RELIABLE P-HUB LOCATION PROBLEMS AND PROTECTION MODELS FOR HUB NETWORK DESIGN

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Graduate School of the Ohio State University

By

Hyun Kim, M.A.

*****

The Ohio State University

2008

Dissertation Committee:

Professor Morton E. O’Kelly, Adviser
Professor Alan T. Murray
Professor Mei-Po Kwan
Professor Ningchuan Xiao

Approved by:

Graduate Program in Geography
This dissertation addresses a series of new hub location problems and network design models, which are termed reliable $p$-hub location problems, and $p$-hub protection problems. The essential motivation to propose these problems arises from the increasing vulnerability in current telecommunication and transportation networks where the networks have converged into more hub-and-spoke type systems. The models presented in the chapters aim to enhance both a network’s performance and its resiliency to the potential disruptions in transportation and telecommunication networks.

The major research objectives and contributions are outlined below.

First, the reliable $p$-hub location problems are addressed in terms of network reliability. The models demonstrate how optimal hub locations are determined under various reliability or capability conditions of network facilities. Geographical closeness of hubs can enhance network performance by increasing efficiency of traffic exchanges among hub facilities. The dispersion of hub facilities can be considered in hub network design which can protect the excessive concentration of flows from particular hub facilities.

Second, the $p$-hub protection problems are focused on hub network design in a survivable network design. Three standard models are addressed based on single,
multiple and survivable routing schemes. As extensions, capacitated and minimum threshold models are presented. The relationship between network costs, redundancy of linkages, and the survivability of networks in order to protect hub activity levels are examined.

Finally, efficient heuristics to solve the reliable hub network design models are presented. In particular, a tabu search heuristic (TABUSA) and a hybrid-type heuristic (TABUEX) are developed to solve the most difficult models which are developed based on single and survivable routing schemes. From the computational results, both search approaches demonstrate their efficiency in solving the problems, offering the viability to apply larger size problems effectively within a reasonable computational time.

As concluding remarks, this dissertation provides good insights for reliable and survivable hub network design. The models are also expected to be utilized as decision tools for geographers, network planners, and policy makers in designing critical network infrastructures.
Dedicated to

My wife, Eunjin,

My daughter Eyrin,

&

My parents
ACKNOWLEDGMENTS

I am gratefully indebted to my adviser, Dr. Morton O’Kelly for his guidance and encouragement throughout my study at the Ohio State University. Not only did he share his knowledge, his patience was greatly appreciated. This dissertation would not been possible without him.

I wish to thank my committee members, Dr. Alan Murray, Mei-Po Kwan, and Ningchuan Xiao for their thoughtful suggestions and comments on my dissertation. I also would like to thank to the faculty members in the Department of Geography in Seoul National University for introducing me to geography.

My special thanks and appreciation go to Eunjin, Eyrin, and two families in Korea for their encouragements and prayers. I thank to God who has guided me throughout my life.


VITA

1999…………………  Bachelor of Arts in Geography, Seoul National University

2001…………………  Master of Arts in Geography, Seoul National University

2002 – Present………  Graduate Teaching and Research Associate, The Ohio State University

PUBLICATIONS

Research publications


Kong-Rae Lee, Junghyop Lee, Chesung Lim, Dongwoo Han, and Hyun Kim, 2002, “Analysis on the Knowledge Clusters of Korea”, Ministry of Science & Technology, Science and Technology policy Institute, Seoul, Korea.


**FIELDS OF STUDY**

Major Field: Geography

(Transportation, Telecommunications, and Urban Geography)
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>VITA</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
</tbody>
</table>

**CHAPTER 1**

INTRODUCTION .................................................................................................................. 1

1.1. Research Background .............................................................................................. 1
1.2. Research Objectives ................................................................................................. 3
1.3. Organization of research .......................................................................................... 5

**CHAPTER 2**

RELIABILITY, SURVIVABILITY AND HUB NETWORK DESIGN ........................................... 8

2.1. Reliable Network Design.......................................................................................... 9
2.1.1. Deterministic approach ..................................................................................... 10
2.1.2. Probabilistic approach ..................................................................................... 11
2.1.3. Reliability optimization problems ................................................................... 19
2.1.4. Summary of reliable network design .............................................................. 21
2.2. Survivability in Network Design ........................................................................... 22
2.2.1. Protective location problems: transmission facility design ......................... 25
2.2.2. Topological design with connectivity .............................................................. 25
2.2.3. Design of self-healing network structure ....................................................... 26
2.2.4. Survivable routing design problems ............................................................... 27
2.2.5. Vulnerability assessments ........................................................................... 31
2.2.6. Summary of survivable network design ....................................................... 34
2.3. Hub Network Design Problems ....................................................................... 34
2.3.1. Hub location problems and network design ............................................... 35
2.3.2. Design perspective of hub-and-spoke network design in telecommunications .................................................................................................................. 37
2.4. Research Statements ....................................................................................... 43

CHAPTER 3
RELIABLE P-HUB LOCATION PROBLEMS FOR HUB NETWORK DESIGN........ 45
3.1. Introduction ..................................................................................................... 45
3.2. Related Research ............................................................................................ 48
3.3. Model Developments ..................................................................................... 50
3.3.1. Routing reliability $R_{ijkm}$ .......................................................................... 50
3.3.2. Routing scheme: reliability factors $\alpha$ and $\gamma$ ........................................ 51
3.3.3. Data: Reliability matrix and interacting flow matrix .................................... 54
3.4. Model Formulations ....................................................................................... 58
3.4.1. MRSA (p-hub Maximum Reliability model with Single Assignment) .......... 58
3.4.2. MRMA (p-hub Maximum Reliability model with Multiple Assignments) .... 60
3.4.3. MDSA (p-hub Mandatory Dispersion model with Single Assignment) ........ 61
3.4.4. MDMA (p-hub Mandatory Dispersion model with Multiple Assignments) .. 62
3.4.5. MRDI (Multiobjective Reliable p-hub DIspersion model) ................................ 63
3.5. Model Experiments ....................................................................................... 64
3.5.1. Model behaviors of the MRSA and MRMA ............................................... 69
3.5.2. Impact of hub dispersion ............................................................................ 74
3.6. Summary ....................................................................................................... 82
CHAPTER 4
P-HUB PROTECTION MODELS FOR RELIABLE HUB NETWORK DESIGN 84
4.1. Introduction ............................................................................................................ 84
4.2. Survivable Network Designs in Telecommunication Systems ......................... 87
4.3. Modeling $p$-Hub Protection Models ................................................................. 90
   4.3.1. Probability of potential traffic loss $L_{ijkm}$ .............................................. 91
   4.3.2. Inter-hub reliability factor $\alpha$ and intra-hub reliability factor $\gamma$ .......... 92
   4.3.3. PROSA ($p$-Hub Protection Model with Single Assignment) ................. 93
   4.3.4. PROMA ($p$-Hub Protection Model with Multiple Assignments) ............ 95
   4.3.5. PROPS ($p$-Hub Protection Model with Primary and Secondary routes) ... 96
4.4. Model Applications ............................................................................................. 98
   4.4.1. Data and solution method .......................................................................... 98
   4.4.2. Basic results ............................................................................................. 101
4.5. Model Extensions ............................................................................................... 107
   4.5.1. Capacitated PROMA (PROMA-CAP) and PROPS (PROPS-CAP) .......... 107
   4.5.2. Minimum threshold model (PROMA-MT and PROPS-MT) ..................... 110
   4.5.3. Minimum threshold on back-up linkage in the PROPS (PROPS-MTB) .... 114
4.6. Summary ............................................................................................................ 120

CHAPTER 5
HEURISTICS FOR RELIABLE $p$-HUB LOCATION PROBLEMS 122
5.1. Introduction .......................................................................................................... 122
5.2. Solution Approaches in Hub Network Design .................................................. 124
5.3. Exhaustive Search and Tabu Search Algorithms .............................................. 127
   5.3.1. Exhaustive search ...................................................................................... 127
   5.3.2. Tabu search .............................................................................................. 131
5.3.2.1. Long-term memory (LTM) ................................................................. 131
5.3.2.2. Location phase (LOC_PAHSE) ......................................................... 132
5.3.2.3. Allocation phase (ALLO_PAHSE) ...................................................... 132
5.3.2.4. Survivable routing phase (SUR_PAHSE) .......................................... 133
5.3.2.5. Tabu lists ......................................................................................... 133
5.4. Tabu Search for the MRSA (TABUSA) ..................................................... 134
  5.4.1. Algorithmic procedure ................................................................. 134
5.5. A hybrid Search for PROPS (TABUEX) ................................................ 140
  5.5.1. Algorithmic procedure ................................................................. 140
5.6. Computational Results ........................................................................ 141
  5.6.1. Data sets ....................................................................................... 141
  5.6.2. Computational experiments ......................................................... 144
5.7. Summary ........................................................................................... 153

CHAPTER 6
CONCLUSIONS AND FUTURE RESEARCH .................................................. 154
  6.1. Conclusions ...................................................................................... 154
  6.2. Future Research ................................................................................ 158
    6.2.1. Model extensions ................................................................. 158
    6.2.2. Extending model applications ............................................. 159
    6.2.3. Developing solution approach ............................................. 159

BIBLIOGRAPHY ..................................................................................... 161
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Types of reliability measurements.</td>
<td>12</td>
</tr>
<tr>
<td>2.2. Definition and interpretation of reliability in telecommunications and transportation.</td>
<td>18</td>
</tr>
<tr>
<td>2.3. Classification of survivable network design and models.</td>
<td>29</td>
</tr>
<tr>
<td>2.4. Representative studies and their models associate with hub location problems.</td>
<td>41</td>
</tr>
<tr>
<td>3.1. Problem sizes for model formulations.</td>
<td>67</td>
</tr>
<tr>
<td>3.2. Testing scheme.</td>
<td>68</td>
</tr>
<tr>
<td>3.3. Routing choices for $p=4$ using SET I.</td>
<td>71</td>
</tr>
<tr>
<td>3.4. Hub locations and facility activity levels for $p=4$ and 5 using SET II.</td>
<td>73</td>
</tr>
<tr>
<td>3.5. Non-inferior solutions of the MRDI for CASE I and CASE IV ($p=4$, SET II).</td>
<td>81</td>
</tr>
<tr>
<td>4.1. Selected results for 3, 4 and 5 hub models with $\alpha$ and $\gamma$ varying from 0.10 to 0.95</td>
<td>102</td>
</tr>
<tr>
<td>4.2. Results for capacitated models: 5-hubs PROMA and PROPS.</td>
<td>109</td>
</tr>
<tr>
<td>4.3. Results for minimum threshold models: 5-hubs PROMA and PROPS.</td>
<td>112</td>
</tr>
<tr>
<td>4.4. Results for back-up minimum threshold models ($p=4$ and 5, $\alpha$ and $\gamma = 0.90, 0.70$ and 0.50)</td>
<td>116</td>
</tr>
<tr>
<td>5.1. Tabu size and maximum iterations of each phase applied in the heuristics.</td>
<td>144</td>
</tr>
<tr>
<td>5.2. Results of the computational experiments of the EX-SA and EX-MA.</td>
<td>146</td>
</tr>
<tr>
<td>5.3. Results of the computational experiments with the 10, 15, and 25 node problems for the MRSA.</td>
<td>148</td>
</tr>
<tr>
<td>5.4. Computational results with the 10, 15, and 20 node problems for PROPS.</td>
<td>152</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.</td>
<td>7</td>
</tr>
<tr>
<td>2.1.</td>
<td>14</td>
</tr>
<tr>
<td>2.2.</td>
<td>32</td>
</tr>
<tr>
<td>2.3.</td>
<td>36</td>
</tr>
<tr>
<td>3.1.</td>
<td>54</td>
</tr>
<tr>
<td>3.2.</td>
<td>56</td>
</tr>
<tr>
<td>3.3.</td>
<td>57</td>
</tr>
<tr>
<td>3.4.</td>
<td>66</td>
</tr>
<tr>
<td>3.5.</td>
<td>70</td>
</tr>
<tr>
<td>3.6.</td>
<td>76</td>
</tr>
<tr>
<td>3.7.</td>
<td>78</td>
</tr>
<tr>
<td>3.8.</td>
<td>79</td>
</tr>
<tr>
<td>4.1.</td>
<td>87</td>
</tr>
<tr>
<td>4.2.</td>
<td>90</td>
</tr>
<tr>
<td>4.3.</td>
<td>100</td>
</tr>
<tr>
<td>4.4.</td>
<td>104</td>
</tr>
<tr>
<td>4.5.</td>
<td>106</td>
</tr>
<tr>
<td>4.6.</td>
<td>113</td>
</tr>
<tr>
<td>4.7.</td>
<td>118</td>
</tr>
</tbody>
</table>

1.1. The structure of the dissertation
2.1. Computations of Origin-Destination reliability
2.2. Variations of network performance on the disruptions of network failures
2.3. Hub-and-spoke network: (a) Single assignment, (b) Multiple assignment
3.1. Route choice based on reliability factor $\alpha$ and reliability factor $\gamma$
3.2. The relationship between city node reliabilities and geographical distance in the $R$ matrix
3.3. 14 U.S city nodes and their telecommunication potential
3.4. The enumerated routing variable set $S$
3.5. Model behaviors: MRSA (left) and MRMA (right) for $p=4$ using SET I
3.6. Comparison of facility activity level for a clustered (5a) and dispersed hub network (5b) at a moderate level of reliability factors ($p=3$, both $\alpha$ and $\gamma=0.7$)
3.7. The solutions of MDSA for CASE I and IV ($p=5$, $D_{man}=0$ and 1100 miles)
3.8. The solutions of MDMA for CASE I and IV ($p=5$, $D_{man}=0$ and 1100 miles)
4.1. The protocols of the p-hub protection models
4.2. The computation of potential traffic loss: a single (a) and a primary and back-up routing scheme (b)
4.3. The enumerated routing variable set $S'$
4.4. PROSA, PROMA and PROPS for 4 hubs with $\alpha, \beta = 0.90$
4.5. Hub activity envelopes for PROSA, PROMA and PROPS ($p=5$)
4.6. Hub activity envelopes for selected capacitated and minimum threshold models
4.7. Comparison of network design between PROPS and PROPS-MTB
4.8. Comparison of hub activity envelopes for the PROPS and PROPS-MTB ($p=4$, $\alpha$ and $\gamma = 0.70$). ................................................................................................................................. 119

5.1. Procedure of exhaustive search for single assignment problem (EX-SA)......... 128

5.2. Procedure of exhaustive search for multiple assignment problems (EX-MA) .... 130

5.3. Procedure of the tabu search TABUSA for the MRSA. ................................. 138

5.4. Procedure of the hybrid search TABUEX for the PROPS. ......................... 142

5.5. List of 25 nodes.............................................................................................. 143

5.6. Optimal hubs and assignments of non-hub nodes for the 15-node problem..... 150
CHAPTER 1

INTRODUCTION

1.1. Research Background

The hub location problem is an important class of location models in geography and location science because hubs and their assignments on networks are identified in many practical networks in current telecommunication and transportation networks (Daskin 1995; Campbell 2002).

‘Hubs’ are critical elements of telecommunication and transportation networks since they play a vital role as a switching or transshipment point, allowing mass traffic movement. In particular, the placement of hubs in telecommunications has been identified as a significant operational and defense strategy because the performance of current networks is highly reliant on hub locations and any malfunction at a hub may cause degradation of the entire network’s ability to transfer flow (Grubesic, O'Kelly, and Murray 2003; NSTAC 2003). Considering the increasing vulnerability of hub facilities to
disruption, designing more reliable hub networks is a critical issue in current network-based infrastructure systems (Klincewicz 2006; Skorin-Kapov et al. 2006).

A couple of recent incidents exemplify the importance of the issue. The degradation of network communications due to the 9/11 attack at the World Trade Center in New York showed how vulnerable the current Internet is to such an unexpected damage. The damage gave rise to not only short-term failures of statewide public online services, but led to temporal degradation of telecommunication service quality on the entire East Coast of the United States (BBC news 26, Nov 2002; Grubesic et al. 2003). As a non-western case, severe virus attacks on the critical hub of one major backbone provider in Korea caused disastrous network malfunctions successively over 70 other ISPs’ networks. These incidents raise the need to examine the location of hubs to ensure a desirable level of overall network reliability as well as sufficient redundancies for network protection (NCA 2004a and b).

Maintaining network reliability at a ‘desired’ level or to ensure survivability from any disruptions at any network component is a major challenge in networks design (O’Kelly and Kim 2006; Grover and Tipper 2005). Although researchers have explored locating hubs optimally with the objectives of maximizing a network’s performance or minimizing the potential loss of interacting flows from possible failure, the issue of reliable or survivable network design has received little attention in hub-and-spoke network design. This is because the concept of a reliable or survivable network system is often recognized as a counter-concept to the principle of a hub-and-spoke network design, which mainly intends to construct network to achieve economies of scale with less-redundant network structure (Klincewicz 1998, 2002; Campbell 2002).
Within this context, this research addresses a new hub location problem referred to as the reliable hub network design problem. More specifically, two sub-problems are addressed: 1) the reliable p-hub location problems (RPHLP) and 2) the p-hub protection problems (PHPRO). The first approach, reliable p-hub location models are proposed in terms of network reliability, which focuses on maximizing the network reliability to carry out flows among origin-destination pairs under normal circumstances. The p-hub protection problems are proposed as a survivable network design, which pay more attention to minimizing the potential loss of traffic in case of disruptions to network components. The key ideas and formulations of both problems are closely related to hub-and-spoke network problems since both approaches basically deal with the design of interactions or movements between an origin location and its destination via hubs.

Both approaches are basically derived from reliability theory. Accordingly, I will conceptualize key elements, reliability measurement, inter- and intra-hub reliability factors, and routing schemes. Proposed models for two approaches are applied for the infrastructure design of telecommunication networks.

1.2. Research Objectives

There are three main objectives with corresponding specific research questions. Each sub-problem intends to answer the questions below.

First, the main concern of the reliable p-hub dispersion problems (RPHLP) is to provide new mathematical models taking into account reliability-relevant components in hub and spoke network design. Then, two investigations, what reliability conditions
influence the geographical dispersion of hubs in order to maximize network reliability, and what impacts of mandatory dispersion of hub facilities on the network occur will be conducted.

Second, *p-hub protection problems* (PHPRO) are more focused on the resilience and capability of flow protection on the network, which are the main themes in survivable network design (Grötschel *et al.* 1995; Soni *et al.* 1999). In this dissertation, *p-hub protection* is defined as the ability of a network to maintain its capability to deliver traffic between nodes by minimizing potential traffic loss against possible failure on network components. Three different types of models are proposed based on the assignment schemes, and the relationship among survivability, network redundancy, and capacities will be examined.

Third, a method to evaluate designed reliable or survivable networks to any disruptions will be explored. A couple of good measures have been proposed to measure the vulnerability of given networks (for example, Grubesic *et al.* 2003). In this dissertation, a new measure which is called the *reliability envelope* will be presented and used to evaluate network performance or the consequence of network resiliency as the degree of failure worsens.

Finally, new heuristics will be developed to solve associated problems. Especially, the models, which are formulated based on single assignment scheme or survivable routing scheme, require an effective solution approach to handle a larger problem in practice. In this dissertation, a tabu-based solution approach and a hybrid type searching algorithm will be provided. The computational experiments will show the effectiveness
of both heuristics compared to the other solution approaches such as LP relaxation and enumeration based exhaustive search.

1.3. Organization of research

This dissertation is organized with the intent of enabling readers to focus on selected chapters (especially chapter 4, 5, and 6) without the need to go through the entire dissertation. However, readers without a background in hub location problems and reliability theory are encouraged to begin with chapter 2. The diagram in Figure 1.1 provides guidance for readers to access the contents of this dissertation. A brief description of each chapter is as follows.

• **Chapter 2**: Reviews the issues of reliability, survivability in network design, and hub location problems. These topics are the direct foundation in addressing the models of each problem. Readers who are already familiar with reliability theory and hub network design can directly go to chapters 3, 4, and 5. Research statements are also provided based on the literature reviews, followed by expected contributions of this dissertation research.

• **Chapter 3**: Presents a new hub location problem, termed the **reliable p-hub location problems (RPHLP)**, which aims to maximize network performance in terms of reliability by locating hubs for delivering flow among city nodes. Two sub-models, the **p-hub maximum reliability model (PHMR)** and the **p-hub mandatory dispersion model (PHMD)** are formulated. As an extension, the
multi-objective reliable p-hub mandatory dispersion model (MRDI) is also presented. In order to develop the series of models in to the subsequent chapters, this chapter focuses on describing how core model components: routing reliability, routing schemes, and reliability factors, are embedded into the models. An empirical analysis is presented using telecommunication networks in the United States.

• Chapter 4: Addresses new survivable network models, called the p-hub protection problems (PHPRO), as a survivable hub and spoke network design. As suggested by the term ‘protection’, the models take a preventive approach in which a network is designed to reduce traffic loss during either normal conditions or in the face of malfunctioning network components. Three standard models followed by three extensions are formulated.

• Chapter 5: Proposes two efficient heuristics, which are called ‘TABUSA’ and ‘TABUEX’. A tabu-heuristic, the TABUSA, aims to solve the problem with single assignment. The latter is a hybrid-type search approach and is prepared for the PROPS, which is the most complicated model among the models addressed in this dissertation.

• Chapter 6: The expected contributions from this research are summarized followed by further research directions.
Figure 1.1. The structure of the dissertation.
CHAPTER 2

RELIABILITY, SURVIVABILITY AND HUB NETWORK DESIGN

Traditionally, the issue of a high level of reliability or survivable in a network system has been well recognized as a critical factor in designing network infrastructure by not only practitioners but also researchers in various fields (Gavish and Neuman 1992; Grover and Tipper 2005). Hub location problems also have seen a great deal of attention in telecommunications and transportation networks design because hubs and their assigned nodes are found in many practical networks in telecommunications and transportation. In general, hub-and-spoke network system require constructing a network, which handles interacting flows among the number of origin-destination demands via a handful of hub facilities and fewer links (Daskin 1995; Campbell 2002).

1 A version of this chapter, “Reliable network design and hub location problems”, is to be submitted to *Korean Geographical Society* (2008).
In this context, this chapter reviews literature on the studies of network design problems with respect to two conceptual scopes – *reliability* and *survivability*. This chapter also covers the literature on hub network design problems, both of which are the foundation of the *reliable hub network design problem*. In section 2.1, I first review the design of reliable network problems. Section 2.2 deals with survivable network design. The nature and characteristics of each perspective are of concern. Section 2.3 provides the characteristics of hub location problems, focusing on hub-and-spoke network design models followed by the research statements of the *reliable hub network design problem* based on the limitations from previous works.

2.1. Reliable Network Design

One of the earliest work on reliable network design was done by Moore and Shannon (1956), who studied reliable circuit systems by the re-configuration of unreliable components in the system. From that time, various concepts of reliability have been developed subject to their research purposes and applications across system engineering, telecommunications and transportation (Ball 1979; Ball *et al.* 1995; Ball 1986).

The broadest definition of reliability is ‘the ability of a network to carry out a desired network operation’ (Colbourn 1987; Shier 1994). In a network, there would be a number of users and interacting movements, such as traffic or information flows among them, which should be delivered successfully. To satisfy this condition, it is required that the underlying network should ensure reliable transferring routes (or paths) between the
nodes. Further, the overall performance of the network is regarded as a function of its ability to satisfy this requirement (Colbourn 1999).

In a broad sense, the key difference between reliability and survivability relies on what circumstance is assumed for network operations. Reliability is more associated with the ability of successful communication or adequate performance with the absence of failures. In the field of system engineering, this concept is generally defined as the probability that a system or device will perform its intended function under a set of conditions when operational probability of network component is assumed (Wakabayashi and Iida 1992; Lewis 1996). Specifically, two approaches, deterministic and probabilistic, are identified in the reliability literature.

2.1.1. Deterministic approach

In the deterministic approach, reliability is defined as connectivity or connectedness, which is primarily focused on assessing the potential availability of network components to a given network, rather than designing an optimal network. In this perspective, the degree of network reliability is expressed as the 0-1 binary function, discrete score, or graph theoretical indices (Taaffe et al. 1996). Garrison (1960) used graph theory to assess the potential availability of nodes in the transportation network of the U.S. Such measures as connectivity and accessibility have been developed as the main analytical tools in geographical network analysis because these measures are recognized as simple but effective in comparing the performance level of network
components as long as the network could maintain the required connection (Kwan et al. 2003; Nicholson et al. 2003).

In recent research, these measures are still used in the spatial analysis of commercial Internet to highlight the level of criticality of city nodes (Malecki and Gorman 2001; O'Kelly and Grubesic 2002; Huh and Kim 2003; Gorman et al. 2004). Although the definition in this approach was conceived simply to examine the status of a system operation, it becomes the fundamental idea to extend into other reliability definitions.

2.1.2. Probabilistic approach

The idea of network reliability has been used often in telecommunications and transportation network literature, due to its conceptual appropriateness for applications. Network reliability is generally defined as the ability of a network to maintain connection between nodes against the probability of failure of components or loss of traffic on nodes or linkages (Colbourn 1987).

The degree of network reliability is expressed in the form of probability for successful communications among specific pairs of nodes. Table 2.1 summarizes the general classification of network reliability measures and its corresponding terms with respect to their purpose of measurements. For example, ‘O-D reliability’ is used to examine the operational probability of a path between origin-destination. Network reliability is not only used to analyze the network performance, but also applied as
requirements (i.e. objective function or constraints) in reliable or survivable network
design models (Kuo and Prasad 2000; Soni et al. 1999).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-D reliability</td>
<td>The probability that there exists at least one origin to destination</td>
</tr>
<tr>
<td>(Origin – Destination)</td>
<td>paths for a pair of nodes in network</td>
</tr>
<tr>
<td>k-terminal reliability</td>
<td>The probability that the network contains paths between each pair of nodes</td>
</tr>
<tr>
<td>(Origin – K destinations)</td>
<td>of the k nodes ( 2 \leq k \leq n )</td>
</tr>
<tr>
<td>All terminal reliability</td>
<td>The probability that the network contains at least one spanning tree for</td>
</tr>
<tr>
<td></td>
<td>every pair of nodes</td>
</tr>
<tr>
<td>Reachability</td>
<td>The probability that there are paths from 'origin' to every other node</td>
</tr>
</tbody>
</table>

**Table 2.1.** Types of reliability measurements.
In telecommunication networks, ‘reliability’ is defined as the probability of successful communication to deliver traffic without congestion or loss between OD pairs (Bell and Iida 1997). Since here the term ‘successful communication’ is open to interpretation, the most relevant definition is ‘the probability that a facility, such as a link and hub, performs its operations of transmitting flows for a period of time without failure’ (Yoo and Deo 1988; Wakabayashi and Iida 1992). Suppose that the reliability between two nodes is known when they are connected with a linkage. Then the reliability of the route which consists of \( n \) links is calculated as follows (see Colbourn 1987; Shier 1991):

\[
R_{OD}(G, r) = \sum_{i=1}^{n} \Pr\{E_i\} \\
\Pr\{E_i\} = \prod_{j=1}^{m} \Pr\{e_{ij}\}
\]

Where

\( R_{OD} \) the reliability for nodes origin \( O \) and destination \( D \)

\( G \) the route of a network with given probability \( r \) for edges

\( \Pr\{E_i\} \) the probability of the disjoint event \( E_i \)

\( e_{ij} \) the linkages \( j \) constituting \( E_i \)

Two simple cases are exemplified in Figure 2.1 to show how to compute reliability \( R_{ijkm} \), which represents the reliability of the flow route from node \( i \) to node \( j \) via hub \( k \) and \( m \) in that order (This term will be named routing reliability in chapters 3 and
4). Suppose that the operational probability of ‘0.9’ is given for links $r_{ik}$, $r_{km}$, $r_{mj}$ and ‘0.7’ is given for $r_{in}$, $r_{nj}$ as a known parameter. In the case of a serial route (Figure 2.1a) for a particular pair from $i$ and $j$, it is clear that only one successful event is enumerated as a set $E_1 = \{r_{ik}, r_{km}, r_{mj}\}$. So, the routing probability of 0.729 is computed by multiplying the probabilities of elements $r_{ik}$, $r_{km}$ and $r_{mj}$ directly. The probability of 0.729 indicates that route $R_{ijkm}$ can deliver 72.9% of traffic from $i$ to $j$ without delay via hub $k$ and $m$.

