	15.1%	0.0%	0.0%	0.0%
20	2	2.0%	0.0%	11.8%	7.0%	0.0%	20.1%
	3	2.3%	0.0%	7.5%	11.7%	2.1%	22.9%
25	2	0.5%	0.0%	2.6%	9.0%	1.0%	22.3%
	3	3.3%	0.0%	9.5%	10.1%	0.5%	24.4%

provides an indication of the relative improvement in the sample average value $\hat{\mathcal{Q}}_M(\boldsymbol{X})$ if we use the stochastic programming approach instead of the deterministic expected value approach. We see that in some instances, the expected value approach obtains similar results to the stochastic programming approach (i.e., 0% values in the 'Minimum' columns in Table 4.3). However, there are some instances where there is a huge benefit to using the stochastic programming approach, with a possible improvement of over 15% and 20% for the CAB and AP data sets, respectively. Hence, there is certainly some benefit to using a stochastic programming approach for this problem.

Finally, we notice that the minimum VSS/EEV values for some $(|N|, p)$ combinations in Table 4.3 are 0%. This indicates that the EEV solution is equal to the SpHCP solution for some problem instances. Then there are those $(|N|, p)$ combinations that have minimum VSS/EEV values close to 0% – indicating that the EEV solution is close to the SpHCP solution. This observation suggests that an improvement heuristic applied to the EEV hub configuration may be able to find the SpHCP solution. We explore this possibility using the Improve2 heuristic described in the previous section on the EEV solution. Using the Improve2 heuristic also allows us to compare the performance of this EEV and Improve2 solution approach to that of the radial and Improve2 heuristic.

Table 4.4 shows the number of problem instances for each $(|N|, p)$ combination where the t-test results in the rejection of the null hypothesis that the sample mean of the optimal solution is equal to that found by the EEV+Improve2 solution at the 5% significance level. The EEV+Improve2 approach works well for the CAB data set (low reject counts), but not as well for the AP data set. However, the EEV+Improve2 is not as good as

the radial with Improve2 heuristic as it has more rejects.

4.5 Large-Scale Stochastic p-Hub Center Problems: A Simulated Annealing Solution Approach

The size of the problem instances used in our experiments in Section 4.4 are considered to be small (up to 25 nodes). Many realistic hub-and-spoke networks have hundreds or thousands of nodes. From computational experience, CPLEX is able to solve the second-stage problem (i.e., the hub center single-allocation problem or HCSAP) for the largest AP data set (200 nodes). We do note, however, that the computer memory requirements to load this problem size as well as to maintain the branch-and-bound tree is significant (close to exhausting the 4GB of available memory in the test computer). Thus, to solve much larger problems, other solution methods are required. In this section, we propose a simulated annealing (SA) approach for solving large-scale instances of the HCSAP.

Annealing is the process where molten material is cooled slowly in a heat bath into a solid with minimal energy (Pirlot, 1996). When applied to an optimization problem, the simulated annealing algorithm takes a proposed feasible solution obtained from the neighborhood of the current solution and calculates its objective function value. If the proposed solution improves the objective function value, the proposed solution is accepted. If, however, the proposed solution degrades the objective function value, the proposed solution is accepted with a probability value inversely proportional to the degradation of the objective function value. The aim of the algorithm is to converge to the globally optimum solution when the process reaches the "cooled state." Metropolis et al. (1953) first introduced the

Table 4.4: Total number of problem instances where the null hypothesis (the sample mean of the optimal solution is equal to the best sample mean found using the EEV+ Improve2 solution approach) is rejected (5% significance level).

$\|N\|$	p	CAB data set	AP data set
10	2	0 out of 12	0 out of 12
	3	1 out of 12	0 out of 12
	4	0 out of 12	3 out of 12
	5	0 out of 12	0 out of 12
15	2	3 out of 12	
	3	1 out of 12	
20	2	2 out of 12	3 out of 12
	3	0 out of 12	12 out of 12
25	2	0 out of 12	6 out of 12
	3	4 out of 12	8 out of 12
Total		11 out of 120	32 out of 96

concept behind simulated annealing, and Kirkpatrick et al. (1983) and Černý (1985) independently implemented this idea for solving combinatorial optimization problems such as the traveling salesman problem.

Implementing the SA algorithm requires, at a minimum, a specification of the objective (or evaluation) function and the neighborhood of the current solution. In the HCSAP, the objective function is simply the longest transmission path in the network. A neighbor for the current solution (i.e., the allocation of nodes to hubs) can be obtained by swapping the assignment of a node from its current hub to another hub. The evaluation of a proposed solution can be straightforwardly performed in $\mathcal{O}(|N|^2 log|N|)$ time (Fredman and Tarjan, 1987). A more efficient approach with better worst-case complexity can be devised by simply keeping track of the farthest node connected to each hub. Consider a current solution to the HCSAP. For each hub H_k, let $i_k^{\max}$ be the node connected to hub k that is farthest away from it, i.e.,

$$i_k^{\max} = \arg\max_{i \in N} \{t_{ik} : Z_{ik} = 1\}.$$

Suppose we re-assign node j from hub H_1 to hub H_2. If node $j \neq i_1^{\max}$ and $j \leq i_2^{\max}$, then such a swap will not affect the objective function value β. If $j = i_1^{\max}$ or $j > i_2^{\max}$, then we first need to update the values of $i_1^{\max}$ or $i_2^{\max}$. Once the updating is completed, we then need to evaluate

$$\max_{k \in H} \left\{ t_{i_1^{\max}, H_1} + \alpha t_{H_1, H_k} + t_{i_k^{\max}, H_k} \right\} \quad \text{and} \quad \max_{k \in H} \left\{ t_{j, H_2} + \alpha t_{H_2, H_k} + t_{i_k^{\max}, H_k} \right\}$$

to obtain the length of the longest path $\hat{\beta}$ in the proposed network. Each of the two equations requires $|H| = p$ evaluations. This step dominates the entire evaluation process, and

thus, the evaluation of a proposed feasible solution takes $\mathcal{O}(p)$ time.

