A Lagrangean heuristic and GRASP for the hub-and-spoke network design with economies of scale and congestion

By

Faisal Alkaabneh

A Thesis Presented to the

Masdar Institute of Science and Technology

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

In

Engineering Systems and Management

© 2014 Masdar Institute of Science and Technology

All rights reserved

A Lagrangean heuristic and GRASP for the hub-and-spoke network system with economies of scale and congestion

By Faisal Alkaabneh

A Thesis Presented to the Masdar Institute of Science and Technology in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering Systems and Management

July 2014

© 2014 Masdar Institute of Science and Technology

All rights reserved

AUTHOR'S DECLARATION

I understand that copyright in my thesis is transferred to Masdar Institute of Science and Technology.

Author ______Faisal Alkaabneh__________________

RESEARCH SUPERVISORY COMMITTEE MEMBERS

Dr. Ali Diabat, Chair, ______________________________

Masdar Institute of Science and Technology

Prof. Samir Elhedhli, ______________________________

University of Waterloo

Dr. I-Tsung Tsai, ______________________________

Masdar Institute of Science and Technology

Abstract

In the current fiercely competition environment, cost and time efficient distribution strategies provide companies with a competitive advantage. In a distribution network, direct shipments are neither cheap nor practical between each origin and destination pair. The cost efficiency of these networks can be improved by inserting hub points (switching/sorting centers) onto the link connecting origins and destinations and reducing direct links concentrates overall flow between fully interconnected hub points, thus creating economies of scale.

The hub-and-spoke network design problem is a strategic logistics planning problem with applications for airlines, telecommunication companies, computer networks, postal services, and trucking companies, for example. Basically, the problem in all these applications is that for a given set of nodes (airports, computers, depots, etc.), goods must be transported between possibly each pair of nodes. Direct connections between every origin-destination of nodes would result in $n*(n-1)$ connections which is impractically high and economically non-profitable. Consider, for instance, an airline company that serves several airports worldwide. Offering nonstop flights between every pair of airports would require a huge amount of planes and crews and would result in empty seats on board for many flights.

This thesis offers deeper investigation into the hub-and-spoke design network for the airline industry; and it considers a hub-and-spoke network design problem with a concave cost function representing the economies of scale on the interhub links and a convex function representing congestion at hubs in the objective function of the model. The problem is modeled as a nonlinear mixed integer program that is difficult to solve directly. A Lagrangean approach is proposed to obtain tight upper and lower bounds. Computational experiments are reported for the Civil Aeronautics Board(CAB) dataset with a number of nodes –up to 25– that are solved efficiently by the Lagrangean heuristic. The solution methodology provides a high quality solution for

the reported instances with reasonable time for all the tested instances. Moreover, a Greedy Randomized Adaptive Search Procedure(GRASP) is developed to solve large instances of the developed mathematical model which proved to provide high quality solutions when compared to the Lagrangean heuristic.

This research was supported by the Government of Abu Dhabi to help fulfill the vision of the late President Sheikh Zayed Bin Sultan Al Nahyan for sustainable development and empowerment of the UAE and humankind.

Acknowledgments

All praise and thanks is to Allah who gave me the ability to finish this thesis. Secondly, I would like to thank all the people who supported me in this work. I have been lucky to work under the supervision of Dr. Ali Diabat who provided me with continuous support, help and guidance. Also, I grateful to Prof. Samir Elhedhli who shared with me his experience in this field of knowledge and and provided me with valuable insights.

Furthermore I would like to thank Masdar Institute for Science and Technology for providing me with full scholarship to study in this wonderful research institute. Also, I am very thankful to Masdar faculty members and employees for their constant presence and the constant support they provided in an atmosphere of academic accomplishment. Special thanks for Mrs. Glenda El Gamal for checking the English writing of the thesis.

I would like to express my ultimate gratitude to my mother, father and my brothers for their constant support and unconditional love for me. Special thanks for my mother and father for their prayers and encouragement, which I believe is the main reason for my success. Thanks for my brothers Hamzah, Hashem and Zaid for taking care of our parents and their understanding while I was working on this thesis.

Faisal Alkaabneh,

Masdar City, May 19, 2014.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Hub location problem (HLP) is an extension of classical facility location problems. Hubs are facilities that work as consolidation, connecting, and switching points for flows between stipulated origins and destinations.

In order to satisfy a demand, HLP involves the movement of people, commodities, or information between required origin-destination pairs. Hubs are applied to decrease the number of transportation links between origin and destination nodes. For example, a fully connected network with k nodes and with no hub node has $k(k-1)$ origin destination links. However, if a hub node is selected to connect all other nodes (nonhub nodes, also known as spokes) with each other, there will be only $2(k-1)$ connections to serve all origin-destination pairs. This idea can be extended to a network with more than a hub node, called a multiple-hub network. Thus, by using fewer resources, demand pairs can be served more efficiently with a hub network than with a fully connected structure [50].

Though- the telecommunication industry is originally one of the oldest users of the hub network concept [50], hub location also has been an active research area for interdisciplinary researchers from operations research, transportation sciences, geography, network design, telecommunications, regional science, economics, etc. In logistical systems, the airline industry and postal companies are one of the main users of this concept [20]. Today, there are many other areas that can take advantage of the hub concept such as the maritime industry, freight transportation companies, public transit, and message delivery networks. Applications of HLP included the areas of airlines and airports as found in Aykin [7], post delivery services and rapid delivery packing systems [40], supply chain and logistics [21], telecommunication services and message delivery networks [25], transportation systems for tracking companies as in [32].

1.1 Structure of the hub-and-spoke network

What can be classified as a hub varies among industries. ”For example, as observed by Wu [104] ”in urban traffic networks, a hub represents a terminal or a transit stop for a number of routes; in trucking systems, a hub denotes a warehouse or a center facility; in air transportation systems, a hub is not only defined as a transit point, but also as a geographical area whose volume of air passengers exceeds a certain level; in express delivery service networks, a hub denotes a center for switching and sorting operations; and in telecommunication systems, a hub refers to a server for receiving, processing, and sending information”. Basically, the problem in all these applications is that for a given set $V = \{1,...,n\}$ of nodes (airports, computers, post offices, depots,...) goods must be transported between possibly each pair of nodes. Direct connections between each pair of nodes would result in $n(n-1)$ linkages which is impractically high and economically non-profitable. Consider, for instance, an airline that serves several airports worldwide. Offering nonstop flights between each pair of airports would require a huge amount of planes and crews and many empty seats on board would result for many connections”.

Studies on the hub location problem often assume three things: that the hub network is complete with a link between every hub pair; that there are economies of scale incorporated by a discount factor (α) for using the inter-hub connections; and that no direct service (between two non-hub nodes) is allowed. However, these assumptions are relaxed in some studies. Hub location problems involve the determination of hubs facilities to be located through which flows (e.g., passengers) are to be routed from origins to destinations (e.g., airports). Origin to destination (OD) connections from node i to node j are associated with OD pairs, which may be connected with direct trips (not via hubs) or via paths that visit hubs Figure 1.1. Figure 1.1 illustrates three possible paths from origin node i to destination node j: in the first path (1) a direct connection (the direct point-to-point link from i to j); in the second path (2) there is a one stop trip at hub h; and the third path (3) there is a two stop trip at hubs $h1$ and $h2$. Note that the flow (e.g., passengers) transiting through h, or through $h1$ and $h2$, has switching options at the hubs where it can be joined by flow (e.g., other passengers) arriving from other origins (the in-arrows) and can depart to other destinations (the out-arrows) [20]. Note that the third path is the path allowing the use of economies of scale due to higher flows carried on that link; by consolidating and disseminating demand and flows, hub inter-linkages are able to achieve a much lower unit transportation cost than other linkages, which is often embodied by a discount rate α. Thus, economies of scale are exploited on hub inter-linkages and total operational costs are reduced.

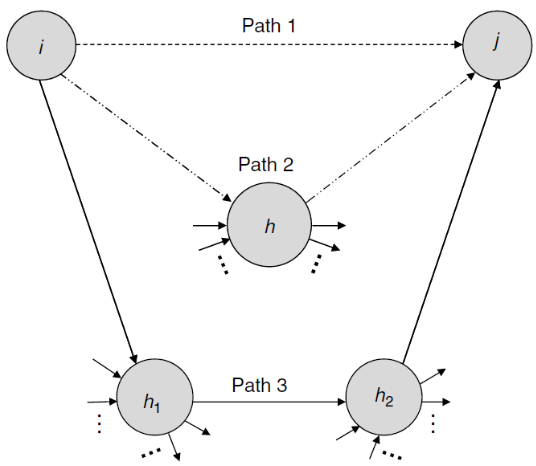

Figure 1.1: Alternative Paths Between Origin i and Destination j.

OKelly (1986) [80] was the first to consider the formulation of the hub location problem. By studying the passenger network in the airline industry he presented the first mathematical formulation for a hub location problem. His formulation is referred to as the single allocation p-hub median problem. Given (n) demand nodes, flow between origin-destination pairs and the number of hubs (p) to be located, the objective is to minimize the total transportation cost (time, distance, etc.) to serve the given set of flows. However, assigning the nearest location for each demand city to a hub might not give the optimal solution for the hub problem; thus Aykin [6] defined a procedure to optimally allocate a set of nodes to a set of hubs based on the difference, for example of assigning node i to hub k instead of hub l.

The data sets used by researchers in the HLP field are mainly two datasets, the first one is the dataset proposed by O'Kelly [81] based on the airline passenger flow traveling between 25 US major cities in 1970 evaluated by the Civil Aeronautics Board (CAB), and this dataset in particular has been used by almost all of the hub location researchers and will be referred to as the CAB data set [3]. The second is the Australia Post (AP) data set (introduced by Ernst and Krishnamoorthy [45]). AP data set is based on a postal delivery in Sydney, Australia and consists of 200 nodes representing postal districts. The AP data set differs from the CAB data set in that the flow matrix of the AP data set is not symmetrical and they-also they differ in the number of nodes.

1.2 Advancements in the hub location studies

Great advancements have been achieved to extend the basic model of HLP, and these advancements contributed to the basic model of HLP by the addition of new mathematical formulations to capture the real life aspect which was not introduced in the basic model of these advancements are the hub capacity (capacitated/uncapacitated), allocation of nonhub nodes to hub nodes (multiple and single assignment), alternative topologies,integrating cost and service, dynamic hub location, stochastic elements, congestion consideration, nonlinear cost calculation, reliability considerations, solution domain (network, discrete, and continuous), criterion (Mini-Max (the maximum transportation cost from origin nodes to destination nodes is minimized) and Mini-Sum (the total cost incurred by locating hub nodes and allocation of non-hub nodes to hub nodes is minimized), source determining the number of hubs to locate (endogenous and exogenous), and cost of installing a hub (no cost, fixed cost, and variable cost)), and the ability to solve large scale problems by developing heuristics.

1.2.1 Economies Of Scale (EOS) in the hub location studies

As illustrated above, the economies of scale concept is one of the main motivations for installing hub-and-spoke systems; however, the way in which costs of economies of scale are modeled does not reflect the real-life aspects of the economies of scale. Some authors have noted that applying a discount factor α to the costs on arcs between hubs while disregarding the flow on these arcs contradicts the motivation of this discount factor. Consequently, Podnar et al. [90] introduced flow thresholds that must be reached in order to gain the discount α. Campbell et al. [22] state that ”...the basic assumption in hub median models that flow costs are discounted on hub arcs to reflect high volumes leads to a possible mismatch between the abstracted model and the underlying motivations of the model”. So, they consider models where so-called hub arcs, i.e. arcs that link two hubs and on which costs being discounted by a factor α, are to be explicitly selected which means that the cost for a flow between two hubs may or may not be discounted by a factor a depending on the selection. The total number of such hub arcs to be chosen is pre-specified in their models, because otherwise every arc that links two hubs would be selected as a hub arc.

O’Kelly and Bryan [83] are the first to note that hub location models that utilize a constant discount factor (α) represent something of an oversimplification. In some cases, traffic on a particular interhub link in the optimal solution is rather small. Therefore, if the model were used to study realistic communications or transportation networks, the model would apply a discount that would not be warranted. Thus, they proposed a FLOWLOC model where the constant discount factor on the inter-hub links is replaced by a piecewise-linear concave function. Thus, the amount of the discount will depend on the actual flow of traffic on the link. Bryan [12] presented four extensions for the FLOWLOC model namely,(1) a capacitated network model; (2) a minimum threshold model; (3) a model that endogenously determines the number of open hubs for the network; and (4) a model that incorporates a flow-dependent cost function for the spokes as well as the interhub links.

1.2.2 Congestion in the hub location studies

Economies of scale is the major advantage of hub-and-spoke networks. However, this network structure has side effects. Congestion is a major one, Mayer and Sinai [77] examined two factors that might explain the extent of air traffic delays in the United States: network benefits due to hubbing and congestion externalities. In their study they concluded that: the benefits due to hubbing is actually the dominant factor leading to congestion and the consolidation and dissemination of demand flows imply a higher tendency of congestion at hub airports than non-hub ones. They also found that uncertainty of demand flows is another potential trigger for congestion and when the demand flow climbs unexpectedly up to a peak volume within a short time, it results in congestion.

The hub network congestion problem was first considered by Grove and O Kelly [57] in 1986. Given hub locations, they did a simulation for the daily operation of a single assignment hub-and-spoke air network. They concluded that the extent of schedule delays depends largely on the size of hub flows. Later, schedule delays at hub airports were well studied by Kara and Tansel [65], As post-design analysis, both papers contributed to identify the critical factors affecting hub congestion behavior [103]. Aykin [6] introduced the capacitated hub-and-spoke network design problem in which hubs have limited capacity for channelling flows between the nodes served by the system.