![Diagram](image)

**Figure 2.1.** Computations of Origin-Destination reliability.

In a parallel arrangement (Figure 2.1b), the Inclusion-Exclusion method is generally used to exclude double counting derived from unions among all successful events so that all

---

2 The operational probability ‘0.9’ and ‘0.7’ are simply given to explain the computation of reliability, and much of the literature commonly use this simple figure when testing or comparing the performance of network reliability.
successful events $E_i$ are treated as mutually disjointed. The general expression of the Inclusion-Exclusion method is as follows (Bell and Iida 1997, p.185):

$$R_{OD}(G, p) = \sum_{i=1}^{n} \Pr\{E_i\} - \sum_{i=1}^{n} \sum_{i \neq j}^{n} \Pr\{E_i \cap E_j\} + \sum_{i=1}^{n} \sum_{i \neq j}^{n} \sum_{i \neq j, l \neq m}^{n} \Pr\{E_i \cap E_j \cap E_l\} - \ldots$$

$$+ (-1)^{n-1} \Pr\{\bigcap_{i=1}^{n} E_i\}$$

(Equation 2.3)

In Figure 2.1b, two different routes are identified in pair from $i$ to $j$. First, $R_{ijkm}$ enumerated from the primary route set $E_1$, and secondary route $R_{ijn}$ from $i$ via $n$ to $j$ is enumerated from another set $E_2 = \{r_{in}, r_{nj}\}$. Reliability of this secondary route is computed as $r_{in} \cdot r_{nj} = 0.7 \cdot 0.7 = 0.49$. Therefore, the reliability for $i$ to $j$ is computed as:

$$R_{OD} = \sum_{i=1}^{2} \Pr\{E_i\} - \sum_{i=1}^{2} \sum_{i \neq j}^{2} \Pr\{E_i \cap E_j\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1\} \cap \Pr\{E_2\}$$

$$= 0.729 + 0.49 - (0.729 \cdot 0.49) = 0.8618$$

(Equation 2.4)

The reliabilities of primary route 0.729 as well as the secondary route 0.49 are summed since different routes are identified while the probability of double counting of $Pr\{E_1\}$ and $Pr\{E_2\}$ should be considered to make $E_1$ and $E_2$ be disjoint events to each other.

These computations can be more efficiently done if utilizing the Boolean algebra method. It is known that the Boolean algebra method generates the same result as
reduction both in enumerating disjoint events and in computation time (Fratta and Montanari 1973). Using this method, for the case of Figure 2.1b, first successful events, \( E_1 \) and \( E_2 \) are identified and then disjoint events \( D_i \) are generated by Boolean logic; \( D_i = E_i \cap \overline{E}_j \). The reliability then is computed as below.

\[
R_{OD} = \sum_{i=1}^{n} \Pr\{ D_i \} = \Pr\{ D_1 \} + \Pr\{ D_2 \} = \Pr\{ E_1 \} + \Pr\{ \overline{E}_1 \} \cdot \Pr\{ E_2 \} \\
= 0.729 + (1 - 0.729) \cdot 0.49 = 0.8618
\]

(Equation 2.5)

Where

\( \Pr\{ D_i \} \) the disjoint event computed by Boolean algebra

\( \Pr\{ \overline{E}_i \} \) the probability of the complement of event \( E_i \)

This research utilizes Boolean algebra method for reliability computation (see chapter 3 and 4 in more detail). Generally, reliability between two nodes on IP-based networks decreases with the length of the path because the transmission time of packet data or traffic delay rate is highly correlated with geographical distance and congestion level in the route (Agrawal 1997; Choi et al. 2004; Crovella and Krishnamurthy 2006).³

In transportation, the concept of reliability is differently applied because the characteristics of transportation networks are inherently different in terms of operation. In telecommunications, ‘congestion’ implies that the delay of transmission of flow or the potential loss of flow (i.e. packet loss) to reach the destination. In contrast, congestion in

³ The more detailed definition for this research will be provided in the chapter 3 and 4.
transportation means the length of travel time, or delaying time spent to reach the
destination. The interpretation is also different. For example, if the reliability between an
origin and destination is 0.9, it is defined as the probability that a trip flow from origin
can reach its destination without encountering congestion 90 times out of 100 (Iida 1999).
Another variant, time reliability defined as a successful probability of a trip from the
congestion within a specified time interval is used in transportation studies, and capacity
reliability is conceived to evaluate how well a network can accommodate a specified
demand level (Chen et al. 1999; D'Este and Taylor 2003). Table 2.2 summarizes the
definitions and interpretations used in previous literatures.

A number of studies on reliability have paid more attention to evaluating the
performance of a given network under simplified assumptions. As practical applications,
the work by Kansal et al. (1995) applied network reliability measure into water
distribution system to analyze the overall performance under uncertainty. In geography,
O’Kelly et al. (2006) employ network reliability to evaluate vulnerabilities of
telecommunication networks of U.S. commercial Internet backbones. By taking into
account the peering arrangements between Internet Backbone Providers, they compared
the O-D reliabilities among the U.S. city nodes and explored how each city node’s
reliability would be degraded in the face of potential malfunctions of hubs.
<table>
<thead>
<tr>
<th>Type</th>
<th>Definition and interpretation</th>
<th>Literature</th>
</tr>
</thead>
</table>
| Routing (path) reliability | The probability of successful communication to deliver traffic without congestion or loss between O-D pairs | • Wakabayashi and Iida (1992)  
• Yoo and Deo (1988) |
| s-t reliability      | The probability that nodes are connected to reach the destination from origin                   | • Wakabayashi and Iida (1992)  
• Bell and Iida (1997)  
• Iida (1999) |
| Travel time reliability | The probability that a trip can successfully reach within a specified time interval (or less than a specified cost) | • Asakura (1999)  
• Iida (1999)  
• Lam and Xu (1999)  
• Du and Nicholson (1997) |
| Capacity reliability  | The probability of the network that can accommodate a specific demand level (terminal reliability taking into account of capacity restriction of links) | • Chen et al. (1999)  
• D'Este and Taylor (2003) |
| Flow-decrement Reliability | The probability that the reduction in flow as a result of supply-demand interaction is not less than a threshold | • Nicholson and Du (1997) |

**Table 2.2.** Definition and interpretation of reliability in telecommunications and transportation.
The computation of reliability in a given network has been a critical issue since no one has yet discovered a polynomial-time technique for computing reliability. Although a body of literature has focused on developing better solution techniques (see Agrawal and Barlow 1984; Resende 1986; Ball et al. 1995 for a review of computational issues), the complete enumeration of all disjoint paths in a given network, or even an O-D, is a class of \(NP\)-hard (Fratta and Montanari 1973; Dotson and Gobien 1979; Yoo and Deo 1988). Given a small size problem, an exact method can be used using the inclusion-exclusion method and the Boolean-algebra method, both of which are the best known exact methods (Shier 1991; Ball et al. 1995).

**2.1.3. Reliability optimization problems**

The computational complexity of reliability measurements makes reliability optimization problems intrinsically difficult to get an optimal solution as well as to formulate the practical reliable network design. In other words, many reliability optimization problems cannot be solved successfully unless solid solution techniques are established (Ball 1979, 1986). Although a number of reliability optimization problems have been addressed in computing network reliability during the recent decades, however, the computational difficulty is still untreated, thus an optimal solution is not necessarily possible except for special cases. The main concern of most reliability optimization problems is to maximize system reliability \(R(x)\) by allocating redundant components or modifying the topology of a given network subject to given resource constraints such as system cost and network capacity (Colbourn 1999; Medhi 1999; Kuo
and Prasad 2000; Mohamed et al. 1992). In general, reliability optimization problems can be classified into three sub-areas.

First, the classical redundancy allocation problem aimed to design a system which can maximize overall system reliability by constructing serial, parallel, or both serial-parallel structures. For example, Mizukami (1968) addressed the model that maximizes system reliability with parallel redundancy in a serial system considering system failure. The comparison of reliability performance to the types of system configurations was conducted by Lim and Koh (1997), using the reliability function (referred to as MTBF) and system failure rate.

The design of reliable telecommunication networks has drawn a lot of attention from many researchers in the past. In particular, a mathematical model for reliable network design of a large telecommunication system was addressed by Gavish et al. (1989). (see also Gavish and Neuman 1992). The model aimed to construct a network that minimizes cost for building circuits available for all O-D pairs under a number of possible links, nodes, switching points, and traffic circuit requirements. Gavish (1992) also presented a model which minimizes traffic delay rates by taking into account of the probability of system breakdown and predetermined second paths in a given topology. However, the model takes multiple steps, searching to find the optimal value and network topology, beginning from random node selections. Thus, results produce a relatively high variance between feasible and optimal values.

From a different perspective, Raghavendra and Hariri (1985) addressed the problem to maximize $O-D$ reliabilities for a set of specific nodes. The objective is to determine the optimal level of investment on each link in the network to satisfy the
desired level of reliability. In terms of topological design, Chu et al. (2000) proposed a reliable telecommunication network for the DDS (Digital Data Service) network design. Their model aimed to find the minimum cost tree connecting a set of nodes using the Steiner tree structure. While most techniques and reliability models assumed that network connectivity is the only determining factor in network reliability, Belovich (1995) suggested a more practical network design problem which intends to enhance all-terminal network performance by modifying network topology in the given network. In a similar vein, Shao and Zhao (1997) addressed the optimal network design problem to identify some possible links to be added to the backbone network in order to improve overall network reliability to a certain desired level while minimizing total link costs.

Finally, in facility location problems, recent studies introduced the concept of reliability into a location model. Snyder and Daskin (2005, 2007) applied reliability theory in both the $p$-median problem and the uncapacitated fixed-charge location problem (UFLP). In their work, the concept of reliability was defined differently as the ability of performance when a part of the system failed. The objective function of the model is the same as the traditional one, i.e., to minimize fixed cost and transportation cost for assigned customers to facilities, but they also embed the expected failure cost to the possible failure of facilities into the models.

2.1.4. Summary of reliable network design

For the development of reliable hub network design models, three observations are noteworthy as summarized in this section. First, for a number of reliability
optimization problems, it is often hard to find the optimal or feasible solution even in the case of simple system structure because constructing networks under certain reliability constraints entails huge combinatorial problems. Due to this computational complexity, a number of studies have tried to develop or evaluate network reliability in the existing network or given topology, rather than developing optimization models. Second, many reliable network design models are only applied to simple and small networks. As results, these models were often regarded as impractical or limited in applying to real-world systems.

The development of solution techniques in reliability optimization is important. Reliability design problems in the literature usually use heuristics and approximation techniques to compute reliability efficiently, and further to find a desired level of solution quality or near optimal solution with a reasonable time (Colbourn 1999; Kuo and Prasad 2000 for an extensive survey on this issue). The design strategy generally consists of two steps to find optimal/near optimal solutions: 1) selecting a sufficient size of linkages to interconnect given nodes under a given cost, then 2) examine whether model could achieve the desired level of reliability or not.

2.2. Survivability in Network Design

Many survivable network design models have a tendency to incorporate reliability measurements related to network performance such as packet latency rate, quality of transmission, and/or traffic loss rate into their design procedures. The concept of reliability has an inherently similar root to survivability in the sense that both ideas
mainly concern network operation (Gavish and Neuman 1992). However, in contrast to reliability, survivability deals with the ability of a network to maintain its communicative capability in the face of failures or interdiction, rather than on how well it performs operations under normal conditions (Soni et al. 1999).

In a broad sense, survivability refers to a network’s ability to provide services continuously despite damage to network components such as nodes or links. Thus, the design of survivable networks has been a significant issue in network-based infrastructure across transportation, electric power systems, and telecommunications (Stavroulakis 2003; Klincewicz 2006; Murray and Grubesic 2007). The literature emphasizes the need for less-disruptable systems in critical infrastructure in order to maintain high levels service as well as for building networks resilient to potential incidents, e.g., natural disasters, electric outrages, and even intentional attacks (Cowie et al. 2003). It is obvious that there is a trade-off between network efficiency or cost and the level of robustness of a network (Wu 1992). As fewer routes can deliver a large volume of traffic, the network structure is simpler and less-costly, whereas the impact of any disruption becomes larger.

Two distinctive approaches are generally known in survivable network design (Pióro and Medhi 2004; Nucci et al. 2001). First, the protection approach refers to a preventive strategy put in place before a failure happens in a system. In contrast, the restoration approach focuses on how a network’s function recovers after the failure impacts the network. Thus, the former approach focuses on constructing a network that can minimize potential flow loss and reduce traffic latency by pre-specifying routes at normal conditions. Also, the structural or topological design needed to enhance network resiliency is of interest. In contrast, the latter approach deals with designing networks that
can keep functioning at tolerable levels despite malfunctions in the network components. Thus, techniques such as dynamic routing for traffic re-assignment and topological design for quick path-restoration in response to incidents are of particular concern (Newport and Varshney 1991).

The original idea of survivable networks was implicitly recognized in the middle 1960s by Baran (1964) who realized the importance of redundancy and how networks could be vulnerable to the possible random disruptions. A number of models on survivability thereafter were developed in such fields as computer engineering, operations research, electronic power systems, and telecommunications. In recent years, the design of survivable network is highlighted as a significant issue particularly in telecommunication systems due to the increasing and heavy social and business reliance on Internet, as well as the vulnerable characteristics of current telecommunication network topologies (especially Internet backbones).

Recent studies show that even small losses of traffic or services in current telecommunication systems would result in a disastrous impact on business, the public sector and even our daily lives (see Walkowiak 2004; O'Kelly et al. 2006 for examples). The literature of survivable network design problems can be generally categorized as five types according to the characteristics of the problems: 1) protective location problems, 2) topological design with connectivity, 3) design of self-healing network structure, 4) survivable routing design problems, and 5) vulnerability assessments of networks. Detailed literature of each category is summarized in Table 2.3.
2.2.1. Protective location problems: transmission facility design

The study by Baybars and Edahl (1988) is recognized as the first to address a classical protective model to design of locations of transmission facilities for telecommunication networks. They concerned where transmission facilities should be located in a given network with minimizing installation and link cost. They conceived the idea of alternative circuits for certain pairs of nodes, which can ensure the network keep its operations. Based on this initial idea, Agarwal (1989) and Medhi (1992, 1994) suggested similar models, so called facility survivable network models, by finding alternative shortest paths in a given topology under constraints of capacities and the number of circuits.

2.2.2. Topological design with connectivity

Another dimension of classical survivability problem was concerned with ‘connectivity’ to survive any incident cut on links. As a protective strategy, in order to avoid disconnecting from a network, a node should have at least two linkages connecting other nodes. Monma and Shallcross (1989) suggest the idea of minimum connectivity requirements to protect certain nodes from linkage failures. They designed a two-connected topological network, which is the lowest level of connectivity. Followed by

---

4 Before the models shown in previous literature, a couple of good models were suggested in the context of survivability in the past (e.g. Boesch and Thomas 1970).
this work, cost-minimizing network design satisfying required connectivity has been an important issue (see Cardwell et al. 1989; Grötschel et al. 1992; Koh and Lee 1995; Clarke 1992; Clarke and Anandalingam 1995). Their common idea was to identify an optimal network in a given topology, or optimize the network where some special nodes or all nodes can survive any single link cut. According to the degree of connectivity, some models require higher connectivity; greater than two linkages are connected to particular nodes. The objective of most models is to minimize the building costs for links, interconnections, and transmission facilities. In more recent works, the problem of path protection with varying the connectivity levels was addressed by Xin and Rouskas (2004).

2.2.3. Design of self-healing network structure

As a restoration approach, the idea of self-healing network design was proposed by Kawamura et al. (1994), and this concept has been extended into the survivability techniques to keep the network with an acceptable level of performance in the face of any possible failures (Louca et al. 1999; Redman and Warren 2005; Ellison et al. 1999). The main concern is to find the condition of self-healing in particular network structures (Armony et al. 2000; Grover and Doucette 2001; Sreenath et al. 2001) so that a network can be restorable to a possible failure. The models employing this technique were extensively studied in such particular networks as self-healing ring (SH) design, mesh-restorable topology, and ATM (Asynchronous Transfer Mode) or WDM (Wavelength Division-Multiplexing) networks. The objective of these studies still focused on cost-
effective networks with limited resources such as bandwidths, capacity, and allowable paths.

A dual-homing telecommunication network is also classified as a self-healing survivable network design (Wu 1992). As Lee and Koh (1997) addressed the ring-chain architecture, offices in a chain are connected to the ring network in dual-homing fashion to increase its resiliency to possible malfunction. This technique has been well applied in most hub-and-spoke type telecommunication networks (Klincewicz 2006).

2.2.4. Survivable routing design problems

The issue of routing has been found to be an important dimension of survivability in the literature. The main focus is based on finding an optimal set of disjoint paths or circuits in a given network while minimizing the total cost of paths. Constraints usually consider how to ensure key paths (or primary routes) by finding other back-up paths within the capacity of the links over the given network. Torrieri (1992) suggested a model to maximize the number of shortest disjoint-paths for particular set of nodes. Pirkul and Narasimhan (1994) and Amiri and Pirkul (1996) also proposed flow protection models in terms of minimizing the total queuing delay cost in carrying traffic using pre-determined best alternative routes to the primary paths in the event of primary link failures. Following these works, various re-routing strategies to protect flows and heuristics have been proposed. For example, Louca et al. (1999) and Ghashghai and Rardin (2002) studied a survivable path problem for a special topological network, and Walkowiak (2004) addressed another type of survivable network design model to
determine alternative routes to the failure of main routes under a node capacity constraint. It should be noted that their models basically used re-routing strategies as pre-determined back-up paths to protect the primary paths, thus the network can maintain its route survivability.

A final type of survivable network design is hop-constrained network design. Generally, the problem can be stated as identification of the optimal paths for a commodity for particular nodes according to a specified limited number of hops and link costs. Balakrishnan and Altinkemer (1992) first addressed this problem, and a new mathematical formulation was presented by Pirkul and Soni (2003) by relaxing the integer programming into a linear problem. The optimal procedure of most survivable network models was limited to solving smaller size problems. As presented in Table 2.3, of particular concern of most survivable network design models is to minimize the network cost while at the same time satisfying given survivability requirements. However, like reliable network design, most models generally require various heuristic solution techniques to solve the problems efficiently (Keirivin and Mahjoub 2005).
<table>
<thead>
<tr>
<th>Type</th>
<th>Problems</th>
<th>Literature</th>
<th>Model description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protective Network design</strong></td>
<td>• Locating transmission facility in a given network</td>
<td>• Baybars and Edahl (1988)</td>
<td>• Min cost of facility and link installation/ circuit capacity, the number of circuits</td>
</tr>
<tr>
<td></td>
<td>• Finding Alternative circuits</td>
<td>• Agarwal (1989)</td>
<td>• Min cost of routes and facilities/ link capacity, the number of trunk lines</td>
</tr>
<tr>
<td></td>
<td>• Transmission &amp; facility</td>
<td>• Medhi (1994, 1992)</td>
<td>• Min cost of both traffic and facility networks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Topological design with connectivity</strong></td>
<td>• Designing network from a given topology satisfying connectivity requirements</td>
<td>• Monma and Shallcross (1989)</td>
<td>• Min cost under low connectivity constraints for certain nodes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cardwell et al. (1989)</td>
<td>• Min link costs under high connectivity constraints for certain nodes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Grötschel et al. (1992)</td>
<td>• 2-connectivity with capacity constraints</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clarke (1992)</td>
<td>• Min total number of wavelengths to find protective path and the least number of links assignment and connectivity constraints</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Clarke and Anandalingam (1995)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Koh and Lee (1995)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Xin and Rouskas (2004)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Topological design with particular structure</strong></td>
<td>• Designing Self-Healing network structure</td>
<td>• Kawamura et al. (1994)</td>
<td>• Min cost of Self-healing network based on virtual paths</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Louca et al. (1999)</td>
<td>• Min cost of Self-healing network by finding k-best paths in K-Trellis graph</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Armony, Klinewicz, Luss, and Rosenwein (2000)</td>
<td>• Min interconnection cost of Self-healing Ring network structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Grover and Doucette (2001)</td>
<td>• Min fixed cost for edge building and capacity cost based on mesh-based network structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lee and Koh (1997)</td>
<td>• Min cost with ring-chain network with dual-homing for special nodes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sreenath et al. (2001)</td>
<td>• Min total fiber requirement to build primary and restorative network via virtual paths</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Max spanning trees to the given fully meshed networks</td>
</tr>
</tbody>
</table>

**Table 2.3.** Classification of survivable network design and models.
Table 2.3. Continued

<table>
<thead>
<tr>
<th>Type</th>
<th>Problems</th>
<th>Literature</th>
<th>Model description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survivable routing design</strong></td>
<td><em>Alternative path or route problem</em></td>
<td>• Torrieri (1992)</td>
<td>• Max the number of shortest disjoint-paths for particular nodes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Pirkul and Narasimhan (1994)</td>
<td>• Min queuing message delay in a given topology/ link capacity, link disjoint constraint</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Amiri and Pirkul (1996)</td>
<td>• Min link costs for finding k-best paths, node disjoint paths</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Louca <em>et al.</em> (1999)</td>
<td>• Min the lost traffic from link failure based on pre-defined local destination re-routing scheme/ node capacity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ghashghai and Rardin (2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Walkowiak (2004)</td>
<td></td>
</tr>
<tr>
<td><strong>Hop-constrained network design</strong></td>
<td><em>Designing network with hop-constrained nodes or links set</em></td>
<td>• Balakrishnan and Altinkemer (1992)</td>
<td>• Min fixed cost and variable cost for network design with a given hop constraint set</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Pirkul and Soni (2003)</td>
<td>• Min cost under traffic flow constraints and capacity within a two-hop constraints in case of link failures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ouveysi <em>et al.</em> (2003)</td>
<td></td>
</tr>
</tbody>
</table>
2.2.5. Vulnerability assessments

Evaluating potential vulnerability including system robustness, survivability, and reliability to possible failures is one of the most important themes in network-based infrastructure systems (Grubesic et al. 2008). While a number of survivable network designs and models are found in the literature, effort to examine the vulnerability of network has been recently taken in terms of reliability, survivability and network interdiction (Grubesic et al. 2003; O'Kelly and Kim, 2007; Grubesic and Murray 2006; Church et al. 2003). In terms of survivability, Grubesic et al. (2003) examined how survivability of the commercial Internet in the United States would change if some selected city nodes were disrupted. In their research, survivability was defined as the smallest amount of damage to cause a network disconnection. However, they employed all-or nothing assumption which implies either 100% nodal and linkage operation or total loss of its functionality in the given networks. O’Kelly and Kim (2007) simulated the degradation of the complicated but realistic characteristics of telecommunication networks in the United States by taking into account of the critical role of hubs with peering arrangements between Internet service providers.

Figure 2.2, which is named as ‘reliability envelope’, reflects how network performance could be degraded in the event of disruptions (i.e. unintended failure, interdiction, and attack. etc). When a disruption impacts $r$ out of $n$ components at $r$ stage where the number of $r$ failure occurs ($r = 1 \rightarrow n$, $n$: the total number of linkages), there is a range of consequences among the $r$ chosen failures, from the relatively unimportant
(best-scenario) to the highly damaging (worst-scenario). The chances that $r$ failures produce minimal damage decline with an increase of $r$.

Figure 2.2. Variations of network performance on the disruptions of network failures.

(source: O’Kelly and Kim 2007)
The ranges between upper and lower bounds at each of \( r \) stages show how vulnerable the network is. If the network can keep the range of impact narrow even as disrupted components \( r \) increase, then this indicates that the network has a good resiliency against disruptions. The larger the range, however, the more susceptible the network is to failure.

As discussed by Kim and O’Kelly (2006), the desired level (or tolerant level) can be discussed in terms of network design. The most susceptible network components whose failure may cause significant degradation of network can be examined using this envelope by allowing for the exploration of such pertinent issues as determining which components should be fortified to improve the survivability of the network. The basic idea of the envelope is simply applied in other measures. For example, the \( Y \)-axis can be replaced with flow-activity level or connectivity of a network according to what network design approach is taken. The \( X \)-axis also can indicate the level of other interdictions on particular network components. This idea will be discussed in detail in chapter 4.

As a same vein of this simulation-based approach, optimization models are also addressed in order to identify the worst- or the best-scenario to the interdictions using mixed integer programming (Scaparra and Church 2008; Murray et al. 2007; Grubesic et al. 2008; Church et al. 2007). The advantage of this approach is not only to identify the worst-scenarios but also to provide what resources need protected based on the both identified scenarios. This clearly informs network designers and planners to design more resilient and survivable networks (see Snyder et al. 2006 for recent issues).
2.2.6. Summary of survivable network design

The characteristics of survivable network problems are summarized as follows. First, most survivable models focus on finding the optimal topological design to minimize network costs while locating telecommunication facilities and links in a given network. As survivability requirements, connectivity and alternative paths are mainly considered, often special restorable network structures such as ring, mesh, disjoint paths are modeled. Second, many survivability models consider the network performance in the case of possible failure on any topological component; however, due to the complexity of the problem, most classical survivable models are developed under simplified assumptions with limited a simple survivable requirement. For this reason, substantial characteristics in real telecommunication network such as interactions among nodes are often ignored in many applications.

2.3. Hub Network Design Problems

This section reviews the literature of hub-and-spoke models directly relevant to telecommunication network design. The next section provides the fundamental hub location models with general characteristics, and section 2.3.2 describes hub location models which are applied in telecommunication network designs. Finally, the characteristics and limitations of previous researches are discussed in section 2.3.3. The solution approach for hub location problems will be dealt with in Chapter 5.
2.3.1. Hub location problems and network design

Since the hub location problem was formulated by O’Kelly (1987), a number of models and variants have been developed. Most hub location models deal with optimizing two levels of objectives, 1) locating hubs, and 2) allocating non-hub nodes to hubs for specifying routes for given origin-destination flows by utilizing inter-hub links (Campbell 1994).

The problem of locating hubs is classified as two different problems - 1) to find optimal location of $p$-hubs, or 2) to determine the number of hubs endogenously with given constraints and cost structure. Assignment schemes are also classified based on the routing scheme, which is how non-hub nodes are assigned to the hubs.

As illustrated in Figure 2.2, single assignment requires that each non-hub node should be connected to a single hub. All traffic for a given non-hub node must travel through the single hub to which it is connected. In contrast, multiple assignment allows each node to be connected more than one hub. Each O-D pair should be routed via at least one hub to carry out its demand from origin to destination. A route with more than one hub should utilize hub to hub linkages so called as inter-hub links experiencing the advantage of economy of scale by aggregating of flows. The hubs are generally assumed fully connected to each other unless there is a capacity constraint to limit the amount of traffic on inter-hub linkages or in the hubs themselves.