The SA algorithm also requires an initial solution, an initial temperature value T_0, the loop length L, and the temperature reduction rate τ. There is much literature on how to set these parameter values such as Johnson et al. (1989), Kirkpatrick et al. (1983), Pirlot (1996), and van Laarhoven and Aarts (1987). In our implementation of the SA algorithm, we construct an initial solution to the assignment problem by simply assigning each node to its closest hub. We then use this initial solution to derive the initial temperature T_0 as follows: from this initial solution, we randomly generate a neighbor using the process described in the previous paragraph. If the objective function value $\hat{\beta}$ of this neighbor is greater than that of the initial solution (i.e., the neighbor is a worse solution), we keep track of the difference ϵ between the objective function values. If the neighbor is an improving solution (i.e., lower $\hat{\beta}$ value), we toss it out. We repeat this process of finding bad neighbors of the initial solution until we have 50 of these neighbors. The initial temperature T_0 is then calculated by

$$T_0 = \frac{\hat{\epsilon}}{\ln\left(\chi^{-1}\right)},$$

where $\hat{\epsilon}$ is the average deterioration from the initial starting solution (i.e., the difference in objective function values), and χ is the probability of accepting deteriorating moves at the start of the simulated annealing algorithm (van Laarhoven and Aarts, 1987). In our experiments, we set $\chi = 0.8$. At each temperature level T, we evaluate $L = 25000$ neighbors to the current solution, and we reduce the temperature by $100(1-\tau)\%$ following the completion of the loop with τ set at 0.95. We terminate the SA process when the best solution found has not changed in 100 temperature reductions. Figure 4.9 gives the

Generate an initial solution Z_0 by assigning all nodes to its closest hub.

Calculate β from Z_0, and set current solution $Z \leftarrow Z_0$.

Calculate initial temperature T_0 using Z_0, and initialize temperature $T \leftarrow T_0$.

While not yet converged, For 1 to loop size L

Generate a random neighbor $\hat{Z}$.

Let $\Delta = \hat{\beta} - \beta$.

If $\Delta \leq 0$, set $Z \leftarrow \hat{Z}$, else set $Z \leftarrow \hat{Z}$ with probability $e^{-\Delta/T}$.

Set $T \leftarrow \tau T$.

Return the best solution found.

Figure 4.9: Simulated Annealing.

pseudocode for our implementation of the simulated annealing algorithm.

We tested the SA algorithm using both the CAB and AP data sets. For each $(|N|, p)$ combination, we first generate 10 different hub configurations. For each of these 10 hub configurations, we run the algorithm 10 times. Table 4.5 shows the performance of the SA algorithm for the 200-node AP data set with $p = 5$. (Similar tables of results for the smaller-sized CAB and AP data sets are provided in Appendix A.) The column headings in the tables are:

- Config.: Configuration number (1 to 10)
- Opt.: The optimal objective function value from CPLEX
- Min.: The minimum objective function value found by the simulated annealing algo-

rithm over the 10 runs

- Min. % gap: The percentage optimality gap between the minimum objective function value found by the simulated annealing algorithm over the 10 runs and the optimal objective function value
- Avg.: The average objective function value over the 10 runs of the simulated annealing algorithm
- Avg. % gap: The percentage optimality gap between the average objective function value and the optimal objective function value, and
- St. Dev: The standard deviation of the objective function value over the 10 runs of the simulated annealing algorithm.
- Run-time (CPLEX): Run-time for CPLEX in seconds
- Run-time (SA): Total run-time of the SA algorithm in seconds (10 runs)

The results show that the SA algorithm was able to find the optimal solution for 2 out of the 10 different hub configurations, with varying degrees of success on the remaining 8 hub configurations. We have not been able to identify a systematic reason for why the SA works well for some problem instances but not for others. As stated earlier, CPLEX requires a significant amount of memory to solve this 200-node problem. The SA algorithm, on the other hand, uses only 2% to 3% of available memory.

Table 4.5: Simulated annealing results for the 200-node AP data set with $p = 5$.

	Objective function value $\hat{\beta}$						Run-time (sec)	
Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.	CPLEX	SA
1	78276	87684	12.0%	87684	12.0%	0	507	1307
2	71732	92457	28.9%	92457	28.9%	0	147	1308
3	70498	70498	0.0%	70498	0.0%	0	463	1304
4	70484	72808	3.3%	72808	3.3%	0	1481	1304
5	76208	97418	27.8%	97418	27.8%	0	216	1310
6	75472	79831	5.8%	93043	23.3%	4642	156	1307
7	78859	84496	7.1%	84496	7.1%	0	4534	1313
8	71732	76694	6.9%	76694	6.9%	0	528	1304
9	75292	87174	15.8%	87174	15.8%	0	157	1312
10	83950	83950	0.0%	83950	0.0%	0	124	1310

4.6 Conclusion

In this chapter, we introduce the two-stage stochastic p-hub center problem motivated by the communications industry. The SpHCP seeks to find the best hub locations such that the expected value of the longest transmission time in the network is minimized. From observations of the optimal solution to the SpHCP, we notice that the optimal configuration of the p hubs tends to have a hub located centrally with the remaining $p-1$ hubs scattered around this central hub. We thus considered using the radial heuristic, which generates solutions that exhibit this particular structure, to direct the search for the optimal hub locations. Our implementation of the radial heuristic in this chapter includes a simple improvement heuristic. Computational results show that our radial heuristic performs well in that it provides solutions that are not significantly different from the optimal solutions. We use the sample average approximation approach to evaluate the recourse function. The second-stage problem is solved via CPLEX, and we propose a simulated annealing approach for solving large problem instances.

CHAPTER 5
CONCLUSIONS

In this dissertation, we study three hub location problems motivated by their use in different application areas: the air travel, small package delivery, and communication networks. The hub location problem we study assumes that the hub nodes are fully connected, and that the non-hub nodes are connected to exactly one hub (single-assignment rule) or to at least one hub (multiple-assignment rule).

In Chapter 2, we introduce the hub covering flow problem (HCFP), which is motivated by the air travel industry. We observe that travel on the spokes (i.e., the link between the hub and non-hub nodes) are often performed using small-sized aircrafts, which have limited maximum flying ranges. These maximum flying ranges are commonly referred to as coverage constraints in the location literature. Furthermore, there are often many different aircraft types operating out of a hub, each with different operating costs and maximum flying ranges. The choice of which aircraft type to use on a link is an important consideration when planning or designing an airline's hub-and-spoke network. The hub covering flow problem looks at finding the best location of hubs and the assignment of non-hub nodes to hubs while satisfying the coverage constraints so as to minimize the total cost of opening hubs and transferring demand through the network.