Congestion weakens the performance of hub-and-spoke networks. Hence, strategies to prevent congestion effects become important to improve the performance of hub-and-spoke systems. Enlarging hub capacity is an intuitive solution. However, a larger hub capacity implies higher fixed costs. Furthermore, with the same demand flow, a larger hub capacity implies a lower average utilization ratio, a negative indicator of airport performance. Therefore, simply enlarging hub capacity is not a genuine solution to the congestion problem. One way to mitigate the effect of congestion is the capacity consideration on the incoming flow for the hub, this solution already used in a number of previous studies, where the constraint is formulated as that the hub flow is less or equal to the hub capacity, for example Aykin [6] introduced the capacitated hub-and-spoke network design problem in which hubs have limited capacity for channelling flows between the nodes served by the system. However, this does keep the amount of hub flow below the given hub capacity but is not regarded as a thorough solution to the congestion problem. The reason is that congestion is often proportional to the relative difference between the hub flow and the hub capacity. The smaller the difference, the larger is the congestion. This relationship can not be accurately reflected by a simple capacity constraint [43]. Elhedhli and Hu [42] are the first to include congestion costs in the objective function, along with transportation and fixed costs, and developed a heuristic Lagrangean solution algorithm. Via comparison with the non-congestion problem on the CAB data set, the authors stated that the congestion model results in a more balanced distribution of flows through hubs.

1.3 Motivation

In a recent review conducted by Campbell and O'Kelly [20] they especially encouraged models that incorporate both more realistic transportation costs and service measures along with other relevant aspects. One of the realistic aspects in the HLP is to model the economies of scale as a nonlinear function depending on the amount of flow rather than the a constant discount factor, and the consideration of congestion to account for incoming flow directed to major hubs in the objective function of the HLP is another real-life aspect. However, till the present time, there is no work in the literature, to the best of our knowledge, to tackle the economies of scale for the single allocation case nor the integration of congestion and economies of scale effects on the design of hub-and-spoke networks. Previous work considered the congestion effect or the economies of scale separately. This is because of the great complication resulting from the incorporation of those two effects in one single model, moreover for the single allocation case the allocation decision variables are binary taking either 0 or 1, unlike the multiple allocation version of HLP where the allocation decision variables are linear taking values from 0 to 1, which presents a more challenging task to be solved.

In addition to this complicated model that considers the effect of nonlinear representation of the Economies Of Scale and congestion together, the previous work done by OKelly and Bryan [83] and Klincewicz [70] demonstrated that interhub traffic flow tended to be concentrated on a few interhub arcs. That is, some interhub arcs had carry relatively large traffic flows that were highly discounted, while others had very small amounts of traffic. As mentioned by Bryan [12] analysts should be aware of this potential imbalance in the interhub network (at least as witnessed in these CAB problems) when applying the model to real world networks. One way to mitigate the imbalance effect is to consider the congestion effect to count for the incoming flow on the hubs, which is the solution proposed by Elhedhli and Hu [42]. In our work we integrated the congestion and economies of scale in the same model, therefore, we provided more insights and justifications for the use of integrated model in terms of impact of the structure of the network as well as the routing of flows in the network.

1.4 Contribution

The two main contributions of this work can be summarized in terms of the model formulation and the solution methodologies developed. Regarding the model formulation, the integration of the congestion effect and the nonlinear economies of scale in the design of the hub-and-spoke network which have mainly been dealt with independently in the literature and in terms of solution methodologies, two solution methodologies are developed rather than one to tackle the difficulty of the model.

An important aspect, besides the new formulation of the integrated model, is the solution methodology. The contribution of this work lies in the implementation and comparison of two different algorithms. Lagrangean relaxation is an exact technique that can yield an optimal solution, or at least very useful information about the bounds of the optimal solution when implemented as part of a heuristic algorithm. On the other hand, the GRASP is a robust heuristic approach with the main advantage of low computational time. The efficient design of GRASP heuristic implemented in this study was shown by the comparison with the solution obtained by the Lagrangean approach, the GRASP heuristic proved its ability to obtain high quality solutions when compared with Lagrangean relaxation.

The use of the proposed two solution methodologies will be justified in Chapter 6 based on the solution quality and computational time required to find the solution for the proposed integrated model. Specifically, the leading edge best commercial solver GAMS required several days to solve instances that took less than 1 hour to be solved using the Lagrangean relaxation heuristic and less than 1 second for the GRASP, also note that this is for small size instances, for larger instances it was prohibited for GAMS to find a solution even given more that 15 days of running time, while the developed solution methodologies were able to find near-optimal solutions effectively in less than one hour.

Another contribution in this study is the insights and the detailed analysis provided by analyzing the integrated model when compared to models developed previously. As will be shown in Chapter 6, the integrated model was able to achieve a balance in the network design in terms of flow for the hub nodes as well as a more realistic presentation for the economies of scale that

is the main motivation for installing hub-and-spoke networks.
In the computational experiments conducted by OKelly and Bryan [83] and Klincewicz [70] they found that interhub traffic flow tended to be concentrated on a few interhub arcs. That is, some interhub arcs had to carry relatively large traffic flows that were highly discounted, while others had very small amounts of traffic. As mentioned by Bryan [12] analysts should be aware of this potential imbalance in the interhub network (at least as witnessed in these CAB problems) when applying the model to real world networks. One way to mitigate the imbalance effect is to consider the congestion effect to count for the incoming flow on the hubs, which is the solution proposed by Elhedhli and Hu [42]. In our work we integrated the congestion and economies of scale in the same model, therefore, we provided more insights about this issue. As shown in the numerical results the imbalance effect is clear as the nonlinearity factor in the EOS function decreases (i.e. more discount is provided) which is the imbalance effect mentioned by Bryan [12] in his study; however, for the same discount rate considering the congestion effect eliminates the imbalance effect found in the network. Thus, the use of the integrated model is justified when it comes to considering real-life aspects of the design of hub-and-spoke network.

In this work we presented a Lagrangean relaxation heuristic to tackle the model of Uncapacitited Single Allocation p Median Hub Location Problem (USA-p-MHLP) incorporating the congestion effect and the economies of scale together. The use of Lagrangean relaxation enabled us to decompose the model into two sub-problems that are solved to optimality at each iteration. Moreover GRASP algorithm is developed to solve larger instances.

1.5 Relevance to the Abu Dhabi 2030 Vision

The Government of Abu Dhabi has set guidelines and priorities for the Emirate's socio-economic progress in its Policy Agenda. The Abu Dhabi Economic Vision 2030 has been developed by the Government, in consultation with the private sector, as a 22-year strategy to achieve these aims, and to ensure that all stakeholders in the economy are moving in concert, with a clear view of the long-term goals. The mission of the Masdar Institute is in line with the development goals of the Abu Dhabi Economic Vision 2030 and its long-term strategies.

Two of the main pillars in Abu Dhabi Economic Vision 2030, are developments in the Aviation, Aerospace, & Defense and Transportation, Trade, & Logistics sectors. These sectors are a major contributor to this vision of diversification of Abu Dhabis economy.

The UAE is at the forefront of global aviation growth, with UAE airlines and airports setting the benchmark for global excellence. This growth has also created a necessity to design efficient hub and spoke networks for the airports in Abu Dhabi and UAE in total. The decisions made today regarding the design of the airports network will have a long-term impact on the aviation industry in Abu Dhabi and UAE. In Abu Dhabi International Airport, for example, development work has started on a new passenger terminal, the main building and centerpiece of the new airport, to be between the two runways and known as the Midfield Terminal. Upon completion in 2017, the Midfield Terminal will increase the airports passenger capacity to more than 20 million per year, with options for this to double in capacity to 40 million. The decisions related to expansion at airports to avoid congestion and to utilize the economies of scale can benefit from the study we have conducted, as we provided a solid foundation to consider an integrated model with congestion effect and economies of scale.

Abu Dhabi has a long and proud tradition as a trading economy. Given its primary strategic location, this is a tradition that Abu Dhabi will expand, becoming a logistics hub for businesses and industries in the region. Over the past 40 years, Abu Dhabi has developed its port infrastructure, its road network, and more recently its airports to ensure it is well-connected to trading partners in the region and beyond. Logistics as a profitable industry will make an important contribution to the Emirate's economic diversification, in the direction of becoming entirely sustainable and not dependent on any finite resource. Maintaining an efficient transport system and connections with global markets is also an important component in the establishment of other sectors in the economy.

1.6 Thesis Organization

This thesis is organized as follows: Chapter 2 reviews the literature of the hub location problem, Chapter 3 details the mathematical model, Chapters 4 and 5 describe the solution methodologies, Lagrangean heuristic and GRASP, respectively, Chapter 6 gives the implementation and testing results and finally Chapter 7 presents final conclusions and discusses future work.

CHAPTER 2

Literature Review

Studies in the optimization of Hub Location Problems (HLP) have received considerable attention due to the nature of the HLP and its wide applications. And in order to maximize the efficiency of the systems being served by the hub facilities. As these studies progress, new mathematical models for HLP have been introduced to serve a specific purpose or application. Therefore, a large number of variations from the original HLP exist.

This chapter aims to describe the different mathematical formulations and variants of the HLP. The following sections discuss the background material, which consists of the basic description of HLP and its origins. After introducing the basic definition of the HLP and its motivation in Section 2.1, Section 2.2 will provide the various formulations for the HLP in each section the mathematical model corresponding to this variant. Solution methodologies found in the literature to tackle each formulation is presented. Section 2.3 then provides successful real-life applications of HLP research and Section 2.4 provides some of the challenges in the HLP model. Sections 2.5 and 2.6 provides more insights into the Economies Of Scale (EOS) and congestion aspects in the HLP, and finally Section 2.7 presents our work and compares it to these previous work in the literature in order to highlight my contribution to the field.

2.1 Origins of HLP and motivations

This section begins with a broad definition of hub location problems and discusses the relationship between hub location problems and network design problems. We will describe the components of hub networks based on the description provided by [20] and define fundamental hub location problems that have been the focus of much research for the past quarter-century.

2.1.1 Origins and motivations of hub location problem as a research area

Several factors played an important role for the emergence of research on hub location and of these factors include the development in the 1970s and 1980s regarding deregulation of transportation, the rise of express delivery services, the development of large telecommunications networks, and advances in modeling [20].

Campbell and OKelly [20] observed ”that in the start of the late 1970s, the swift spread of deregulation of transportation in the United States across all modes, including air cargo in 1977 (the Air Cargo Deregulation Act), passenger airlines in 1978 (the Airline Deregulation Act), and trucking companies and railroads in 1980 (the Motor Carrier Act of 1980 and the Staggers Rail Act, respectively) affected the emergence of research on hub location. In the addition of relaxing price controls and deregulation reduced barriers to entry, expansion, mergers, and acquisitions. This led to considerable focus by transportation firms (especially airlines and trucking companies) to design more efficient and larger-scale transportation service networks, including optimizing the terminal (or hub) locations because carriers could now use their networks and services to achieve competitive advantages. Hub transportation networks were deployed in short order throughout much of the passenger and cargo airline and Less Than Truck Load trucking industries”, Campbel and O’Kelly [20].

Another important practical motivation for hub location research was the advent of express delivery firms, most notably Federal Express (FedEx). Air express has existed in some form since the beginning of air transport, and several national postal networks had hub-like structures.

The hub-and-spoke problem was first studied by OKelly ([80], [81]) who developed a quadratic programming model and proposed two enumeration heuristics. In this work, the network is modeled as a quadratic integer program, whose objective function contains a product term of two decision variables. Each non-hub node is assigned to a single hub. A set of two-dimensional binary variables is defined for hub assignments with each variable indicating whether a hub node is assigned to a non-hub node or not. Demand flows between origin-destination pairs are routed through paths formed as "origin $\rightarrow$ origin hub $\rightarrow$ destination hub $\rightarrow$ destination. The "origin hub" and "destination hub" are hubs assigned respectively to the origin and destination nodes. A route is denoted by a quadratic term the product of two hub-assignment variables. Accordingly, the unit travel cost is composed of three parts: the travel cost from the origin to the origin hub, from the origin hub to the destination hub, and from the destination hub to the destination. A discount factor determined exogenously is applied to capture the economies of scale on hub interlinkages. This discount factor is assumed uniform on all inter linkages, independent on the amount of demand flows routed through them. The fixed number of hubs p is given as a parameter. The objective function minimizes the total transportation costs. This model is referred to as the "p-hub location model".

Later, Campbell [17] proposed multiple mathematical formulations for the HLP to consider similar objective functions as several classical facility location problems. He presented integer programming formulations for four types of discrete hub location problems: the p-hub median problem, the uncapacitated hub location problem, p-hub center problems and hub covering problems. Since then there has been a great advancement in the Hub Location Problem (HLP). Those advancements have contributed to the basic model of HLP by the addition of new mathematical formulations variants to capture real life aspects which were not introduced in the basic model; specifically the consideration of hub capacity (capacitated/incapacitated), allocation of nonhub nodes to hub nodes (multiple and single assignment), alternative topologies, integrating cost and service, dynamic hub location, stochastic elements, congestion consideration, nonlinear cost calculation, reliability considerations, and etc. In terms of solution methodologies, different studies were conducted to solve the different mathematical models using exact methods, meta-heuristics, and local searches.

2.1.2 Distinguishing Features for Hub Location Problems

The key distinguishing features for hub location problem are as follows:

1. "Demand is associated with flows between Origin Destination (OD) pairs (not with individual points)" Campbell and O'Kelly [20].

2. Flows are allowed to go through hub facilities.

3. Hubs are facilities to be located.

4. There is benefit in routing flows via hubs (or a requirement to route flows via hubs). The bundle of flows at the hubs increases the traffic on the hub inter-connection, enabling the use of more efficient and higher volume carriers, which then result in lower per unit transportation costs.