A number of hub location models aim to minimize total flow-costs or network costs, but maximizing network performance is often considered as an important objective
function. The hub location problem was originally formulated as a quadratic assignment program utilizing integer variable $X_{ik}$, which represents the link between none-hub node $i$ to a hub at node $k$. By multiplying three quadratic such terms as $X_{ik} \cdot X_{jm} \cdot C_{km}$, the routing $i \rightarrow k \rightarrow m \rightarrow j$ is completed with the cost $C_{km}$. Since the linearized form of the routing variable $X_{ijkm}$ was proposed, a variety of hub location problems based on linear programming to find an optimal integer solution have been addressed (Campbell 1994; Klincewicz 1994; Bryan and O’Kelly 1998).

![Hub-and-spoke network: (a) Single assignment, (b) Multiple assignment.](image)

**Figure 2.3.** Hub-and-spoke network: (a) Single assignment, (b) Multiple assignment.
2.3.2. Design perspective of hub-and-spoke network design in telecommunications

Telecommunication network design before the advent of fiber-optic technology in the 1990s were targeted for copper-based networks having inherent limitations such as extremely small bandwidth and capacity (Grötschel et al. 1995). In a copper-based network, a greater number of switching points and links are required in order to minimize the costs, because the ability of hubs and physical links for transmission was quite limited within a small distance. However, with IP-based network design, the network performance is highly dependent upon the capability of a handful number of hubs and backbones. The importance of hub location models is more vital in current telecommunications network design (Gourdin et al. 2002; Klincewicz 2006).

Hub location models for telecommunication networks differ from those of transportation network design models in several ways. First, the model employing single assignment scheme makes good sense for cost-effective telecommunication network design (O'Kelly and Bryan 2002; Klincewicz 1998), whereas multiple assignment models is considered if reliability of the system is more of an issue (Daskin 1995).

Second, the cost for building hubs and linkages (i.e. physical facilities) is the most important concern in telecommunication networks because the flows on the inter-hub links or spokes does not directly influence to the network costs. In contrast, in air-transportation, flow-based costs are the main concern of the model.

Third, more importantly, current models are required to take the measures representing the QoS (Quality of Service) such as delay and loss of traffic into account in
the model formulations, coupled with topological requirements such as connectivity and multiple assignments to ensure a reliable or survivable network system (Sherali et al. 2000; Campbell 1996; Daskin 1995).

In a number of studies dealing with telecommunications in other fields, hub location problems are often recognized as backbone and tributary network design problem (Carello et al. 2004; Klincewicz 1998). From this perspective, the network is constructed with two-levels, and traffic from access nodes of a tributary network must be routed through the backbone network via transit nodes to reach other desired access nodes. An access node sometimes simply represents a kind of tributary network such as ring/star/tree networks. Here, access node, transit node, and backbone links are analogous to the non-hub node, hub and inter-hub links in hub location models, respectively. The assignment scheme is also named differently as single-homing and double-homing, instead of using single/multiple assignments.

As one of the earliest work on this issue, Monma and Sheng (1986) presented the design of a packet switch system in terms of hubs (switching center) and nodes. The design technique consisted of two phases – first determining the number of hubs, then evaluating network performance in terms of transmission delay. Then a search for a better solution with iterative heuristics is considered where clustering procedures are modeled to select hubs for each clustered node set. With respect to a dual-hierarchical structure, the design problem of establishing a fully connected backbone and assigning of tributary networks to hubs was studied in by Chung et al. (1992) and Kim et al. (1995) based on
single-homing and multiple-homing, respectively. However, the flows over nodes pairs were not explicitly included in their models.

The design of computer network systems can be regarded as a hub location problem in a broad sense. Pirkul et al. (1988) studied the design of computer network system where each non-hub node should be connected to a concentrator via primary and secondary homing. Lee (1993) also addressed the similar but different type of assignment problem related to the computer network system that determines the location of multi-types of concentrators, subject to various capacity requirements and costs conditions.

The topological design of two level of telecommunication systems with double homing to satisfy a sufficient level of redundancy was considered by Kim et al. (1995). This work was more focused on the connectivity requirement of each access node in terms of network survivability. As an iterative network design technique, Chamberland and Sansò (2000) addressed the network expansion problem for the metropolitan area network (MAN) by updating the configuration of switches, additional access networks, and expansion of the backbones. A multiple-ring topology is considered for the backbone network in the model to protect the backbone network against single-link or switch failure.

As a variant of the hub and spoke network, the design problem of DDS (Digital Data Service) networks was addressed by Lee et al. (1994, 1996). The differences from the conventional hub-and-spoke problem are that the inter-hub linkages are replaced by the degree-constrained Steiner tree that connect the hubs, and the cost reflecting the tariff structure of telecommunication systems was used in the objective instead of a flow-
dependent costs structure. Finally, the cost-allocation problem based on game theory in the hub-and-spoke telecommunication system was also studied by Podnar et al. (2002) and Matsubayashi et al. (2005). They focused on a cost-allocation game according to cost structures in uncapacitated hub-and-spoke problem with single allocation. Table 2.4 summarizes representative hub location problems and related models, which are applied in telecommunication network design.
<table>
<thead>
<tr>
<th>Problem</th>
<th>Addressed models</th>
<th>Literature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Assignment (SA)</td>
<td>▪ SA with Quadratic Assignment Problem (QAP)</td>
<td>▪ O'Kelly (1987)</td>
<td>▪ Single assignment with quadratic formulation</td>
</tr>
<tr>
<td></td>
<td>▪ SA with Quadratic Assignment Problem (QAP)</td>
<td>▪ O'Kelly (1986a)</td>
<td>▪ Heuristic approach to examine hub facility location with activity level</td>
</tr>
<tr>
<td></td>
<td>▪ SA with Quadratic Assignment Problem (QAP)</td>
<td>▪ O'Kelly (1986b)</td>
<td>▪ SA problem in continuous space</td>
</tr>
<tr>
<td></td>
<td>▪ SA with linearization formulation</td>
<td>▪ Klincewicz (1991)</td>
<td>▪ Applying Exchange heuristic</td>
</tr>
<tr>
<td></td>
<td>▪ SA with linearization formulation</td>
<td>▪ Klincewicz (1992)</td>
<td>▪ Applying GRASP heuristic</td>
</tr>
<tr>
<td></td>
<td>▪ SA with linearization formulation</td>
<td>▪ Skorin-Kapov and Skorin-Kapov (1994)</td>
<td>▪ Applying Tabu search</td>
</tr>
<tr>
<td></td>
<td>▪ Advanced heuristics</td>
<td>▪ Campbell (1996)</td>
<td>▪ Applying ALLFLO, MAXFLOW</td>
</tr>
<tr>
<td></td>
<td>▪ Advanced heuristics</td>
<td>▪ O'Kelly et al.(1996)</td>
<td>▪ Applying exact solution with Tabu and variable reduction technique</td>
</tr>
<tr>
<td></td>
<td>▪ Advanced heuristics</td>
<td>▪ Pirkul and Schilling (1998)</td>
<td>▪ Applying Lagrangian &amp; Sub-gradient relaxation</td>
</tr>
<tr>
<td></td>
<td>▪ SA with variable number of hubs</td>
<td>▪ O'Kelly (1992)</td>
<td>▪ Fixed cost, the number of hubs are determined as endogenously</td>
</tr>
<tr>
<td></td>
<td>▪ SA with variable number of hubs</td>
<td>▪ Aykin (1994)</td>
<td></td>
</tr>
<tr>
<td>Multiple Assignment (MA)</td>
<td>▪ MA with fixed cost with heuristic</td>
<td>▪ O'Kelly et al.(1996)</td>
<td>▪ Exact solution with Tabu and variable reduction technique</td>
</tr>
<tr>
<td>Multiple Assignment (MA)</td>
<td>▪ MA with fixed cost with heuristic</td>
<td>▪ Skorin-Kapov et al. (1996)</td>
<td>▪ Tight LP relaxation</td>
</tr>
<tr>
<td>Multiple Assignment (MA)</td>
<td>▪ MA with fixed cost with heuristic</td>
<td>▪ Ernst &amp; Krishnamoorthy (1998)</td>
<td>▪ Exact method for uncapacitated MA</td>
</tr>
<tr>
<td>Multiple Assignment (MA)</td>
<td>▪ MA with variable number of hubs</td>
<td>▪ Bryan (1998)</td>
<td>▪ MA the number of hubs are determined as endogenously</td>
</tr>
</tbody>
</table>

Continued

Table 2.4. Representative studies and their models associate with hub location problems.
<table>
<thead>
<tr>
<th><strong>Multiple Assignment (MA)</strong></th>
<th><strong>Other applications in telecommunication</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ MA with flow dependent cost function</td>
<td>▪ Backbone and tributary network design</td>
</tr>
<tr>
<td>▪ O’Kelly and Bryan (1998)</td>
<td>▪ Design of distributed FTN system</td>
</tr>
<tr>
<td>▪ Embedding flow-dependent cost function in the objective (FLOWLOC model)</td>
<td>▪ Monma and Sheng (1986)</td>
</tr>
<tr>
<td>▪ Flow dependant cost &amp; extension with capacitated interlink</td>
<td>▪ Chung <em>et al.</em> (1992)</td>
</tr>
<tr>
<td>▪ Capacity restriction on a hub</td>
<td>▪ Kim <em>et al.</em> (1995)</td>
</tr>
<tr>
<td>▪ Capacity restriction on inter-hub linkages</td>
<td>▪ Yoon <em>et al.</em> (1998)</td>
</tr>
<tr>
<td>▪ Capacity restriction on a hub with heuristic with shortest paths</td>
<td>▪ Two-hierarchical structure design, determine the location of switching centers and assigning node to nearest hubs/considering single/dual homing</td>
</tr>
<tr>
<td>▪ Design of computer network system</td>
<td>▪ Min network costs to build Fiber Transport Network for locating hubs and conduits and linkages</td>
</tr>
<tr>
<td>▪ Local access transport area problem</td>
<td>▪ Lee (1993)</td>
</tr>
<tr>
<td>▪ Network expansion problem</td>
<td>▪ Pirkul <em>et al.</em> (1988)</td>
</tr>
<tr>
<td>▪ Multiple ring topology design to assign additional access network to the expansion of metropolitan area network</td>
<td>▪ The location of multi-type concentrators in computer system/primary and secondary homing</td>
</tr>
<tr>
<td>▪ Min total building cost for locating Point Of Presences allocation of clients</td>
<td></td>
</tr>
</tbody>
</table>
2.4. Research Statements

This research can be highlighted by three major differences compared to previous researches. First, *reliability* theory is new in location modeling, and even in geography. In particular, this research would introduce the first hub location models to incorporate the concepts and measurement of both *reliability* and *survivability* into a hub location problem.

Second, the idea of *reliable-hub and spoke models* has many advantages compared to previous reliable or survivable network models, which are characterized as being under *topology-oriented approach*. As mentioned, the particular concern of the topological approach models is often limited to design topologically robust networks, thus the substantial characteristics of real-world network systems are often ignored or simplified. It should be noted that current IP-based telecommunication networks are very close to the hub-and-spoke type network, consisting of *hubs*, interacting flows among *O-D* nodes, and *backbones*. Accordingly, the approach of reliable network design can be refined by incorporating the concepts of reliability and survivability into hub location problems. In addition, from this point of view, this research aims to provide insight for network design with more realistic solutions from geographic perspective.

Third, as pointed by many researchers (Soni *et al.* 1999; Koh and Lee 1995; Grötschel *et al.* 1995), few previous models addressed what conditions could influence topological change (location of hubs) of networks. Both reliable *p*-hub location problems and *p*-hub protection models are expected to explain what conditions influence the geographic dispersion or closeness of hubs, or how the optimal configurations of network
are changed in terms of survivable network design. In a similar vein, virtually no research in geography addresses the factors that make hubs more geographically dispersed or clustered in terms of network reliability although the potential advantage of geographical dispersion of hub facilities has been implicitly stressed as a defensive strategy in the literature.
CHAPTER 3

RELIABLE P-HUB LOCATION PROBLEMS FOR HUB NETWORK DESIGN

This chapter addresses reliable p-hub location problems (RPHLP). As mentioned in chapter 1, the RPHLP can be classified as a reliable hub network design. The main purpose of this chapter is to examine the relationship between network performance and hub location, and impact of dispersion of hub facilities on the network through the proposed models. Hypothetical and empirical analyses are presented using telecommunication networks in the United States.

3.1. Introduction

Early telecommunication networks were as decentralized as possible to achieve a defensible outcome. For example, a defensive locational philosophy underpinned the

---

5 A version of this chapter, “Reliable p-hub location problems in telecommunication networks”, co-authored with Dr. Morton E. O’Kelly, has been accepted in Geographical Analysis (July 21, 2008).
earliest designs for ARPANet where important hubs were strategically dispersed to protect against probable attacks or failures (Boehm and Baran 1964). With increasing commercialization, however, networks have evolved into more centralized systems. As a result, there is an intense concentration of key functions at select hubs. Moreover, Internet hubs often are located close to each other for reasons of efficiency (Pastor-Satorras and Vespignani 2004).

As discussed extensively in Murray and Grubesic (2007), this functional and geographical centralization of hubs comes at a time when the need to protect high-speed and broadband technologies and their networks is a critical issue. Even though geographical proximity among hubs reduces the possibility of traffic loss in transmitting aggregated flows, this benefit could be compromised by the increased concentration of hub activities, which could potentially degrade network performance in the face of disruptions (Wagner 2002; BBC News 2002a, 2002b; NCA 2004).

The potential advantage of geographical dispersion of hubs is stressed by O’Kelly and Kim (2007), who analyze the recent cascading failure of Internet services in Korea in terms of network reliability. They demonstrate that the damage to just a single hub can cause severe impacts on the entire network when hub facilities are highly clustered with geographical closeness (see also O'Kelly, Kim and Kim 2006). Another practical example is provided by the recent disruption of Internet service due to the 2006 earthquake in Taiwan, which occurred in an area where undersea cables are highly concentrated, subsequently causing severe telecommunications problems between continents. This highlights the importance of dispersing telecommunication components to maintain the reliability of telecommunication systems at a desired level (Heiskanen 2006).
Within this context, the RPHLP mainly focuses on locating $p$-hubs on the network in order to improve network reliability to deliver interacting flows among the set of origin-destination nodes. As mentioned in the chapter 1, the RPHLP comprises two-sub models depending on whether the model considers the rationale of dispersed facility location into the model, and each sub model has two different mathematical models based on the assignment schemes.

The first sub-model is the $p$-hub maximum reliability model ($PHMR$). This model optimizes a hub network by focusing solely on maximizing the completed flows among the set of nodes via $p$-hubs based on the level of reliability, which is imposed on such network facilities as hubs and inter-hub links. Of particular concern to the PHMR aims to explore the model behaviors when reliability-relevant components are embedded in the hub location problem.

Second, based on the PHMR model, the $p$-hub mandatory dispersion model ($PHMD$), is developed. The PHMD determines the optimal hub locations in order to improve network reliability while retaining the mandatory dispersion of hubs that requires hubs to be farther apart than a certain minimum separation. Although various types of facility dispersion problems have been addressed in the literature, the rationale for dispersion of hubs has rarely been explored (see Curtin and Church 2006 for a recent review of dispersion models). The main purpose of the PHMD is therefore to investigate the impacts of hub facility dispersion with respect to network performance, the flow pattern, and flow activity levels of hub facilities. The PHMD ultimately aims to mitigate a potential damage from a possible malfunction on hub facilities with excessive activity levels. As an extension of the PHMD model, I also provide the mutiojective reliable $p$-
hub dispersion model (MRDI), where both objectives, network reliability and the level of hub separation, are determined simultaneously.

The term $p$-hub mandatory dispersion used in this chapter is initially different from that of ‘the $p$-dispersion problem’ addressed by Kuby (1987). Conventionally, the $p$-dispersion problem focuses on locating $p$-noxious or undesirable facilities among $n$ candidate nodes. This characteristic suggests that the facilities should be located as far away as possible from the nearest one so that an accident at one of the facilities will not impact the others. Moreover, the goal of $p$-hub mandatory dispersion models is to avoid concentrating strategic assets in one area, so as to reduce the growing vulnerability that may arise from probable congestion or disruptions in a network.

This chapter is organized as follows. The relevant research is reviewed in the next section, and sections 3.3 and 3.4 develop the models and provide formulations, respectively. The computational results and analysis are presented in section 3.5 followed by summary of findings.

3.2. Related Research

The issue of a reliable system has been recognized as a critical factor in the design of network infrastructure (Gavish and Neuman 1992; McGrath 2003; Herder and Thissen 2003; Konak and Smith 2006). In general, reliability planners have developed models to maximize system reliability by allocating redundant components to systems with a given set of resources (Mohamed, Leemis, and Ravindran 1992). Many telecommunication network models seek either to maximize network performance for the particular set of
origin and destination nodes by increasing the availability of links, or to minimize the building costs of a network subject to reliability-relevant requirements (Gavish et al. 1989; Forsgren and Prytz 2006). Because most reliability optimization problems are difficult to solve due to their computational complexity (see Kuo and Prasad 2000), often practical design approaches are proposed. For example, Shao and Zhao (1997) present a reliable telecommunications network design model that focuses on finding the best linkages to add to the existing backbones so that a network can reach a certain level of reliability while minimizing network costs.

Many reliable network models in telecommunications, however, are more focused on the topological optimization of networks, and thus complicated network properties are often ignored in the models (Soni, Gupta, and Pirkul 1999; Carello et al. 2004). For example, the main characteristic of a hub, which is defined as a set of interacting facilities for the switching and consolidation/concentration of flows, often used to be characterized as simply the more important node out of a number of nodes in a system (Campbell 1994a).

In recent research, the idea of a reliable system has been applied in traditional location problems such as the p-median problem (Snyder and Daskin 2007). Even though the need to address a reliable network model in terms of a hub-and-spoke network design has been stressed, few models are found for this in the literature (Klincewicz 1998, 2006). In particular, incorporating performance-related measures (e.g., reliability, traffic loss rate) in a hub network design is recognized as the most important issue because ensuring good performance for transferring interacting traffic becomes a critical concern
in current telecommunication networks due to its delay-sensitive nature (Klincewicz 2006; Skorin-Kapov et al. 2006).

3.3. Model Developments

The two main variants in hub network design, which are generally referred to as single assignment and multiple assignment models are considered in the RPHLP. As reviewed in chapter 2, the single assignment model (SA) restricts each non-hub node to interact with only one hub so that all flows from an origin must travel to the same hub. In contrast, the multiple assignment model (MA) is more flexible in routing by allowing each node to interact with more than one hub. In order to formulate the RPHLP, we first introduce two reliability-relevant components: (1) the routing reliability $R_{ijkm}$ and (2) the reliability factors $\alpha$ and $\gamma$, both of which are different compared to conventional hub location models. The data that are used in our model experiments are described in section 3.3.3.

3.3.1. Routing reliability $R_{ijkm}$

In telecommunication networks, reliability is defined as the probability of successful communication to deliver traffic without congestion or loss between OD pairs (Bell and Iida 1997). Because here the term ‘successful communication’ is open to interpretation, the most relevant definition; i.e., the probability that a facility, such as a link and hub, performs its operations of transmitting flows for a given period of time
without failure, is employed (Wakabayashi and Iida 1992). Based on this definition, the routing reliability \( R_{ijkm} \) is defined as a probability of successful delivery of flow for the routing variable \( X_{ijkm} \), which is generally used in many hub location models to represent the path that flow from origin \( i \) to destination \( j \) is to be routed via hubs \( k \) and \( m \) \((i \rightarrow k \rightarrow m \rightarrow j)\).

As mentioned in chapter 2, reliability between two nodes decreases with the length (distance) of a path since the transmission time of packet data or the traffic delay rate is highly correlated with physical distance and the congestion levels in the route (Murnion and Healey 1998; Choi et al. 2004; Pióro and Medhi 2005; Crovella and Krishnamurthy 2006). Thus, of great concern is determining the routes whose reliabilities maximize interacting flows delivered among a number of sources and destinations on a network (Stavroulakis 2003).

### 3.3.2. Routing scheme: reliability factors \( \alpha \) and \( \gamma \)

As indicated by \( X_{ijkm} \), the routing reliability \( R_{ijkm} \) is basically calculated by sequentially multiplying the reliabilities of each link \( r_{ik}, r_{km}, \) and \( r_{mj} \). However, performance-related facility conditions need to be reflected in telecommunication networks because the physical aspects of network facilities are related to network performance. For instance, a link with a larger capacity or bandwidth can transmit traffic for a longer distance without the loss of flow, noise, and the attenuation of signals, all of which are expected to increase with geographical distance. The capacity and technological level of hub facilities are designed to keep the reliability of traffic at a
desired level (Pióro and Medhi 2004; Zheng et al. 2005). In order to reflect performance-related facility conditions in telecommunication networks, the reliability factors $\alpha$ and $\gamma$ are introduced into the RPHLP.

The *reliability factor* $\alpha$ represents the degree of benefit from enhancing the reliability of *inter-hub* links when traffic utilizes inter-hub links. This factor is set as a power parameter of the term $(r_{km})^{1-\alpha} \ (0 \leq \alpha \leq 1, k \neq m)$. Thus, the computation of $R_{ijkm}$ is written as $R_{ijkm} = r_{ik} \cdot r_{km}^{(1-\alpha)} \cdot r_{mj}$. For example, for known link reliabilities ‘$r_{ik}=0.9$’, ‘$r_{km}=0.8$’ and ‘$r_{mj}=0.9$’ with $\alpha=0.9$, the $R_{ijkm}$ is computed as $0.792 \ (= r_{ik} \cdot r_{km}^{0.10} \cdot r_{mj})$, indicating that the route $X_{ijkm}$ can deliver 79.2% of traffic between $i$ and $j$ via hub $k$ and $m$ without delay or congestion (see Figure 3.1a) where the value of $\alpha$ is closer to 1.0, the value of $R_{ijkm}$ improves since the reliability of the inter-hub link $(r_{km}^{1-\alpha})$ is enhanced. In contrast, a smaller $\alpha$ indicates that a smaller benefit is imposed when flows utilize the inter-hub link.

In the case of the one-hub stop route $X_{ijkk} \ (i \rightarrow k \rightarrow j)$, the routing reliability is calculated as $r_{ik} \cdot r_{kk}^{1-\alpha} \cdot r_{kj}$ according to the notational order. In these routings, the reliability of a hub $(r_{kk})$ plays a significant role in transmitting the flows to the destination. To reflect this routing characteristic, we introduce the *reliability factor* $\gamma \ (0 \leq \gamma \leq 1)$ for routing computations. Specifically, the *reliability factor* $\gamma$ is defined as the ability of a hub to transmit traffic without delay or congestion, which could represent the performance level of *intra-hub* communication. This factor is the value of $r_{kk}$ or $r_{mm}$, which is the diagonal element in a reliability OD matrix, and the factor is applied when a routing reliability such as $R_{ijkk}$ or $R_{iji}$ is calculated. A larger $\gamma$ indicates a better capability for transmission facilities, including the router, repeater and amplifier in the hub itself.
Note that the route decision for an $ij$ pair is endogenously made depending on $R_{ijkm}$ whose value is calculated based on the magnitude of reliability factors. As demonstrated in Figure 3.1, if the same level of $\alpha$ is assumed for both routing schemes (for instance, $\alpha = 0.9$) and $\gamma$ is less than 0.8 (for instance, $\gamma = 0.7$ in b1) for a one-hub stop route, then utilizing the inter-hub (or two-hub stop) route $X_{ijkm}$ (Figure 3.1a) is the more reliable route. However, if $\gamma$ is greater than 0.8 (for instance, $\gamma = 0.9$ in b2) as in Figure 1b, then the one-hub stop route $X_{ijkk}$ becomes a better routing strategy ($R_{ijkm} < R_{ijkk}$). Since the optimal route is chosen from a number of possible routes, numerous solutions which maximize the sum of traffic flows over OD pairs can be explored for various values of both factors and based on the assignment type.

---

6 It might be possible to apply the intra-hub factor $\gamma$ in computing the reliability of two-hub stop routes. However, in this paper, $\gamma$ is considered only if the $R_{ijkm}$ involves $r_{kk}$ or $r_{mm}$ according to notational order because considering $\gamma$ in computing two-hub routes might mitigate the effect of $\alpha$. Technically, the reliability of two-hub stop routes in telecommunication networks is more influenced by the transferability of the inter-hub links where the transferability of inbound flows via hubs can be enhanced, whereas a one-hub stop only depends on the reliability of a hub (Pióro and Medhi 2004).
\[ R_{ijkm} = r_{ik}(r_{km})^{1-\alpha} r_{mj} \]
\[ = 0.9(r_{km})^{1-\alpha} \cdot 0.9 \]

\[ R_{ijkk} = r_{ik}(r_{kk})^{1-\alpha} r_{kj} \]
\[ = 0.9(\gamma)^{1-\alpha} \cdot 0.9 \]

\( a \) Inter-hub route \( R_{ijkm} \)

\( b \) One-hub stop route \( R_{ijkk} \)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Reliability of ( X_{ijkm} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) ( \alpha = 0.9, r_{km} = 0.8 )</td>
<td>( R_{ijkm} = 0.792 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Reliability of ( X_{ijkk} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>b1) ( \alpha = 0.9, \gamma = 0.7 )</td>
<td>( R_{ijkk} = 0.782 )</td>
</tr>
<tr>
<td>b2) ( \alpha = 0.9, \gamma = 0.9 )</td>
<td>( R_{ijkk} = 0.802 )</td>
</tr>
</tbody>
</table>

**Figure 3.1.** Route choice based on reliability factor \( \alpha \) and reliability factor \( \gamma \).

### 3.3.3. Data: Reliability matrix and interacting flow matrix

Two OD matrices, the reliability and the interacting flow matrix (hereafter \( R \)-matrix and \( W \)-matrix, respectively) are needed for model experiments. We construct the \( R \)-matrix utilizing the empirical traffic loss rate among 14 United States city nodes, which have been monitored by the SAVVIS network (http://ipsla.savvis.net/) from March to October of 2005. Traffic loss rate is a standard measure to indicate how stably a network can provide Quality of Service (QoS) on telecommunication networks, and it can be easily converted into reliability measures (AT&T 2003; Ciabattoni *et al.* 2003). As
illustrated by Figure 3.2, the reliabilities for 14 city node pairs (91 cases) in the empirical R-matrix are not strong but considerably correlated with their geographical distances.

For the W-matrix, two data sets are prepared for testing purposes. SET I uses equal flows for all ij pairs (i.e., $W_{ij}=1.0$). The purpose of SET I is to investigate model behaviors by suppressing size effects of interacting flows. In SET II, the potential interacting flows are estimated based on the amount of telecommunications demand and supply between $i$ and $j$. The $W_{ij}$ in SET II is calculated as follows, using the data of Zook (2000) and Atkinson and Gottlieb (2001):

$$W_{ij} = k \times (D_i \cdot S_j + D_j \cdot S_i),$$

where

- $W_{ij}$ the amount of flow traveling between city nodes $i$ and $j$,
- $D_i$ the amount of online population in city node $i$,
- $S_j$ the number of domains in city node $j$, and
- $k$ a scaling factor ($10^{-9}$).