We propose two different formulations for the HCFP. The first formulation models the coverage requirement as a constraint – a similar approach to the set covering problem in the location theory literature. The second formulation incorporates the coverage requirement as a penalty function in the objective function. This second formulation is

then extended to include the multiple aircraft type issue, which makes the model more closely reflect reality. It also has an additional benefit of providing a solution to the fleet assignment problem via the optimal assignments of non-hub nodes to hubs, and it aids in decision-making by quantifying the costs and listing the characteristics of optimal network designs for different fleet combinations.

Chapter 3 considers hub-and-spoke networks in the small package delivery industry. Expedited delivery services form a significant revenue stream for many companies in this industry. These time-definite delivery service requires companies to ensure delivery of a package within a specified time-frame. With possible delays in transporting a package from its origin to destination, a company must ensure that their delivery network accounts for this possibility. We introduce the chance-constrained stochastic p-hub center problem, which seeks to configure a network that minimizes the longest transportation time in the network for a specified service level in delivery time.

We find that the optimal location of the p-hubs tend to form a certain structure, where one hub is located in the center of the service region with the remaining hubs located around this central hub. Also, when the level of uncertainty in the network increases (e.g., an increase in the variability in the transportation times), the surrounding hubs exhibit a clustering effect where they tend to be located closer to the central hub. We propose our radial heuristic that seeks to exploit this particular structure, and use the best hub configuration found by this heuristic as a starting point for the one-opt best-improvement heuristic proposed by Teitz and Bart (1968). We compare the performance of this solution approach to the optimal solution where possible, and our computational results show that this solu-

tion methodology returns hub configurations that are optimal or close to optimal. We then use this heuristic to solve larger problem instances.

The final application area that we study in this dissertation is communication networks. Two service quality metrics used in the communications industry are robustness and response time. Robustness relates to the ability of the network to continue functioning when a portion of the network is unavailable due to failure. Response time refers to the amount of time required to transfer information through the network to the final destination. Chapter 4 discusses the two-stage stochastic p-hub center problem, which accounts for the possible unavailability of network components or delays experienced during data transmission.

We found that the optimal hub locations often exhibit a structure where there is a central hub with the remaining $p-1$ hubs surrounding the central hub. This is similar to what was observed in the chance-constrained stochastic p-hub center problem. We use the same radial heuristic to guide the search for the optimal hub locations. We approximate the recourse function using a Monte Carlo technique called sample average approximation (Shapiro, 2003). We solve the second stage assignment problem using CPLEX, and we propose a simulated annealing algorithm for large data sets.

There are several possible avenues for future work arising from this dissertation. With regards to the hub covering flow problem, we envision using it to study current trends in the airline industry; in particular, competition between the legacy airlines and low-cost carriers. Examples of this are the works of Adler (2005) and Adler and Smilowitz (2007).

For the chance-constrained stochastic p-hub center problem, an interesting corre-

sponding problem is that of what maximum service level can be achieved for a given maximum path length value β for the β-path. This can be stated as the following optimization problem:

$$\max \quad \gamma \tag{5.1a}$$

$$\text{s.t.} \quad P\left(\sum_{k\in N}(t_{ik}(\omega)+\alpha t_{kl}(\omega))Z_{ik}+t_{jl}(\omega)Z_{jl}\leq\beta\right)\geq\gamma,\forall i,j,l\in N, \tag{5.1b}$$

$$\sum_{k\in N}Z_{kk}=p, \tag{5.1c}$$

$$Z_{ik}\leq Z_{kk},\forall i,k\in N, \tag{5.1d}$$

$$\sum_{k\in N}Z_{ik}=1,\forall i\in N, \tag{5.1e}$$

$$Z_{ik}\in\{0,1\},\forall i,k\in N. \tag{5.1f}$$

Problem (5.1) is basically the stochastic p-hub covering problem with chance constraints.

Another research project closely related to the chance-constrained stochastic p-hub center problem is the (absolute) robust p-hub center problem, which finds the best location of the p hubs and assignment of nodes to hubs so that the length of the longest path β is minimal over *all* scenarios. By setting $\gamma = 1$ in the SpHCP-CC (problem (3.1)), constraints (3.1b) must hold almost surely. In problem SpHCP-CEP, constraints (3.6b) are now redundant and constraints (3.6a) become

$$\sum_{k\in N}(t^s_{ik}+\alpha t^s_{kl})Z_{ik}+t^s_{jl}Z_{jl}\leq\beta,\forall i,j,l\in N,s\in S.$$

Problem SpHCP-CEP then simplifies to

$$\min \quad \beta$$

$$\text{s.t.} \quad \sum_{k\in N}(t^s_{ik}+\alpha t^s_{kl})Z_{ik}+t^s_{jl}Z_{jl}\leq\beta,\forall i,j,l\in N,s\in S, \tag{5.2a}$$

$$\sum_{k \in N} Z_{kk} = p, \tag{5.2b}$$

$$Z_{ik} \leq Z_{kk}, \forall i, k \in N, \tag{5.2c}$$

$$\sum_{k \in N} Z_{ik} = 1, \forall i \in N, \tag{5.2d}$$

$$Z_{ik} \in \{0, 1\}, \forall i, k \in N. \tag{5.2e}$$

Problem (5.2) is essentially the p-hub center problem with additional constraints on the paths in the network generated by the different scenarios $s \in S$. Therefore, if we can solve pHCP, we can solve problem (5.2). Further discussion on robust optimization problems can be found in Kouvelis and Yu (1996), including the two variants of the minimax regret problems: the *robust deviation* problem and the *relative robust* problem.

Finally, further work is required on the simulated annealing algorithm, which we proposed for the hub center single-assignment problem. In particular, robust parameter settings should be derived so that the algorithm will perform well for any problem instance. We will use the design of experiments approach suggested by Coy et al. (2000).