5. There is an objective that depends on the locations of hub facilities and the routing of the flows.

It is assumed that a directed complete graph $G = (V, A)$ is given, with $|V| = N$, where nodes correspond to origins, destinations, and potential hub locations. Associated with its arcs $(i, j), \quad i \neq j$, are lengths $d_{ij} > 0$.For convenience, d_{ii} is defined as 0. It is assumed that these lengths satisfy the triangle inequality. Also given are $W_{ij} \geq 0$, representing the amount of flow of a certain product to be sent from node i (origin) to node j (destination) for all i and j in V. In this general problem statement, flows are allowed, though not required, to go through hubs, so the incentive to use the hubs is provided by item 4. The constraints of this problem are that all flows must be satisfied by visiting one or at maximum two hubs with the objective function of minimizing the transportation cost, and that maximum profit models might yield unsatisfied demand nodes (not served). With this description about HLP, the main decisions to be taken are the location of the hub facility among the set nodes and the routing of flows to serve the demand. Routing flow decisions define the structure of hub network consisting of (i) origin and destination nodes, (ii) hub nodes, (iii) connections between hubs and non-hub nodes, and (iv) hub arcs that connect two hubs together.

Campbell and OKelly [20] found that ”two large classes of network design problems that can be viewed from a hub location perspective are (1) multicommodity network design problems, where commodities are associated with flows between OD pairs and the network design includes locating intermediate nodes, and (2) network design problems that require connectivity via intermediate nodes to be located. This includes many two-level network design problems with location, where the upper level is the hub facilities and hub arcs and the lower level consists of the demand points and access arcs. In telecommunication networks, the upper level is often termed the backbone level and the lower level is the tributary or access network”.

By considering this classification, various mathematical models for hub network design can be formulated. The most common formulations, which have been widely applied by the literature, are introduced in this section.

2.2 Mathematical models and formulations of HLP

The notation will be used throughout this thesis is as follows:

- x_{ijkm}= Fraction of flow from location (origin) i to location (destination) j that is routed via hubs at locations k and m in that order.
- z_{kk}=1 if location k is a hub and 0 otherwise.
- z_{ik} = 1 if location i is allocated to the hub at location k and 0 otherwise.
- W_{ij} = Flow from location i to location j.
- c_{ij} = Standard cost per unit from location i to j.
- $C_{ijkm} = c_{ik} + c_{mj} + \alpha c_{km}$.

Decision variables x_{ijkm} and z_{ik} determine the allocation. Decision variable z_{kk} indicates hub locations. Usually c_{ij} is proportional to the distance between i and j. C_{ijkm} is the cost per unit from origin i to destination j via hubs k and m, in that order. In the remainder of this thesis i and j are used to index origins and destinations respectively, and k and m are used to index potential hub locations.

2.2.1 Single-HLP

O'Kelly [81] represented this problem as a quadratic integer program. The parameters for this problem are the amount of flow between nodes i and j, the unit cost of transferring from non-hub node i to hub node j. The outputs (decision variables) are z_{ik} if node i is allocated to a hub located at node k (and 0, otherwise). If z_{kk} is equal to one, it means that node k is a hub node. The mathematical model is as follows:

$$\min_{z} \quad \sum_{i}\sum_{k}\sum_{k} W_{ij}(C_{ik}+C_{kj})z_{ik}z_{jk} \tag{2.1}$$

$s.t.$

$$\sum_{k} z_{kk} = 1, \tag{2.2}$$

$$z_{ik} - z_{kk} \leq 0, \ \forall i,k \tag{2.3}$$

$$z_{ik}, \in \{0,1\} \quad \forall i,k \tag{2.4}$$

$$\tag{2.5}$$

The objective function minimizes the total transfer cost via the hub. Constraint (2.2) stipulates that there is only one hub. Constraint (2.3) stipulates that node i can only be linked to a hub node at k. Constraint (2.4) defines decision variable type to be binary.

2.2.2 *p*-HLP

In the p-hub median problem the objective is to minimize the total transportation cost to serve the flow among the nodes in the network (time, distance, etc.), given the number of nodes n in the network, the flow between origin-destination pairs and the number of hubs to locate (p). Two different ways are used to analyze the p-hub median problem single allocation and multiple allocation.

Single Allocation *p*-HLP

The p-median problem has been a fundamental problem in locational research since its inception. Campbell [18] defined a p-hub median analogous to a p-median.

The basic formulation as defined by Campbell [18] for the p-hub median problem is as follows:

$$\min_{x,z} \quad \sum_i \sum_j \sum_k \sum_m W_{ij} x_{ijkm} C_{ijkm} \tag{2.6}$$

$$s.t.$$

$$\sum_k z_{kk} = p, \tag{2.7}$$

$$0 \leq z_{kk} \leq 1, \; and integer \forall \quad k \tag{2.8}$$

$$0 \leq x_{ijkm} \leq 1, \; \forall \quad i,j,k,m \tag{2.9}$$

$$\sum_k \sum_m x_{ijkm} = 1, \; \forall \quad i,j \tag{2.10}$$

$$x_{ijkm} \leq z_{kk}, \; \forall \quad i,j,k,m \tag{2.11}$$

$$x_{ijkm} \leq z_{mm}, \; \forall \quad i,j,k,m \tag{2.12}$$

$$\tag{2.13}$$

The objective function in the p-hub median sums the transportation cost over all origin-destination pairs. Constraint (2.7) precisely establishes p hubs. Constraint (2.8) restricts z_{kk} to be zero or one. Constraint (2.9) limits the range of x_{ijkm}. Constraint (2.10) assures that the flow for every OD pair is routed via some hub pair. Constraints (2.11) and (2.12) assure that flows are routed via locations that are hubs. Constraints (2.11) and (2.12) may alternately be expressed as:

$$\sum_i \sum_j \sum_m x_{ijkm} \leq n(n-p+1) z_{kk} \quad \forall \quad k \tag{2.14}$$

$$\sum_i \sum_j \sum_k x_{ijkm} \leq n(n-p+1) z_{mm} \quad \forall \quad m \tag{2.15}$$

In the absence of capacity constraints on the links, an optimal solution will have all x_{ijkm} equal to zero or one since the total flow for each OD pair should be routed via the least cost hub pair.

Campbell [17] noted that the inter-hub discount factor α plays an important role in determining both the hub locations and the number of hubs to which each demand point is allocated, i.e. the number of spokes in the hub-and-spoke system. As α decreases, hubs tend to spread further apart and the number of spokes decreases, since a lower inter-hub transportation rate favors allocation to the nearest hub. In the extreme case of $\alpha = 0$ there is no cost for inter-hub transportation, so each demand point is allocated to exactly one hub: the nearest (least cost) hub. In this case, the p-hub median problem reduces to the p-median problem and there is no interaction between hubs. For large α, hub interactions are expensive and hubs are drawn closer together to reduce the significant inter-hub transportation costs. In the extreme case of $\alpha = 1.0$, there is no discount for inter-hub transportation and each demand point is allocated to all p-hubs (assuming non-zero flow between each demand point and hub location). This may result in some spokes carrying very small flows and may lead to somewhat unrealistic configurations with large numbers of spokes.

Since the CAB dataset is symmetrical, O'kelly et al. [84] proposed a formulation that utilized the symmetric flow data found in the CAB dataset, thus further decreasing the size of the problem. This reduced formulation still finds integer solutions to the LP relaxation most of the time. The results showed that the integer-friendliness of the formulation depends on the value of α. A new result in their paper was a determination of the optimal number of hubs as the fixed costs and inter hub discount factors change.

In an attempt to solve larger problems, Ernst and Krishnamoorthy [45] proposed a different linear integer programming formulation which requires fewer constraints and variables. They modeled the inter-hub transports as a multicommodity flow problem by representing each commodity as the traffic flow originating from a particular node. The authors observed and modeled how Australia Post uses different discount factors for collection and distribution. Then, Ebery [39] proposed a formulation for the single allocation p-median HLP that requires $O(n^2)$ variables and $O(n^2)$ constraints.

In terms of solution methodologies and heuristics, the first solution methodology to tackle this model was a two heuristics for the single allocation p-hub median problem, proposed by OKelly [81]. These two solution methodologies were based on the principle of enumerating all possible choices of p hub locations. Klincewicz [68] described exchange heuristics that work with an incumbent set of hubs and systematically substitute other nodes for the incumbents based on local improvement measures. Based on the results he obtained, the heuristics developed in his work showed that these heuristics are superior to a clustering heuristic and to the heuristics proposed in OKelly [81].

Klincewicz [69] presented two heuristics based on Tabu search design and a Greedy Randomized Adaptive Search Procedure (GRASP) heuristic; in both of these heuristics non-hub nodes are allocated to their nearest hubs. Skorin-Kapov and Skorin-Kapov [98] developed a heuristic method based on Tabu search for the problem of locating p interacting hub facilities among n interacting nodes in a network. Their method treats equally the problem of locating hub facilities, as well as the problem of allocating the nodes to one and only one hub.

Ernst and Krishnamoorthy [45] developed a simulated annealing heuristic . They used the simulated annealing heuristic to provide an upper bound to develop an Linear Programming-based branch-and-bound solution method. They were able to solve instances up $n = 50$, efficiently. Ernst and Krishnamoorthy [46] proposed another branch-and-bound algorithm which solves shortest-path problems to obtain lower bounds. In contrast to the traditional branch-and-bound algorithms, their algorithm starts with a set of root nodes instead of a single root node. Marin et.al. [76] generalized the basic model by providing tighter LP bounds and introduced a formulation for the problem.

It is worth noting that in the case of using decomposition techniques (e.g., Lagrangean relaxation and benders decomposition), this formulation would be a better choice as it provides tighter linear programming relaxation bounds. Pirkul and Shilling [89] used the Lagrangean Relaxation Model with a cut constraint to find tight upper and lower bounds in a reasonable amount of time. Abdinnour [1] proposed a hybrid heuristic based on Genetic Algorithms and Tabu Search to solve the uncapacitated HLP.

Labbe et al. [71] investigated polyhedral properties of p-hub median problem problems to develop a branch and cut algorithm. Contraras et. al. [28] considered the capacitated hub location problem with single assignment. They proposed using Lagrangean relaxation to obtain tight upper and lower bounds. The Lagrangean function that they formulated exploits the structure of the problem and can be decomposed into smaller subproblems that can be solved. They tested the heuristic on instances (ranging from 10 to 200 nodes), the upper and lower bounds, never exceed 3.4%.

Elhedhli and Hu [42] considered, for the first time, the congestion at the hubs and modeled it as a non-linear convex cost function in the objective function of the single allocation p-hub median model. They proposed a Lagrangean relaxation. The authors stated that the addition of congestion cost in the objective function results in a more balanced distribution of flows through hubs via comparison with the non-congestion problem on the CAB data set. Great advancement was achieved by Ilić et al. [62] where they presented a general variable neighborhood search approach for the uncapacitated single allocation p-hub median problem in networks. They used three neighborhoods and efficiently updated data structures for calculating total flow in the network. In addition to the usual sequential strategy, a nested strategy is proposed in designing a deterministic variable neighborhood descent local search and they were able to solve instances up to up to 1,000 nodes and 20 hubs. Table 2.1 summarizes the single allocation p-median HLP literature.

Authors	Notes	Solution Methodology
OKelly [81]	Quadratic integer program	HEUR1, HEUR2
Campbell [17]	First linear integer formulation	Exact solution algorithm
O'Kelly et al. [84]	Reduce the size of the formulation by assuming symmetric flow data	Mixed Integer Linear Programming
Ernst and Krishnamoorthy [45]	Linear integer programming	-
Ebery [39]	New formulations	-
Klincewicz [68]	Heuristic for Single p median HLP	Tabu search and GRASP heuristics
Skorin-Kapov and Skorin-Kapov [98]	developed heuristic method	tabu search
Ernst and Krishnamoorthy [45]	developed heuristic method	simulated annealing heuristic
Ernst and Krishnamoorthy [46]	proposed another heuristic	Branch-and-bound algorithm based on shortest-path problems
Marin et.al. [76]	Generalizing the basic model and providing heuristic	tighter LP bounds
Pirkul and Shilling [89]	New heuristic	Lagrangean relaxation
Abdinnour [1]	New heuristic	Hybrid of Genetic Algorithms (GAs) and Tabu Search (TS)
Labbe et. at. [71]	Investigated polyhedral properties	A branch and cut algorithm
Contraras et. at. [28]	The capacitated hub location	Lagrangean relaxation
Elhedhli and Hu [42]	Considered the congestion at the hubs	Lagrangean relaxation
Ilić et al. [62]	A new general variable neighborhood search approach	Neighborhood search approach

Table 2.1: Summary of Single Allocation p-HLP literature.

Multiple Allocation p-HLP median

One of the variants of p-median problem has been a fundamental problem in location research since its inception. In the p-hub median nodes can be allocated to more than one hub, and unlike single allocation p-HLP problem the demand nodes can be assigned to more than one hub.

Campbell [16] was the first to formulate the basic formulation for the p-hub median (or Multiple Allocation p-HLP) problem is as follows:

$$\min_{x,z} \quad \sum_i \sum_j \sum_k \sum_m W_{ij} x_{ijkm} C_{ijkm} \tag{2.16}$$

$$s.t. \quad 2.7, 2.8$$

$$x_{ijkm} \geq 0, \ \forall \quad i, j, k, m \tag{2.17}$$

$$\sum_k \sum_m x_{ijkm} = 1, \ \forall \quad i, j \tag{2.18}$$

$$x_{ijkm} \leq z_{kk}, \ \forall \quad i, j, k, m \tag{2.19}$$

$$x_{ijkm} = z_{mm}, \ \forall \quad i, j, k, m \tag{2.20}$$

Skorin-Kapov et al. [99] developed a mixed linear formulations with tight linear programming relaxations and the solution methodology was tabu search. Ernest and Krishnamoorthy [46] presented a MILP formulations for the multiple allocation p-hub median problem. They proposed an efficient heuristic algorithm, based on shortest paths. LP based solution methods as well as an explicit enumeration algorithm were developed to obtain exact solutions in their work.

Clearly, the objective function that results in solving the multiple allocation p-hub median provide a lower bound on the optimal solution of the SA-p-MHLP Campbell [18]. Based on this idea, Campbell [18] proposed two heuristics for the SA-p-MHLP, namely MAXFLO & ALLFLO, to find solutions to the SA-p-MHLP from the solutions provided by to the multiple allocation p-hub median problem. Ernest and Krishnamoorthy [46] proposed a Linear Programming based on the B-n-B method. They strengthened the lower bound by identifying violated inequalities and adding them to the LP. In the same year, the same authors presented another study [46] where they developed another more effective (in terms of CPU time requirement) branch-and-bound algorithm.