According to this calculation, the estimated telecommunication potentials for 14 city nodes are illustrated in Figure 3.3.
Figure 3.2. The relationship between city node reliabilities and geographical distance in the $R$ matrix.
Figure 3.3. 14 U.S city nodes and their telecommunication potentials.
3.4. Model Formulations

As mentioned previously, the PHMR has two different variants, which are named the MRSA and MRMA, based on what assignment scheme is considered in the model formulation. Both models have the same objective function, but different constraints are used. Sections 3.4.1 and 3.4.2 provide the model formulation of the MRSA and the MRMA, respectively.

3.4.1. MRSA (p-hub Maximum Reliability model with Single Assignment)

Maximize

$$\Omega = \sum_{i} \sum_{j} \sum_{k} \sum_{m} W_{ij} R_{ikm} Z_{ikm}$$  \hspace{1cm} (3-1)

Subject to

$$\sum_{k} Z_{kk} = p \quad (2 \leq p)$$  \hspace{1cm} (3-2)

$$\sum_{k} Z_{ik} = 1 \quad \forall \ i$$  \hspace{1cm} (3-3)

$$Z_{ik} - Z_{kk} \leq 0 \quad \forall \ i, k \ (i \neq k)$$  \hspace{1cm} (3-4)

$$\sum_{m} X_{ikm} - Z_{ik} = 0 \quad \forall \ j > i; \ k$$  \hspace{1cm} (3-5)

$$\sum_{k} X_{ikm} - Z_{jum} = 0 \quad \forall \ j > i; \ m$$  \hspace{1cm} (3-6)

$$Z_{ik} \in \{0, 1\}$$  \hspace{1cm} (3-7)

$$0 \leq X_{ikm} \leq 1$$  \hspace{1cm} (3-8)
where

\( p \)  \hspace{1cm} \text{the number of hubs to be located}

\( W_{ij} \)  \hspace{1cm} \text{the amount of flow to travel between } i \text{ and } j

\( X_{ijkm} \)  \hspace{1cm} \text{the fraction of flow from origin } i \text{ to destination } j \text{ via hub } k \text{ and } m \text{ in that order } (i\rightarrow k \rightarrow m \rightarrow j)

\( R_{ijkm} \)  \hspace{1cm} \text{the routing reliability for the route } X_{ijkm}

\( Z_{ik} \)  \hspace{1cm} 1 \text{ if node } i \text{ is allocated to hub } k; \text{ 0 otherwise}

\( Z_{kk} \)  \hspace{1cm} 1 \text{ if node } k \text{ is a hub; } 0 \text{ otherwise}

\( \alpha \)  \hspace{1cm} \text{Inter-hub reliability factor } (0 \leq \alpha \leq 1)

\( \gamma \)  \hspace{1cm} \text{Intra-hub reliability factor } (0 \leq \gamma \leq 1), \gamma = r_{kk} \text{ or } r_{mm}

The objective function (3-1) maximizes the total network flow that can be transported based on computed reliability \( R_{ijkm} \) for route \( X_{ijkm} \) over each OD pair.

Constraint (3-2) requires \( p \) hubs to be open, and constraint (3-3) forces each node to be assigned to only a single hub. Constraint (3-4) requires a hub to be opened before a node is allocated to hub \( k \) denoted as \( Z_{kk} \). Taken together, constraints (3-5) and (3-6) guarantee that flow \( i \) to \( j \) should not be routed via hubs \( k \) and \( m \) unless origin \( i \) is linked to hub \( k \) and \( j \) is linked to hub \( m \). In constraint (3-7), \( Z_{ik} \) is the integer variable that prevents partial facility location.
3.4.2. MRMA (p-hub Maximum Reliability model with Multiple Assignments)

Maximize

\[ \Omega = \sum_{i} \sum_{j} \sum_{k} \sum_{m} W_{ij} R_{ijkm} X_{ijkm} \]  \hspace{1cm} (3-9)

Subject to

\[ \sum_{k} Z_{k} = p \]  \hspace{1cm} (2 \leq p)  \hspace{1cm} (3-10)

\[ \sum_{k} \sum_{m} X_{ijkm} = 1 \]  \hspace{1cm} \forall j > i  \hspace{1cm} (3-11)

\[ \sum_{m} X_{ijkm} - Z_{k} \leq 0 \]  \hspace{1cm} \forall j > i; k  \hspace{1cm} (3-12)

\[ \sum_{k} X_{ijkm} - Z_{m} \leq 0 \]  \hspace{1cm} \forall j > i; m  \hspace{1cm} (3-13)

\[ Z_{k} \in \{0, 1\} \]  \hspace{1cm} (3-14)

\[ 0 \leq X_{ijkm} \leq 1 \]  \hspace{1cm} (3-15)

In the MRMA, \( Z_{k} \) represents the facility location variables instead of the variables \( Z_{kk} \) in the MRSA because the variables \( Z_{ik} \) are not needed in the MRMA. The objective function value from the MRMA is always greater or equal to that of the MRSA because the flexibility of assignment to hubs allows more strategic options when determining the optimal route. In other words, each origin \( i \) can route its interactions with nodes \( j \) through different hubs to give the most reliable route. Constraint (3-10) ensures that \( p \) hubs are selected. Constraint (3-11) ensures that all flows from \( i \) to \( j \) should travel through hub(s) \( k \)
and $m$, and $k$ can be equal to $m$. Constraints (3-12) and (3-13) together prevent the flow between $i$ to $j$ from being routed through non-hub nodes.

In the following sections, two PHMD models (i.e., MDSA and MDMA) are developed based on the MRSA and the MRMA. Section 3.4.5 provides the formulation of the MRDI, which considers network reliability and mandatory dispersion as a bi-objective model.

### 3.4.3. MDSA (p-hub Mandatory Dispersion model with Single Assignment)

Maximize

$$\Omega = \sum_i \sum_j \sum_k \sum_m W_j R_{ijkm} X_{ijkm}$$

Subject to constraints (3-2) – (3-8) and

$$D_{man} \leq d_{km} + M(1-Z_{kk}) + M(1-Z_{mm}) \quad \forall \ m > k$$

where

- $d_{km}$ geographical distance (the length of inter-hub link) between potential hub $k$ and $m$
- $D_{man}$ mandatory separation distance between $p$-hubs ($0 < D_{man} \leq D_{max}$)$D_{max}$
- the maximin distance from any pair of nodes, $D_{max} = \text{Max} \{\text{Min} \{d_{ij}\}\}$
- $M$ a large number greater than $\text{Max} \{d_{km}\}$
The MDSA has the same objective function and employs the constraints (3-2) to (3-8) used in the MRSA as well as constraints (3-17) which are named as *mandatory dispersion constraints (MDC)* are newly added. These last constraints ensure that two hubs that are closely located within $D_{man}$ cannot both be members of a feasible $p$-hubs set. The value $D_{man}$ is exogenously given, but it should be bounded by $D_{max}$ to avoid an infeasible solution. The value of $D_{max}$ can be pre-obtained by running Kuby’s $p$-dispersion model based on the given nodes and the distances among them (Kuby 1987, p. 319).

### 3.4.4. MDMA (p-hub Mandatory Dispersion model with Multiple Assignments)

Maximize

$$\Omega = \sum_{i} \sum_{j} \sum_{k} \sum_{m} W_{ij} R_{jkm} X_{ijm}$$

Subject to constraints (3-10) to (3-15) and

$$D_{man} \leq d_{km} + M(1-Z_k) + M(1-Z_m) \quad \forall \ m > k$$

The MDMA is a simple extension of the MRMA, in which adding MDC constraints (3-19) are designed to keep a mandatory separation level among potential hub facilities. Note that the distance constraints MDC in the PHMD has a similar characteristic with $p$-
hub center or covering problems, where distance-based constraints (maximum distance) are used for either an entire route, each link in a route, or only the spoke links (Campbell 1994b). However, the PHMD applies the minimum distance constraints only to the inter-hub links for mandatory dispersion of hubs.

3.4.5. MRDI (Multiobjective Reliable p-hub DIspersion model)

A multi-objective type of model provides a trade-off among conflicting goals with non-commensurate units that cannot be simply incorporated into a single objective model (Cohon 1978). Instead of setting the level of mandatory separation as constraints, the MRDI combines two problems directly in order to solve both problems simultaneously: the maximization of network reliability and the maximization of geographical separation among hubs. In this chapter, the MRDI is formulated based on multiple assignments:

Maximize $\Omega_1 = \sum_i \sum_j \sum_k \sum_m W_{ijkm} R_{ijkm} X_{ijkm}$ \hspace{1cm} (3-20)

Maximize $\Omega_2 = D$ \hspace{1cm} (3-21)

Subject to constraints (3-10) to (3-15) and

$$D \leq d_{km} + M(1-Z_k) + M(1-Z_m) \hspace{1cm} \forall m > k$$ \hspace{1cm} (3-22)

$$0 \leq D \leq D_{\text{max}}$$ \hspace{1cm} (3-23)
where

\[ D \] the smallest separation distance among any pair of \( p \) hubs

\[ D_{\text{max}} \] the \textit{maximin} distance pre-obtained by the \( p \)-dispersion model

Using the weighting method, the objective of the MRDI is re-stated as:

\[
\text{Maximize } \Omega = w\Omega_1 + (1 - w)\Omega_2 \tag{3-24}
\]

The objective function (3-21) is essentially a \textit{maximin} dispersion objective, which is to maximize \( D \), the minimum distance between \textit{open} hub facilities (Kuby 1987). Constraints (3-22) are written similarly to constraints (3-19) in the MDMA, but \( D \) plays the role of an upper bound where the distance separating any pair of hubs should be greater than the maximum value of \( D \). Constraint (3-23) is optional, but this constraint helps to reduce the computational effort because the value \( D_{\text{max}} \) plays the role of upper bound of the solution space \( D \). Note that in the objective function (3-24), given the weight \( w = 1 \), this model is the same as the MRMA, but when \( w = 0 \) is assumed, the model is analogous to the classical \( p \)-dispersion model, which finds only the smallest separation distance among opened facilities.

\subsection*{3.5. Model Experiments}

Generally, reducing the problem size is an important issue in hub location models. Especially, solving a problem to optimality is of concern for a clear examination of
model results. Thus, we make two assumptions to reduce the problem size. First, the amount of interacting flows and reliabilities between two nodes is assumed as symmetric; i.e., $W_{ij} = W_{ji}$ and $R_{ij} = R_{ji}$. As mentioned in section 3.3.3, the assumption of symmetric flows is due to the way telecommunication potentials are calculated between $i$ and $j$ by considering the amount of demand and supply in both nodes. Second, excluding impractical routes from the set of routing variables $X_{ijkm}$ (for instance, $i \rightarrow k \rightarrow i \rightarrow j$ or $i \rightarrow j \rightarrow m \rightarrow j$) results in a practical reduction in the number of variables as well as constraints. As illustrated in Figure 3.4, the enumerated routing variables are expressed as a set, $S = \{(i,j,k,m) | (j > i) \cap (k = i) \cup (k = m = j) \cup (k \neq i \cap k \neq j \cap m \neq j)\}$. Accordingly, the number of routing variables $X_{ijkm}$ reduces to $n(n-1)(2n^2-2n+3)/2$ from the $n^4$ that are generated in the full-size model according to the combination of indices $(i,j,k,m)$. 
<table>
<thead>
<tr>
<th>Subset of $S$</th>
<th>Routing variable $s(i, j, k, m) \in S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(j &gt; i) \cap (k = i)$</td>
<td>* $(i, j, i, m)$</td>
</tr>
<tr>
<td>$N_1 = n^3(n-1)/2$</td>
<td><img src="image1" alt="Diagram" /></td>
</tr>
<tr>
<td>$(j &gt; i) \cap (k \neq i \cap k = m=j)$</td>
<td>* $(i, k, k, k)$ or equivalently $(i, j, j, j)$, $(i, m, m, m)$</td>
</tr>
<tr>
<td>$N_2 = n(n-1)/2$</td>
<td><img src="image6" alt="Diagram" /></td>
</tr>
<tr>
<td>$(j &gt; i) \cap (k \neq i \cap k \neq j \cap m \neq j)$</td>
<td>* $(i, j, k, m)$</td>
</tr>
<tr>
<td>$N_3 = n(n-1)^2(n-2)/2$</td>
<td><img src="image8" alt="Diagram" /></td>
</tr>
</tbody>
</table>

**Figure 3.4.** The enumerated routing variable set $S$.

**Note:** Total number of variables $X_{ijkm}$ is $n(n-1)(n^2-2n+3)/2$, which is the sum of each subset $N_1$, $N_2$ and $N_3$. 
<table>
<thead>
<tr>
<th>Models</th>
<th>The number of Variables</th>
<th>The number of Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRSA</td>
<td>((n^4 - 3n^3 + 7n^2 - 3n)/2 )</td>
<td>(n^3 + 1)</td>
</tr>
<tr>
<td>MRMA</td>
<td>((n^4 - 3n^3 + 5n^2 - n)/2)</td>
<td>((2n^3 - n^2 + 2)/2)</td>
</tr>
<tr>
<td>MDSA</td>
<td>((n^4 - 3n^3 + 7n^2 - 3n)/2)</td>
<td>((2n^3 + n^2 - n + 2)/2)</td>
</tr>
<tr>
<td>MDMA</td>
<td>((n^4 - 3n^3 + 5n^2 - n)/2)</td>
<td>(n^3 - n + 1)</td>
</tr>
<tr>
<td>MRDI</td>
<td>((n^4 - 3n^3 + 5n^2 - n + 2)/2)</td>
<td>(n^3 - n + 2)</td>
</tr>
</tbody>
</table>

**Table 3.1.** Problem sizes for model formulations.

The problem sizes for \(n\) nodes according to each model formulation are summarized in Table 3.1. As demonstrated by O’Kelly *et al.* (1996), these variable reduction techniques are effective when solving the hub location problem efficiently by reducing the computational complexity, while generating the same results as the original problems.

The model experiments are organized into two sub-sections according to testing purposes. The first section is designed to examine the behaviors of the MRSA and MRMA models. While various parametric combinations of reliability factors \(\alpha\) and \(\gamma\) are possible, larger and smaller values of \(\alpha\) and \(\gamma\) in Table 3.2 are reported that allow the model behaviors to be examined. CASE I has very small reliability factors (\(\alpha=0.001\) and \(\gamma=0.10\)). In contrast, CASE IV assumes highly reliable conditions on both facilities (\(\alpha=0.99\) and \(\gamma=0.99\)). CASEs II and III represent the conditions of extreme differences between the two factors.
Note that the combination of reliability factors reflects a level of technology in telecommunication systems. For example, CASE I represents traditional telecommunication systems with “copper-based technologies”, which raised frequent disruptions of service due to less-reliable facilities and limited performance. In contrast, CASE IV is associated with “fiber-optic technologies” in current IP-based telecommunication systems, where highly-reliable network facilities are established (Grötschel, Monma, and Stoer 1995; Pastor-Satorras and Vespignani 2004). All models were solved using CPLEX 10.1 on a Dual Intel Pentium III Xeon 2.0 GHz with 1 GB RAM with the Windows NT operating system. The models were solved to optimality and the solution times range from a few seconds for most problems to a few minutes for some MRDI problems.

<table>
<thead>
<tr>
<th>Reliability factor ( \alpha )</th>
<th>Small ( \alpha ) (( \alpha = 0.001 )) (low transmission ability of inter-hub links)</th>
<th>Large ( \alpha ) (( \alpha = 0.99 )) (high transmission ability of inter-hub links)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small ( \gamma ) (( \gamma = 0.10 )) Low intra-hub reliability</td>
<td>CASE I</td>
<td>CASE III</td>
</tr>
<tr>
<td>Large ( \gamma ) (( \gamma = 0.99 )) High intra-hub reliability</td>
<td>CASE II</td>
<td>CASE IV</td>
</tr>
</tbody>
</table>

*Table 3.2. Testing scheme.*
3.5.1. Model behaviors of the MRSA and MRMA

Figure 3.5 presents the results of optimal configurations of hub networks for \( p=4 \) using \( W \)-matrix SET I. As expected, an increase in the objective function values is found for both models as one (or both) reliability factors increase. The objective function value for the MRMA is always greater than that for the MRSA, although the difference in network reliability between the two models decreases as both reliability factors become larger. Notice that the MRMA has more linkages than the MRSA due to its routing flexibility. With respect to network cost, this property indicates that the MRMA would be a less cost-effective design approach than the MRSA if budgetary resources to construct the linkages are considered in the model.\(^7\) As implicitly discussed by O’Kelly (1986) and O’Kelly et al. (1996), the spatial pattern of hubs is related to the distance of routes and the level of benefit, which is imposed on a facility (for example, inter-hub link discounts in air transportation). In this context, Figure 3.5 demonstrates that the distinct change of hub locations and allocations relies on the different levels of the reliability factors. In CASE I, hubs are located in close proximity, but become dispersed as one of the reliability factors improves (CASEs II and III). For both larger reliability factors in CASE IV, the model response is to locate hubs widely spaced with a regular separation.

\(^7\) The MRSA always constructs a hub network with the fixed number of linkages \((n-p+p(p-1)/2)\) for \( n \) nodes and \( p \) hubs, but the MRMA will have at least the same number of linkages as the MRSA, and up to \((p(n-p)+p(p-1)/2)\) linkages, which is the case when every non-hub node is allocated to all \( p \) hubs. Thus, the assumption of multiple assignments might not be appropriate for telecommunication network designs depending on construction costs. In our results of the MRMA, however, the multiple assignments between a non-hub to hubs generally are limited to a few hubs to various levels of both factors.
Figure 3.5. Model behaviors: MRSA (left) and MRMA (right) for $p=4$ using SET I. 
Note: The size of hub is proportionally depicted by the amount of the flows.
Table 3.3. Routing choices for $p=4$ using SET I.

Note: 1) “Direct route” include the “non-hub to hub” and “hub to hub” routes which are expressed as $X_{ijj}$, $X_{ijii}$, and $X_{iiij}$ in Figure 3.4. Notice that the flow in the category of direct routes looks similar in that the route delivers traffic directly without any intermediate hub, but the origin and destination is different in notation. For instance, the destination of the route $X_{ijj}$ is a hub while the origin is a hub in the route $X_{ijii}$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Scheme</th>
<th>two-hub stop</th>
<th>one-hub stop</th>
<th>direct route$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRSA</td>
<td>CASE I</td>
<td>60</td>
<td>25</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>CASE II</td>
<td>55</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>CASE III</td>
<td>62</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>CASE IV</td>
<td>31</td>
<td>44</td>
<td>16</td>
</tr>
<tr>
<td>MRMA</td>
<td>CASE I</td>
<td>91</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CASE II</td>
<td>66</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>CASE III</td>
<td>91</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CASE IV</td>
<td>31</td>
<td>45</td>
<td>15</td>
</tr>
</tbody>
</table>

The model behaviors in Figure 3.5 can be explained by the types of routings used in each phase. Table 3.3 shows how routing choices vary as the reliability factors change, in accordance with results in Figure 3.5. As mentioned previously, if no reliability factors are considered in a model, each interacting flow between $i$ and $j$ tends to utilize the shortest route via hub(s) because the coefficient $R_{ijkm}$ declines as a function of the length (distance) of a route. However, more weight on a reliability factor produces a tendency to use the route whose $R_{ijkm}$ is enhanced by that weight. A large $\alpha$ on inter-hub links increases the chance to utilize two-hub stop routes, and the larger coefficient of one-hub stop ($R_{ijkk}$) or direct routes ($R_{ikkk}$) is produced as the weight $\gamma$ increases. As shown in the
routing choices of CASEs I and III, the use of two hub-stop routes is encouraged more if an extremely small $\gamma$ value is applied in a routing scheme because the route choice simply tends to avoid the routings in which the small $\gamma$ is involved. However, using one-hub stop or direct route becomes a better routing choice than a two-hub stop route when incentive $\gamma$ is larger. This tendency of model behaviors is more clearly observed in the MRMA due to more flexible routing choices (see the result of CASE II and IV).

The behavior of hub selection and facility activities also is influenced by the different size of interacting flows for OD pairs. Table 3.4 summarizes the optimal hub locations as well as facility activity levels of the MRSA and the MRMA for $p=4$ and 5 using SET II for the testing scheme and three additional phases, which apply moderate levels of reliability factors (i.e., MOD I, II, and III with $\alpha$ and $\gamma = 0.5, 0.7$ and 0.9, respectively). As shown in Table 3.4, the hub locations with SET II are less responsive to the change of reliability factors compared with the response of hub locations with identical flows (SET I). The impact of different sizes of flows is more clearly observed in the MRMA. As shown in the results of testing scheme in Table 3.4, the MRMA open hubs frequently at city nodes such as New York and Philadelphia that have larger interactions with other nodes, whereas a different set of hub locations is found in the MRSA models with distinct changes of reliability factors. As indicated in columns 4 and 5, if those hubs are placed close to each other, then the largest concentration of interactions takes place at their inter-hub links and hubs, producing large variances in flow activity levels to other facilities.
<table>
<thead>
<tr>
<th>MRSA</th>
<th>Objective value</th>
<th>Hubs</th>
<th>The largest Inter-hub flow</th>
<th>The largest Intra-hub flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p=4$</td>
<td>CASE I</td>
<td>18912.15</td>
<td>PHX – MIA</td>
<td>MIA</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>28481.02</td>
<td>CHI PHL</td>
<td>PHL</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>28730.20</td>
<td>NY PHL</td>
<td>NY</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>28871.42</td>
<td>PHX DEN PHL</td>
<td>PHL</td>
</tr>
<tr>
<td>MOD I</td>
<td>$^{19}$</td>
<td>23608.05</td>
<td>PHX MIN SEA</td>
<td>MIA</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>25894.19</td>
<td>PHX MIA MIN SEA</td>
<td>MIA – MIN</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>27962.03</td>
<td>PHX PHL MIN SEA</td>
<td>PHL – PHX</td>
</tr>
<tr>
<td>$p=5$</td>
<td>CASE I</td>
<td>18484.27</td>
<td>PHX DENT MIN SEA</td>
<td>MIN – SEA</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>28489.98</td>
<td>PHX CHI PHL MIA SEA</td>
<td>PHX – PHL</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>28782.25</td>
<td>PHX MIA NY PHL SEA</td>
<td>NY – PHL</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>28901.97</td>
<td>PHX DEN PHL MIA SEA</td>
<td>PHL – PHX</td>
</tr>
<tr>
<td>MOD I</td>
<td></td>
<td>23001.83</td>
<td>PHX DENT MIA MIN SEA</td>
<td>MIN – SEA</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>25425.06</td>
<td>PHX DENT MIA MIN SEA</td>
<td>PHX – MIA</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>27815.28</td>
<td>PHX DENT MIA PHL SEA</td>
<td>PHX – PHL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MRMA</th>
<th>Objective value</th>
<th>Hubs</th>
<th>The largest Inter-hub flow</th>
<th>The largest Intra-hub flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p=4$</td>
<td>CASE I</td>
<td>28453.43</td>
<td>NY CHI PHL SEA</td>
<td>NY – PHL</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>28537.44</td>
<td>NY PHL MIA SEA</td>
<td>NY – PHL</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>28815.71</td>
<td>PHX PHL MIA SEA</td>
<td>PHL – MIA</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>28873.56</td>
<td>NY PHL MIA SEA</td>
<td>NY – PHL</td>
</tr>
<tr>
<td>MOD I</td>
<td></td>
<td>28628.48</td>
<td>CHI MIA PHL SEA</td>
<td>SEA – PHL</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>28703.25</td>
<td>CHI MIA PHL SEA</td>
<td>SEA – PHL</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>28778.32</td>
<td>CHI MIA PHL SEA</td>
<td>SEA – PHL</td>
</tr>
<tr>
<td>$p=5$</td>
<td>CASE I</td>
<td>28510.98</td>
<td>CHI MIA NY PHL SEA</td>
<td>NY – PHL</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>28560.92</td>
<td>PHX MIA NY PHL SEA</td>
<td>NY – PHL</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>28863.55</td>
<td>PHX CHI MIA PHL SEA</td>
<td>PHL – MIA</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>28905.54</td>
<td>DEN MIA NY PHL SEA</td>
<td>NY – PHL</td>
</tr>
<tr>
<td>MOD I</td>
<td></td>
<td>28675.81</td>
<td>CHI MIA NY PHL SEA</td>
<td>NY – PHL</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>28745.85</td>
<td>PHX CHI MIA PHL SEA</td>
<td>PHL – MIA</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>28826.04</td>
<td>PHX CHI MIA PHL SEA</td>
<td>PHL – MIA</td>
</tr>
</tbody>
</table>

Table 3.4. Hub locations and facility activity levels for $p=4$ and 5 using SET II.

Note: 1) MOD I, II and III apply the moderate level of reliability factors in the model (both $\alpha$ and $\gamma = 0.5$, 0.7, and 0.9, respectively).

The peripheral hub (for example Seattle) can be selected as a hub with a high concentration of flows if the number of non-hub nodes to be served increases. Note that the MRMA does not produce much difference in locations of hubs and facility activity levels for the moderate levels of reliability factors. The different behavior between the MRSA and the MRMA is due to the allocation scheme. Like other multiple assignment hub location models, the MRMA tends to change allocations of non-hub nodes to hubs to find optimal routes between $i$ and $j$, rather than to move the location of hubs (O’Kelly et al. 1996).

### 3.5.2. Impact of hub dispersion

The benefits from operating hub networks could be compromised by increased congestion levels for particular hub and inter-hub facilities (O’Kelly 1986). The basic rationale of hub dispersion is to redistribute flows by separating facilities appropriately if an excessively amassed flow traveling on particular inter-hub link(s) or hub(s) is identified in the PHMR models. Figure 3.6 demonstrates how the inter- and intra-hub activity levels are changed when clustered hubs are dispersed geographically, based on the results of a MRMA model. These results are obtained using SET II for $p=3$, and a moderate level of reliability factors (both $\alpha$ and $\gamma=0.7$) is applied. The clustered hub network (Figure 3.6a) is the optimal solution of the MRMA model, and the dispersed network (Figure 3.6b) is made by relocating one of the hubs (Philadelphia) based on the network (Figure 3.6a). To compare the flow activity levels, we devised two indicators, the *inter-hub* and the *intra-hub dependency rate* (hereafter INTERD and INTRAD),
which are defined as the ratio of the largest inter-hub flow or the largest intra-hub flow to the total sum of inter- or intra-hub flows, respectively. A larger value indicates that a higher concentration of flow exists on the particular inter- or intra-hub.

As shown in Figure 3.6, even though the gap of objective function values between the two networks is small, the differing flow activity levels on the facilities are quite evident. In terms of both indicators, a large concentration of flows is observed in particular hubs and their inter-hub links in a clustered hubs network configuration (Figure 3.6a). However, dispersing hubs from the clustered configuration (Figure 3.6b) results in the decrease of both dependency rates, from 0.50 to 0.41 (INTERD) and from 0.47 to 0.38 (INTRAD), respectively, implying that an appropriate separation of hubs could help mitigate any excess concentration of flows occurring in particular hub facilities.

This result is due to the non-hub node re-allocations in response to the dispersed hub(s), resulting in a more equitable flow distribution on the network, although it might not be the optimal hub location in terms of network reliability. If the hubs are placed close together on the network so that excessive flow dependencies are identified in terms of both indicators, the benefit from mandatory dispersion of hubs is clearly observed.