APPENDIX A
SIMULATED ANNEALING RESULTS

The tables on the next few pages show the computational results from the simulated annealing experiments using the CAB and AP data sets. For each $(|N|, p)$ problem instance, where $|N|$ are the number of nodes and hubs, respectively, we randomly generated 10 different hub configurations. For each of the 10 hub configurations, we implement 10 runs of the simulated annealing heuristic. The column headings in the tables that follow are:

- Config.: Configuration number (1 to 10)
- Opt.: The optimal objective function value from CPLEX
- Min.: The minimum objective function value found by the simulated annealing algorithm over the 10 runs
- Min. % gap: The percentage optimality gap between the minimum objective function value found by the simulated annealing algorithm over the 10 runs and the optimal objective function value
- Avg.: The average objective function value over the 10 runs of the simulated annealing algorithm
- Avg. % gap: The percentage optimality gap between the average objective function value and the optimal objective function value, and
- St. Dev: The standard deviation of the objective function value over the 10 runs of the simulated annealing algorithm.

Table A.1: Simulated annealing results for the 10-node CAB data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	1864.4	1864	0.0%	1864	0.0%	0
	2	1819.8	1820	0.0%	1820	0.0%	0
	3	2287.6	2288	0.0%	2288	0.0%	0
	4	1814.9	1815	0.0%	1815	0.0%	0
	5	2160.7	2161	0.0%	2161	0.0%	0
	6	1579.8	1580	0.0%	1580	0.0%	0
	7	1814.9	1815	0.0%	1815	0.0%	0
	8	1496.1	1496	0.0%	1496	0.0%	0
	9	1739.6	1740	0.0%	1740	0.0%	0
	10	1377.4	1377	0.0%	1401	1.7%	74
5	1	1814.9	1815	0.0%	1815	0.0%	0
	2	1814.9	1815	0.0%	1815	0.0%	0
	3	2160.7	2161	0.0%	2161	0.0%	0
	4	1864.4	1864	0.0%	1864	0.0%	0
	5	1814.9	1815	0.0%	1815	0.0%	0
	6	1479.2	1479	0.0%	1479	0.0%	0
	7	1496.1	1496	0.0%	1496	0.0%	0
	8	1496.1	1496	0.0%	1499	0.2%	8
	9	1522.4	1522	0.0%	1522	0.0%	0
	10	1323.6	1324	0.0%	1555	17.5%	180

Table A.2: Simulated annealing results for the 15-node CAB data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	3483.7	3484	0.0%	3484	0.0%	0
	2	3061.1	3061	0.0%	3061	0.0%	0
	3	2628	2628	0.0%	2628	0.0%	0
	4	2716.4	2716	0.0%	2716	0.0%	0
	5	2461.2	2461	0.0%	2461	0.0%	0
	6	2499.5	2500	0.0%	2500	0.0%	0
	7	2086.1	2086	0.0%	2154	3.3%	78
	8	2461.2	2461	0.0%	2461	0.0%	0
	9	2348.2	2348	0.0%	2392	1.9%	25
	10	2456	2456	0.0%	2456	0.0%	0
5	1	2516.7	2517	0.0%	2517	0.0%	0
	2	2716.4	2716	0.0%	2716	0.0%	0
	3	2563.1	2563	0.0%	2563	0.0%	0
	4	2628	2628	0.0%	2628	0.0%	0
	5	2437.1	2437	0.0%	2437	0.0%	0
	6	2499.5	2500	0.0%	2500	0.0%	0
	7	2064.3	2212	7.2%	2367	14.7%	102
	8	2461.2	2461	0.0%	2461	0.0%	0
	9	2348.2	2352	0.2%	2394	1.9%	30
	10	2456	2456	0.0%	2456	0.0%	0

Table A.3: Simulated annealing results for the 20-node CAB data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	2716.4	2716	0.0%	2716	0.0%	0
	2	2444.9	2445	0.0%	2563	4.8%	91
	3	2716.4	2716	0.0%	2731	0.5%	31
	4	2716.4	2716	0.0%	2716	0.0%	0
	5	2337.8	2397	2.5%	2397	2.5%	0
	6	2716.4	2716	0.0%	2716	0.0%	0
	7	2364.5	2365	0.0%	2365	0.0%	0
	8	2716.4	2716	0.0%	2716	0.0%	0
	9	2456	2464	0.3%	2464	0.3%	0
	10	2343.4	2606	11.2%	2655	13.3%	32
5	1	2595	2595	0.0%	2595	0.0%	0
	2	2194.4	2286	4.2%	2562	16.7%	101
	3	2716.4	2716	0.0%	2716	0.0%	0
	4	2716.4	2716	0.0%	2716	0.0%	0
	5	2227.6	2287	2.7%	2287	2.7%	0
	6	2716.4	2716	0.0%	2716	0.0%	0
	7	2716.4	2716	0.0%	2716	0.0%	0
	8	2166.7	2167	0.0%	2167	0.0%	0
	9	2456	2464	0.3%	2464	0.3%	0
	10	2077.9	2473	19.0%	2588	24.6%	46

Table A.4: Simulated annealing results for the 25-node CAB data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	4072.3	4072	0.0%	4072	0.0%	0
	2	4329.7	4330	0.0%	4330	0.0%	0
	3	3179.7	3180	0.0%	3180	0.0%	1
	4	3051.1	3211	5.2%	3297	8.1%	30
	5	3471.9	3472	0.0%	3472	0.0%	0
	6	3064	3275	6.9%	3275	6.9%	0
	7	2620.1	2620	0.0%	2638	0.7%	19
	8	2537	2660	4.9%	2660	4.9%	0
	9	3782.3	3782	0.0%	3782	0.0%	0
	10	3012.9	3013	0.0%	3013	0.0%	0
5	1	3270.6	3271	0.0%	3271	0.0%	0
	2	4072.3	4072	0.0%	4072	0.0%	0
	3	3179.7	3180	0.0%	3180	0.0%	0
	4	3051.1	3306	8.4%	3306	8.4%	0
	5	3012.9	3013	0.0%	3013	0.0%	0
	6	3471.9	3472	0.0%	3472	0.0%	0
	7	2923.9	3040	4.0%	3040	4.0%	0
	8	2542.9	2543	0.0%	2543	0.0%	0
	9	2537	2660	4.9%	2660	4.9%	0
	10	3017.4	3309	9.7%	3408	12.9%	62