Boland et al. [10] observed characteristics of optimal solutions for Uncapacititated Multiple Allocation pHMP, Capacititated Multiple Allocation HMP and Uncapacititated Multiple Allocation HMP. They used these characteristics to develop preprocessing techniques and tightening constraints and apply them to the appropriate problems. They found that in all cases that the linear programming formulations have become tighter and the overall running time is reduced. They stated that although Ernest and Krishnamoorthy [46] achieved faster computation times (and required less memory) these formulations suffer from weak lower bounds. De Camargo et.

al. [34] presented efficient Benders decomposition algorithms and they were able to solve some large instances, considered "out of reach" of other exact methods in reasonable time. Great advances in solution methodologies and algorithms were achieved by Contreras et al. [29] who solved multiple allocation uncapacitated hub location problems with up to 500 nodes (250,000 Origin Destination pairs or commodities) using an specialized Benders decomposition algorithm enhanced with some properties based on the solution of the Multiple HLP.

Table 2.2 the multiple allocation p-median HLP literature. ADD Camargo

Authors	Notes	Solution Methodology
Campbell [16] [81]	Presented the first formulation	-
Skorin-Kapov et al. [99]	new mixed linear formulations with tight linear programming relaxations	Tabu search
Ernest and Krishnamoorthy [46]	New MILP formulations	-
Campbell [18]	Proposed a heuristic	Greedy-Interchange
Ernest and Krishnamoorthy [46]	Presented a heuristic	LP based Branch-And-Bound method
Boland et al. [10]	Developed preprocessing techniques and tightening constraints	-
Contreras et al. [29]	Developed a solution algorithm	Enhanced Benders decomposition algorithm

Table 2.2: Summary of Multiple Allocation p-HLP literature.

2.2.3 Capacity Limitation of Hub Location Problem

Capacity limitation of the hub node means that the total flows, incoming or outgoing, must be less than or equal to a fixed value and is called the capacitated hub location problem. This model is represented by Campbell ([17]).The assumptions of this model are similar to the median P-hub model except that the capacities of the hub nodes are limited. Model parameters are the same as the median P-hub model in addition to the capacity of a hub at candidate node k (θ_k). The objective function is similar to the median P-hub model except that the following constraints are added to the median P-hub model's constraints 2.21:

$$\sum_m \sum_i \sum_j W_{ij} X_{ijkm} + \sum_s \sum_i \sum_j W_{ij} X_{ijsk} \leq \theta_k Z_{kk} \quad \forall \quad k. \tag{2.21}$$

Ernst and Krishnamoorthy [47] presented an efficient approach for solving capacitated single allocation hub location problems. The modified version of the previous mixed integer linear programming formulation was developed by them. Their new formulation requires fewer variables and constraints than those traditionally used in the literature. They also developed an effective heuristic algorithms for its solution based on simulated annealing (SA) and random descent (RDH).Randall [94] presented four variations of the ant colony optimization meta-heuristic that explored different construction modelling choices for the capacitated single allocation HLP.

Contreras et. al. [31] presented a branch-and-price algorithm for the capacitated HLP with single assignment, in which Lagrangean relaxation was used to obtain tight lower bounds of the restricted master problem. A lower bound that is valid at any stage of the column generation algorithm was proposed. The process to obtain this valid lower bound was combined with a constrained stabilization method that results in a considerable improvement on the overall efficiency of the solution algorithm. Lin et. al. [72] proposed a general capacitated p-hub median model, with economies of scale and integral constraints on the paths. Their model requires the selection of a specific p among a set of candidate hubs so that the total cost on the resulting pure capacitated hub-and-spoke network is minimized while simultaneously meeting OD demands, operational capacity and singular path constraints. They developed a genetic algorithm using the path for encoding. This algorithm is capable of determining local optimality within less than 0.1% of the Lagrangean relaxation lower bounds on the Chinese air cargo network testing case and has reasonable computational times.

2.2.4 P-Hub Center Location Problem

Minimax center problems are fundamentally different from the minisum median and uncapacitated facility location problems. Center problems are important both because of applications, such as locating emergency service facilities and vehicles, and because of the insight into worst case scenarios, e.g., maximum travel times. The p-hub center problem is analogous to the p-center problem. If an OD pair in hub location problem is viewed as analogous to a demand point in a p-center problem, then the natural definition of a hub center is a set of hubs such that the maximum cost for any OD pair is minimized, and the maximum time from an origin-to-destination is of interest.

The three version of center HLP provided by Campbell [17] is as follows:

1. The first type the interest is to minimize the maximum time from an origin-to-destination.

2. The second type of hub center is a set of hubs that minimizes the maximum cost for movement on any single link.

3. The third type of hub center is in which the set of hubs minimizes the maximum cost for movement between a hub and an origin/destination.

As stated by Campbell in [17] the first type is important for a hub system involving perishable or time sensitive items, in which cost refers to time and α is a time discount factor due to higher speed on the inter-hub links. The second type of hub center would be of interest in transportation systems in which cost refers to time and the maximum time for any one link is important. One example would be items that require some preserving or rejuvenating processing, such as heating or cooling, which is only available at the hub locations. Another application would be for vehicle drivers or pilots who are subject to a time limit on continuous service, so the length of any single link is important. For the third hub center type similar examples to the second type can be given considering that hub-to-hub links may have some special attributes.

The basic formulation for the first type is as follows:

$$\min \max\{x_{ijkm}C_{ijkm}\} \tag{2.22}$$

$$s.t.$$

$$\sum_k z_{kk} = p, \; \forall \quad k \tag{2.23}$$

$$0 \leq z_{kk} \leq 1, \text{ and integer} \forall \quad k \tag{2.24}$$

$$\sum_k \sum_m X_{ijkm} = 1, \quad \forall \quad i,j \tag{2.25}$$

$$x_{ijkm} \leq z_{kk}, \; \forall \quad i,j,k,m \tag{2.26}$$

$$x_{ijkm} \leq z_{mm}, \; \forall \quad i,j,k,m \tag{2.27}$$

$$0 \leq x_{ijkm} \leq 1, \quad \forall \quad i,j,k,m \tag{2.28}$$

The constraints are almost identical to those of the p-hub median problem. The only difference is that constraint (2.28) replaces constraint (2.9), since x_{ijkm} must be integer. Note that if α = 0, this does not reduce to a vertex p-center problem, since the cost C_{ijkm} is the sum of the cost from i to k and m to j for α=0.

Kara and Tansel [64] presented different linear formulations for the single allocation p-hub center problem. They provided three different linearizations of the Campbell [17] type (1) model with a new formulation that they presented. Ernst et al. [49] developed new mathematical formulations for p-hub center location problem and demonstrated the superiority of this formulation to the one by Kara and Tansel [64] for single allocation p-hub center location problems. Based on the study conducted by Ernst, Hamacher, Jiang, Krishnamoorthy, and Woeginger (2002) (Unpublished Report, CSIRO Mathematical and Information Sciences, Australia), Baumgartner [9] investigated the polyhedral properties of their formulation. She identified some facet-defining inequalities and defined separation procedures. She proposed a branch-and-cut algorithm. It should be noted that a novel improvement of p-hub center location problems was proposed by Campbell et. al. [15]. Yaman and Elloumi [106] also proposed new formulations for star p-hub center location problem to minimize the length of the longest o-d path.

Gavriliouk and Hamacher [54] developed a heuristic based on aggregation for$k$-hub center problems.Meyer et. al. [79] presented an exact 2-phase algorithm where in the first phase they computed a set of potential optimal hub combinations using a shortest path based branch and bound. This was followed by an allocation phase using a reduced sized formulation which returns the optimal solution. In order to obtain a good upper bound for the branch and bound they developed a heuristic for the single allocation p-hub center problem based on an ant colony optimization approach. Table 2.3 summarizes the center HLP literature.

Authors	Notes	Solution Methodology
Campbell [17]	Presented the basic formulation with different criterion	-
Kara and Tansel [64]	Various linearizations for the single allocation	Linear programming
Pamuk and Sepil [85]	solving the p-hub center problems	A single-relocation heuristic with Tabu search
Ernst et at. [48]	New formulations for both single and multiple allocation	heuristic and a Branch and Bound algorithm
Baumgartner [9]	Investigated the polyhedral properties	Valid inequalities and Branch and cut
Yaman and Elloumi [106]	Proposed new formulations	Mixed Integer Programming solver
Gavriliouk [54]	Studied six models including single allocation of center HLP	developed a heuristic based on aggregation
Campbell et al. [15]	presented complexity results and Integer Programming formulations for several versions of the *p*-HLP	-
Ernst et al. [49]	New formulation	shortest path based branch-and-bound
Mayer et at. [79]	Developed a heuristic	Exact 2-phase algorithm

Table 2.3: Summary of center HLP literature.

2.2.5 Hub covering problems

Covering problems have an inverse relationship to center problems. Demand locations are covered if facilities are 'close enough' to serve the demand within specified parameters. Hub covering problems are analogous to facility covering problems, but the notion of hub coverage may have several interpretations, just as the notion of a hub center in the preceding section had several possible interpretations. Campbell [17] defined three coverage criteria for hubs. The origin destination pair *(i,j)* is covered by hubs *k* and *m* if

1. If the OD pairs in a hub covering location problem are viewed as analogous to the demand points in a facility location covering problem, then the natural interpretation of coverage is that OD pair *(i, j)* is covered by hubs *k* and *m* if the cost from *i* to *j* via *k* and *m* does not exceed a specified value, i.e.:

   $C_{ijkm} \leq \gamma_{ij}$ where γ_{ij} is the given maximum cost for o-d pair (i, j).

2. A second type of coverage can be defined such that the OD pair *(i,j)* is covered by hubs

k and m if the cost for each link in the path from i to j via k and m does not exceed a specified value,

$Max\{c_{ik}, c_{mj}, \alpha c_{km}\} \leq \gamma_{ij}$.

3. A third type of coverage can be defined such that the OD pair *(i, j)* is covered by hubs k and m if each of the hub-origin/destination links meets a separate tolerance:

   $c_{ik} \leq \gamma_i$ and $c_{mj} \leq \gamma_j$

Covering problems may prove important in solving hub center problems, just as algorithms for solving p-center problems require solving a sequence of set covering problems.

The basic formulation as defined in Campbell [17] is as follows:

$$\min_{x,z} \quad \sum_k F_k z_{kk} \tag{2.29}$$

$s.t.$

$$0 \leq z_{kk} \leq 1, \text{ and integer} \forall k \tag{2.30}$$

$$x_{ijkm} \leq z_{kk}, \; \forall \quad i, j, k, m \tag{2.31}$$

$$x_{ijkm} \leq z_{mm}, \quad \forall \quad i, j \tag{2.32}$$

$$x_{ijkm} \leq z_{kk}, \; \forall \quad i, j, k, m \tag{2.33}$$

Where V_{ijkm} takes 1 if hubs k and m cover OD pair *(i,j)* and 0 other wise.

Kara and Tansel [66] studied the hardness of the single allocation hub set-covering problem and proved that it is *NP*-hard.

Later, Ernst et al. [44] presented a new formulation for the single allocation hub set covering problem similar to the one that is proposed in Ernst, Hamacher, Jiang, Krishnamoorthy, and Woeginger (2002) (Unpublished Report, CSIRO Mathematical and Information Sciences, Australia), for the p-hub center problem. The new formulation is

$$\min_{x,z} \quad \sum_k z_{kk} \tag{2.34}$$

$s.t.$

$$\sum_k x_{ik} = 1, \quad \forall \quad i \tag{2.35}$$

$$x_{ik} \leq x_{jj}, \ \forall \quad i,k \tag{2.36}$$

$$r_k + r_m + \alpha C_{km} \leq \beta, \ \forall \quad k,m \tag{2.37}$$

$$x_{ik}, \in \{0,1\} \tag{2.38}$$

where β is the cover radius.

This formulation performs better in terms of CPU time requirement than the strengthened Kara and Tansel [66] formulation because Ernst et at. (2005) [44] strengthened Kara and Tansel [66] formulation by replacing a constraint with its aggregate form.

Later, Hamacher and Meyer [58] presented a study where they introduced and compared various formulations for hub covering problems and analyzed the feasibility polyhedron of the most promising one.

In terms of algorithms, heuristics and solution methodologies, Calik et. al. [13] presented an efficient heuristic based on Tabu search for the hub covering problem over incomplete hub networks. Qu and Weng [91] studied the multiple allocation hub maximal covering problem, and designed a new model for the Multiple Allocation Hub Maximal Covering Problem, they also provided an evolutionary approach based on path relinking. Table 2.4 summarizes the literature for covering HLP.

Authors	Notes	Solution Methodology
Campbell [17]	Presented the basic formulation with different criterion	-
Kara and Tansel [66]	Provided dierent linearizations of the original quadratic model as well as presenting a new linear model	-
Wagner [102]	Improved model formulations using multiple and single allocation problems	-
Ernst et at. [44]	New formulation for singl and multiple allocation	Implicit enumerative
Hamacher and Meyer [58]	Comparing various formulation	Facet defining valid inequalities.
Qu and Weng [91]	Multiple allocation hub maximal covering	evolutionary approach based on path relinking.
Calik et. al. [13]	hub covering problem over incomplete hub networks	Tabu search.

Table 2.4: Summary of Cover HLP literature.