---

8 Although beyond the scope of this dissertation, either imposing capacities on links or upgrading hubs with a better transferability would be an effective prescription to respond to possible congestion in classical hub location models. However, to make these capacity constraints effective, the constraint fixing the exact number of \( p \) hubs is generally dropped (Bryan 1998; Campbell, Ernst and Krishmanoorthy 2002).
Figure 3.6. Comparison of facility activity level for a clustered (6a) and dispersed (6b) hub network at a moderate level of reliability factors ($p=3$, both $\alpha$ and $\gamma=0.7$).
Figure 3.7 and 3.8 show the impact of hub dispersion based on the results of the MDSA and the MDMA ($p=5$) for different mandatory dispersion levels with $D_{man} = 0$ and 1100 miles. The experiments are made for CASE I and CASE IV where both dependency rates INTERD and INTRAD are higher than other model results in Table 3.4.

Our interest lies in the variation of the objective function values and dependency rates as the mandatory dispersion level increases. As the mandatory separation $D_{man}$ increases, the objective function value of the model diminishes because the number of feasible hub sets is limited for a given $D_{man}$. Second, mitigation of the high flow dependencies on particular inter- and intra-hubs is observed with an increase in hub dispersion, resulting in a smaller variance among facilities. In particular, as shown in Figure 3.8, the impact of hub dispersion is more clearly identified in the MDMA, where excessive flow dependencies are formed between the hubs (i.e., New York – Philadelphia) due to their close geographical proximity.

The dispersion of hubs does not necessarily indicate the dispersion of the arcs. For example, New York is frequently assigned to Boston if Boston is selected as a hub due to its close proximity on the network, and the separated hubs might use the same inter-hub or spoke links even when the hubs are dispersed. However, as shown in Figure 3.7 and Figure 3.8, the activity level of inter- and intra-hub flows is highly influenced by the dispersion of hubs because each OD would change its optimal route in response to a change in hub locations.
Figure 3.7. The solutions of MDSA for CASE I and IV ($p=5$, $D_{man}=0$ and 1100 miles).

Note: For visual comparison, the size of hub and inter-hub links is proportionally depicted reflecting the amount of flows.
Figure 3.8. The solutions of MDMA for CASE I and IV ($p=5, D_{\text{max}} = 0$ and 1100 miles).

Note: For visual comparison, the size of hub and inter-hub links is proportionally depicted reflecting the amount of flows.
Table 3.5 presents the solution characteristics of the MRDI for the results of CASEs I and IV \((p=4)\). As expected, network reliability declines as the level of optimal facility separation \(D\) increases, signifying that there is an intrinsic trade-off between the same directional problems. By incrementing weights, this multi-objective model identifies a couple of non-inferior solutions for the particular weights. For example, given the weight \(w=1\), where only the importance of the reliability of a network is considered, the results are the same as those found for a very small separation level \((D=80 \text{ miles, New York - Philadelphia})\) in both CASE I and CASE IV. As more weight is imposed on the geographical separation, the hubs are dispersed whenever the model reaches a certain threshold of a given weight.

Interestingly, as shown in Table 3.5, the same separation level \((D=1120 \text{ miles})\) but different optimal hub locations are identified as non-inferior solutions at \(w=0.0\) and \(w=0.01\). This is because the MRDI is forced to identify the best hubs that can maximize network reliability among probable alternative optima once a very small weight is imposed on the objective function \(\Omega_1\). The MRDI would be helpful for decision-makers to compromise between desired levels of network reliability and an appropriate separation of hubs.
### CASE I

<table>
<thead>
<tr>
<th>Weight $w$</th>
<th>Objective Value ($\Omega_1$)</th>
<th>Separation ($D$) (miles)</th>
<th>Hubs</th>
<th>Iterations</th>
<th>Branches</th>
<th>Solution Times (Sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>28453.43</td>
<td>80</td>
<td>CHI NY PHL SEA(^1)</td>
<td>1122</td>
<td>8</td>
<td>16.80</td>
</tr>
<tr>
<td>0.95</td>
<td>28447.88</td>
<td>664</td>
<td>PHL CHI MIA SEA</td>
<td>2481</td>
<td>19</td>
<td>28.17</td>
</tr>
<tr>
<td>0.90</td>
<td>28409.67</td>
<td>1021</td>
<td>PHL PHX MIA SEA</td>
<td>5535</td>
<td>233</td>
<td>47.94</td>
</tr>
<tr>
<td>0.60</td>
<td>28358.87</td>
<td>1114</td>
<td>PHX CHI MIA SEA</td>
<td>9614</td>
<td>327</td>
<td>83.17</td>
</tr>
<tr>
<td>0.01</td>
<td>28246.98</td>
<td>1120</td>
<td>MIN BOS MIA PHX</td>
<td>10339</td>
<td>544</td>
<td>90.42</td>
</tr>
<tr>
<td>0.00</td>
<td>28145.88</td>
<td>1120</td>
<td>MIN BOS MIA SF</td>
<td>1074</td>
<td>35</td>
<td>16.73</td>
</tr>
</tbody>
</table>

### CASE IV

<table>
<thead>
<tr>
<th>Weight $w$</th>
<th>Objective Value ($\Omega_1$)</th>
<th>Separation ($D$) (miles)</th>
<th>Hubs</th>
<th>Iterations</th>
<th>Branches</th>
<th>Solution Times (Sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>28873.56</td>
<td>80</td>
<td>NY PHL MIA SEA</td>
<td>687</td>
<td>0</td>
<td>1.88</td>
</tr>
<tr>
<td>0.95</td>
<td>28873.21</td>
<td>1019</td>
<td>DEN PHL MIA SEA</td>
<td>702</td>
<td>1</td>
<td>2.26</td>
</tr>
<tr>
<td>0.60</td>
<td>28830.68</td>
<td>1114</td>
<td>PHX BOS MIA SEA</td>
<td>2795</td>
<td>132</td>
<td>41.28</td>
</tr>
<tr>
<td>0.01</td>
<td>28824.29</td>
<td>1120</td>
<td>MIN BOS MIA SEA</td>
<td>2379</td>
<td>95</td>
<td>31.22</td>
</tr>
<tr>
<td>0.00</td>
<td>28803.31</td>
<td>1120</td>
<td>MIN BOS MIA SF</td>
<td>1074</td>
<td>35</td>
<td>17.05</td>
</tr>
</tbody>
</table>

Table 3.5. Non-inferior solutions of the MRDI for CASE I and CASE IV ($p=4$, SET II).

3.6. Summary

This chapter proposes a new reliable hub location problem and its formulations and describes the relationship between network performance and hub locations. The results from these models could give useful insights into telecommunication network design. First, the PHMR demonstrates how optimal hub locations can be determined under various reliability conditions on hubs and inter-hub links. Second, like the well recognized important design aspect in telecommunication networks (Medhi 1999; Ball 2001), the dispersion of hub facilities can be considered in hub network design that can protect the excessive concentration of interacting flows from particular hub facilities. This prescription could be helpful by achieving a more equitable flow distribution on networks facilities not only to prevent networks from potential congestion and degradation that can arise from clustered hubs, but also, ultimately, to reduce damage when any malfunctions or attacks occur in highly active hubs or inter-hub links.

Given that this chapter focuses on models that provide accurate behavior using a small data set, further research needs to expand upon the general problem addressed here. First of all, future models could take into account other constraints or objectives that are commonly addressed in hub network design such as costs, the capacity of facilities, and the minimum threshold of flows on links.

By incorporating these constraints, more generalized and realistic conclusions for model behavior can be drawn. For example, considering the importance of network cost for constructing hubs and links, the RPHLP models can be extended to fixed cost hub
location problems, or uncapacitated hub location problems where the number of hubs is determined implicitly as a part of optimization in a model.

Second, in this chapter, the RPHLP are presented based on simple assumptions that are used in standard hub location models. However, the RPHLP can be developed considering other characteristics of telecommunication networks where inter-hub links may not be established to connect all hubs, or different assignment scheme (i.e., direct routes between non-hub nodes) are required. In terms of reliability factors, the RPHLP can employ flow-dependent reliability factors whose reliability levels are determined based on the activity level at hub facilities (Campbell, Ernst and Krishmanoorthy 2002).

Another avenue for expanding the RPHLP entails exploring model behaviors by applying other network-performance-related measures in network design. For example, traffic latency rate, one of the standard measures of network reliability, is known to respond more sensitively to the length of a route (Forsgren and Prytz 2006; Crovella and Krishnamurthy 2006). The model behavior with different reliability measures could provide a good insight in developing other network design models.
In the previous chapter, the RPHLP models aim to enhance a network’s performance. In this chapter, I present the p-hub protection models which are developed from a survivable network design. The survivable network design approach is different from reliable network design in that the models consider the prevention of the potential loss of flows on network.

4.1. Introduction

In telecommunication networks, hubs and backbones are the most critical assets to be protected from failure because a large amount of network flows use these facilities, with intensive concentrations of flows on these facilities as a result. Conversely,

---

9 A version of this chapter, “p-Hub protection models for reliable hub and spoke network design”, has been submitted for publication in *Journal of Geographical Systems* (JGSY-D-08-00024, 2008).
adequately protecting these facilities is an effective strategy for maintaining the survivability of an entire network with minimum effort (Lewis 2006).

As mentioned in chapter 2, two distinctive approaches are known in survivable network design (Pióro and Medhi 2004). The protection approach refers to a preventive strategy put in place before a failure happens in a system, whereas the restoration approach focuses on the recovery techniques after the failure happened in the network.

The primary purpose of this chapter is to address a series of new hub and spoke network models of the survivable network design type, called p-hub protection problems (PHPRO). As suggested by the term “protection”, the models are close to a preventive approach in which a network is designed to reduce traffic loss during either normal conditions or in the face of malfunctioning network components.

The design approach based on hub and spoke network design has many advantages compared to other survivable network models. First, the structure of current IP-based telecommunication systems is very similar to hub and spoke-type networks, consisting of three main components: a handful of hubs, backbones for the delivery of massive interacting flows, and local access linkages. Second, as stressed by Klincewicz (1998; 2006), hub and spoke-type models can easily reflect the substantial characteristics of real telecommunication systems such as telecommunication potentials of nodes, interactions among origin-destinations, and routing schemes (see also Soni et al. 1999; Gourdin et al. 2002).

The PHPRO models aims to build networks which minimize the total potential loss of interacting flows over the set of origin-destination nodes, based on different routing assumptions. More specifically, PHPRO optimizes the network by 1) locating p-
hubs on a network, 2) inter-connecting hubs, and 3) allocating non-hub nodes to hub(s). PHPRO is classified into three sub-models depending on which routing schemes are considered in constructing the network.

First, the p-hub protection model with single assignment (PROSA) constructs a hub network based on the single assignment scheme. Second, the p-hub protection model with multiple assignments (PROMA) constructs the network based on the multiple assignments scheme. Third, the p-hub protection model with primary and survivable route (PROPS) defines a network with a more aggressive protective routing scheme.

Figure 4.1 illustrates the different protocols among three models. PROSA (Figure 4.1a) can be categorized as a protection network model for normal circumstances in which the model simply aims to optimize the network in order to minimize the potential traffic loss of interacting flows among OD pairs although it might not satisfy the minimum connectivity requirement (i.e., degree of a node $k \geq 2$) in some survivable network design models. PROMA (Figure 4.1b) allows multiple-assignments between non-hub nodes to hubs, which are intended to diversify the routes of interacting flows between OD pairs. PROPS (Figure 4.1c) constructs a primary routes network based on single assignment, while at the same time considering disjoint paths for OD pairs as survivable routes.
This chapter is organized as follows: The next section explores the literature associated with protective network design along with hub and spoke network models. Section 4.3 presents methodology and mathematical formulations. The model applications and results are presented in section 4.4. In section 4.5, three model extensions are provided followed by the conclusions in the final section.

4.2. Survivable Network Designs in Telecommunication Systems

In general, many approaches to protecting network infrastructure in telecommunication systems do not attempt to build a loss-free packet delivery system because it is impossible to operate a network without any failures. The first classical survivable network problem addressed by Baybars and Edahl (1988) was focused on
identifying the best alternative circuits for certain pairs of nodes while minimizing the cost for locating transmission facilities and links. The models by Agarwal (1989) and Medhi (1994) focused a different survivable strategy to find optimal alternative shortest paths in a given topology for the set of facilities enabling a network to continue operation. Based on this initial idea, the main focus of survivable network problems aimed to protect flows on the network with the optimal set of disjoint paths in a given network at the same time the problems minimize cost of paths or alternative routes (Torrieri 1992).

As another protective scheme, Pirkul and Narasimhan (1994) proposed flow protection models in terms of minimizing the total queuing delay cost in carrying traffic using pre-determined the best alternative routes to the primary paths in the event of primary link failures. Following the work, various re-routing strategies to protect the flows and heuristics have been proposed (Louca et al. 1999; Ghashghai and Rardin 2002; Walkowiak 2004).

In terms of topological survivable network design, Monma and Shallcross (1989) suggested the idea of $k$-minimum connectivity requirements to protect certain nodes from linkage failures. To avoid being isolated from a network, a node requires connections with more than two linkages to other nodes. If all nodes in a network are fully connected to each other, then it could be the most survivable network in terms of network connectivity. However, it has been pointed out that simply increasing redundancies of a network would be cost-prohibitive in most network designs (Deeter and Smith 1998). Thus, many survivable network models consider both network cost in a specific context and a certain level of connectivity (see Grötschel et al. 1995; Clarke 1992).
In terms of fault-tolerant routing systems in IP-based networks, multiple-layered network designs have been also addressed (Ball and Vakhutibsky 2001; Xin and Rouskas 2004). Recent literature stresses the need to consider performance-related measures such as reliability, packet loss rate, and traffic latency in telecommunication network design due to the delay-sensitive nature of current telecommunication systems (Skorin-Kapov, Skorin-Kapov and Boljunčić 2006; Klincewicz 2006; Soni et al. 1999).

Finally, optimization approaches which aim at identifying the network components most critical to disruption have been proposed (see Church et al. 2004, for a recent review of interdiction models). These models can be regarded as protection models in that they help to identify highly vulnerable facilities in existing network systems and evaluate the network’s survivability although the models do not directly address the issue of survivable network design (Grubesic et al. 2008; Murray et al. 2007).

Ultimately, the models of PHPRO in this chapter optimize a hub network by reflecting complicated relationships among the elements such as the interactions among origin-destinations, distance, geographical location of nodes, and network costs. In previous research, however, few survivable network models have reflected these substantial, and locational characteristics. For example, many survivable models sometimes oversimplify the function of network components such as hub and route, especially when topological aspects of a network are stressed. As results, a hub— which is defined as a facility acting as a switching point for consolidating and interacting flows among OD pairs— is often recognized as a collection point connected to other nodes in the network hierarchy. A route is also referred to as a set of arcs, rather than understood
as a channel where a specific flow moves from origin to destination based on the direction (Campbell 1994).

### 4.3. Modeling $p$-Hub Protection Models

In order to formulate the PHPRO, two model components are discussed: 1) the probability of potential traffic loss ($L_{ijkm}$) according to routing schemes, and 2) reliability factors $\alpha$ and $\gamma$. We provide the model formulations based on these components.

**Figure 4.2.** The computation of potential traffic loss: a single (a) and a primary and back-up routing scheme (b).

**Note:** The dotted arrows line represent back-up route.

\[
L_{ijkm} = (1 - R_{ijkm})
= 1 - (0.95 \cdot 0.99 \cdot 0.95)
= 0.11
\]

\[
L_{ijkm}^p = L_{ijkm} + L_{ijkm}^{p} \cdot L_{ijkm}^{s}
= (1 - R_{ijkm}^p) + (1 - R_{ijkm}^p) \cdot (1 - R_{ijkm}^s)
= (1 - 0.89) + (1 - (1 - 0.89)) \cdot (1 - 0.903)
= 0.1963
\]
4.3.1. Probability of potential traffic loss $L_{ijkm}$

Figure 4.2(a), an example based on Medhi (1999), shows how the probability of traffic loss of a route is calculated in a telecommunication system. Let $X_{ijkm}$ denote the route where traffic is delivered between nodes $i$ and $j$ via hubs $k$ and $m$ in the following direction $i \rightarrow k \rightarrow m \rightarrow j$. The term $R_{ijkm}$ represents the coefficient of reliability for the route $X_{ijkm}$ and is calculated by multiplying the reliabilities of each link $r_{ik}$, $r_{km}$, and $r_{mj}$. We define the probability of potential traffic loss $L_{ijkm}$ as the complementary probability of $R_{ijkm}$ (i.e., $L_{ijkm} = 1 - R_{ijkm}$). Thus, if $R_{ijkm}$ is 0.89, for instance, $L_{ijkm}$ is 0.11, indicating that route $X_{ijkm}$ will experience 0.11 (11%) loss or delay of total traffic. The lower $L_{ijkm}$, the more protected the transfer flow between $i$ and $j$. As shown in Figure 4.2(b), if we consider the back-up routing scheme where one-stop route $X_{ijkmn}$ ($i \rightarrow n \rightarrow j$) is the path-disjoint route to the primary route $X_{ijkm}$, the probability of potential traffic loss is computed based on the Boolean algebra method as below (Colbourn 1987):

$$
L_{ijkm} = L_{ijkm}^P + L_{ijkm}^S \cdot L_{ijkmn}^S
\quad = (1 - R_{ijkm}^P) + (1 - R_{ijkm}^P) \cdot (1 - R_{ijkmn}^S)
$$

(Equation 4-1)

Where

$L_{ijkm}$ Probability of potential traffic loss from $i$ to $j$ by primary and back-up route

$L_{ijkm}^P$ Probability of potential traffic loss by primary route
\( \overline{L}_{ijklm}^p \) Complement probability of \( L_{ijklm}^p \)

\( L_{ijklmn}^s \) Probability of potential traffic loss by back-up route

\( R_{ijklm}^p \) Reliability of primary route \( (= r_{ik} \times r_{km} \times r_{mj}) \)

\( R_{ijklmn}^s \) Reliability of back-up route \( (= r_{in} \times r_{nj}) \)

Note that the potential loss probability of scheme in Figure 4.2(b) is always greater than that of a single routing scheme (Figure 4.2(a)) because the loss probability of \( L_{ijklmn}^s \) is involved in the computation. However, considering both the primary and back-up routing scheme can deliver the flow unless both routes are simultaneously disrupted.

4.3.2. Inter-hub reliability factor \( \alpha \) and intra-hub reliability factor \( \gamma \)

This chapter also employs two reliability factors in the hub location models, the inter-hub reliability factor \( \alpha \) and intra-hub reliability factor \( \gamma \). As explained in previous chapter, Inter-hub reliability factor \( \alpha \) \( (0 \leq \alpha \leq 1) \) is devised to reflect a degree of capability of inter-hub links in transmitting flows. This factor is imposed on the inter-hub link term in the form of \( (r_{km})^{1-\alpha} \). Intra-hub reliability factor \( \gamma \) \( (0 \leq \gamma \leq 1) \) is the probability of the hub itself transmitting traffic without delay or signal attenuation of traffic. From a practical point of view, both factors are set in the model in order to reflect the reliable level of physical facilities which can affect the capability of the route to transfer the traffic on the network. Note that the higher value of \( \alpha \) or \( \gamma \) encourages use of inter-hubs or intra-hubs, respectively. More importantly, determining the best route for each OD pair depends on
what levels of these two factors are assumed in the model. Given both factors and considering the amount of traffic to be delivered between O and D, equation 1 can be restated as below:

\[
L_{OD} = W_{ij} \cdot L_{ykm} \\
= W_{ij} \left[ L^p_{ykm} + L^s_{ykm} \cdot L^s_{ikmn} \right] \\
= W_{ij} \cdot (1 - R^p_{ykm}) + W_{ij} \cdot (1 - R^p_{ykm}) \cdot (1 - R^s_{ykm}) \\
= W_{ij} \left( 1 - r_{ik} \cdot r^{1-\alpha}_{km} \cdot r_{nj} \right) + W_{ij} \left[ 1 - (1 - r_{ik} \cdot r^{1-\alpha}_{km} \cdot r_{nj}) \right] \cdot (1 - r_{in} \cdot \gamma^{1-\alpha} \cdot r_{nj}) \\
\]  
(Equation 4-2)

Where

- \( L_{OD} \) the amount of traffic loss
- \( W_{ij} \) the amount of flow traveling between \( i \) and \( j \)
- \( \alpha \) Inter-hub reliability factor (0 ≤ \( \alpha \) ≤ 1)
- \( \gamma \) Intra-hub reliability factor (0 ≤ \( \gamma \) ≤ 1)

Based on these two components, sections 4.3.3, 4.3.4, and 4.3.5 present the formulations of the PROSA, PROMA and the PROPS, respectively.

4.3.3. PROSA (p-Hub Protection Model with Single Assignment)

As mentioned previously, the purpose of the PROSA is to present a basic model which embeds reliability-relevant components in a classical hub-location model with single assignment. The PROSA can be regarded as a protection model for normal
circumstances in that it simply optimizes the network in order to minimize the potential traffic loss of interacting flows among OD pairs. The model’s formulation is detailed as below.

Minimize \[ Z = \sum_i \sum_j \sum_k \sum_m W_{ijkm} X_{ijkm} \] (4-1)

Subject to

\[ \sum_k Z_{ik} = p \quad (2 \leq p) \] (4-2)

\[ \sum_k Z_{ik} = 1 \quad \forall \ i \] (4-3)

\[ Z_{ik} - Z_{kk} \leq 0 \quad \forall \ i, k \ (i \neq k) \] (4-4)

\[ \sum_m X_{ijkm} - Z_{ik} = 0 \quad \forall \ j > i; k \] (4-5)

\[ \sum_k X_{ijkm} - Z_{jm} = 0 \quad \forall \ j > i; m \] (4-6)

\[ Z_{ik} \in \{0, 1\} \] (4-7)

\[ 0 \leq X_{ijkm} \leq 1 \] (4-8)

Where

\( p \) the number of hubs to be located

\( X_{ijkm} \) the fraction of flow from origin \( i \) to destination \( j \) via hub \( k \) and \( m \) in that order (\( i \rightarrow k \rightarrow m \rightarrow j \))

\( Z_{ik} \) 1 if node \( i \) is allocated to hub \( k \); 0 otherwise

\( Z_{kk} \) 1 if node \( k \) is a hub; 0 otherwise
The objective of PROSA is to minimize the amount of potential traffic loss for the set of OD pairs. Constraint (4-2) specifies the number of hubs to be open, and constraint (4-3) requires each node to be assigned to only a single hub. Constraint (4-4) requires a hub to be opened before a node is allocated to hub \( k \) denoted as \( Z_{kk} \). Constraints (4-5) and (4-6) guarantee that flow \( i \) to \( j \) should not be routed via hubs \( k \) and \( m \) unless origin \( i \) is linked to hub \( k \) and \( j \) is linked to hub \( m \). Constraint (4-7) impose an integer property on the links and hub facility variables \( Z_{ik} \).

**4.3.4. PROMA (p-Hub Protection Model with Multiple Assignments)**

The PROMA allows origin \( i \) to route its interactions with other destinations \( j \) by selecting different hub pairs if different hub selections can give the better less-flow loss routes. This characteristic can help a network be more resilient to possible disruptions of inter-hub links or hubs via diversified routes for interacting flows between OD pairs.

Minimize \( Z = \sum_{i} \sum_{j} \sum_{k} \sum_{m} W_{ijkm} X_{ijkm} \) \hspace{1cm} (4-9)

Subject to

\[
\sum_{k} Z_{ik} = p \hspace{1cm} (2 \leq p) \hspace{1cm} (4-10)
\]

\[
\sum_{k} \sum_{m} X_{ijkm} = 1 \hspace{1cm} \forall j > i \hspace{1cm} (4-11)
\]

\[
\sum_{m} X_{ijkm} - Z_{ik} \leq 0 \hspace{1cm} \forall j > i; k \hspace{1cm} (4-12)
\]
\[
\sum_{k} X_{ijk} - Z_{m} \leq 0 \quad \forall j > i; m \quad (4-13)
\]

\[
Z_{ik} \in \{0, 1\} \quad \forall k \quad (4-14)
\]

The objective function is the same with the PROSA. Constraints (4-11) to (4-14) are used the same as in classical MA hub location models (Bryan and O’Kelly, 1999). Constraints (4-11) ensure that all flow from \( i \) to \( j \) should travel through hub(s) \( k \) and \( m \) (\( k \) can be equal to \( m \)). Constraints (4-12) and (4-13) prevent the flow from \( i \) to \( j \) being routed through non-hub nodes. Constraints (4-14) impose binary integer restrictions on decision variables.

### 4.3.5. PROPS (p-Hub Protection Model with Primary and Secondary routes)

The PROPS is the strongest version of the three models because it optimizes a network by determining the primary as well backup routes. Consider the notations used in Equation 4-1 and 4-2 with an additional notation for the formulation of PROPS.

\[
\text{Minimize} \quad Z = \sum_{i} \sum_{j} \sum_{k} \sum_{m} W_{ij} \left( L_{ijk} X_{ijk} + \sum_{n} L_{ijn} X_{ijn} \right)
\]

Subject to

\[
\sum_{k} Z_{ik} = p \quad (3 \leq p) \quad (4-16)
\]

\[
\sum_{k} Z_{ik} = 1 \quad \forall i \quad (4-17)
\]

\[
Z_{ik} - Z_{kk} \leq 0 \quad \forall i \neq k \quad (4-18)
\]
\[ \sum_{m} X_{ijkm} - Z_{ik} = 0 \quad \forall \ j > i; \ k \] (4-19)

\[ \sum_{k} X_{ijkm} - Z_{jm} = 0 \quad \forall \ j > i; \ m \] (4-20)

\[ X_{ijkm} - Z_{nm} \leq 0 \quad \forall \ j > i; \ k, \ m, \ n \] (4-21)

\[ \sum_{n} X_{ijkm} - X_{ijkm} = 0 \quad \forall \ j > i; \ k, \ m \] (4-22)

\[ Z_{ik} \in \{0, 1\} \quad \forall \ i, \ k \] (4-23)

\[ 0 \leq X_{ijkm} \leq 1 \] (4-24)

Where

\( X_{ijkm} \) the fraction of flow traveling origin \( i \) to destination \( j \) via hub \( n \) (\( i \rightarrow n \rightarrow j \))

The objective function (4-15) is to minimize the amount of potential traffic loss via primary as well as back-up routes for OD pairs. The first term in the objective function is to determine the best primary route between \( i \) and \( j \) (\( X_{ijkm} \)), and the other terms are the possible disjoint route for \( X_{ijkm} \). Note that the PROPS can be seen as a dual-layer network design problem because the model incorporates two different routing networks (i.e., the layer for primary and back-up routes, respectively). Constraint (4-16) requires no more than three hubs be open in order for back-up routes. As shown in constraints (4-17) to (4-20), the network for primary routes is designed based on the single assignment scheme. Constraints (4-17) force each node to be assigned only a single hub and (4-18) require a hub to be opened before a node is assigned to hub \( k \) (\( Z_{ik} \)). Constraints (4-19) and (4-20) guarantee that flow \( i \) to \( j \) should travel via hubs \( k \) and \( m \) only if each is linked
through hub $k$ and $m$. Constraints (4-21) require that each back-up route should be
disjointed with primary route except OD nodes by using a hub $n$, which is not used in the
primary route. Constraints (4-22) impose that each primary route should have one back-
up route. According to this formulation, PROPS also satisfies a minimum connectivity
requirement of a node ($k \geq 2$) as a survivability requirement.