Table A.5: Simulated annealing results for the 10-node AP data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	38063	38063	0.0%	38063	0.0%	0
	2	40382	40382	0.0%	40382	0.0%	0
	3	41478	41478	0.0%	41478	0.0%	0
	4	35537	35537	0.0%	35537	0.0%	0
	5	41591	41591	0.0%	41591	0.0%	0
	6	50850	50850	0.0%	50850	0.0%	0
	7	46494	46494	0.0%	46494	0.0%	0
	8	46094	46094	0.0%	46094	0.0%	0
	9	52206	52206	0.0%	52206	0.0%	0
	10	40382	40382	0.0%	40382	0.0%	0
5	1	34567	34567	0.0%	35554	2.9%	525
	2	35537	35537	0.0%	35537	0.0%	0
	3	40382	40382	0.0%	40382	0.0%	0
	4	36024	36024	0.0%	36024	0.0%	0
	5	35537	35537	0.0%	35537	0.0%	0
	6	50850	50850	0.0%	50850	0.0%	0
	7	46094	46094	0.0%	46094	0.0%	0
	8	46494	46494	0.0%	46494	0.0%	0
	9	46494	46494	0.0%	46494	0.0%	0
	10	40382	40382	0.0%	40382	0.0%	0

Table A.6: Simulated annealing results for the 20-node AP data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	45592	51864	13.8%	51864	13.8%	0
	2	72214	72214	0.0%	72214	0.0%	0
	3	53346	53824	0.9%	53824	0.9%	0
	4	51390	52537	2.2%	52537	2.2%	0
	5	56443	56443	0.0%	56443	0.0%	0
	6	54646	54646	0.0%	54646	0.0%	0
	7	56059	56059	0.0%	56059	0.0%	0
	8	52010	52010	0.0%	52010	0.0%	0
	9	49206	49206	0.0%	49206	0.0%	0
	10	65608	65608	0.0%	65608	0.0%	0
5	1	45591	49014	7.5%	50250	10.2%	452
	2	72214	72214	0.0%	72214	0.0%	0
	3	52010	52010	0.0%	53577	3.0%	598
	4	45868	45952	0.2%	45952	0.2%	0
	5	56443	56443	0.0%	56771	0.6%	375
	6	54646	54646	0.0%	54646	0.0%	0
	7	46174	46174	0.0%	46174	0.0%	0
	8	56059	56059	0.0%	56059	0.0%	0
	9	49206	49206	0.0%	49206	0.0%	0
	10	65608	65608	0.0%	65608	0.0%	0

Table A.7: Simulated annealing results for the 25-node AP data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	57338	65964	15.0%	65964	15.0%	0
	2	66058	66058	0.0%	68304	3.4%	1092
	3	58921	58921	0.0%	58921	0.0%	0
	4	51954	55856	7.5%	55856	7.5%	0
	5	71358	71358	0.0%	73769	3.4%	904
	6	75416	75416	0.0%	75416	0.0%	0
	7	58384	58384	0.0%	59592	2.1%	1314
	8	58384	58795	0.7%	58795	0.7%	0
	9	69749	69749	0.0%	69985	0.3%	316
	10	75154	75154	0.0%	75154	0.0%	0
5	1	57338	63039	9.9%	63039	9.9%	0
	2	57893	57958	0.1%	57958	0.1%	0
	3	57631	57631	0.0%	57631	0.0%	0
	4	51954	54470	4.8%	55717	7.2%	438
	5	71358	71358	0.0%	71480	0.2%	172
	6	64350	64350	0.0%	65549	1.9%	1186
	7	75416	75416	0.0%	75416	0.0%	0
	8	58384	58384	0.0%	58384	0.0%	0
	9	58384	58919	0.9%	59113	1.2%	68
	10	66828	66828	0.0%	66828	0.0%	0

Table A.8: Simulated annealing results for the 40-node AP data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	65459	73261	11.9%	75436	15.2%	1688
	2	63050	71150	12.8%	78050	23.8%	3181
	3	66275	66275	0.0%	66553	0.4%	359
	4	62383	67402	8.0%	67402	8.0%	0
	5	82966	82966	0.0%	85036	2.5%	1989
	6	88568	90639	2.3%	90639	2.3%	0
	7	79094	80954	2.4%	80954	2.4%	0
	8	84502	85596	1.3%	88870	5.2%	1893
	9	73884	73884	0.0%	73884	0.0%	0
	10	67592	67592	0.0%	70172	3.8%	1851
5	1	62582	71455	14.2%	71455	14.2%	0
	2	53590	71357	33.2%	71747	33.9%	137
	3	66270	66270	0.0%	66476	0.3%	327
	4	62383	67402	8.0%	67402	8.0%	0
	5	82966	82966	0.0%	83061	0.1%	153
	6	80697	87940	9.0%	87940	9.0%	0
	7	79094	79094	0.0%	80582	1.9%	784
	8	84502	85596	1.3%	86560	2.4%	1085
	9	67058	67058	0.0%	67385	0.5%	315
	10	68080	68080	0.0%	68080	0.0%	0

Table A.9: Simulated annealing results for the 50-node AP data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	86990	91906	5.7%	95193	9.4%	1550
	2	67429	76052	12.8%	76052	12.8%	0
	3	70876	72249	1.9%	72249	1.9%	0
	4	68870	68870	0.0%	68870	0.0%	0
	5	69120	69120	0.0%	69120	0.0%	0
	6	70968	70968	0.0%	70968	0.0%	0
	7	70968	70968	0.0%	70968	0.0%	0
	8	80692	80692	0.0%	80692	0.0%	0
	9	90013	90013	0.0%	90013	0.0%	0
	10	85742	93234	8.7%	93234	8.7%	0
5	1	72177	79099	9.6%	84365	16.9%	3187
	2	65649	65649	0.0%	65649	0.0%	0
	3	72153	72153	0.0%	72666	0.7%	630
	4	66354	73039	10.1%	75381	13.6%	1521
	5	78369	78369	0.0%	78369	0.0%	0
	6	78754	78847	0.1%	78847	0.1%	0
	7	59894	69894	16.7%	70124	17.1%	600
	8	77729	77729	0.0%	77729	0.0%	0
	9	86642	86642	0.0%	88069	1.6%	1527
	10	90854	90854	0.0%	90854	0.0%	0