2.2.6 Multi-objective hub location problems

Costa et al. [33] developed a different approach to the capacitated single allocation hub location problem is presented. They found that ”Instead of using capacity constraints to limit the amount of flow that can be received by the hubs, the introduction a second objective function to the model (besides the traditional cost minimizing function), can be used which tries to minimize the time to process the flow entering the hubs. Their model was based on the formulation presented by Ernst and Krishnamoorthy [47]. They defined the following variables: y_i^{km} is the total amount of flow from location i(origin) that is routed via hubs k and m”. The input data are given as: n-number of locations; flow from location i to location j ($O_i = \sum_j W_{ij}$ and $D_i = \sum_j W_{ij}$); d_{ij}-distance between node i and node j; χ coefficient of the collection cost (per unit flow) from any non-hub node to any hub node; δ coefficient of the distribution cost from any hub node to any non-hub node; $\alpha(0 \leq \alpha \leq 1; \alpha < \chi and \alpha < \delta)$ coefficient of the transfer cost between any two hubs; F_k fixed cost of establishing a hub at node k; Γ_k-capacity of collecting flow at hub k (flow into hub k).

2.2.7 Fixed cost HLP

In the p-hub location model, it is usual to ignore the fixed costs of opening facilities. In contrast, the simple plant location problem includes fixed facility costs and thereby makes the number of facilities one of the decision variables. O'Kelly [82] introduced fixed facility costs into a hub location model, thereby making the number of hubs a decision variable. He used a two-step procedure to solve the formulation.

He formulated this problem as:

$$\min \sum_i \sum_k x_{ik} C_{ik} (O_i + D_i) + \sum_j \sum_m x_{jm} \sum_i \sum_k x_{ik} (\alpha W_{ij} C_{km}) + \sum_j z_{jj} F_k \tag{2.39}$$

Subject to 2.35, 2.38, and 2.36

Where F_k is the fixed cost of opening a hub at node k. Since the number of hubs is not predefined in advance, it is possible to have uncapacitated/capacitated hub location problems with fixed costs, in addition of the possibility of having single and multiple allocation. Campbell [17] presented the first linear programming formulations for multiple/single allocation uncapacitated/capacitated HLP. The uncapacitated hub location problem with fixed cost was formulated as:

$$\min \sum_i \sum_j \sum_k \sum_m W_{ij} x_{ijkm} C_{ijkm} + \sum_k F_k z_{kk} \tag{2.40}$$

Subject to 2.38,2.10, 2.11, 2.12.

$$0 \leq x_{ijkm} \leq 1 \quad \forall \quad i, j, k, m \tag{2.41}$$

Abdinnour-Helm and Venkataramanan [2] proposed a new quadratic integer formulation for the Uncapacitated fixed cost HLP. The new formulation lends itself well for using a branch-and-bound procedure to find optimal solutions. The branch-and-bound procedure was not implemented in a traditional fashion,instead, a more sophisticated approach was used where bounds are obtained by employing the underlying network structure of the problem. In addition, an artificial intelligence-based technique (Genetic Search) is designed to find solutions quickly and efficiently. Abdinnour-Helm [1] proposed a new heuristic method based on a hybrid of Genetic Algorithms (GAS) and Tabu Search (TS) by simultaneously finding the number of hubs, the location of hubs, and the assignment of spokes to the hubs. The results of this study clearly demonstrate that using TS in combination with GAS leads to much better solutions than if only GAS are used based on the comparison with the previous study [2].

Topcuoglu et al. [100] presented a new and robust solution based on a genetic search framework. To present its effectiveness, they compared the solutions of their GA-based method with Abdinnour-Helm [1]. Cuhna and Silva [32] presented a heuristic based on genetic algorithms to the problem of configuring hub-and-spoke networks for trucking companies that operate less-than-truckload (LTL) services in Brazil. The proposed formulation differs from similar formulations in the sense that it allows variable scale-reduction factors for the transportation costs according to the total amount of freight between hub terminals, as occurs to less-than-truckload (LTL) freight carriers in Brazil.

Chen [27] presented two approaches to determine the upper bound for the number of hubs along with a hybrid heuristic based on the simulated annealing method, Tabu list, and proposed improvement procedures are proposed. Computational results by Chen have demonstrated that the proposed heuristic outperforms the algorithms presented by Topcuoglu et al. [100] in terms of run time and solution quality. Mayer and Wagner [78] suggested HubLocator a new branch-and-bound procedure for the uncapacitated multiple allocation HLP. HubLocator considers an aggregated model formulation enabling them to significantly tighten the lower bounds.

Canovas et al. [23] considered the dual problem of a four-indexed formulation and a heuristic method, based on a dual-ascent technique. This heuristic, which is reinforced with several specifical subroutines and does not require any external linear problem solver, is the core tool embedded in an exact branch-and-bound framework. In addition, their heuristic provides the branch-and-bound algorithm with good lower bounds for the nodes of the branching tree. The results of the computational experience showed the profound effectiveness of this approach: instances with up to 120 nodes were solved.

Aykin [6] introduced the capacitated fixed cost HLP in which hubs have limited capacity for directing demand between the nodes served by the system. He provided a formulation to the problem under two networking policies. In the first policy shipments are allowed to go directly from the origin node to its its destination (nonstop, direct shipment) and in the second policy the shipments are allowed to be routed through the hubs connection (one hub stop and two hub stop). He presented two solution algorithms, a branch and bound algorithm and a heuristic procedure partitioning the set of solutions on the basis of hub locations. Then, Aykin [7] analyzed a similar problem with fixed costs and a given number of hubs to locate. He compared two hubbing policies which he named as strict and non-strict (direct connections are allowed). He proposed an enumeration algorithm and a simulated annealing-based greedy interchange heuristic.

Ernst and Krishnamoorthy [47] modified version of a previous mixed integer linear programming formulation developed by them for p-hub median problems. The new formulation requires fewer variables and constraints than those traditionally used in the literature. They also developed effective heuristic algorithms for its solution based on simulated annealing (SA) and random descent (RDH), and their formulation is as follows:

$$\min_{x,y} \quad \sum_i \sum_k C_{ik} x_{ik} (\chi O_i + \delta D_i) + \sum_i \sum_k \sum_l \alpha C_{kl} y^i_{kl} + \sum_k F_k x_{kk} \tag{2.42}$$

$$s.t. \quad 2.38, 2.35, 2.36$$

$$\sum_l y^i_{kl} - y^i_{lk} = O_i x_{ik} - \sum_j W_{ij} x_{jk}, \ \forall \quad i,k \tag{2.43}$$

$$y^i_{kl} \geq 0, \ \forall \quad i,k,l \tag{2.44}$$

$$\sum_i O_i x_{ik} \leq \Gamma_k x_{kk}, \ \forall \quad k \tag{2.45}$$

Where Γ_k is the capacity of hub k. The capacity limits are only applied to the traffic arriving at the hub directly from non-hub nodes. This capacity definition is usually used in postal service applications in order to represent the sorting capacity of hubs.

Lebbe et al. [71] investigated polyhedral the properties of single allocation capacitated fixed cost HLP and developed a branch and cut algorithm based on the results. Ebery et al. [40] present a new Mixed Integer Linear Program formulation for capacitated Multiple Allocation HLP. Their formulation used fewer variables and constraints than that reported in Campbell [17] or the closely related to the formulation by Ernst and Krishnamoorthy [46]. They also described a new heuristic method based on shortest paths and incorporated the upper bound obtained from this heuristic in an LP-based branch-and-bound solution procedure. Sasaki and Fukushima [95] considered a one-stop capacitated hub location problem in which both hubs and arcs have capacity constraints. They formulated the problem as a mixed $\{0,1\}$ integer programming problem and solved the problem using branch-and-bound method with Lagrangean relaxation bounding strategy. Table 2.5 summarizes the literature on fixed cost HLP.

2.2.8 Competitive Hub location

Campbell and OKelly [20] observed that ”although some hub-based transportation systems operate without competition (e.g., public sector systems, such as transit or postal operations), for most private sector hub-based transportation systems there is competition between multiple service providers. When two or more firms compete for customers (e.g., passengers or freight), the optimal hub location, networks, and traffic patterns may differ greatly from those without competition. Thus, competitive hub location problems provide a rich source of realistic and difficult problems”.

Marianov et al. [75] offered a formulation that locates hubs on a network in a competitive environment; that is, customer capture is sought, which happens whenever the location of a new hub results in a reduction of the current cost (time, distance) needed by the traffic that goes from the specified origin to the specified destination. The formulation presented reduces the number of variables and constraints as compared to existing covering models. Eiselt and Marianov [41] considered the competitive hub location problem with the feature that the customer choice function is probabilistic and uses a gravity model as a utility function, and customers choose an airline depending on a combination of functions of flying time and fare. The problem was formulated and a version of the principle of heuristic concentration was described for its solution. Gelareh et al [55] proposed a mixed integer programming formulation for hub-and-spoke network design in a competitive environment. It addressed the competition between a newcomer liner service provider and an existing dominating operator, both operating on hub-and-spoke networks.

2.2.9 Models with Stochastic Elements and Reliability Considerations

As in logistic systems the models can be used to describe deterministic and stochastic issues, so for the HLP, the stochastic nature can be presented. For instance, Sim et al. [97] introduced the stochastic p-hub center problem with service-level constraints, which seeks to configure a

Authors	Notes	Solution Methodology
O'Kelly [82]	Introduced fixed facility costs for HLP	Two step heuristic
Campbell [17]	Presented the first linear programming formulations	-
Abdinnour-Helm and Venkataramanan [2]	A new quadratic integer formulation	Branch-and-bound with the design of artificial intelligence-based method
Abdinnour-Helm [1]	Proposed a new heuristic method	Hybrid of Genetic Algorithms (GAS) and Tabu Search (TS)
Topcuoglu et al. [100]	A new and robust solution methodology	A genetic search framework
Cuhna and Silva [32]	Presented a heuristic	Genetic algorithms
Chen [27]	Presented a heuristic	Hybrid heuristic based on the simulated annealing method, tabu list, and improvement procedures
Mayer and Wagner [78]	Presented a solution procedure	A new branch-and-bound
Cnovas et al. [23]	Considered the dual problem of a four-indexed formulation and a heuristic method	Design of dual-ascent technique
Aykin [6]	Introduced the capacitated fixed cost HLP with limited capacity for channelling flows	Branch and bound and heuristic based on partitioning the set of solutions
Aykin [7]	Compared two hubbing policies	An enumeration algorithm and a simulated annealing-based greedy interchange heuristic.
Ernst and Krishnamoorthy [47]	New modified version of a previous mixed integer linear programming and heuristic algorithms	Simulated annealing (SA) and random descent (RDH)
Lebb et al. [71]	Investigated polyhedral the properties of single allocation capacitated fixed cost HLP	Developed a branch and cut algorithm based on the results.
Ebery et al. [40]	Present a new Mixed Integer Linear Program formulation for capacitated Multiple Allocation HLP	Heuristic based on shortest paths and an LP-based branch-and-bound solution procedure
Sasaki and Fukushima [95]	A one-stop capacitated hub location problem	Branch-and-bound method with Lagrangean relaxation

Table 2.5: Summary of Fixed cost HLP

network that minimizes the longest transportation time in the network for a specified service level in delivery times. Chance constraints were used to model the service-level constraints with variability in travel time. Contreras et al. [30] studied stochastic uncapacitated hub location problems in which uncertainty is associated with demands and transportation costs. They described a Monte-Carlo simulation-based algorithm that integrates a sample average approximation scheme with a Benders decomposition algorithm to solve problems having stochastic independent transportation costs. Yang [108] introduced a stochastic programming model to address the air freight hub location and flight routes planning under seasonal demand variations. The model is separated into two decision stages. The first stage, which is the decision not affected by randomness, determines the number and the location of hubs. The second stage, which is the decision affected by randomness, determines the flight routes to transport flows from origins to destinations based upon the hub location and realized uncertain scenario. Hult et al. [61] solved the single allocation *p*-hub center problem with stochastic travel times using cutting planes and Benders decomposition.

Hub location models with reliability considerations have also been recently analyzed. Reliability issues may arise in several forms in hub networks, including failures of edges and nodes. Kim and OKelly [67] presented a new hub location problem, termed the reliable p-hub location problem, which focuses on maximizing network performance in terms of reliability by locating hubs for delivering flows among city nodes. An et al.[5] presented a study on reliable single and multiple allocation hub-and-spoke network design problems where disruptions at hubs and the resulting hub unavailability can be mitigated by backup hubs and alternative routes. They built nonlinear mixed integer programming models and present linearized formulas. To solve those difficult problems, Lagrangean relaxation and Branch-and-Bound methods were developed to efficiently obtain optimal solutions.

2.2.10 Time definite and service levels in HLP

Due to the competitive environment in cargo and delivery services in the transportation market, companies pay more attention to service levels. In cargo delivery, service is primarily measured via delivery time. Cargo companies offer different delivery time promises, such as next day or second day delivery, to their customers.

Campbell [19] was the first study minimizing transportation costs subject to a constraint on the service level in hub location problems. Yaman et al. [107] addressed this issue and proposes models incorporating transportation costs and service levels for complete hub networks.

2.3 Case studies and real life applications for HLP

In 1994, Australia Post (AP) requested that the operations research group at the Commonwealth Scientific and Industrial Research Organization (CSIRO) model their mail sorting and distribution center (SDC) network in Sydney to assess whether the continued use of a single large SDC in central Sydney was efficient. Mohan Krishnamoorthy began this work at CSIRO and quickly realized that the AP problem was a version of the newly published p-hub median (where SDCs are hubs). Although much of the literature was concerned with solving problems with 25 nodes (cities), AP was interested in solving larger problems of about 200 nodes (postcodes) and about 10 hubs (SDCs). CSIRO developed a simulated annealing heuristic solution algorithm and results were used by AP in an internal report (and later in Ernst and Krishnamoorthy [46]). Following the contract work for AP, CSIRO researchers (Krishnamoorthy and Andreas Ernst) undertook efforts to solve larger problems to optimality using the AP data set (with up to 200 nodes and 10 hubs). A key contribution by Ernst and Krishnamoorthy was the development of smaller formulations with three-subscript variables that track flows by origin on each arc.

Cetiner et. al. [26] considered the combined hubbing and routing problem in postal delivery systems and developed an iterative two-stage solution procedure for the problem. A case study discussing Turkish postal delivery system data was utilized. As the case study data referred to was applied on a road network, a final stage, seeking improvements based on special structures in the routed network, is appended to the two-stage solution procedure.