4.4. Model Applications

4.4.1. Data and solution method

For model applications, this chapter uses the $R$-matrix and $W$-matrix for 14 nodes,
both of which are used in the last chapter. The $R$-matrix is made based on an empirical
traffic latency rate for 14 United States cities, which is provided by the SAVVIS network
provider as of 2005.

Note that the reduction of the back-up routes is very important to solve the
problems. The number of back-up route $X_{ijkmn}$ would be prohibitive for a large size of
problems because the number of routing variables is following the complexity of $O(n^5)$,
in theory. In this chapter, an additional reduction method in the number of impractical
routing variables of $X_{ijkmn}$ and $X_{ijkm}$ is considered to solve the problems efficiently. The
reduction is carried based on the enumerated practical routes of $X_{ijkm}$ which are expressed
as a set $S=\{(i,j,k,m)|(j>i)\cap[(k=i)\cup(k=m=j)\cup(k\neq i\cap k\neq j\cap m\neq j)]\}$ (see Figure 3.3 for more
detail). Accordingly, the set $S'$ for back-up routing variables $X_{ijkmn}$ is expressed as $S'=\{(i,j,k,m,n)|[(j>i)\cap(n\neq j\neq m)]\cap[(k=i)\cup[(k=m=j)\cap(n\neq k)]\cup(k\neq i\cap k\neq j\cap m\neq j)]\}$. 98
As illustrated Figure 4.3, the back-up routes for each exclusive case are all a one-hub stop disjoint-routes. Based on this reduction, all problems in this chapter were solved to optimality using CPLEX 10.1 on a Dual Intel Pentium III Xeon 2.0 GHz with 2 GB RAM.

In the following sections, the experiments focus on analyzing the multi-facet behaviors of the models, where primary concern is on examining the relationship between hub locations, redundancies of the network, and hub activity levels under various parameter conditions. The experiments were made for three-, four-, and five hub-models under the selective parameter settings of $\alpha$ and $\gamma$, which range from 0.10 to 0.95 with appropriate increments.
<table>
<thead>
<tr>
<th>Subset of $S'$</th>
<th>Routing variable $s'$ $(i, j, k, m, n) \in S'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(j \neq i) \cap (k = i)$ &amp; $(n \neq j) \cap (n \neq m)$</td>
<td><img src="image1.png" alt="Image 1" /> <img src="image2.png" alt="Image 2" /> <img src="image3.png" alt="Image 3" /></td>
</tr>
<tr>
<td>$N_1' = n(n-1)^2(n-2)/2$</td>
<td><img src="image4.png" alt="Image 4" /></td>
</tr>
<tr>
<td>$(j \neq i) \cap (n \neq k) \cap (n \neq m)$</td>
<td><img src="image7.png" alt="Image 7" /></td>
</tr>
<tr>
<td>$(j \neq i) \cap (k \neq i \cap k \neq j \cap m \neq i)$ &amp; $(n \neq j) \cap (n \neq k) \cap (n \neq m)$</td>
<td><img src="image10.png" alt="Image 10" /></td>
</tr>
</tbody>
</table>

**Figure 4.3.** The enumerated routing variable set $S'$.

**Note:** Total number of variables $X_{ijkmn}$ is $n(n-1)(n-2)(n^2-4n+5)/2$, which is the sum of each subset $N_1'$, $N_2'$ and $N_3'$. 
4.4.2. Basic results

The computational results of the three models for \( p=3, 4, \) and 5 with selected ranges of \( \alpha \) and \( \gamma \) are summarized in Table 4.1. It should not be surprising that the objective function of all models decreases as both reliability parameters increase because the enhanced transferability of the facilities, hubs and inter-hubs links result in improved network performance for handling interacting flows. Three observations are noteworthy.

First, other things being equal, the objective function value of the PROMA is smallest among three models, which implies the PROMA is the best with regard to network performance. The main reason is due to its flexible routing scheme. In general, the multiple assignments scheme allows more possible routes than the single assignment (O’Kelly et al. 1996).

Second, the optimal hubs are not necessarily located in nodes that have large telecommunication potential such as San Francisco, Los Angeles and New York. Rather, small potential nodes such as Seattle, Phoenix, Minneapolis, and Denver can be frequently selected as optimal hub locations over various levels of both factors.

Third, the optimal hubs are selected differently among the models (horizontal comparison in Table 4.1), but simply increasing \( p \) within a model with the same level of factors would not influence the change of hub locations dramatically (vertical comparison of each model in Table 4.1).
<table>
<thead>
<tr>
<th>$\alpha$ &amp; $\gamma$</th>
<th>PROSA</th>
<th></th>
<th></th>
<th>PROMA</th>
<th></th>
<th></th>
<th>PROPS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hubs</strong></td>
<td><strong>Obj.value</strong></td>
<td><strong>Hubs</strong></td>
<td><strong>Obj.value</strong></td>
<td><strong>Hubs</strong></td>
<td><strong>Obj.value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three hubs model ($p=3$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>PHX MIA MIN</td>
<td>10679.98</td>
<td>PHX ATL MIA</td>
<td>1128.29</td>
<td>PHX MIA MIN</td>
<td>11429.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>PHX MIA MIN</td>
<td>8045.08</td>
<td>PHX ATL MIA</td>
<td>1009.60</td>
<td>PHX MIA MIN</td>
<td>8881.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>PHX MIA MIN</td>
<td>5344.10</td>
<td>PHX MIA PHL</td>
<td>866.66</td>
<td>PHX MIA MIN</td>
<td>6259.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.70</td>
<td>PHX MIA MIN</td>
<td>3003.15</td>
<td>MIA PHL SEA</td>
<td>730.96</td>
<td>PHX MIA MIN</td>
<td>3977.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td>PHX MIN PHL</td>
<td>1209.94</td>
<td>MIA PHL SEA</td>
<td>599.69</td>
<td>PHX MIN PHL</td>
<td>2122.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>PHX PHL SEA</td>
<td>781.40</td>
<td>MIA PHL SEA</td>
<td>538.23</td>
<td>PHX PHL SEA</td>
<td>1697.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four hubs model ($p=4$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>PHX MIA MIN SEA</td>
<td>10283.02</td>
<td>CHI NY PHL SEA</td>
<td>816.01</td>
<td>PHX MIA MIN SEA</td>
<td>10893.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>PHX MIA MIN SEA</td>
<td>8068.84</td>
<td>CHI MIA PHL SEA</td>
<td>746.56</td>
<td>PHX MIA MIN SEA</td>
<td>8742.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>PHX MIA MIN SEA</td>
<td>5693.21</td>
<td>CHI MIA PHL SEA</td>
<td>672.77</td>
<td>PHX MIA MIN SEA</td>
<td>6421.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.70</td>
<td>PHX MIA MIN SEA</td>
<td>3407.04</td>
<td>CHI MIA PHL SEA</td>
<td>598.00</td>
<td>PHX MIA MIN SEA</td>
<td>4179.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td>PHX MIN PHL SEA</td>
<td>1339.26</td>
<td>CHI MIA PHL SEA</td>
<td>522.92</td>
<td>PHX MIN PHL SEA</td>
<td>2100.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>PHX MIN PHL SEA</td>
<td>826.58</td>
<td>PHX MIA PHL SEA</td>
<td>485.14</td>
<td>PHX MIN PHL SEA</td>
<td>1562.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five hubs model ($p=5$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>PHX DEN MIA MIN SEA</td>
<td>10739.36</td>
<td>CHI MIA NY PHL SEA</td>
<td>758.48</td>
<td>PHX DEN MIA MIN SEA</td>
<td>11282.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>PHX DEN MIA MIN SEA</td>
<td>8630.60</td>
<td>CHI MIA NY PHL SEA</td>
<td>693.60</td>
<td>PHX DEN MIA MIN SEA</td>
<td>9237.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>PHX DEN MIA MIN SEA</td>
<td>6299.39</td>
<td>CHI MIA NY PHL SEA</td>
<td>625.44</td>
<td>PHX DEN MIA MIN SEA</td>
<td>6956.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.70</td>
<td>PHX DEN MIA MIN SEA</td>
<td>3876.20</td>
<td>PHX CHI MIA PHL SEA</td>
<td>555.38</td>
<td>PHX ATL MIA MIN SEA</td>
<td>4566.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td>PHX DEN MIN PHL SEA</td>
<td>1486.02</td>
<td>PHX CHI MIA PHL SEA</td>
<td>475.20</td>
<td>PHX MIN PHL SEA</td>
<td>2187.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>PHX DEN MIN PHL SEA</td>
<td>876.81</td>
<td>PHX MIA NY PHL SEA</td>
<td>444.04</td>
<td>PHX DEN MIA PHL SEA</td>
<td>1548.71</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1. Selected results for 3, 4 and 5 hub models with $\alpha$ and $\gamma$ varying from 0.10 to 0.95

Figure 4.4 illustrates one of the snapshots of the 4-hubs model configurations of three models at $\alpha$ and $\gamma=0.90$. As expected, PROMA diversifies the routes in terms of each node compared to PROSA; and the network of PROPS has more linkages than the other two models. Notice that each non-hub node in PROPS is linked with more than two other nodes, which indicates that the network redundancy of PROPS directly helps a network be more resilient as well as keeps the flow loss of hubs at a minimum in the face of possible failures on the network. PROSA might not be a survivable network model due to its limited assignment scheme, but the model itself optimizes the traffic loss-minimizing network at normal condition.

The expected capability of flow protection on the network can be examined in terms of the change of flow activity levels of hubs. As mentioned in the previous section, how effectively a hub network can maintain flows of hubs is an important dimension especially in hub and spoke networks because most of the flows for OD pairs are encouraged to use inter-hub links via hubs for flow economies of scale. In this chapter, hub activity envelope is devised to depict the possible outcomes of a network to increased failures on inter-hub links, which are the most important assets in hub and spoke-type networks because a certain combination of malfunctions of inter-hub links could have a critical impact on the loss of flow activity in hub facilities (O’Kelly 1986).
Figure 4.4. PROSA, PROMA and PROPS for 4 hubs with $\alpha, \beta = 0.90$. 
Figure 4.5 illustrates the hub activity envelopes for three models with selected levels of $\alpha$ and $\gamma (= 0.30, 0.70, \text{and } 0.90 \text{ for } p=5)$. For better comparison among the models, the envelope of the PROSA is shown on the left, while both envelopes of the PROMA and PROPS are illustrated together on the right panel. In each envelope, a series of computations of hub activity levels defines a range of losses from best- to worst-case scenarios of hub activity level (i.e. total amount of flows using $p$ hubs) as possible failures impact $r$ out of total $pC_2$ inter-hub links increases.

As mentioned in previously, losing flows on the network, even to a small disruption, is inevitable, however, Figure 4.5 highlights that the difference of flow protection among the models is quite evident. First, the PROPS always keep the hub activity levels better than the other models, and the PROMA protects hub flow activity compared to the PROSA. More importantly, the PROPS delivers the interacting flow at a certain level even with complete disruption of inter-hub links at the final phase ($10C_{10}$) with the help of the back-up routes, whereas the PROMA and PROSA often lost its capability to deliver flows.

Second, the models whose envelope has a wider gap between lower and upper bounds, indicates that some inter-hub links are highly over-utilized, at the same time highly underutilized inter-links exist. Theoretically, various different shapes of hub activity envelopes can be formed based on the type of network structure (see O’Kelly and Kim 2007). Note that an envelope which can keep the narrow gap between the best and worst scenarios is regarded as a more robust network, less potentially prone to any disruptions.
Figure 4.5. Hub activity envelopes for PROSA, PROMA and PROPS ($p=5$).
Related to the flow distribution, we present three extensions with regard to constraints on inter-hub links: 1) the capacitated PROMA and PROPS models (PROMA-CAP and PROPS-CAP); 2) the minimum threshold model (PROMA-MT and PROPS-MT); and 3) the minimum threshold on back-up routes model (PROPS-MTB).

4.5. Model Extensions

4.5.1. Capacitated PROMA (PROMA-CAP) and PROPS (PROPS-CAP)

One of the critical questions in survivable network design is how to assign the capacity of linkages. An excessive large volume of flows concentrating on particular inter-hub links may entail a critical consequence of network degradation if any disruption occurred to those links. It therefore is important for network designers to explore what capacity is to be assigned to the inter-hub links to accommodate hub activity level (Soni, Gupta and Pirkul 1999).

Imposing appropriate level of capacities on inter-hub links could produce a balanced flow distribution on the network. Thus, it is worthwhile to explore how impact on hub activity levels differs if standard models are capacitated. PROPS-CAP or PROMA-CAP models are formulated simply by adding the following constraints (4-25) to the standard formulation of the PROMA ((4-9) to (4-14)) and PROPS ((4-15) to (4-24)), respectively:

\[
\sum_{j} \sum_{k} W_{ij} R_{ijkn} X_{ijkm} \leq \Omega_{km} \quad \forall \ j > i; \ k, \ m \ (k \neq m) \quad (4-25)
\]
Where

\( \Omega_{km}^U \) the capacity of flow imposed on inter-hub links connecting hubs \( k \) and \( m \)

Constraint (4-25) imposes \( \Omega_{km}^U \), which places a maximum amount of flow (CAP) on inter-hub links \( k \) and \( m \). The amount of interacting flows cannot exceed a given level of \( \Omega_{km}^U \). Table 4.2 summarizes the results of capacitated models for five hub models under two selected levels of factors (both \( \alpha, \gamma = 0.70 \) and 0.90). Note that all the models experience the increase of objective function value as the forced capacity level increases, which shows that the network performance could be affected by the imposed capacity level. Increasing capacity level also impacts the change of hub location. Note that the hubs in the PROMA-CAP are less sensitive to capacity changes compared to PROPS-CAP. The reason of the result can be explained by the model behavior in determining optimal hubs location according to the assignment scheme. The first response of PROMA is to change its allocations from a hub to other hubs for interacting flows for each OD taking advantage of its routing flexibility, whereas PROPS prefers to move the hub from one to another until the best combination of a hub set is identified (O’Kelly and Bryan 2002).
### Table 4.2. Results for capacitated models: 5-hubs PROMA and PROPS.

**Note:** Infeasible solutions are found below the capacity level of 500.


<table>
<thead>
<tr>
<th>Hub location (α and γ = 0.90)</th>
<th>CAP level</th>
<th>Objective value</th>
<th>Hub location (α and γ = 0.70)</th>
<th>CAP level</th>
<th>Objective value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>No cap</td>
<td>475.20</td>
<td>PHX CHI MIA PHL SEA</td>
<td>No cap</td>
<td>555.38</td>
</tr>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>3000</td>
<td>476.60</td>
<td>PHX CHI MIA PHL SEA</td>
<td>3000</td>
<td>557.40</td>
</tr>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>2000</td>
<td>494.87</td>
<td>PHX CHI MIA PHL SEA</td>
<td>2000</td>
<td>575.06</td>
</tr>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>1000</td>
<td>537.13</td>
<td>PHX CHI MIA PHL SEA</td>
<td>1000</td>
<td>614.90</td>
</tr>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>500</td>
<td>561.16</td>
<td>PHX CHI MIA PHL SEA</td>
<td>500</td>
<td>636.10</td>
</tr>
<tr>
<td>PHX MIA MIN PHL SEA</td>
<td>No cap</td>
<td>2187.22</td>
<td>PHX ATL MIA MIN SEA</td>
<td>No cap</td>
<td>4566.25</td>
</tr>
<tr>
<td>PHX MIA MIN PHL SEA</td>
<td>3000</td>
<td>2190.99</td>
<td>PHX ATL MIA MIN SEA</td>
<td>3000</td>
<td>4566.25</td>
</tr>
<tr>
<td>PHX MIA MIN PHL SEA</td>
<td>2000</td>
<td>2264.89</td>
<td>PHX ATL MIA MIN SEA</td>
<td>2000</td>
<td>4612.10</td>
</tr>
<tr>
<td>PHX DEN MIA MIN SEA</td>
<td>1000</td>
<td>2348.88</td>
<td>PHX ATL DEN MIA MIN</td>
<td>1000</td>
<td>4773.60</td>
</tr>
<tr>
<td>PHX ATL DEN MIA MIN</td>
<td>500</td>
<td>2512.05</td>
<td>PHX ATL DEN MIA MIN</td>
<td>500</td>
<td>4784.18</td>
</tr>
</tbody>
</table>
4.5.2. Minimum threshold model (PROMA-MT and PROPS-MT)

The idea of inter-hub linkages is that they are intended to carry agglomerated flows so that a network can enjoy economies of scale. Thus, it is unrealistic to expect inter-hub links to carry only low volumes, especially in hub network design of telecommunications in which the costs for establishing physical inter-hub links are quite expensive (Bryan 1998; Campbell et al. 2002). The minimum threshold model is stressed if extremely underutilized inter-hub links are found on the network. To formulate the PROMA-MT or PROPS-MT, the constraints (4-26) to (4-28) are added in the standard formulation of the PROMA with the constraints (4-9) to (4-14) and the PROPS, from (4-15) to (4-24), respectively.

\[
\sum_{i} \sum_{j} W_{ij} \cdot R_{ijkm} X_{ijkm} \geq \Omega_{km}^{L} Y_{km} \quad \forall j > i; k, m (k \neq m) \quad (4-26)
\]

\[
X_{ijkm} - Y_{km} \leq 0 \quad \forall j > i; k, m (k \neq m) \quad (4-27)
\]

\[
Y_{km} \in \{0, 1\} \quad \forall k, m \quad (4-28)
\]

Where

\[
\Omega_{km}^{L} \quad \text{the minimum threshold level over inter-hub link between hubs } k \text{ and } m
\]

\[
Y_{km} \quad 1 \text{ if inter-hub link } km \text{ is open; } 0 \text{ otherwise}
\]
Constraint (4-26) requires that the inter-hub link $Y_{km}$ cannot be open unless the amount of inter-hub flow is greater than the minimum amount threshold, $\Omega_{L_{km}}$. Constraint (4-27) ensures that no route utilizing inter-hub link $Y_{km}$ is used without opening the inter-hub link $Y_{km}$. Constraint (4-28) imposes binary integer restrictions.

Table 4.3 summarizes the results of PROMA-MT and PROPS-MT. Similar to the result of CAP models, the objective function value increases as a higher level of minimum threshold level is imposed. Note that inter-hub links which cannot satisfy the MT level are suppressed in the MT models. For instance, to a higher MT level (=3000), the PROMA-MT optimizes the network that suppress four inter-hub links out of 10 links (40%).

More importantly, the impact from a forced flows level from CAP or MT models is much clearly shown through hub activity envelopes. Figure 4.6 compares the hub activity envelopes of the CAP and MT models with standard models. As the dotted upper and lower envelope curves show, the CAP and MT models form a more narrow range of envelope compared to the standard model, which complete a more balanced flow distribution among inter-hub links due to the given CAP or MT level. Based on our results, the effect of minimum threshold is more clearly shown in MT models, and reduced once any of inter-hub links are suppressed.
Table 4.3. Results for minimum threshold models: 5-hubs PROMA and PROPS.

<table>
<thead>
<tr>
<th>Hub location (α and γ = 0.90)</th>
<th>MT level</th>
<th>Objective value</th>
<th>Hub location (α and γ = 0.70)</th>
<th>MT Level</th>
<th>Objective value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>No MT</td>
<td>475.20</td>
<td>PHX CHI MIA PHL SEA</td>
<td>No MT</td>
<td>555.38</td>
</tr>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>500</td>
<td>475.23</td>
<td>PHX CHI MIA PHL SEA</td>
<td>500</td>
<td>555.38</td>
</tr>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>1000</td>
<td>475.96</td>
<td>PHX CHI MIA PHL SEA</td>
<td>1000</td>
<td>556.95</td>
</tr>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>2000</td>
<td>484.73</td>
<td>CHI MIA NY PHL SEA</td>
<td>2000</td>
<td>563.53</td>
</tr>
<tr>
<td>PHX CHI MIA PHL SEA</td>
<td>3000</td>
<td>496.96</td>
<td>PHX CHI MIA PHL SEA</td>
<td>3000</td>
<td>573.18</td>
</tr>
</tbody>
</table>

PROMA-MT

<table>
<thead>
<tr>
<th>Hub location (α and γ = 0.90)</th>
<th>MT level</th>
<th>Objective value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHX MIA MIN PHL SEA</td>
<td>NO MT</td>
<td>2187.22</td>
</tr>
<tr>
<td>PHX MIA MIN PHL SEA</td>
<td>500</td>
<td>2195.51</td>
</tr>
<tr>
<td>PHX MIA MIN PHL SEA</td>
<td>1000</td>
<td>2255.10</td>
</tr>
<tr>
<td>PHX ATL MIA PHL SEA</td>
<td>1500</td>
<td>2485.33</td>
</tr>
</tbody>
</table>

PROPS-MT
Figure 4.6. Hub activity envelopes for selected capacitated and minimum threshold models
4.5.3. Minimum threshold on back-up linkage in the PROPS (PROPS-MTB)

It is no surprising that the PROPS can protect flows on the network better than other models due to its increased redundancy with back-up routes. However, such prescription generally generates a complex network with extremely high levels of connectivity (Grötschel et al. 1995; Klincewicz 2006). Further, back-up linkages sometimes are created only to carry a small amount of flow, which is unrealistic in telecommunication networks (Wu 1992). As an extension of the PROPS, the minimum threshold on back-up linkages can be an important issue. The formulation is extended from the standard PROPS with the following constraints from (4-29) to (4-35):

\[
\sum_{j} \sum_{k} \sum_{m} \left( W_{pq} R_{pqmk}^{s} X_{pqmk}^{s} + W_{qj} R_{qjkm}^{s} X_{qjkm}^{s} \right) = F_{pq} \quad \forall p, q \quad (4-29)
\]

\[
\sum_{i} \sum_{k} \sum_{m} \left( W_{iq} R_{iqkm}^{s} X_{iqkm}^{s} + W_{iq} R_{iqkm}^{s} X_{iqkm}^{s} \right) = F_{iq} \quad \forall p, q \quad (4-30)
\]

\[
F_{pq} + F_{qp} \geq \Omega_{pq}^{L} Y_{pq}^{b} \quad \forall p, q \quad (4-31)
\]

\[
X_{pqmk} - Y_{pq}^{b} \leq 0 \quad \forall j > p; k, m \quad (4-32)
\]

\[
X_{iqkm} - Y_{pq}^{b} \leq 0 \quad \forall q > i; k, m \quad (4-33)
\]

\[
Y_{pq}^{b} = Y_{qp}^{b} = 0 \quad \forall j > i \quad (4-34)
\]

\[
Y_{pq}^{b} \in \{0, 1\} \quad \forall p, q \quad (4-35)
\]
where

\[ Y_{pq}^b = \begin{cases} 1 & \text{if back-up link between node } p \text{ and } q \text{ is open; 0 otherwise} \\ \Omega_{pq}^L & \text{the minimum threshold of flow for back-up link } pq \end{cases} \]

Constraints (4-29) and (4-30) calculate the amount of flow traveling on back-up link \( pq \) based on the flow direction \( p \rightarrow q \) and \( q \rightarrow p \), respectively. Constraints (4-31) ensure that no back-up linkage can be open unless the sum of total flows on the back-up link \( pq \) keep the flow greater than the minimum threshold level \( (\Omega_{pq}) \). Constraints (4-32) to (4-33) ensure that back-up route associated with link \( pq \) can be used only if link \( Y_{pq}^b \) is open \( (Y_{pq}^b = 1) \). Constraints (4-34) are optional, but used to specify the link \( Y_{pq} \) as the non-directional same physical link with \( Y_{qp} \).

Table 4.4 shows the relationship between total length of linkages and objective function values as the levels of back-up minimum threshold increase. In order to explore the direct comparison of the impact of the varied MTB levels, the tests are designed with a set of fixed hubs which are optimal hub locations in the un-capacitated models in Table 4.1 for three different reliability factor levels \( (\alpha \text{ and } \gamma \text{ levels } = 0.90, 0.70, \text{ and } 0.50) \) with \( p = 4 \text{ and } 5 \) problems. Since the network cost can be simply translated from the total length of the linkages in the networks (Chamberland and Sansò 2001), the total linkage length in terms of total and back-up linkages is calculated.
Table 4.4 (a)

<table>
<thead>
<tr>
<th>4 hub models (p = 4)</th>
<th>Linkage length (miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>α and γ MTB level</strong></td>
<td><strong>Obj.value</strong></td>
</tr>
<tr>
<td>0.90 No-MTB</td>
<td>2100.87</td>
</tr>
<tr>
<td>100</td>
<td>2114.90</td>
</tr>
<tr>
<td>200</td>
<td>2116.46</td>
</tr>
<tr>
<td>300</td>
<td>2120.30</td>
</tr>
<tr>
<td>400</td>
<td>2122.05</td>
</tr>
<tr>
<td>0.70 No-MTB</td>
<td>4566.25</td>
</tr>
<tr>
<td>100</td>
<td>4582.78</td>
</tr>
<tr>
<td>200</td>
<td>4583.06</td>
</tr>
<tr>
<td>300</td>
<td>4583.67</td>
</tr>
<tr>
<td>400</td>
<td>4586.58</td>
</tr>
<tr>
<td>0.50 No-MTB</td>
<td>6956.22</td>
</tr>
<tr>
<td>100</td>
<td>6971.39</td>
</tr>
<tr>
<td>200</td>
<td>6972.62</td>
</tr>
<tr>
<td>300</td>
<td>6866.93</td>
</tr>
</tbody>
</table>

Table 4.4 (b)

<table>
<thead>
<tr>
<th>5 hub models (p = 5)</th>
<th>Linkage length (miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A and γ MTB level</strong></td>
<td><strong>Obj.value</strong></td>
</tr>
<tr>
<td>0.90 Non-MTB</td>
<td>2187.22</td>
</tr>
<tr>
<td>100</td>
<td>2187.22</td>
</tr>
<tr>
<td>200</td>
<td>2187.22</td>
</tr>
<tr>
<td>300</td>
<td>2188.31</td>
</tr>
<tr>
<td>400</td>
<td>2190.49</td>
</tr>
<tr>
<td>500</td>
<td>2192.90</td>
</tr>
<tr>
<td>0.70 Non-MTB</td>
<td>4179.37</td>
</tr>
<tr>
<td>100</td>
<td>4194.75</td>
</tr>
<tr>
<td>200</td>
<td>4195.92</td>
</tr>
<tr>
<td>300</td>
<td>4196.45</td>
</tr>
<tr>
<td>400</td>
<td>4198.18</td>
</tr>
<tr>
<td>0.50 Non-MTB</td>
<td>6421.39</td>
</tr>
<tr>
<td>100</td>
<td>6434.30</td>
</tr>
<tr>
<td>200</td>
<td>6434.85</td>
</tr>
</tbody>
</table>

Table 4.4. Results for back-up minimum threshold models (p=4 and 5, α and γ = 0.90, 0.70 and 0.50)
As shown in Table 4.4, the total length of the linkages generally decreases as minimum threshold levels increase. Based on the numerical results, the total length of the linkages generally decreases as minimum threshold levels increase because a back-up linkage which carries a very small amount of flow is suppressed to be closed unless the amount of flows utilizing the link is greater or equal than a given minimum threshold level. Reducing the total length of linkages may indicate the more saving of total network costs. Table 4.4 also highlights that there is a trade-off between the objective function value and its minimum threshold levels. Increasing minimum threshold could help to reduce the network costs, but the network performance to deliver the traffic would be degraded.