Table A.10: Simulated annealing results for the 100-node AP data set.

p	Config.	Opt.	Min.	Min. % gap	Avg.	Avg. % gap	St. Dev.
4	1	85814	90870	5.9%	91828	7.0%	336
	2	89678	92886	3.6%	92886	3.6%	0
	3	73366	89858	22.5%	89858	22.5%	0
	4	76726	93535	21.9%	93535	21.9%	0
	5	70146	79083	12.7%	79083	12.7%	0
	6	70146	80359	14.6%	82785	18.0%	1162
	7	87581	87581	0.0%	87581	0.0%	0
	8	84456	84456	0.0%	84456	0.0%	0
	9	97784	97784	0.0%	97784	0.0%	0
	10	92719	92719	0.0%	92719	0.0%	0
5	1	74364	82396	10.8%	82396	10.8%	0
	2	72360	72360	0.0%	72360	0.0%	0
	3	71508	73172	2.3%	73172	2.3%	0
	4	75929	76639	0.9%	76639	0.9%	0
	5	70146	70146	0.0%	70146	0.0%	0
	6	70846	87037	22.9%	87742	23.8%	248
	7	94016	106460	13.2%	107159	14.0%	509
	8	101030	112270	11.1%	116064	14.9%	1664
	9	77108	83864	8.8%	83864	8.8%	0
	10	87604	87604	0.0%	87604	0.0%	0

REFERENCES

Abdinnour-Helm, S. (1998). A hybrid heuristic for the uncapacitated hub location problem. *European Journal of Operational Research 106*, 489–499.

Abdinnour-Helm, S. and M. A. Venkataramanan (1998). Solution approaches to hub location problems. *Annals of Operations Research 78*, 31–50.

Adler, N. (2005). Hub-and-spoke network choice under competition with an application to Western Europe. *Transportation Science 39*, 58–72.

Adler, N. and K. Smilowitz (2007). Hub-and-spoke network alliances and mergers: Price-location competition in the airline industry. *Transportation Research Part B: Methodological 41*, 394–409.

Ahmed, S. and A. Shapiro (2002). The sample average approximation method for stochastic programs with integer recourse. http://www2.isye.gatech.edu/sahmed/saasip.pdf.

Ahmed, S., M. Tawarmalani, and N. V. Sahinidis (2004). A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming 100*, 355–377.

Aykin, T. (1994). Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *European Journal of Operational Research 79*, 501–523.

Beasley, J. E. (1990). OR-Library: Distributing test problems by electronic mail. *Journal of the Operational Research Society 41*, 1069–1072.

Beasley, J. E. (2004). OR-Library. http://www.brunel.ac.uk/depts/ma/research/jeb/info.html.

Benders, J. F. (1962). Partitioning procedures for solving mixed variables programming problems. *Numerische Mathematik 4*, 238–252.

Birge, J. R. and F. Louveaux (1997). *Introduction to Stochastic Programming*. Springer-Verlag, New York.

Brush, G. and N. Marlow (1990). Assuring the dependability of telecommunications networks and services. *IEEE Network Magazine 4*, 29–34.

Bryan, D. L. (1998). Extensions to the hub location problem: Formulations and numerical examples. *Geographical Analysis 30*, 315–330.

Bryan, D. L. and M. E. O'Kelly (1999). Hub-and-spoke networks in air transportation: An analytical review. *Journal of Regional Science 39*, 275–295.

Butler, R. V. and J. H. Huston (1990). Airline service to non-hub airports ten years after deregulation. *Logistics and Transportation Review 26*, 3–16.

Campbell, A. M., T. J. Lowe, and L. Zhang (2007). The p-hub center allocation problem. *European Journal of Operational Research 176*, 819–835.

Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research 72*, 387–405.

Campbell, J. F., A. T. Ernst, and M. Krishnamoorthy (2002). Hub location problems. In Z. Drezner and H. W. Hamacher (Eds.), *Facility Location: Applications and Theory*, pp. 373–407. Springer-Verlag, Berlin.

Cánovas, L., S. García, and A. Marín (2007). Solving the uncapacitated multiple allocation hub location problem by means of a dual-ascent technique. *European Journal of Operational Research 179*, 990–1007.

Carøe, C. C. and J. Tind (1997). A cutting-plane approach to mixed 0-1 stochastic integer programs. *European Journal of Operational Research 101*, 306–316.

Carøe, C. C. and J. Tind (1998). L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming 83*, 451–464.

CBC News Online (2005, June 20). Air canada timeline. http://www.cbc.ca/news/background/aircanada/timeline.html. Accessed on March 14, 2006.

Central Intelligence Agency (2006). The World Factbook. http://www.cia.gov/cia/publications/factbook. Accessed on June 20, 2006.

Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications 45*, 41–51.

Charnes, A. and W. W. Cooper (1959). Chance-constrained programming. *Management Science 6*, 73–79.

Charnes, A. and W. W. Cooper (1963). Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations Research 11*, 18–39.

Chen, H., A. M. Campbell, B. W. Thomas, and A. Tamir (2007). Minimax flow tree problems. Submitted for publication.

Chen, J. F. (2007). A hybrid heuristic for the uncapacitated single allocation hub location problem. *Omega 35*, 211–220.

CNN.com (2004, September 27). Low-cost carriers under new pressure. http://www.cnn.com/2004/TRAVEL/09/27/bt.low.cost.carriers.reut/index.html. Accessed on October 5, 2004.

CNNMoney.com (2005, September 14). More airline turbulence ahead. http://money.cnn.com/2005/09/13/news/fortune500/airlines_future/index.htm. Accessed on March 14, 2006.

Coy, S. P., B. L. Golden, G. C. Runger, and E. A. Wasil (2000). Using experimental design to find effective parameter settings for heuristics. *Journal of Heuristics 7*, 77–97.

Cunha, C. B. and M. R. Silva (2007). A genetic algorithm for the problem of configuring a hub-and-spoke network for a LTL trucking company in brazil. *European Journal of Operational Research 179*, 747–758.