Campbell [19] provided models for multiple allocation p-hub median problems and hub arc location problems for time definite Less Than Truck Load service system. He imposed service levels by limiting the maximum travel distance via the hub network for each origindestination pair. Computational results were presented to demonstrate the effects of the time definite service levels on practical network design for truck transportation in North America.

Barla and Constantatos [8] provided a study that explains the reasons of adoption of the hub-and-spoke network structure in the airline industry. They showed that when an airline has to decide on its capacity before the demand conditions are perfectly known, a hub-and-spoke network structure by pooling passengers from several markets into the same plane helps the firm to lower its cost of excess capacity in the case of low demand and to improve its capacity allocation in the case of high demand.

Campbell [19] presented models for time definite hub location and network design motivated by time definite trucking in the US. The results documented the impact on network design of different levels of service, and highlighted the relative importance of different terminals or geographic regions in achieving high levels of service. The models also can help evaluate the ability of firms to improve their level of service without significant changes in their network. The results showed that an increase in the service level may require modifying a given hub network by relocating hubs (and hub arcs)or by adding hubs and hub arcs. Adding hubs and hub arcs would increase fixed costs for the new facilities and assets, while decreasing transportation costs.

Alumur et al. [4] introduced a new hub location problem motivated by the network structure of a cargo delivery company that operates a multimodal and hierarchical hub network. The aim of their problem was to design a minimum cost two-level hub network such that each pair of demand nodes receives service within a predetermined time bound. A linear mixed integer programming model was derived and some variable fixing rules and valid inequalities were proposed. Comprehensive computational experiments were presented on the Turkish network data set with the proposed hierarchical multimodal hub location model with time-definite deliveries.

Vasconcenlos et al. [101] presented an integer programming formulation, subject the new model to experiments with an intermodal general cargo network in Brazil, and addressed questions regarding the solution of the problem in practice. The main contribution of their work was to permit the analysis of a hub-and-spoke network operated under decentralized management. In this type of network, various transport companies act independently, and each makes its route choices according to its own criteria, which can include cost, time, frequency, security and other factors, including subjective ones. The results revealed that it is feasible to use the new model as an instrument to support decision making on investments in new cargo terminals.

Ishfaq and Cox [63] integrated the hub operation queuing model and the hub location-allocation model, and they investigated the effect of limited hub resources on the design of intermodal logistics networks under service time requirements. They gained the managerial insights from a study of 25-city road-rail intermodal logistics network that showed the level of available hub resources significantly affects the logistics network structure in terms of number and location of hubs, total network costs, choice of single-hub and interhub shipments and service performance.

2.4 Challenges in Hub Location

Campbell and O'Kelly [20] found that "hub location problems are NP-hard, except in very special cases, and this difficulty stems from the inclusion of elements of both facility location problems and the quadratic assignment problem because of the interhub flows. Because both of these problems are very difficult by themselves, their combination tends to make hub location problems at least as difficult as the analogous regular (non-hub) facility location problem. The majority of hub location research has addressed network and discrete problems, and researchers have employed the full range of standard optimal and heuristic solution approaches to solve hub location problems".

Moreover, Campbell and OKelly stated that "the fundamental hub location models have been extended in many ways, analogous to the extensions in facility location research (e.g., with capacities, competition, reliabilities, stochasticity, etc.), but also with features from network design problems (e.g., hop constraints and restricted topologies). This blend of location and network design aspects gives rise to special challenges, as for example with capacities that may be upper and lower limits on the flows on the arcs, on the flows entering hubs from non-hubs, or on the flows into or through the hubs". A variety of topologies, in addition to complete graphs as discussed earlier, may also be reasonable, including stars, trees, and more general shapes [20].

2.5 Congestion in Hub-and-Spoke Systems

As mentioned earlier, congestion is a major side effect of hub-and-spoke networks, especially if the hub network is designed based on deterministic settings that ignore congestion. Mayer and

Sinai [77] examined two factors that might explain the extent of air traffic delays in the United States: network benefits due to hubbing and congestion externalities. In their study they concluded that: The benefits due to hubbing is actually the dominant factor leading to congestion. The consolidation and dissemination of demand flows imply a higher tendency of congestion at hub airports than non-hub ones. Other findings were that the uncertainty of demand flows is another potential trigger for congestion since when the demand flow climbs unexpectedly up to a peak volume within a short time, it results in congestion.

The hub network congestion problem was first considered by Grove and O Kelly [57]. Given hub locations, they analyzed the relationship between hub and spoke networks and congestion and they did a simulation for the daily operation of a single assignment hub-and-spoke air network. The authors concluded that the extent of schedule delays depends largely on the size of hub flows. Later, Kara and Tansel [65] presented a study where they focused on the minimization of the arrival time of the last arrived item in cargo delivery systems and developed a model that correctly computes the arrival times by taking into account both the flight times and the transient times. They introduced nonlinear and linear integer formulations and the effects of delays on the system performance were analyzed. As post-design analysis, both papers contributed to the identification of the critical factors affect hub congestion behavior.

Several authors attempted to tackle the congestion effect on the hub-and-spoke systems. For example, Pels and Verhoef [87] looked at the congestion effect from an economical perspective and suggested that congestion pricing would be an appropriate response to cope with the growing congestion levels currently experienced at many airports. They developed a model of airport pricing that captures a number of congestion features. The model in particular reflects that airlines typically have market power and are engaged in oligopolistic competition at different sub-markets. It also indicates that part of external travel delays that aircraft impose are internal to an operator and hence should not be accounted for in congestion tolls. Moreover, different airports in an international network will typically not be regulated by the same authority. Raffarin [93] presented a work to deal with Air Traffic Control (ATC) pricing as a means of sorting out the European airspace congestion problem.

Other efforts to mitigate the congestion effect on the hub-and-spoke systems can be found in the work by Aykin [6] in which he introduced the capacitated hub-and-spoke network design problem where hubs have limited capacity for channelling flows between the nodes served by the system. Since then, capacity constraints were largely applied in hub design models as a medium to control congestion. Ebery et. al. [40] presented a mixed integer linear programming formulation for the capacitated multiple allocation hub location problem. They also constructed an efficient heuristic algorithm, using the shortest paths. They incorporated the upper bound obtained from this heuristic in a linear-programming-based branch-and-bound solution procedure.

Marianov and Serra [74] presented models for the optimal location of hubs in airline networks, which take into consideration the congestion effects. Hubs, which are typically the most congested airports, were modeled as M/D/c queuing systems. They derived a formula for the probability of a number of customers in the system, which is later used to propose a capacity constraint. This constraint limits the probability of more than b airplanes in queue, to be smaller than or equal to a given value.

Sasaki and Fukushima [95] presented a formulation of one-stop capacitated hub-and-spoke model as a natural extension of the uncapacitated one-stop model. The model involves arc capacity constraints as well as hub capacity constraints, which enabled them to incorporate some practical factors into the model. They also presented a branch-and-bound based exact solution method with Lagrangean relaxation bounding strategy.Later, capacity constraints were extensively used in the work in HLP, for example Ebery et al. used the capacities for multiple allocation version of HLP [40], Costa et al. [33] used the capacity restriction in their bicriteria approach, Sasaki and Fukushima [95] proposed capacity restrictions on the inter-hub connections, [47], [94], [31], and [28].

However, capacity constraints does not keep the amount of hub flow below the given hub capacity but is not regarded as a thorough solution to the congestion problem. The reason is that congestion is often proportional to the relative difference between the hub flow and the hub capacity. The smaller the difference, the larger is the congestion. This relationship can not be accurately reflected by a simple capacity constraint.

Elhedhli and Hu [42] are the first to include congestion costs in the objective, along with transportation and fixed costs, and developed a heuristic Lagrangean solution algorithm. The reasoning behind this is that congestion related costs increase with an accelerating rate when the flow increases, and this relationship can be properly captured by an exponential function. Hence, the they inserted this congestion cost into the objective of an uncapacitated single assignment p-hub location problem. The optimal solution to this novel formulation is challenged with the coexistence of nonlinearity and large scale. The nonlinear congestion cost function is linearized by a set of piece-wise linear and tangent hyperplanes.

Elhedhli and Wu [43] viewed the hub-and spoke system as M/M/1 queues taking into consideration the congestion effect, they proposed a Lagrangean heuristic to solve the Nonlinear Mixed Integer Program. Elhedhli and Wu (2010) [43] included congestion costs in the objective, along with transportation and fixed costs, and developed a heuristic Lagrangean solution algorithm. The incorporation of nonlinear term in the objective function by Elhedhli & Hu [42] and Elhedhli & Wu [43], could be considered a valuable contribution to the hub location literature.

Later, this concept was adopted by Camargo et al. [36] who addressed the multiple allocation in hub and spoke network with congestion and they developed a general Benders decomposition algorithm to solve the NLMIP formulation. De Camargo and Miranda [38] studied the single allocation HLP with congestion considering two different network perspectives: the owner and the network user and they proposed Benders decomposition algorithm to solve the NLMIP formulation. De Camargo et al. [37] solved up to 200 node network with congestion consideration of HLP using Outer-Approximation and Benders cuts to tackle the single allocation HLP.

2.6 Economies Of Scale in Hub-and-Spoke Systems

Surprisingly enough, although the economies of scale phenomenon is one of the main motivations for installing hub-and-spoke systems, the way costs are modeled has not really been questioned by many authors in this area. Economies of scale, due to the amalgamation of flows, provide a reason to exist for hub systems. Models of hub location, in geography, operations research and transportation, that capture the scale economies that accrue due to the bundling of flows by a constant discount factor α do not adequately capture the scale of economies happening at the inter-hub link. Travel costs on the interhub links in these models are assumed to be independent of the amount of flow traveling across the link.

In most existing models, travel across the interhub links is discounted (relative to the cost of traveling on the spokes or non-interhub links) by an exogenously determined amount (the interhub discount factor, α) and the same discount is assumed to apply to all interhub links in the network regardless of the differences in the flows traveling across them. This is an oversimplification, however, because some interacting pairs enjoy benefits that are not earned or warranted [83].

The introduction of nonlinear term to account for economies of scale started by carried out by O'kelly and Bryan [83] on the use of a constant discount factor α to represent scale economies is an oversimplification of reality. This simplification permits optimal solutions with a rather small flow on inter-hub connections. In many real-life applications, the granted discount depends on the traffic flow on the inter-hub connections. They replaced the constant discount factor by a piece wise linear concave function, their model was named FLOWLOC, then allowing the amount of the discount to depend on the flow on the inter-hub connection.

Bryan [12] presented four extensions to the FLOWLOC model. The aim of his work was two-fold: (a) to investigate whether a change in network design could result in a more equitable distribution of interhub flows and (b) to further the capability of the model to represent actual hub-and-spoke networks. These models are more relevant for air transportation than telecommunication networks. He presented numerical examples for all the extensions. Moreover, Bryan presented the concept of a minimum threshold for the interhub links and the assumption of a completely interconnected hub network can be used to formulate a model that determines the optimal number of hubs to open in a network within a location/allocation framework. Finally, he formulated a model that incorporates a flow-dependent cost function for the spokes in addition to the one for the interhub links.

Klincewicz [70] described a specialized optimal enumeration procedure for the FLOWLOC model, as well as some search heuristics that are based upon Tabu search and Greedy Random Adaptive Search Procedures (GRASP). These require less computation time and can be applied to larger-sized problems. He showed that the FLOWLOC model can be sloved using using the classic Uncapacitated Facility Location problem. Racunica and Wynter [92] presented an optimization model that has been developed to address the problem of increasing the share of rail in inter-modal transport through the use of hub-and-spoke type networks for freight rail. The model defined is a generalization of the hub location problem in that it allows for non-linear and concave cost functions on different segments. A linearization procedure along with two efficient variable-reduction heuristics was developed for its resolution. Then, Camargo et al. [35] presented a tighter formulation for the uncapacitated hub location problem with multiple allocation and scale economies was presented. Their formulation outperformed the FLOWLOC model of OKelly and Bryan. They have also developed a Benders decomposition algorithm for these two versions.

2.7 Summary

The research described in the previous sections shows that the field of hub location is moving in exciting new directions with improved models that better capture important elements of transportation and logistics systems missing from the early fundamental hub location models.

Campbell and O'Kelly [20] stated that ”practical transportation hub networks would seem rarely to be larger than those solvable with the current state of-the-art; furthermore, faster solution methods may be less of a concern for strategic hub network design than for other large operational or tactical combinatorial optimization problems in logistics, such as crew scheduling and vehicle routing. More practical considerations will be favored. Models that integrate both cost and service may provide better insights into practical transportation networks”. Therefore, the aim of this thesis is to consider real life aspects for the HLP. Specifically, the consideration of Congestion and Economies Of Scale aspects on the design of hub-and-spoke networks.

2.7.1 Motivation

In a recent review conducted by Campbell and O'Kelly [20] they especially encouraged models that incorporate both more realistic transportation costs and service measures along with other relevant aspects. One of the realistic aspects in the HLP is the model of economies of scale as a nonlinear function depending on the amount of flow rather than the a constant discount factor, also the consideration of congestion to account for incoming flow directed to major hubs in the objective function of the HLP is another real-life aspect. However, to the best of our knowledge, no studies in the literature have tackled the economies of scale for the single allocation case nor the integration of congestion and economies of scale effects on the design of hub-and-spoke networks. Previous work considered the congestion effect or the economies of scale separately. This is because of the great complication resulting from the incorporation of those two effects in one single model, moreover for the single allocation case the allocation decision variables are binary taking eighther 0 or 1, unlike the multiple allocation version of HLP where the allocation decision variables are linear taking values from 0 to 1, which presents a more challenging task to be solved.