Figure 4.7 visualizes the network configurations of 4-hubs standard PROPS model and the PROPS-MTB model with 400 minimum thresholds, respectively. As expected, the PROPS-MTB model construct less-sparse network structure compared to the standard model.
a) Optimal network: PROPS standard model

b) Optimal network: PROPS-MTB (Minimum Threshold = 400)

**Figure 4.7.** Comparison of network design between PROPS and PROPS-MTB.
Figure 4.8. Comparison of hub activity envelopes for the PROPS and PROPS-MTB (p=4, \(\alpha\) and \(\gamma = 0.70\)).
Figure 4.8 compares the resiliency of the network based on best and worst scenario of hub activity envelopes between 4-hubs standard PROPS and the PROPS-MTB (MTB = 400). The resiliency of the network to probable malfunction from primary-linkages or inter-hub links is also influenced by the reduced number of back-up linkages. In terms of the protection level of flows to the complete malfunctions of inter-hub links, the standard model is more resilient to the possible disruptions of linkages as well as better protective of flows in both the best- and worst-case scenarios.

4.6. Summary

The hub network should be a less-attractive system in terms of survivability (Wu 1992). However, it is the most economical and efficient architecture so that it is widely applied in current telecommunication network systems. In this context, this chapter addresses a survivable hub network design. By considering the characteristics of IP-based network systems, reliability factors are newly employed and a series of models are formulated.

Given that this chapter focuses on the standard hub models in terms of survivable network design, further research is possible on these models. First, a multi-objective type of model which takes into account other aspects such as cost structure of facilities can be addressed. Second, considering that most hub location models are difficult to solve for substantially large problems (O’Kelly et al. 1996), it is necessary to develop efficient heuristics, particularly for the PROPS model. Finally, by embedding other survivable
requirements and performance-related constraints (e.g., tolerance level), more realistic models can be developed.
CHAPTER 5

HEURISTICS FOR RELIABLE \( p \)-HUB LOCATION PROBLEMS

5.1. Introduction

Reliable \( p \)-hub location problems are basically similar to conventional hub location models in that the problem is to determine both the optimal location of hubs and allocations of non-hub nodes to those hub(s) simultaneously.

One of challenging issues of the hub location problem is the solution of larger problems optimally (Daskin 1995; Campbell 1994; Groothedde 2005). Basically, most hub-and-spoke models are known as \( NP \)-Hard, which means that they are unlikely to be solved within a polynomial time in the entire problem class (Cormen et al. 2001; Nievergelt 2000). Therefore, the development of heuristics to get optimal or near solutions efficiently in a reasonable time is an important issue among researchers (O’Kelly 1987; Pham and Karaboga 2000; Xiao 2006).
It is known that single assignment problems are much more difficult to solve optimally compared to multiple assignment problems because the assignment of the non-hub nodes to hubs still remains \(NP\)-hard even once a set of hubs \((p \geq 3)\) is fixed (Skorin-Kapov et al. 1994; Kara and Tansel 1998). In contrast, the multiple assignment problem can be solved in polynomial time for an assignment level once the set of hubs is fixed. For this reason, many heuristics in hub location problems have focused on single assignment.

This chapter thus concentrates on the development of heuristics to solve selected problems for reliable hub network design. The first concern of this chapter is to develop heuristics for such problems that involve a single assignment scheme: the MRSA, and PROPS. The motivation for this is that the MRSA is regarded as a representative single assignment problem, and the PROPS is the most difficult problem to solve among all of the models presented in this dissertation. As mentioned in chapter 4, the PROPS should determine one location phase as well as two allocation phases – the primary routes based on single assignment plus back-up routes based on the disjoint paths of primary routes, simultaneously. In this context, two different heuristic algorithms are proposed to solve the problem: 1) TABUSA, which is a tabu-heuristic for the MRSA, and 2) TABUEX, a hybrid search for the PROPS, which is developed based on the tabu approach coupled with exhaustive search.

The second purpose of this chapter is to test the performance of enumeration-based exhaustive search algorithms for the MRSA and MRMA, which are different in assignment scheme. The results will be compared to the performance of the other approaches, heuristic and LP.
The chapter is organized as follows. The next section briefly reviews the development of solution approaches in hub network design. In section 5.3, the general structures of exhaustive search and tabu heuristics are described. Section 5.4 and 5.5 present the new heuristic search algorithms for the TABUSA and TABUEX, respectively. In section 5.6, computational experiments are provided, followed by conclusions in section 5.7. The model formulations of the MRSA and PROPS are presented in chapter 3 and 4, respectively. Thus, they are not repeated here.

5.2. Solution Approaches in Hub Network Design

Since O’Kelly (1987) first formulated the classical hub location problem with a single-assignment scheme, a number of solution methods have been proposed. For small size problems (less than 10 nodes), the exhaustive search approach based on the enumeration of all possible instances was proposed for single or multiple assignment problems (Aykin 1995; Ernst and Krishnamoorthy 1998). Even though the exhaustive approach guarantees exact optimal solutions, this method becomes quickly intractable as the size of the problem increases. Thus, an enumeration-based heuristic is often combined with other solution techniques.

O’Kelly (1987) proposed two different enumeration-based heuristics for the single assignment problem. Both basically rely on enumerating all possible sets of hub locations, and then different ‘distance-based’ assignments of non-hub nodes to hubs were explored according to the level of discount factor. Klincewicz (1991) proposed the application of exchange-heuristics as well as a clustering heuristic for deciding location
of hubs and allocations of non-hub nodes to hubs in a clustered area. As an advanced heuristic approach, Klincewicz (1992) presented a tabu search (TS) and greedy based search procedure (GRASP) for the case of single allocation problem. The most advanced tabu search heuristic was developed by Skorin-Kapov and Skorin-Kapov (1994), which is known as the best solution technique for single assignment problems to the standard data set (CAB). However, the quality of solutions was not measured because no bounds were determined from that approach. Subsequently, O'Kelly et al. (1995) presented a method to compute the lower bounds for the single hub location problem in order to evaluate the performance of heuristics.

The tight linear programming (LP) relaxation approach is the most common approach, which is proven to be an effective method to optimally solve moderate size problems (see Skorin-Kapov et al. 1996 in more detail). They demonstrated that most reasonable size problems (less than 25 nodes) can be solved effectively as integers by reducing the total number of variables and constraints used in original single and multiple allocation problems. However, the method often produces fractional solutions or its solvability is highly dependent upon the computational capability of machine used.

As a special case, for the fixed two-hub location case with single allocation, linear programming transformed from the previous mixed integer program was presented using minimum cut method (Sohn and Park 1997). This method proved that the 2-hub location problem can be solved in polynomial time. Pirkul and Schilling (1998) developed an efficient procedure to solve single allocation problems utilizing Lagrangian relaxation and sub-gradient optimization to produce tight upper and lower bounds.
In the class of uncapacitated single allocation problems that the number of hubs should be determined endogenously, a genetic algorithm was developed to solve this problem (Abdinnour-Helm and Venkataramanan 1995). Followed by this work, Klincewicz (1996) also provided a heuristic with a dual ascent algorithm for uncapcitated multiple allocation problems, and recently, a hybrid type of heuristic which combines tabu search with a genetic algorithm was presented to improve the quality of solutions for a large problem (Abdinnour-Helm 1998). For the multiple allocation hub location problems, Ernst and Krishnamoorthy (1998) proposed a tight-cutting and branch solution technique based the shortest paths tested on a given set of hubs. Their work formulated a new mixed integer linear programming, and benchmarked with exact enumeration algorithms in order to present a better solvability of their method in terms of solution time.

Recent studies in telecommunication hub network design have concentrated on reducing the computational burden in solving large scale systems by applying different types of search approaches in location and allocation level, respectively. For example, after determining the locations of hubs using tabu search, the phase of assigning non-hub nodes to the hubs is explored using a genetic algorithm (Carello et al. 2004). In sum, both enhancing the quality of solution and developing a way to increase optimal solvability are common concerns in most hub location models.
5.3. Exhaustive Search and Tabu Search Algorithms

5.3.1. Exhaustive search

Exhaustive search is conceptually simple but it is often regarded as a powerful tool to get the exact solution unless it fails to terminate. Recently, as computing power and the storage capacity of computer systems have been enhanced, the exhaustive search algorithm is often used to benchmark the performance of heuristics, even though the problem is in the class of NP-hard (Nievergelt 2000).

As discussed in previous work (Skorin-Kapov and Skorin-Kapov 1994; Ernst and Krishnamoorthy 1998), the complexity of the exhaustive search in the single assignment problem is measured as $O\left(\binom{n}{p}p^{(n-p)}\right)$ for $n$ nodes and $p$ hubs. Note that the complexity of this enumeration heavily relies on the term $p^{(n-p)}$ which accounts for the complexity of assignment level. For example, given $n=10$ and $p=2$, the total enumeration for all network configurations requires just 46,080. However, with only a small increment for $p=4$, the total enumeration increases dramatically to 860,160. This indicates it is clearly intractable even for a moderate size of $n$ and $p$. In other words, the exhaustive search for solving single assignment problem is effective only for the small size problems.

Figure 5.1 illustrates a simple algorithmic structure of exhaustive search for single assignment problem, which is termed EX-SA (Exhaustive search for MRSA).
**Exhaustive search procedure: EX-SA**

**Step 0:** Set Initialize:
- Set Reliable factor: = α, γ
- Set current solution as $obj_c$
- Set incumbent solution as $obj_{inc}$
- Set incumbent assignment vector $X_{inc}$

**Step 1:** Enumerate all possible $p$-hub set $H_p$ based on $^nC_p$
- Create non-hub nodes assignment set $NH_{np}$ for each instance of $H_p$
- Label each $H_p$ and its corresponding $NH_{np}$

**Step 2:** Location Phase (LOC_PHASE)

*Loop 1:*
- Call a set from $H_p$ and its corresponding $NH_{np}$ based in the labeled order

**Step 2.1. Allocation Phase (ALLO_PHASE)**

*Loop 2:*
- Set current assignment vector $X_c$
- Assign $(n-p)$ non-hub nodes to a fixed $p$-hubs

**Step 2.2. Evaluate objective value**

*If $obj_c > obj_{inc}$ then*
- $obj_{inc} ← obj_c$
- $X_{inc} ← X_c$

**Step 2.2.1. All $NH_{np}$ are evaluated?**

*If yes, then end loop2, otherwise, do loop2*

*Endloop 2*

**Step 2.3. All $H_p$ are evaluated?**

*If yes, then end loop1, otherwise, do loop1*

*Endloop 1*

**Step 3:** Terminate procedure and display incumbent solution

---

**Figure 5.1.** Procedure of exhaustive search for single assignment problem (EX-SA).
In multiple assignment problems, the exhaustive search type-search is used often because the problem can be solved quickly for the established set of \( p \)-hubs (Klincewicz 2002; Ernst and Krishnamoorthy 1998; Aykin 1995). For example, Klincewicz (2002) used the enumeration-based solution approach to solve the FLOWLOC model which was developed based on multiple assignment. It was possible to solve this problem by first enumerating all possible instances of \( p \)-hub locations for \( n \) nodes \( (nC_p) \), and then identifying the best route for each \( ij \) pair among all possible routes \( X_{ijkm} \) using shortest paths algorithm. However, employing this approach would be only effective for small size of \( p \) since the number of instances of the hubs set is exponential in \( p \).

As illustrated in Figure 5.2, the structure of exhaustive search, termed the EX-MA (Exhaustive search for the MRMA) is similar to that of the single assignment problem. But the capability of EX-MA to deal with increasing problem size is quite different due to the flexibility in deciding routes for each \( ij \) pair. In other words, each \( ij \) can decide its best route regardless of the other pairs’ route decisions. Note that in the EX-SA, the vector should separate the non-hubs nodes with the selected hub nodes (termed ‘\( NH_{np} \)’ ‘\( H_p \)’ and in Figure 5.1) in order to assign each non-hub node to only one of the candidate hubs. In contrast, the EX-MA does not need this procedure. More specifically, the complexity of location level for the EX-MA is \( O(nC_p) \) for given \( n \) and \( p \), which is the same with that of the EX-SA. In the allocation level, however, the complexity of the allocation phase is only required \( O(nC_2\cdot(p^2)) \), indicating that enumerating all possible routes for all \( ij \) pairs can be completed within a reasonable time unless \( p \) is a large number. The computational experiments and discussion of exhaustive search will be presented in the section 5.5.
Exhaustive search procedure: EX-MA

Step 0: Set Initialize:
    Set Reliable factor: = α, γ
    Set current solution as $obj_c$
    Set incumbent solution as $obj_inc$
    Set incumbent routing vector $X_{ijkm_inc}$

Step 1: Enumerate all possible $p$-hub set $H_c$ based on $nC_p$
    Label all hub set of $H_c$

Step 2: Location Phase (LOC_PHASE)
    Loop 1:
        Call a set from $H_c$ based in the labeled order
    Step 2.1. Allocation Phase (ALLO_PHASE)
        Loop 2:
            Set current assignment vector $X_{ijkm_c}$
            Swap hub $k$ for each ij pair ($k=1 \rightarrow p$)
        Loop 3:
            Swap hub $m$ for each ij pair ($m=1 \rightarrow p$)
            If $X_{ijkm_c}$ is impractical route then do loop 3
        Loop 4:
            Step 2.2. Evaluate objective value $X_{ijkm_c}$
            If $obj_{ijkm_c} > obj_{ijkm_inc}$ then
                $X_{ijkm_inc} \leftarrow X_{ijkm_c}$
                $obj_{ijkm_inc} \leftarrow obj_{ijkm_c}$
                $obj_c = obj_c + obj_{ijkm_inc}$
            EndlLoop 4
            EndLoop 3
    Step 2.2.1. All $X_{ijkm_c}$ are evaluated?
        If yes, then end loop 3, otherwise, do loop 3
    If $obj_c > obj_inc$ then
        $obj_inc \leftarrow obj_c$
    Endloop 2
    Step 2.3. All $H_c$ are evaluated?
        If yes, then end loop 1, otherwise, do loop 1
    Endloop 1

Step 3: Terminate procedure and display incumbent solution

---

Figure 5.2. Procedure of exhaustive search for multiple assignment problems (EX-MA).
5.3.2. **Tabu search**

Tabu search is a meta-heuristic approach to reach near optimal or ultimately global optimality by guiding the local search to continue exploring solution space beyond local optima. This approach is known as a promising solution approach in hub location problems especially for single assignment models (Klincewicz 1992; 2000).

The fundamental principle of tabu search is to overcome local optima using ‘tabu’ lists when the procedure goes from one solution to another. The ‘tabu’ is defined as an element, which is used as an element in a current solution, but forbidden to use while it is listed in the tabu list to avoid the reversal use of the elements in the tabu list for a certain amount of iterations (see Glover 1989 and 1990 for general structure of tabu search).

In this chapter, the heuristics TABUSA (Tabu search for the MRSA) and TABUEX (A hybrid search for the PROPS) are basically developed based on the structure of the tabu search approach. The strategy of tabu heuristic consists of a couple of components: 1) strategy of long-term memory, 2) different levels of heuristics for location, allocation, and back-up route allocation, and 3) tabu lists with aspiration criteria. The role of each component is described as follows.

5.3.2.1. **Long-term memory (LTM)**

The purpose of long term memory aims to find probable best optimal solutions by beginning from a very different, but better configuration of a feasible solution. The invocation of long-term memory is used to construct the new hub sets based on
frequency; the occurrence of a node selected as hub or assigned nodes during the previous search procedure. The short-term memory is implemented in each phase where a new iteration is needed to begin in each phase for given parameters.

5.3.2.2. Location phase (LOC_PHASE)

The location phase (hereafter LOC_PHASE) employs tabu search to find the best candidate $p$ hubs. Basically, the rule is to select good candidate hubs based on each node’s potentiality, which is defined as the amount of the interactions with other nodes. A single hub exchange rule is employed in this PHASE where non-hub nodes are swapped with one of current hubs one by one until the iteration is allowed. After evaluating exchanges, the best one is updated in current hub set and the current objective function is found. The exchanged hub nodes are listed in the locational tabu list ($loc_{tabu}$) to forbid the reversal use as hubs for the future iterations if the seat of the tabu list is also allowed.

5.3.2.3. Allocation phase (ALLO_PHASE)

As mentioned previously, enumerating all possible events in the allocation level of the single assignment scheme becomes quickly intractable due to its complexity. Therefore, the strategy of re-assignment of non-hub nodes to the candidate hubs focuses on narrowing the searching space, which can produce a better solution. In this allocation phase, it needs to distinguish non-hub nodes with the currently selected $p$ candidate hubs set. After the first assignment of each non-hub node to one of $p$ hubs based on the
potentiality of each node, re-allocation is carried out for all non-hub nodes based on the
evaluation process of the ‘move’ when a non-hub node changes its assignment from
current assigned hub to other candidate hubs (see step 3.2. in the section 5.4.1).

5.3.2.4. Survivable routing phase (SUR_PAHSE)

This phase is included in the TABUEX for the PROPS. As mentioned previously,
the back-up route for each $ij$ pair is determined as a disjoint path to the primary route of
it. The searching approach for this phase adapts an exhaustive search algorithm based on
complete enumeration of all possible disjoint routes once a primary route of $ij$ pair is
determined. It is clearly reasonable to use exhaustive search in this phase since the
computational complexity for enumerating all $X_{ijkmn}$ is just $O(n^2p)$. This indicates the
SUR_PAHSE is tractable to find the best survivable routes within a polynomial time.

5.3.2.5. Tabu lists

The purpose of setting up the tabu lists is to prevent from being entrapped in local
optimality by avoiding re-selecting a node as a candidate hub (for LOC_PHASE), and re-
assigning a node to the particular hubs (for ALLO_PHASE) which are in the tabu lists
(termed loc_tabu, allo_tabu, respectively). For example, once a node is exchanged with a
hub from the current hub set $H$, then the node is moved into a tabu list and it is not
allowed to be used again for a certain number of iterations. However, the tabu list is
inactive only if move gets a better solution or pass over the aspiration criteria to the best
known objective function value so far. The aspiration criteria used in the heuristics will be described in the next section.

5.4. Tabu Search for the MRSA (TABUSA)

The mathematical formulation of the MRSA is already presented in chapter 3. Thus, this section is more focused on describing the structure of the tabu-search algorithm, termed TABUSA (a tabu search algorithm for MRSA). The overall structure would be similar to the other tabu-search algorithms, but different strategic rules are considered.

5.4.1. Algorithmic procedure

The procedure for the algorithm presented in Figure 5.3, and detailed techniques are described as follows.

- **Step 0: Initialization of parameters:** Set the size of tabu lists (loc_tabu, allo_tabu) for LOC_PHASE and ALLO_PHASE, respectively. Set the number of iterations for each phase (max_location, max_allocation). Set the number of iterations for long term memory (max_longterm) to invoke long-term memory.

- **Step 1: Long-term memory:** If long-term memory is invoked (i.e. $1 \leq \text{max\_longterm}$), the initial configuration (selection of hubs and non-hubs assignments) of every invocation is determined based on the frequency, which records how frequently
each node is selected as a hub and assigned to a specific hub during the previous procedure. Tabu lists are set as empty whenever long-term memory restarts.

**Step 2: Initial solution:** In this step, a feasible solution configuration is constructed with the current hub set and the set of assignments of non-hub nodes to hubs. Each set is stored as a \((p \times 1)\) vector \(H_c = [h_1, h_2, ..., h_p]\) and \((n \times 1)\) vector \(X_c = [h_1, h_2, ..., h_n]\), respectively. Note that a selected hub node should be assigned to itself in the hub assignment vector \(X_c\) according to the model formulation of the MRSA. The initial hubs are selected based on the evaluation of each node’s potentiality, the sum of potential interacting flows to the other destinations \((\sum_{j=1}^{k} w_{ij} r_{ij} / frequency)\), then select \((p-1)\) nodes from the lowest, and pick one highest node as initial hubs. This rule is made based on the observation of the model experiments in Chapters 3 and 4 where the optimal hub locations are not always located to the largest city nodes. Rather, small size nodes are often preferred as optimal hubs. In contrast, non-hub nodes are generally well allocated to the highest reliability hub the among candidate hubs.

**Step 3.1: Location phase (LOC_PHASE):** The single hub exchange rule in this phase works to exchange a non-hub node of the non-hub nodes set \(NH_{np}\) with one of current hubs one by one. As one non-hub node is changed, then the re-assignment for all non-hub nodes is performed based on a ‘reliability-based’ allocation rule, where each non-hub node is assigned from the highest reliability hub to the lowest ones during the iterations. Among the prescribed number of iterations, the best configuration is stored as the current hub set and current objective function.
The exchanged hub nodes are listed in the *loc_tabu* to forbid the reversal use of this non-hub node for the future iterations. As a forbidding strategy during iterations, the LOC_PHASE performs the best tabu exchange if it improves the incumbent objective value after carrying the best assignment for all *ij* pairs.

**Step 3.2: Allocation phase (ALLO_PHASE):** The most time-consuming procedure is allocation phase. As mentioned, enumerating all possibilities of feasible solution structure of assignment for fixed set of hubs is \( O(p^{(n-p)}) \). Thus, the main strategy of the ALLO_PHASE is to find good assignment sets to the fixed hubs by evaluating moves in the feasible solution structure. After evaluating the first assignments, the exchange of the assignments from one hub to another for all non-hub nodes is performed during the permitted iterations. Note that the rule of the re-assignment is done by evaluating the value of move for each non-hub node ‘z’ from current assigned hub ‘u’ to candidate hub ‘v’. Here, the move is measured as the difference of total potential loss of flow of non-hub node *z* to the other *j* when *z* is re-assigned from candidate hub *u* to *v*. Specifically, it is defined as the function below.

\[
move = \sum_{j=1}^{n} w_{ij}(1 - r_{zu} \cdot r_{um}^{1-z} \cdot r_{mj}) - \sum_{j=1}^{n} w_{ij}(1 - r_{zv} \cdot r_{vm}^{1-z} \cdot r_{mj}) 
\]  
(Equation 5.1)

If a move passes the level of aspiration criteria (a range of small values from 0-50 is given in the tests), then it performs the best tabu exchange of allocation or performs the best non-tabu reallocation (see steps 3.2.1 and 3.2.2 in Figure 5.3).
The ALLO_PHASE also creates allocation tabu list (allo_tabu) to prevent the incursion of the same assignment during the iterations. Perform the best tabu exchange, which improves the incumbent objective value \( obj_{inc} \) after carrying on the best re-allocation followed by updating the new incumbent objective function, best hub set, allocation tabu, and assignment vector.

**Step 4: Terminate and display**: After terminating all procedures, display the saved incumbent objective value \( obj_{inc} \), the best hub set, and assignments vector \( X_{c,inc} \).
Tabu search procedure: TABUSA

Step 0: Set initialize:
- Max_iteration: (max_location, max_allocation)
- Longterm memory size: (max_longterm)
- Set longterm: = 0
- Set Reliable factor: = α, γ

Step 1: Invoke long-term memory (LTM ≥1)
While longterm ≤ max_longterm

Step 2: Initial Solution:
* Choose p hubs based on selection rule and store them into the hub list for next long-term memory procedure, then allocate each non-hub node to hub based on assignment rule.
- Set current solution as obj_c
- Set incumbent solution as obj_inc

Step 3: LOC_PHASE and ALLO_PHASE

Step 3.1: LOC_PHASE * Tabu_location heuristic
- Set loc_iter := 0
- Set loc_tabu := empty
While loc_iter ≤ max_location

- During iterations for exchange between non-hub node and hub node, Evaluate the current solution obj_c with obj_inc
- Exchange single non-hub node q with hub node r
- Evaluate the exchange
  
  \[ \text{obj_temp}(q, r) \] with \[ \text{obj}_c \]
- \[ \text{If } \text{obj_temp}(q, r) > \text{obj}_c \text{ then} \]
  \[ \text{obj}_c \leftarrow \text{obj_temp}(q, r) \]
- loc_iter := loc_iter + 1
- Start ALLO_PHASE

---

**Figure 5.3.** Procedure of the tabu search TABUSA for the MRSA.
Figure 5.3. Continued

Step 3.2 ALLOC_PHASE * Tabu_Allocation heuristic

While $allo_{iter} \leq max_{allocation}$

- Set $iter_{allocation} := 0$
- Set $allo_{tabu} := empty$
  
  Exchange re-allocation of non-hub node $z$ to hub $u$ and $v$, then evaluate every re-assignment $obj_c(z, u)$ with $obj_temp(z, v)$

3.2.1 Perform the best allocation if it improves $obj_{inc}$, then

  - Update $obj_{inc} \leftarrow obj_c$
  - Update $X_{c_{inc}} \leftarrow X_c$
  - Update $H_{inc} \leftarrow H_c$

3.2.2 Otherwise, perform the best non-tabu reallocation. If it improves $obj_{inc}$, then

  - Update $obj_{inc} \leftarrow obj_c$
  - Update $X_{c_{inc}} \leftarrow X_c$
  - Update $H_{inc} \leftarrow H_c$
  - Update $allo_{tabu}$
  - $allo_{iter} = allo_{iter} + 1$

Endwhile

Step 3.3 Perform the best non-tabu exchange or the best tabu exchange if it passes aspiration criteria. If it improves $obj_{inc}$ then

- Update $obj_{inc} \leftarrow obj_c$
- Update $X_{c_{inc}} \leftarrow X_c$
- Update $H_{inc} \leftarrow H_c$
- Update $loc_{freq}$
- Update $loc_{tabu}$
- $loc_{iter} = loc_{iter} + 1$

Endwhile

$longterm := longterm + 1$
5.5. A hybrid Search for PROPS (TABUEX)

In general, a hybrid type search algorithm is preferably prescribed in network design if there exists a polynomial time phase of the entire problem structure, or a special need to diversify searching spaces to finding a optimal/approximate optimal solution effectively (Ghashghai and Rardin 2002).

TABUEX is a hybrid searching algorithm which combines two different searching strategies: the tabu-search for the location and allocation phases, plus the enumeration-based exhaustive search for the survivable routes phase (SUR_PHASE). The structures for two search phases (LOC_PHASE, ALLO_PHASE) with tabu search algorithms are same with those of TABUSA. Thus, this section is more focused on describing the algorithmic procedure of SUR_PHASE. Detailed description is presented below.

5.5.1. Algorithmic procedure

Having the set of hubs and assignments fixed, the first step of the SUR_PHASE is to enumerate all probable survivable routes for each $ij$ according to the routing scheme $S'$ (see Figure 4.3. for more detailed illustration of the set $S'$). With help of the routing scheme $S'$, the complexity of this phase is just $O(n^2(2p))$. Identifying disjoint paths is a simple procedure. That is to search and select one of the hubs which is not used in primary route of $ij$. Thus, this procedure is not computationally burdensome when it is compared to the other phases. The second step is to compare the objective values of all
enumerated instances. Once the best back-up routes for each \( ij \) pair (termed \( Z_s \)) are identified, then they are added to the current objective (\( obj_c \)) as the next step. If the \( obj_c \) satisfies the aspiration criteria, then the a non-tabu or the best tabu exchange procedures are performed.