Czyzyk, J., M. P. Mesnier, and J. J. Moré (1998). The NEOS Server. *IEEE Journal on Computational Science and Engineering 5*, 68–75.

Dash Optimization (2004). http://www.dashoptimization.com.

Daskin, M. S. (1995). *Network and Discrete Location: Models, Algorithms and Applications*. John Wiley, New York.

Dolan, E. (2001). The NEOS Server 4.0 Administrative Guide. Technical Memorandum ANL/MCS-TM-250, Mathematics and Computer Science Division, Argonne National Laboratory.

Domínguez, E., J. Muñoz, and E. Mérida (2003). A recurrent neural network model for the p-hub problem. *Lecture Notes in Computer Science 2687*, 734–741.

Dyer, M. and A. Frieze (1985). A simple heuristic for the p-center problem. *Operations Research Letters 3*, 285–288.

Ernst, A. T., H. W. Hamacher, H. Jiang, M. Krishnamoorthy, and G. Woeginger (2006a). Heuristic algorithms for the uncapacitated p-hub center single allocation problem. Forthcoming in the *European Journal of Operational Research*.

Ernst, A. T., H. W. Hamacher, H. Jiang, M. Krishnamoorthy, and G. Woeginger (2006b). Uncapacitated single and multiple allocation p-hub center problems. Forthcoming in *Operations Research*.

Ernst, A. T. and M. Krishnamoorthy (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problems. *Location Science 4*, 139–154.

Ernst, A. T. and M. Krishnamoorthy (1998). An exact solution approach based on shortest-paths for p-hub median problems. *INFORMS Journal on Computing 10*, 149–162.

Ernst, A. T. and M. Krishnamoorthy (1999). Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research 86*, 141–159.

Farley, A. M., P. Fragopoulou, D. W. Krumme, A. Proskurowski, and D. Richards (2000). Multi-source spanning tree problems. *Journal of Interconnection Networks 1*(1), 61–71.

FedEx (2007). FedEx Express U.S. Terms and Conditions. http://www.fedex.com/us/services/terms/us.html?link=4#guarantee. Accessed on June 6, 2007.

Fiorino, F. (2002, September 30). Carty to analysts: AA aims to survive. *Aviation Week & Space Technology 157*, 47–48.

Fredman, M. L. and R. E. Tarjan (1987). Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM 34*, 596–615.

Gendreau, M., G. Laporte, and R. Séguin (1995). An exact algorithm for the vehicle routing problem with stochastic demands and customers. *Transportation Science 29*, 143–155.

Gropp, W. and J. J. Moré (1997). Optimization environments and the NEOS Server. In M. D. Buhmann and A. Iserles (Eds.), *Approximation Theory and Optimization: Tributes to M. J. D. Powell*, pp. 167–182. Cambridge University Press.

Hakimi, S. L. (1964). Optimum location of switching centers and the absolute centers and medians of a graph. *Operations Research 12*, 450–459.

Hakimi, S. L. (1965). Optimum location of switching centers in a communications network and some related graph theoretic problems. *Operations Research 13*, 462–475.

Hannon, D. (2005, March 17). Small package market changes could benefit buyers. http://www.purchasing.com/article/CA510904.html. Accessed on September 20, 2005.

Iberia Airlines (2004). Press Release July 2nd 2004. http://grupo.iberia.es/content/GrupoIberia/Hechos%20Relevantes/HR020704Miami%20Hub.pdf. Accessed on June 18, 2006.

Johnson, D. S., C. R. Aragon, L. A. McGeoch, and C. Schevon (1989). Optimization by simulated annealing: An experimental evaluation; Part I, Graph partitioning. *Operations Research 37*, 865–892.

Kall, P. and S. W. Wallace (1994). *Stochastic Programming*. John Wiley & Sons, Chichester.

Kara, B. Y. and B. Ç. Tansel (2000). On the single-assignment p-hub center problem. *European Journal of Operational Research 125*, 648–655.

Kara, B. Y. and B. Ç. Tansel (2001). The latest arrival hub location problem. *Management Science 47*, 1408–1420.

Kara, B. Y. and B. Ç. Tansel (2003). The single-assignment hub covering problem: Models and linearizations. *Journal of the Operational Research Society 54*, 59–64.

Kim, H.-J., S.-H. Chung, and D.-W. Tcha (1995). Optimal design of the two-level distributed network with dual homing local connections. *IIE Transactions 27*, 555–563.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi (1983). Optimization by simulated annealing. *Science 220*, 671–680.

Klein Haneveld, W. K. and M. H. van der Vlerk (1999). Stochastic integer programming: General models and algorithms. *Annals of Operations Research 85*, 39–57.

Kleywegt, A. J., A. Shapiro, and T. Homem-De-Mello (2001). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization 12*, 479–502.

Klincewicz, J. G. (1991). Heuristics for the p-hub location problem. *European Journal of Operational Research 53*, 25–37.

Klincewicz, J. G. (1992). Avoiding local optima in the p-hub location problem using tabu search and GRASP. *Annals of Operations Research 40*, 283–302.

Klincewicz, J. G. (1996). A dual algorithm for the uncapacitated hub location problem. *Location Science 4*, 173–184.

Klincewicz, J. G. (1998). Hub location in backbone/tributary network design: A review. *Location Science 6*, 307–335.

Kouvelis, P. and G. Yu (1996). *Robust discrete optimization and its applications*. Kluwer Academic Publishers, Netherlands.

Kratica, J., Z. Stanimirović, D. Tošić, and V. Filipović (2007). Two genetic algorithms for solving the uncapacitated single alocation p-hub median problem. *European Journal of Operational Research 182*, 15–28.

Labbé, M. and H. Yaman (2004). Projecting the flow variables for hub location problems. *Networks 44*, 84–93.

Labbé, M., H. Yaman, and E. Gourdin (2005). A branch-and-cut algorithm for hub location problems with single assignment. *Mathematical Programming 102*, 371–405.

Laporte, G., F. V. Louveaux, and H. Mercure (1994). A priori optimization of the probabilistic traveling salesman problem. *Operations Research 42*, 543–549.

Laporte, G., F. V. Louveaux, and L. van Hamme (2002). An integer L-shaped algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research 50*, 415–423.