In addition of having a complicated model that considers the effect of nonlinear representation of the Economies Of Scale and congestion together, the previous work done by OKelly and Bryan [83] and Klincewicz [70], they found that interhub traffic flow tended to be concentrated on a few interhub arcs. That is, some interhub arcs had to carry relatively large traffic flows that were highly discounted, while others had very small amounts of traffic. As mentioned by Bryan [12] analysts should be aware of this potential imbalance in the interhub network (at least as witnessed in these CAB problems) when applying the model to real world networks. One way to mitigate the imbalance effect is to consider the congestion effect to account for the incoming flow on the hubs, which is the solution proposed by Elhedhli and Hu [42]. In our work we integrated the congestion and economies of scale in the same model, therefore, we provided more insights and justifications for the use of integrated model in terms of impact of the structure of the network as well as the routing of flows in the network.

2.7.2 Contribution

As shown in sections 2.5 and 2.6, few authors have addressed the congestion issue in the design of hub-and-spoke network model as nonlinear in the objective function. Similarly few authors addressed the use of nonlinear term to utilize the inter-hub connection economies of scale. None of the previous work considered both the congestion and nonlinear term for inter-hub connection economies of scale, as stated by Klincewicz [70], Elhedhli and Hu [42] and Camargo et al. [35] considering a hub-and-spoke network with both economies of scale captured by a concave cost function and congestion modeled by a convex cost function is a challenging issue although combination of both is needed due to the fact that congestion is responsible for degradation of economies of scale.

This thesis serves as a deeper investigation into hub-and-spoke design networks for the airline industry. With a formulation for hub-and-spoke network taking into account congestion effects and accounting for Economies Of Scale occurring on the interhub links. The resulting Mixed Integer Nonlinear Program model integrates congestion costs together with EOS on the interhub links and network transportation costs. A Lagrangean heuristic and GRASP algorithm are also proposed to generate reliable solutions. Chapter 3 details the mathematical model including the economies of scale and congestion effects. The Lagrangean and GRASP heuristics are discussed in Chapters 4 and 5, respectively. Finally Chapter 7 presents the results and test implementation.

CHAPTER 3

Problem Formulation

3.1 Introduction

This chapter presents a formulation for a Hub-and-Spoke System Design with Congestion and Economies Of Scale. First, the model for Single Allocation p-Median Hub Location Problem is presented in section 3.2. Then, the economies of scale and congestion costs are modeled and a linearization scheme is proposed for each of of them individually in sections 3.3 and 3.4. Section 3.5 presents the formulation for the Single Allocation-p-Median Hub Location Problem with congestion and economies of scale as Mixed Integer Programming (MIP) model and finally section 3.6 presents the summary of the chapter.

3.2 The Uncapacitated Single Allocation p-Median Hub Location Problem

Skorin et al. [99] developed a mixed binary linear formulations with tight linear programming relaxations. They proved that this approach to be very effective for the Uncapacitated Single Allocation p-Median Hub Location Problem (USA-p-MHLP).

The p-hub median problem can be described as follows: Suppose there are n nodes that should interact, and p of those should be designated as hubs. The objective is to facilitate interactions between nodes of the network via a set of hubs. The hubs are assumed to be fully interconnected, but the non-hub nodes can interact only via hubs. The hubs are uncapacitated, i.e., there is no restriction on the number of nodes allocated to a given hub. If there is no restriction on the number of hubs to which a non-hub node can be allocated, we have the multiple allocation version of the problem. A connectivity protocol in which each non-hub node is allocated to exactly one hub corresponds to the single allocation p-hub median problem.

Let us define a pair of origin-destination nodes (i, j) with a flow W_{ij} that should be sent from i to its final destination j, and it should pass through 2 hubs at maximum, namely hubs k and m. The model should determine the optimal set on nodes to be located as hubs and the optimal routes $i \to k \to m \to j$, where nodes $k\&m$ are hubs.

(USA-p-MHLP)

$$\min_{x,z} \quad \sum_i \sum_j \sum_k \sum_m F_{ijkm} x_{ijkm} \tag{3.1}$$

$$s.t.$$

$$\sum_k z_{ik} = 1, \ \forall \quad i \tag{3.2}$$

$$z_{ik} \leq z_{kk}, \ \forall \quad i, k \tag{3.3}$$

$$\sum_k z_{kk} = p, \tag{3.4}$$

$$\sum_m x_{ijkm} = z_{ik}, \ \forall \quad k, i, j \neq i \tag{3.5}$$

$$\sum_k x_{ijkm} = z_{jm}, \ \forall \quad m, i, j \neq i \tag{3.6}$$

$$z_{ik}, x_{ijkm}, \in \{0, 1\} \tag{3.7}$$

$$\tag{3.8}$$

The decision variables are:

1. x_{ijkm} a binary variable that equals 1 if the demand flow from node i to j is routed through path $i \to k \to m \to j$; equals 0 otherwise;

2. z_{ik} a binary variable that equals 1 if node i is assigned to hub k; equals 0 otherwise;

3. z_{kk}a binary variable that equals 1 if node k is selected as a hub; equals 0 otherwise. Note that z_{kk} is a special case of z_{ik} when $i = k$ (if a node k is assigned to itself as a hub, (i.e.

$z_{ik} = z_{kk} = 1$) this node must be selected as a hub).

The parameters are:

1. $F_{ijkm} = W_{ij}(\delta c_{ik} + \alpha c_{km} + \gamma c_{mj})$: the total transportation cost of route $i \rightarrow k \rightarrow m \rightarrow j$:

   (a) c_{ik} is the transportation cost between nodes i,k which is relative to the distance between the nodes.

   (b) c_{km} is the transportation cost between nodes k,m which is relative to the distance between the nodes.

   (c) c_{mj} is the transportation cost between nodes m,j which is relative to the distance between the nodes.

   (d) W_{ij} is the amount of flow to be sent from node i to node j.

   (e) $\delta, \alpha, \& \gamma$ are the discount factor for flow on the links for collection, inter-hub flow, and distribution amount, respectively, which are assumed to be constant through the network.

2. p is the number of hubs to be located in the network.

The model constraints are:

1. The objective function (3.1) minimizes the total cost of routing flows through one or two hubs.

2. Constraints (3.2) ensure that every node is allocated to exactly one hub.

3. Constraints (3.3) guarantee that a node will not be assigned to a hub unless that hub is opened.

4. Constraint (3.4) requires p hubs be opened.

5. Constraints (3.5) and (3.6) ensure that for any flow (i, j) that uses interhub link (k,m), nodes i and j must be assigned to hubs k and m, respectively.

Model USA-p-MHLP has been used broadly in the literature due to its strong linearity properties as shown in [99] and its structure which is appealing for the use of decomposition techniques, ([89], [42], [35], and [28]) to name few.

The objective function in the original formulation does not adequately model the economies of scale flow that accrue due to the agglomeration of flows. Travel costs on the inter-hub links in this model are assumed to be independent of the amount of flow traveling across the link. To deal with this deficiency, O'Kelly and Bryan [83], generalized the definition of the interhub cost term to be a concave increasing function. Travel across the inter-hub links is discounted (relative to the cost of traveling on the spokes or non-inter-hub links) by a determined amount (the inter-hub discount factor, α) and the same discount is assumed to apply to all interhub links in the network regardless of the differences in the flows traveling across them. This is an oversimplification, and a more realistic expression for hub network costs would allow costs to increase at a decreasing rate as flows increase [83]. Moreover, it does not account for congestion which occurs at the hubs due to the high flow of traffic being directed to the hubs.

3.3 Economies Of Scale

O'Kelly and Bryan [83] noted that hub location models that utilize a constant discount factor (α) represent something of an oversimplification. In some cases, traffic on a particular interhub link in the optimal solution is rather small. Therefore, if the model were used to study realistic communications or transportation networks, the model would apply a discount that would not be warranted. Thus, they proposed a FLOWLOC model where the constant discount factor on the inter-hub links is replaced by a piecewise-linear concave function. Thus, the amount of the discount will depend on the actual flow of traffic on the link.

The objective function (3.1) chooses the p hubs, and the routing of traffic for each directed pair (i, j), is to minimize total cost. There is assumed a cost c_{ik} per unit of traffic sent along the link between a location i and a hub k, similarly, a cost c_{jm} per unit of traffic sent along the link between a hub m and location j and due to symmetric $c_{jm} = c_{mj}$. Thus, for a directed pair (i, j), there is a cost $W_{ij}c_{ik}$ incurred on the link between i and k and a cost $W_{ij}c_{jm}$ incurred on the link between m and j. There is also a cost incurred on the inter-hub links. In the FLOWLOC model, the latter cost is a piece-wise-linear concave cost, which is a function of the total traffic utilizing each directed inter-hub link. Since x_{ijkm} equals 1 if traffic between directed pair (i, j) utilizes hubs (k,m) and 0 otherwise. The total traffic on the directed link from k to m is then $\sum_i \sum_j W_{ij}x_{ijkm}$. The concave cost function can be denoted as $f(V_{km})$where $V_{km} = \left(\sum_i \sum_j W_{ij}x_{ijkm}\right)$.

In the nonlinear cost function, then, costs are increasing at a decreasing rate as flows increase. The size of the discount given to flows traveling across the interhub link depends on the total interhub flow. As the interhub flow increases, the discount also increases. This discount is expressed as:

$$\alpha = \theta \left(\frac{\sum_i \sum_j W_{ij} x_{ijkm}}{\sum_i \sum_j W_{ij}} \right)^{\eta} \tag{3.9}$$

and ranges from 0 to 1, with 0 being the discount earned if no flow is traveling across the interhub link and 1 being the largest possible earned discount (the cost on the interhub link would be free). By varying the parameters, θ and η, different cost functions can be formulated.

Incorporating the economies of scale cost function in the objective of USA-p-MHLP, we obtain the uncapacitated single assignment hub location problem with economies of scale:

$$\min_{x,z} \quad \sum_{j \neq i} \sum_{k} \sum_{m} W_{ij}(c_{ik} + c_{mj}) x_{ijkm} + \sum_{k} \sum_{m} c_{km} (\sum_{i} \sum_{j \neq i} W_{ij} x_{ijkm})^{b} \tag{3.10}$$

$$s.t.$$

$$\sum_{k} z_{ik} = 1, \; \forall \quad i \tag{3.11}$$

$$z_{ik} \leq z_{kk}, \; \forall \quad i,k \tag{3.12}$$

$$\sum_{k} z_{kk} = p, \tag{3.13}$$

$$\sum_{m} x_{ijkm} = z_{ik}, \; \forall \quad k,i,j \neq i \tag{3.14}$$

$$\sum_{k} x_{ijkm} = z_{jm}, \; \forall \quad m,i,j \neq i \tag{3.15}$$

$$z_{ik}, x_{ijkm}, \in \{0,1\} \tag{3.16}$$

$$\tag{3.17}$$

Since this concave cost function $f(V_{km})$ on the inter-hub links is piece-wise linear,O'Kelly and Bryan [83] and Klincewicz [70] represented this function in a way that allows the problem to be solved using linear integer programming. Specifically, they viewed the piece-wise linear cost function as the "lower envelope" of a set of linear functions. Each of the linear functions in the set corresponds to one of the "piece-wise" lines. He used q to index the linear functions in the set; each such function can be described in terms of an intercept $f_{qkm}c_{km}$ and a slope

$a_{qkm}c_{km}$. For any given solution, the linear function that determines the value of the piecewise-linear concave cost function is the one that minimizes $\left(f_{qkm} + a_{qkm}\left(\sum_i \sum_j W_{ij}x_{ijkm}\right)\right)c_{km}$.

Figure 3.1 represents a piece-wise-linear concave cost function and the concave function.

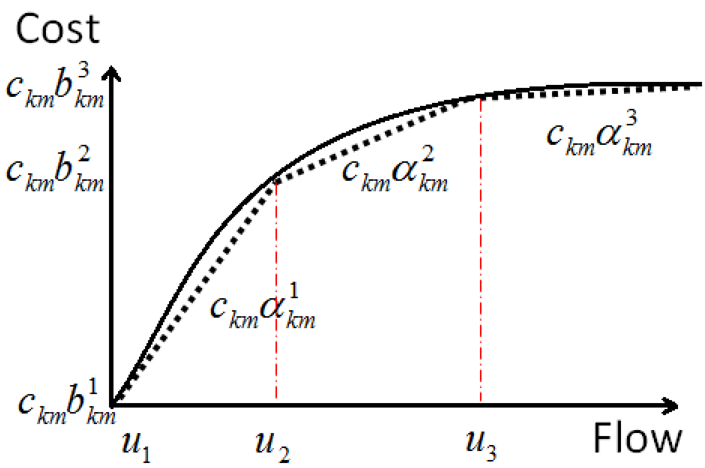

Figure 3.1: Piecewise linear concave cost function on interhub connection.

Following their linearization scheme and saving the structure of the single allocation case, we obtain the following as a mixed integer program:

$$\min_{x,y} \quad \sum_i \sum_j \sum_k \sum_m W_{ij}(c_{ik} + c_{mj})x_{ijkm} + \sum_q \sum_k \sum_m c_{km}(a_q r_{qkm} + f_q y_{qkm}) \tag{3.18}$$

$s.t.$

$$\sum_k z_{ik} = 1, \ \forall \quad i \tag{3.19}$$

$$z_{ik} \leq z_{kk}, \ \forall \quad i,k \tag{3.20}$$

$$\sum_k z_{kk} = p, \tag{3.21}$$

$$\sum_m x_{ijkm} = z_{ik}, \ \forall \quad k,i,j \neq i \tag{3.22}$$

$$\sum_k x_{ijkm} = z_{jm}, \ \forall \quad m,i,j \neq i \tag{3.23}$$

$$\sum_q r_{qkm} = \sum_i \sum_j W_{ij}x_{ijkm}, \ \forall k,m : k \neq m \tag{3.24}$$

$$r_{qkm} - y_{qkm} \sum_i \sum_j W_{ij} \leq 0, \ \forall q,k,m : k \neq m \tag{3.25}$$

$$\sum_q y_{qkm} = x_{ijkm}, \ \forall k,m : k \neq m \tag{3.26}$$

$$x_{ijkm}, y_{qkm}, z_{ik}, \in \{0,1\} \tag{3.27}$$

Where y_{qkm} = 1 if the flow on interhub link (k,m) will use cost parameters f_q and a_q, 0 otherwise;r_{qkm} the total flow of interhub link (k,m) to which a_q will be applied.