5.6. Computational Results

5.6.1. Data sets

For computational experiments, the 25-node data set which was obtained from the IP network performance statistics of QWEST Internet service provider (http://stat.qwest.net) as of 2005 is used. This data includes average reliabilities among 25 city nodes where their POP servers are resided in the United States. Similar to the SAVVIS data set used in Chapters 3 and 4, reliabilities among nodes in the data are also considerably correlated with geographical distance. That is, the reliability between two nodes is expected to decrease with the length of its route. The potential interactions \( W_{ij} \), is simply estimated by multiplying the population size of node \( i \) and \( j \). To compare the performance of the other searching approaches (MIP and exhaustive search), different sizes of data sets, 10-node, 15-node, and 20-node are created as a subset of the 25 nodes set which is illustrated in Figure 5.5. Considering the solvability of each approach, the number of hubs \( p \) ranging from 3 to 6 based on the problem size, and two intermediate levels of reliability factors, both \( \alpha \) and \( \beta = 0.60 \) and 0.90, are tested.
Hybrid search procedure: TABUEX

**Step 0: Initialization of parameters:** Set the number of maximum iterations (max_sur_iter) for survivable route phase (SUR_PHASE). Given a large number will perform a complete enumeration of all possible routes.

**Step 1: Long-term memory:** Same with TABUSA.

**Step 2: Initial solution:** Add the initial disjoint survivable route for determined primary route for each $ij$ pair. Save this current location-allocation vector as $X_c$, and survivable route vector $[i,j,k,m,n]$ as $X_{c_ijklmn}$.

**Step 3: LOC_PHASE, ALLO_PHASE, and SUR_PHASE heuristics**

**Step 3.1: Location Phase (LOC_PHASE):** Same with TABUSA.

**Step 3.2: Allocation Phase (ALLO_PHASE):** Same with TABUSA.

**Step 3.3: Survivable route Phase (SUR_PHASE)**

Set $\text{sur}_\text{iter} = 0$

While $\text{sur}_\text{iter} \leq \text{max}_\text{sur}_\text{iter}$ (or complete enumeration is allowed)

* Given $H_{inc}$ and $X_{c_{inc}}$, perform exhaustive search to find the best survivable routes fixed $ij$ pair. Calculate the objective value ($Z_c$) of the best survivable route for $ij$ pairs and add to $\text{obj}_c$

$\text{sur}_\text{iter} = \text{sur}_\text{iter} + 1$

Endwhile

* Perform the best tabu exchange if it improves the $\text{obj}_{inc}$ after carrying the best survivable routes for all $ij$ pairs.

Update solution: $\text{obj}_{inc} \leftarrow \text{obj}_c$

Update $X_{c_{inc}} \leftarrow X_c$

Update $X_{inc_{ijklmn}} \leftarrow X_{c_{ijklmn}}$

**Step 4: Terminate and display:** Display $\text{obj}_{inc}$, $H_{inc}$, $X_{inc}$ and $X_{inc_{ijklmn}}$

---

**Figure 5.4.** Procedure of the hybrid search TABUEX for the PROPS.
Figure 5.5. List of 25 nodes.

Note: The members of 10, 15, 20-node set are as follows.

5.6.2. Computational experiments

All exhaustive search algorithms and heuristics are implemented in FORTRAN 77 and run on a Dual Intel Pentium III Xeon, 2.0 GB RAM on the Windows NT. In order to benchmark heuristics for various sets of problems, the Mixed Integer Programming (hearafter MIP) approach for all instances was performed using commercial optimization software CPLEX 10.1 MIP solver under the same platform.

Table 5.1 shows that the parameter setting of tabu size, maximum number of iterations, and maximum invocation of long term memory allowed in each phase of the TABUSA and TABUEX used in the tests. The CPU time (sec.) is measured when the best solution is found in heuristics. That is, if the best solution is obtained with a smaller iteration levels, then it is regarded as the best CPU time.

<table>
<thead>
<tr>
<th>Problem size</th>
<th>Phase</th>
<th>tabu size</th>
<th>Max. iteration</th>
<th>Max. Ltm</th>
<th>tabu size</th>
<th>Max. iteration</th>
<th>Max. Ltm</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/15</td>
<td>LOC_PHASE</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ALLO_PHASE</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>SUR_PHASE</td>
<td>complete enumeration</td>
<td></td>
<td></td>
<td>complete enumeration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20/25</td>
<td>LOC_PHASE</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ALLO_PHASE</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>SUR_PHASE</td>
<td>complete enumeration</td>
<td></td>
<td></td>
<td>complete enumeration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1. Tabu size and maximum iterations of each phase applied in the heuristics.
In order to compare the quality of solutions among different solution approaches, a measure $\lambda$, which is suggested by Skorin-Kapov et al. (1994) is used. The $\lambda$ is defined as follows.

$$\lambda = \frac{[(\text{Objective value} - \text{Best known objective value})]}{\text{Best known objective value}} \quad \text{(Equation 5.2)}$$

Of the first concern is to examine the solving capability of the EX-SA and EX-MA. As shown in Table 5.2, both exhaustive searches solved the 10-node problem quickly. Note that the exhaustive search for the EX-MA performs always better than the EX-SA due to its better computational complexity as well as the reduction of impractical routes which are not needed to be evaluated. The EX-MA solved all cases of the 10-node problem in a second. For example, the case of $p = 4$, the EX-SA enumerates 860,160 possible instances, whereas the EX-MA just needs the evaluation for the total 129,780 instances (15% of the EX-SA).

For the 15-node set, the capability of the exhaustive search to solve the problem is much different. The EX-SA took more than 4 hours for $p=3$, and even 2 days for $p=4$ due to the dramatic increase of instances. This fact implies that solving larger problems is clearly extremely time-consuming in terms of the current platform, and becomes intractable. In contrast, the EX-MA demonstrates its good solvability to the problem for the 25-node set. The solution time of the EX-MA is influenced with the increase of $p$, but it solves the problems within a few minutes in our tests.
### Table 5.2. Results of the computational experiments of the EX-SA and EX-MA.

**Note:** 1) BUR (Burbank), DAL (Dallas), ECK (Eckington), SAC (Sacramento), SLC (Salt Lake City), TMP (Tampa), KAS (Kansas City), WAS (Washington D.C.).

2) The MRSA problem for 25 nodes set is not tested since the total enumerations is beyond the capability of the current machine.

3) The solution times using CPLEX for the MRMA ranges from a few seconds to 22 seconds. All problems are solved to optimality.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha$, $\beta$ level</th>
<th>Problem</th>
<th>Objective</th>
<th>Solution Time (sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.6</td>
<td>EX-SA</td>
<td>9,107,247</td>
<td>7.7</td>
<td>ATL SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EX-MA</td>
<td>9,902,769</td>
<td>&lt; 0.1</td>
<td>ATL BOS DAL</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>EX-SA</td>
<td>9,883,886</td>
<td>7.5</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EX-MA</td>
<td>10,028,997</td>
<td>&lt; 0.1</td>
<td>ATL BOS DAL</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
<td>EX-SA</td>
<td>8,917,977</td>
<td>27.4</td>
<td>ATL SLC BOS TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EX-MA</td>
<td>9,960,629</td>
<td>0.1</td>
<td>ATL BOS DAL DEN</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>EX-SA</td>
<td>9,836,545</td>
<td>25.3</td>
<td>ATL SLC BOS TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EX-MA</td>
<td>10,082,222</td>
<td>0.1</td>
<td>ATL BOS DAL DEN</td>
<td>0.0</td>
</tr>
</tbody>
</table>

#### 10-node problems

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha$, $\beta$ level</th>
<th>Problem</th>
<th>Objective</th>
<th>Solution Time (sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6</td>
<td>EX-SA</td>
<td>14,645,587</td>
<td>15,870</td>
<td>MIN SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EX-MA</td>
<td>15,988,868</td>
<td>0.28</td>
<td>KAS MIN WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>EX-SA</td>
<td>15,979,871</td>
<td>15,856</td>
<td>KAS SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EX-MA</td>
<td>16,147,037</td>
<td>0.28</td>
<td>KAS SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
<td>EX-SA</td>
<td>14,658,212</td>
<td>2.14 (days)</td>
<td>ATL MIN SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EX-MA</td>
<td>16,065,629</td>
<td>1.48</td>
<td>DAL DEN KAS WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>EX-SA</td>
<td>15,937,175</td>
<td>2.06 (days)</td>
<td>KAS SLC TMP WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EX-MA</td>
<td>16,208,217</td>
<td>1.48</td>
<td>DAL KAS SLC WAS</td>
<td>0.0</td>
</tr>
</tbody>
</table>

#### 15-node problems

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha$, $\beta$ level</th>
<th>Problem</th>
<th>Objective</th>
<th>Solution Time (sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.6</td>
<td>EX-MA</td>
<td>19,397,766</td>
<td>38.0</td>
<td>BUR CHI ECK NY</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>EX-MA</td>
<td>19,565,378</td>
<td>38.1</td>
<td>BUR CHI NY SAC</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>0.6</td>
<td>EX-MA</td>
<td>19,444,444</td>
<td>244.8</td>
<td>BUR CHI ECK NY TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>EX-MA</td>
<td>19,623,068</td>
<td>247.0</td>
<td>BUR CHI ECK NY TMP</td>
<td>0.0</td>
</tr>
</tbody>
</table>
The main reason for the good performance of the EX-MA can be explained by the reduction methods, although the effect of the reduction varies with given $n$ (see Figure 3.1 in detail) as well as the dramatic enhancement of the computing of current machines.

As the second issue, computational results of three different approaches for the MRSA are presented in Table 5.3. To compare the solution time among the solution approaches, the results of the exhaustive search (the EX-SA) are included. Two observations are noteworthy from this table. First, in 10-node problems, all three approaches obtained the optimal solution very quickly within several seconds. However, in the 15- and 25-node sets, various levels of solution times are identified among the approaches, clearly showing the merits of each approach. Note that the Mixed Integer Programing (MIP) approach using CPLEX solves well in all problems, producing optimal solutions. However, the performance is dependent upon the size of problem, the number of $p$, and the level of reliability factors. As the reliability factors get lower, the solution time of the MIP approach generally increases with producing a greater number of branch and bounds. Finally, as shown in the Table 5.3, the TABUSA performs well compared to the LP relaxation solution approach since it is not influenced by the level of reliability factors, but by the level of parameters. Although the TABUSA shows a similar performance level to the others in solving the small size problems, as the problem size increases, it performs very well in obtaining the known-optimal solutions for the 25-node problems, just requiring 10-20 seconds.
<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha, \beta$ level</th>
<th>Approach</th>
<th>Objective</th>
<th>Solution Time (sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.6</td>
<td>MIP</td>
<td>9,107,247</td>
<td>2.0</td>
<td>ATL SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exhaustive</td>
<td>9,107,247</td>
<td>7.7</td>
<td>ATL SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>9,107,247</td>
<td>0.4</td>
<td>ATL SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>9,883,886</td>
<td>0.1</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exhaustive</td>
<td>9,883,886</td>
<td>7.5</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>9,883,886</td>
<td>0.9</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
<td>MIP</td>
<td>8,917,977</td>
<td>1.1</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exhaustive</td>
<td>8,917,977</td>
<td>27.4</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>8,917,977</td>
<td>0.9</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>9,836,545</td>
<td>0.1</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exhaustive</td>
<td>9,836,545</td>
<td>25.3</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>9,836,545</td>
<td>0.8</td>
<td>ATL SLC BOS</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**15-node problems**

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha, \beta$ level</th>
<th>Approach</th>
<th>Objective</th>
<th>Solution Time (sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.6</td>
<td>MIP</td>
<td>14,645,587</td>
<td>1046.1</td>
<td>MIN SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exhaustive</td>
<td>14,645,587</td>
<td>15870.0</td>
<td>MIN SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>14,645,587</td>
<td>2.9</td>
<td>MIN SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>15,979,871</td>
<td>0.6</td>
<td>KAS SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exhaustive</td>
<td>15,979,871</td>
<td>15856.0</td>
<td>KAS SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>15,979,871</td>
<td>2.7</td>
<td>KAS SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
<td>MIP</td>
<td>14,658,212</td>
<td>184.8</td>
<td>ATL MIN SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exhaustive</td>
<td>14,658,212</td>
<td>2.14 (days)</td>
<td>ATL MIN SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>14,658,212</td>
<td>2.1</td>
<td>ATL MIN SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>15,937,175</td>
<td>1.1</td>
<td>KAS SLC TMP WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exhaustive</td>
<td>15,937,175</td>
<td>2.06 (days)</td>
<td>KAS SLC TMP WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>15,937,175</td>
<td>4.9</td>
<td>KAS SLC TMP WAS</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**25-node problems**

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha, \beta$ level</th>
<th>Approach</th>
<th>Objective</th>
<th>Solution Time (sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.6</td>
<td>MIP</td>
<td>18,781,577</td>
<td>14886.7</td>
<td>BUR DEN ECK STR(^1)</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>18,781,577</td>
<td>22.7</td>
<td>BUR DEN ECK STR</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>19,475,905</td>
<td>23.0</td>
<td>BUR ECK NY TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>19,475,905</td>
<td>9.2</td>
<td>BUR ECK NY TMP</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>0.6</td>
<td>MIP</td>
<td>18,906,513</td>
<td>1786.0</td>
<td>BUR DEN ECK SEA TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>18,906,513</td>
<td>23.6</td>
<td>BUR DEN ECK SEA TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>19,492,846</td>
<td>18.1</td>
<td>BUR ECK NY SEA TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>19,492,846</td>
<td>12.8</td>
<td>BUR ECK NY SEA TMP</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 5.3. Results of the computational experiments with the 10, 15, and 25 node problems for the MRSA.

Note: 1) STR (Sterling).
The good performance of the TABUSA can be explained in two ways: 1) the selection rule of initial candidate hubs, and 2) a ‘reliability-based exchange’ rule for the allocation level. Both rules are reflected based upon the observations which are commonly found in the results in previous chapters 3 and 4.

Figure 5.6 is one of the snapshots showing the characteristics of the optimal locations of hubs and their assignments of nodes with the 15-node problem \((p=4, \alpha\text{ and } \beta=0.9, 0.6)\). As shown in both optimal solutions, the first observation is that optimal hubs are not always located in the cities with a large amount of interactions such as New York (1\textsuperscript{st}) and Chicago (2\textsuperscript{nd}). Rather, the city nodes with small amounts of interactions such as Salt Lake City (15\textsuperscript{th}), Tampa (14\textsuperscript{th}), Kansas City (11\textsuperscript{th}), and Washington D.C. (7\textsuperscript{th}), are frequently selected as optimal hubs (see Table 5.2 and 5.3) in entire results. As explained in algorithmic procedure of the TABUSA, this characteristic is important in constructing the initial solution. By beginning from a highly probable hub set, the entire procedure can improve its solution quality to reach optimal solutions. Second, the assignment of a non-hub node is highly dependent on its reliabilities to the candidate hub nodes.

As indicated in Figure 5.6 (a), all non-hub nodes are assigned to its highest reliability hub among candidate hubs. Of course, this rule is not always applied for all problems as shown in Figure 5.6 (b) where Dallas, Boston, and Washington D.C. are not assigned to the hub with its second highest reliability. The reason for this property is that the assignment is also jointly influenced by both the amount of interactions and the reliability factor’s level.
Figure 5.6. Optimal hubs and assignments of non-hub nodes for the 15-node problem.
In general, the smaller reliability factors given, the less possible it is to be assigned to a closer node. The logic of both rules is simple, but the combinations of these rules pose a good solution status, and ultimately guide the search for optimal locations and allocations effectively.

Third, the efficiency and solvability of the heuristic is well demonstrated to solve the PROP. As summarized in Table 5.4, the MIP approach obtained the optimal solutions for the 10-node problems for the PROPS. For the 15-node problems, all instances tested are solved to optimality, but most instances require at least 40 minutes, and even a few hours in the worst case (for example, $p=3$, $\alpha$ and $\beta = 0.6$) in solution time. However, the solvability of this approach is expected to be limited considerably as the problem size increases. For 20-node problems, the solution time is considerably increased, and the optimal solution could not be verified due to memory limit termination for some cases (for example, $\alpha$ and $\beta = 0.6$ with $p = 5$). This is mainly because the number of constraints and variables associated with survivable routes is dramatically increased so that it impacts the capability of the tested machine to handle the data as a result. In contrast, the hybrid type-search, TABUEX performs well by solving all problems sets to optimality within a minute. In the additional tests using hypothetical data, which is more than 25-node problems (not shown in the Table 5.4), the heuristic TABUEX produced a good solution within a few minutes. When it compared to the TABUSA, which has the similar structure of tabu search, the solution time slightly increase because the procedure of complete enumerations of survivable routes for each $ij$ pair is involved, but the experiments also prove that an exhaustive search, combined with other searching methods cannot be a computational burden in the TABUEX.
<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha$, $\beta$ level</th>
<th>Approach</th>
<th>Objective</th>
<th>Solution Time(sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.6</td>
<td>MIP</td>
<td>1,863,720</td>
<td>158.1</td>
<td>ATL SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>1,863,720</td>
<td>0.6</td>
<td>ATL SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>934,019</td>
<td>19.6</td>
<td>ATL BOS DAL</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>934,019</td>
<td>0.8</td>
<td>ATL BOS DAL</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
<td>MIP</td>
<td>1,851,289</td>
<td>4.4</td>
<td>ATL BOS SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>1,851,289</td>
<td>1.5</td>
<td>ATL BOS SLC TMP</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>898,144</td>
<td>2.4</td>
<td>ATL BOS DAL SLC</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>898,144</td>
<td>2.0</td>
<td>ATL BOS DAL SLC</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**15-node problems**

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha$, $\beta$ level</th>
<th>Approach</th>
<th>Objective</th>
<th>Solution Time(sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.6</td>
<td>MIP</td>
<td>2,786,144</td>
<td>3145.1</td>
<td>ATL KAS MIN</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>2,786,144</td>
<td>2.1</td>
<td>ATL KAS MIN</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>1,451,679</td>
<td>456.2</td>
<td>KAS SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>1,451,679</td>
<td>2.9</td>
<td>KAS SLC WAS</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
<td>MIP</td>
<td>2,648,667</td>
<td>575.4</td>
<td>ATL KAS MIN SLC</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>2,648,667</td>
<td>4.7</td>
<td>ATL KAS MIN SLC</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>1,342,728</td>
<td>257.4</td>
<td>ATL BOS KAS SLC</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>1,342,728</td>
<td>5.1</td>
<td>ATL BOS KAS SLC</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**20-node problems**

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\alpha$, $\beta$ level</th>
<th>Approach</th>
<th>Objective</th>
<th>Solution Time(sec)</th>
<th>Optimal hubs</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.6</td>
<td>MIP</td>
<td>4,232,370</td>
<td>5351.3</td>
<td>ATL KAS OMH SLC</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>1,865,816</td>
<td>40.1</td>
<td>ATL BUR MIN NW SUV</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>867,261</td>
<td>3990.0</td>
<td>ATL BUR NW OMH SLC</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>571,758</td>
<td>36.3</td>
<td>BUR KAS NW SLC</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.6</td>
<td>MIP</td>
<td>2,365,892</td>
<td>5217.1</td>
<td>BUR MIN NW OMH SLC TMP</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>1,896,170</td>
<td>53.1</td>
<td>ATL MUR MIN NW SLC</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>MIP</td>
<td>883,419</td>
<td>3760.2</td>
<td>BUR KAS MIN NW SLC TMP</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tabu</td>
<td>618,794</td>
<td>53.8</td>
<td>BUR KAS NW SLC</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Table 5.4.** Computational results with the 10, 15, and 20 node problems for PROPS.

**Note:**
1) Optimality is not proven due to memory limit termination (the gap was measured 60.6, 21.4, 19.4, 18.4 %, respectively).
2) NW (Newark), SUV (Sunnyvale), OMH (Omaha).
5.7. Summary

In this chapter, two heuristics are proposed, both of which are ‘reliability-based’ searching approaches. Although it would be possible to obtain exact solutions for various reliable $p$-hub location problems using the MIP as shown in previous chapter 3 and 4, it is obvious that heuristic approaches are required to handle a larger problem whose size is more than 20-node problems for the PROPS, and 25-node for the other models based on the computational experiments.

Even though exhaustive search can be applicable for small size problems, it is prohibitive in many instances. The results from both the MRSA and PROPS clearly demonstrate that both heuristics are very efficient in solving the problems in a time-efficient manner, as well as producing exact solutions for the cases where the MIP approach can solve to optimality. Although a generalized structure is provided from both heuristics, it is also needed to develop heuristic for the extended models, which may embed different conditions such as constraints or objective function in the model.
6.1. Conclusions

Hub location problem have comprised an important area of study in geography and location science because location of hubs and their connection to the other demand nodes are important to analyze for quality of network reliability and survivability in current telecommunications and transportation infrastructure (Lewis 2006). In this dissertation, I address a series of new hub location problems and network designs models, which are termed the reliable p-hub location problems (RPHLP), and the p-hub protection problems (PHPRO).

The essential motivation to propose these models arises from the increasing vulnerability in current telecommunication and transportation network structure where the networks have converged into more hub-and-spoke type systems.
As a major contribution, the models are developed by incorporating three different dimensions into hub location network design decisions; facility location (i.e., dispersion and closeness), reliability theory, and survivable strategies for network design. The major contributions of this dissertation are as follows.

In chapter 2, I reviewed current network design issues related to network reliability and survivability. Although many reliable network design models addressed construction of a better system in terms of traffic delivery, network topology, and routing schemes, the models are often created under simple assumptions that ignore the substantial aspects of current network systems where locations demanding service, their interactions, the role of hubs and inter-hub links are important components to be reflected in network design. Regarding hub location models, many hub location models rarely focused on the design aspect of reliability or survivability which becomes critical network infrastructure design. Based on this research background, I highlight the needs of new hub location models in approaching practical issues.

The contributions from chapter 3 are highlighted in three ways. First, the reliable $p$-hub location problems are new hub network problems in that the models reflect the requirements of current delay-sensitive and network performance-driven network systems. Second, the models demonstrated how optimal hub locations are determined under various reliability or capability conditions of network facilities by exploring the relationship between network performance and hub locations.

In term of hub locations, it is observed that hubs are dispersed in order to maximize the network reliability if more reliable conditions of the hubs and inter-hub links are assumed. In contrast, for the less-reliable hub facilities, the model locates the hubs closer
to the each other to enhance network performance. This is because the geographical
proximity among hubs increases efficiency of massive traffic exchanges. Third, and most
important, the potential advantage of mandatory dispersion of hub facilities is stressed in
hub network design to avoid the excessive concentration of interacting flows from
particular hub facilities. Ultimately, this prescription could be a significant design aspect
to reduce the potential damage when any malfunctions or attacks are targeted in highly
active hubs or inter-hub links.

In chapter 4, the $p$-hub protection problems are addressed as a survivable network
design. In contrast to the RPHLP, the models are focused on designing the network which
minimizes potential loss of flows among nodes. As the first step, I developed three
different types of protection models (PROSA, PROMA, and PROPS) based on routing
schemes. As an analytical framework, the hub activity envelope is used to identify the
impact of disruptions scenarios. The relationship between the capability of flow-
protection on hubs and resiliency to the inter-hub links disruptions among three standard
models are examined. Second, as extensions, capacitated (CAP), minimum threshold on
inter-hub links (MT), and minimum threshold on back-up links (MTB) models are
proposed. One major finding is that the resiliency and capability of flow-protection in a
hub-and-spoke type network can be differentiated from the imposed capacitated or
minimum threshold levels. In terms of network resiliency, the PROPS which employs a
back-up disjoint routing scheme as a defense strategy, demonstrated its ability to
maintain its operations in the face of possible disruption scenarios, but it entails the
increased network cost. Both models are expected to be utilized as decision tools to
prescribe the optimal level of dispersion or protection for better reliable or survivable network design.

In chapter 5, this research also proposed efficient solution approaches. The reduction techniques to reduce the problems size are employed. Reducing the number of constraints or variables without influencing the problem structure is important to increase solvability for a LP relaxation approach. The reduction strategy involves removing impractical routes from all possible routing variables. In particular, the solvability of the PROPS, which might be prohibitive for a moderate size problem due to the dramatic increase of possible back-up routes variables, increases by removing all impractical routes variables. Due to this technique, all of the problems tested in chapters 3 and 4 were solved optimally within a reasonable solution time.

As a solution approach in order to solve large size problems effectively, two heuristics are developed. A tabu-based search heuristic is developed to solve the problem with single assignment (MRSA) and a hybrid type search which combines an exhaustive and tabu-based search algorithm is also developed to tackle the single and back-up routes scheme (PROPS). Of course, these search algorithms are tested for the 25-nodes set, thus it is still needed to test to a larger problems (more than 25 nodes). However, using the characteristics that the reliability of routes which are jointly related to physical distance and the amount of interactions among nodes, the ‘reliability-based’ searching approach performed well, offering the viability to apply a large size problems which a commercial LP solver is not likely to solve exactly.
6.2. Future Research

Although this dissertation research provides a broad background for a new class of location problems and network design, many directions of future research are suggested. The future research topics are summarized in three ways as follows.

6.2.1. Model extensions

Given that the chapters focused on models that provide accurate behavior using a relatively small data set, further research needs to expand upon these models. The models addressed in the dissertation are standard models. There are many possibilities of extension for practical applications. First of all, the $p$-reliable hub location models can be approached differently from a ‘$p$-defense’ or ‘$p$-center’ hub problem (Kuby 1987). In other words, the class of dispersion of hub facilities can be reclassified, and thus different model responses according to the different requirements could be an important research agenda. More importantly, the principle of facility dispersion can be combined into $p$-hub protection models. Exploring various model behaviors to the different model requirements could provide good insight in developing other network design models. In terms of survivable network design, a facility cost-type model can be addressed. In this dissertation, the relationship between network resiliency and network performance is explored in terms of capacity and network linkage length which can be used as a proxy for network cost. However, considering more specific network cost structure such as
flow-dependent linkage cost and facility installation cost, the models can provide a more explicit relationship between them.

6.2.2. Extending model applications

Many recent papers stress the extension of the hub network models to the more complex dimensions in real-world networks. In telecommunication networks, the cost structure of the physical facilities such as hubs, backbones, routers, backbones, and local linkages hubs are the most important components for network design. As indicated in chapters 3 and 4, it is obvious that there is a trade-off between network cost and network performance (reliability or survivability). Applying a more detailed cost structure, the models can be more sophisticated and as a result the models are used as more realistic infrastructure design tools. In applications for transportation network design, the models consider different performance-measures into the model formulations because the concept of reliability in transportation is defined differently, by focusing more on trip-time and congestion levels on the road network.

6.2.3. Developing solution approaches

Considering that very few literatures on real-world applications have been found in the hub location models, and the lack of solvability of hub network design in obtaining exact solutions using commercial solvers, it is essential to develop better algorithmic approaches to tackle real-world cases. While the TABUSA and TABUEX algorithms
demonstrated their capability in applying a more complex problem, the algorithms can be refined for the other classes of hub location problems. For example, the heuristics in this research are only focused on *uncapacitated* $p$-hub location problems where the number of $p$-hubs is specified. However, if the problem requires determination of the number of $p$ as an internal optimal process, then it is more difficult to solve, thus developing a better algorithm is necessary. Combinations of other different methods (for example, a hybrid method incorporating genetic algorithm with tabu search) would provide more effective solution methods for different classes of hub location problems.

Another possible avenue to improve solvability of the problems is to develop reduction techniques, especially for the LP approach. In this research, two reduction techniques are employed which are exploring types of routes and removing impractical routes. As a further reduction technique, removing unrealistic routes, which are determined based on the length of a route, can be studied. In particular, if the network cost to establish a physical linkage is the primary concern in the network design, allocating a non-hub node to the hub which is far away with a certain level (i.e. threshold) of distance is not likely to happen in practice. In this context, GIS can be used to capture the practical routes especially for large complex problems.