Linderoth, J., A. Shapiro, and S. Wright (2006). The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research 142*, 215–241.

Marianov, V. and D. Serra (2003). Location models for airline hubs behaving as $M/D/c$ queues. *Computers & Operations Research 30*, 983–1003.

McMahan, H. B. and A. Proskurowski (2004). Multi-source spanning trees: algorithms for minimizing source eccentricities. *Discrete Applied Mathematics 137*, 213–222.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21*, 1087–1092.

Musicant, D. R. (2000). Matlab/cplex mex-files. http://www.cs.wisc.edu/~musicant/data/cplex/.

Nickel, S., A. Schobel, and T. Sonneborn (2000). Hub location problems in urban traffic networks. In M. Pursula and J. Niittymäki (Eds.), *Mathematical Methods on Optimization in Transportation Systems*, pp. 1–12. Kluwer Academic Publishers, Netherlands.

Norkin, V. I., Y. M. Ermoliev, and A. Ruszczyński (1998). On optimal allocation of indivisibles under uncertainty. *Operations Research 46*, 381–395.

Norkin, V. I., G. C. Pflug, and A. Ruszczyński (1996). A branch and bound method for stochastic global optimization. http://www.mat.univie.ac.at/~neum/glopt/mss/NorPR96.pdf.

O'Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research 32*, 393–404.

O'Kelly, M. E. (1992). Hub facility location with fixed costs. *Papers in Regional Science: The Journal of the RSAI 71*, 293–306.

O'Kelly, M. E. and D. L. Bryan (1998). Hub location with flow economies of scale. *Transportation Research B 32*, 605–616.

O'Kelly, M. E. and H. J. Miller (1991). Solution strategies for the single facility minimax hub location problem. *Papers in Regional Science: The Journal of the RSAI 70*, 367–380.

Ott, J. (1993). New entrants, majors differ on hub-and-spoke networks. *Aviation Week & Space Technology 139*, 70–71.

Pamuk, F. S. and C. Sepil (2001). A solution to the hub center problem via a single-relocation algorithm with tabu search. *IIE Transactions 33*, 399–411.

Pérez, M., F. A. Rodríguez, and J. M. Moreno-Vega (2004). On the use of path relinking for the p-hub median problem. *Lecture Notes in Computer Science 3004*, 155–164.

Pirlot, M. (1996). General local search methods. *European Journal of Operational Research 92*, 493–511.

Rodíguez, V., M. J. Alvarez, and L. Barcos (2007). Hub location under capacity constraints. *Transportation Research Part E: Logistics and Transportation Review 43*, 495–505.

Shapiro, A. (2003). Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro (Eds.), *Handbooks in OR&MS, Vol. 10: Stochastic Programming*, Chapter 6, pp. 353–425. Elsevier Science, Netherlands.

Skorin-Kapov, D. and J. Skorin-Kapov (1994). On tabu search for the location of interacting hub facilities. *European Journal of Operational Research 73*, 502–509.

St. Louis Business Journal (2002, August 13). Local impact of american airlines moves unclear. http://www.bizjournals.com/stlouis/stories/2002/08/12/daily17.html. Accessed on March 16, 2006.

Subramanian, R., R. P. Scheff, Jr., J. D. Quillinan, D. S. Wiper, and R. E. Marsten (1994). Coldstart: Fleet assignment at Delta Air Lines. *Interfaces 24*, 104–120.

Swan, W. M. and N. Adler (2006). Aircraft trip cost parameters: A function of stage length and seat capacity. *Transportation Research Part E 42*, 105–115.

Tan, P. Z. and B. Y. Kara (2007). A hub covering model for cargo delivery systems. *Networks 49*, 28–39.

Teitz, M. B. and P. Bart (1968). Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research 16*, 955–961.

The Mathworks Inc. (1997). *MATLAB: The Language of Technical Computing*. The Mathworks Inc.

The New York Times (2003, June 7). Airlines' unwanted fleet grows in the desert. http://www.nytimes.com/2003/06/07/business/07AIR.html. Accessed on March 14, 2006.

Toh, R. S. and R. G. Higgins (1985). The impact of hub and spoke network centralization and route monopoly on domestic airline profitability. *Transportation Journal 24*, 16–27.

Topcuoglu, H., F. Coruta, M. Ermisb, and G. Yilmaza (2005). Solving the uncapacitated hub location problem using genetic algorithms. *Operations Research 32*, 967–984.

Toregas, C., R. Swain, C. ReVelle, and L. Bergman (1971). The location of emergency service facilities. *Operations Research 19*, 1363–1373.

United Parcel Services (2006). UPS Annual Report 2006. http://investor.shareholder.com/ups/. Accessed on July 7, 2007.

United Parcel Services (2007). UPS tariff/terms and conditions of service for small package shipments in the United States. http://www.ups.com/media/en/terms_service_12012007.pdf. Accessed on May 1, 2007.

van Laarhoven, P. J. M. and E. H. L. Aarts (1987). *Simulated Annealing: Theory and Applications*. Kluwer Academic Publishers, MA.

Van Slyke, R. M. and R. Wets (1969). L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics 17*, 638–663.

Verweij, B., S. Ahmed, A. J. Kleywegt, G. Nemhauser, and A. Shapiro (2003). The sample average approximation method applied to stochastic routing problems: A computational study. *Computational Optimization and Applications 24*, 289–333.

Wagner, B. (2004). A note on "The Latest Arrival Hub Location Problem". *Management Science 50*, 1751–1752.

Wagner, B. (2007). Model formulations for hub covering problems. Forthcoming in the *Journal of the Operational Research Society*.

Wu, B. Y. (2004). An improved algorithm for the k-source maximum eccentricity spanning trees. *Discrete Applied Mathematics 143*, 342–350.

Yaman, H. (2005). Polyhedral analysis for the uncapacitated hub location problem with modular arc capacities. *SIAM Journal of Discrete Mathematics 19*, 501–522.

Yaman, H. and G. Carello (2005). Solving the hub location problem with modular link capacities. *Computers & Operations Research 32*, 3227–3245.

Yaman, H., B. Y. Kara, and B. Ç. Tansel (2007). The latest arrival hub location problem for cargo delivery systems with stopovers. *Transportation Research Part B: Methodological 41*, 906–919.