Constraint (3.24) determines the amount of flow on the interhub link (k,m). Constraint (3.25) ensures that this interhub flow, and the associated slope a_q, is matched with the correct intercept f_q. Constraint (3.26) ensures that, for every interhub link (k,m), exactly one y_{qkm} is equal to one, i.e., exactly one piece of the piecewise linear function is used. A great improvement can be implemented based on observing that any interhub flow can be presented as a convex combination between the two points where the value of flow lies. Based on this, we can model the above formulation based on SOS Type 2 linearization, the commercial solvers have a special algorithms to solve SOS type problems faster than the above formulation. Therefore, we will propose a linearization scheme based on SOS Type 2 technique as follows:

$$\min \quad \sum_i \sum_j \sum_k \sum_m W_{ij}(c_{ik}+c_{mj})x_{ijkm} + \sum_k \sum_m c_{km}h_{km}(w_{km}) \tag{3.28}$$

$$s.t.$$

$$\sum_k z_{ik} = 1, \ \forall \quad i \tag{3.29}$$

$$z_{ik} \leq z_{kk}, \ \forall \quad i,k \tag{3.30}$$

$$\sum_k z_{kk} = p, \tag{3.31}$$

$$\sum_m x_{ijkm} = z_{ik}, \ \forall \quad k,i,j \neq i \tag{3.32}$$

$$\sum_k x_{ijkm} = z_{jm}, \ \forall \quad m,i,j \neq i \tag{3.33}$$

$$\sum_k \sum_m x_{ijkm} = 1, \ \forall i,j : i \neq j \tag{3.34}$$

$$w_{km} = \sum_l \omega^l \lambda_{km}^l, \ \forall k,m \tag{3.35}$$

$$h_{km}(w_{km}) = \sum_l h_{km}(\omega^l)\lambda_{km}^l, \ \forall k,m \tag{3.36}$$

$$\sum_l \lambda_{km}^l = 1, \ \forall k,m \tag{3.37}$$

$$w_{km} = \sum_i \sum_{j \neq i} W_{ij}x_{ijkm}, \ \forall k,m \tag{3.38}$$

$$\lambda_{km}^l \in \quad SOS2, \ \forall k,m,l \tag{3.39}$$

$$x_{ijkm}, z_{ik} \in \{0,1\}, \tag{3.40}$$

Where w_{km} is the amount of flow carried by interhub link (k,m), $h(w_{km})$ is the corresponding value of the flow based on the economies of scale function, and ω is the point of discontinuity.

3.4 The congestion cost function

The objective function 3.1 only minimizes the transportation cost of traveling through the network while satisfying demand flows. As observed by Elhedli and Wu [43] ”One of the major characteristics of USApHLP it tends to overuse cheaper hubs while leaving other hubs rarely exploited when a large volume of demand and flows are directed into a hub”. Grove and O'Kelly [57] stated that when more flows are directed into a hub, congestion increases accordingly and causes a dramatic increase in costs. Elhedhli and Hu [42] proposed a power-law cost function to depict the relationship between flow and congestion at hubs of the form:

$f(v) = \sigma v^{\tau}$,

where v is the flow at a hub; σ and τ are positive constants with $\tau \geq 1$.

Thus, using the notation in USA-p-HLP , $\sum_i \sum_{j \neq i} W_{ij} x_{ijkm}$,the flow through hub k is:

$$\sigma \left(\sum_i \sum_{j \neq i} \sum_m W_{ij} x_{ijkm}\right)^{\tau} = \sigma \left(\sum_i \sum_{j \neq i} W_{ij} z_{ik}\right)^{\tau}$$

And based on Constraint (3.5) the last equality can be written as:

$$\sum_i \sum_{j \neq i} \sum_m W_{ij} x_{ijkm} = \sum_i \sum_{j \neq i} W_{ij} \sum_m x_{ijkm} = \sum_i \sum_{j \neq i} W_{ij} z_{ik} \quad \forall \quad k$$

Figure 3.2 plot the congestion cost functions for different values of the parameters σ and τ.

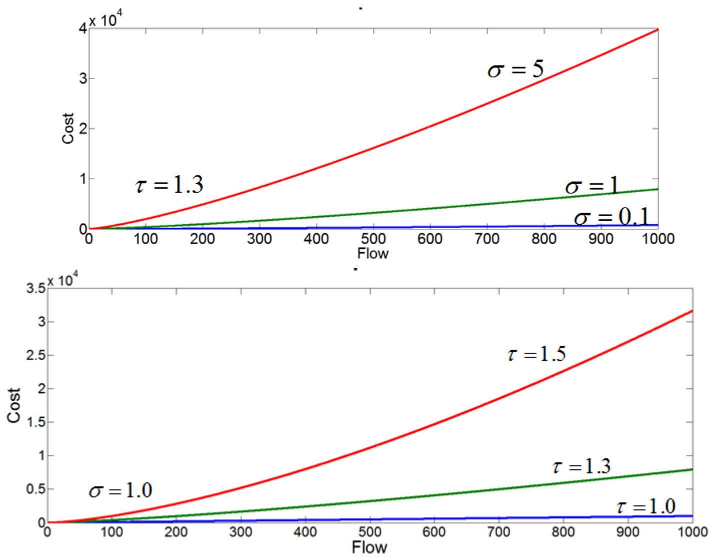

Figure 3.2: The congestion cost functions for different values of σ and τ.

Incorporating the congestion cost function in the objective of USA-p-HLP, results in the uncapacitated single assignment hub location problem with congestion

$$\min_{x,z} \quad \sum_i \sum_{j \neq i} \sum_k \sum_m F_{ijkm} x_{ijkm} + \sum_k \sigma \left(\sum_i \sum_{j \neq i} W_{ij} z_{ik} \right)^{\tau} \tag{3.41}$$

$s.t.$

$$\sum_k z_{ik} = 1, \ \forall \quad i \tag{3.42}$$

$$z_{ik} \leq z_{kk}, \ \forall \quad i,k \tag{3.43}$$

$$\sum_k z_{kk} = p, \tag{3.44}$$

$$\sum_m x_{ijkm} = z_{ik}, \ \forall \quad k,i,j \neq i \tag{3.45}$$

$$\sum_k x_{ijkm} = z_{jm}, \ \forall \quad m,i,j \neq i \tag{3.46}$$

$$z_{ik}, x_{ijkm}, \in \{0,1\} \tag{3.47}$$

$$\tag{3.48}$$

The USApHLPC is a non-linear mixed integer problem that might be difficult to solve directly. In what follows we will propose a linearization as shown in Elhedhli and Hu in [42].

Being convex, the congestion cost function can be approximated by the maximum of a set of piece-wise linear and tangent hyper planes. The approximation allows us to formulate the problem as a linear mixed integer program, but with an infinite number of constraints. For ease of exposition, let v_k denote the flow at hub k. At a given flow v_k^h , the tangent to $f(v_k)$ is given by: $f(v_k^h) + f^{'}(v_k^h)(v_k - v_k^h) = \sigma(1-\tau)(v_k^h)^{\tau} + \sigma\tau(v_k^h)^{\tau-1} v_k$

As the upper envelope of the tangent lines follows the shape of the function, the congestion cost function $f(v_k)$ can be linearized as: $f(v_k) = \max_{h \in H_k} \{\sigma(1-\tau)(v_k^h)^{\tau} + \sigma\tau(v_k^h)^{\tau-1} v_k\}$

where $v_k^h, h \in H_k$, is an infinite set of points. The congestion function and its approximation are shown in figure 3.3.

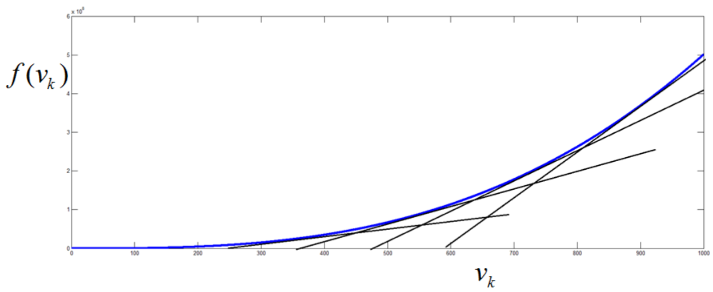

Figure 3.3: The congestion cost function and its approximation.

Replacing $f(v_k)$ by its value, the hub-and-spoke network model with congestion can be written as:

$$\min \sum_i \sum_{j \neq i} \sum_k \sum_m F_{ijkm} x_{ijkm} + \sum_k \max_{h \in H} \{ \sigma (1-\tau) \left(\sum_i \sum_{j \neq i} W_{ij} z_{ik}^h \right)^{\tau} + \sigma\tau \left(\sum_i \sum_{j \neq i} W_{ij} z_{ik}^h \right)^{\tau-1} \sum_i \sum_{j \neq i} W_{ij} z_{ik} \} \tag{3.49}$$

$s.t.$

$$\sum_k z_{ik} = 1, \; \forall \quad i \tag{3.50}$$

$$z_{ik} \leq z_{kk}, \; \forall \quad i,k \tag{3.51}$$

$$\sum_k z_{kk} = p, \tag{3.52}$$

$$\sum_m x_{ijkm} = z_{ik}, \; \forall \quad k,i,j \neq i \tag{3.53}$$

$$\sum_k x_{ijkm} = z_{jm}, \; \forall \quad m,i,j \neq i \tag{3.54}$$

$$\Omega_k - \tau \left(\sum_i \sum_{j \neq i} w_{ij} z_{ik}^h \right)^{\tau-1} \left(\sum_i \sum_{j \neq i} w_{ij} z_{ik} \right) \geq (1-\tau) \left(\sum_i \sum_{j \neq i} w_{ij} z_{ik} \right), \; \forall h \in H_k \tag{3.55}$$

$$z_{ik}, x_{ijkm}, \in \{0,1\} \tag{3.56}$$

$$\Omega_k \geq 0, \quad \forall \quad k, \tag{3.57}$$

Which is equivalent to:

$$\min_{x,z} \quad \sum_i \sum_{j \neq i} \sum_k \sum_m F_{ijkm} x_{ijkm} + \sigma \sum_k \Omega_k \tag{3.58}$$

$s.t.$

$$\sum_k z_{ik} = 1, \; \forall \quad i \tag{3.59}$$

$$z_{ik} \leq z_{kk}, \; \forall \quad i,k \tag{3.60}$$

$$\sum_k z_{kk} = p, \tag{3.61}$$

$$\sum_m x_{ijkm} = z_{ik}, \; \forall \quad k,i,j \neq i \tag{3.62}$$

$$\sum_k x_{ijkm} = z_{jm}, \; \forall \quad m,i,j \neq i \tag{3.63}$$

$$\Omega_k - \tau \left(\sum_i \sum_{j \neq i} w_{ij} z_{ik}^h \right)^{\tau-1} \left(\sum_i \sum_{j \neq i} w_{ij} z_{ik} \right) \geq (1-\tau) \left(\sum_i \sum_{j \neq i} w_{ij} z_{ik} \right), \; \forall h \in H_k \tag{3.64}$$

$$z_{ik}, x_{ijkm}, \in \{0,1\} \tag{3.65}$$

$$\Omega_k \geq 0, \quad \forall \quad k, \tag{3.66}$$

The later is a linear mixed integer program that has an infinite number of constraints.

3.5 A hub-and-spoke network model with congestion and economies of scale

Incorporating the congestion cost and economies of scale function in the objective of USA-p-MHLP, we obtain the uncapacitated single assignment hub location problem with congestion and economies of scale: C-EOS-USA-p-MHLP:

$$\min \quad \sum_k \sigma(\Omega_k)^\tau + \sum_i \sum_{j\neq i} \sum_k \sum_m W_{ij}(c_{ik}+c_{mj})x_{ijkm} + \sum_k \sum_m c_{km} h_{km}(w_{km}) \tag{3.67}$$

$s.t.$

$$z_{ik} \leq z_{kk}, \ \forall i,k \tag{3.68}$$

$$\sum_k z_{kk} = p \tag{3.69}$$

$$\sum_m x_{ijkm} = z_{ik}, \ \forall i,j,k : i \neq j \tag{3.70}$$

$$\sum_k x_{ijkm} = z_{jm}, \ \forall i,j,m : i \neq j \tag{3.71}$$

$$\sum_k \sum_m x_{ijkm} = 1, \ \forall i,j : i \neq j \tag{3.72}$$

$$w_{km} = \sum_l \omega^l \lambda^l_{km}, \ \forall k,m \tag{3.73}$$

$$h_{km}(w_{km}) = \sum_l h_{km}(\omega^l)\lambda^l_{km}, \ \forall k,m \tag{3.74}$$

$$\sum_l \lambda^l_{km} = 1, \ \forall k,m \tag{3.75}$$

$$w_{km} = \sum_i \sum_{j>i} W_{ij} x_{ijkm}, \ \forall k,m \tag{3.76}$$

$$\Omega_k - \tau\left(\sum_i \sum_{j\neq i} w_{ij} z^h_{ik}\right)^{\tau-1} \left(\sum_i \sum_{j\neq i} w_{ij} z_{ik}\right) \geq (1-\tau)\left(\sum_i \sum_{j\neq i} w_{ij} z_{ik}\right), \ \forall h \in H_k \tag{3.77}$$

$$\lambda^l_{km} \in \quad SOS2, \ \forall k,m,l \tag{3.78}$$

$$x_{ijkm}, z_{ik} \in \{0,1\}, \tag{3.79}$$

In the C-EOS-USA-pMHLP the parameters $\sigma \& \tau$ control the effect of convex function for the congestion contribution to the objective function and parameters $a \& b$ control effect of non-linearity in the economies of scale to discount the flow on the interhub links. The C-EOS-USA-p-MHLP model is a nonlinear mixed integer programming model that is difficult to be solved directly, in what follows I propose a Lagrangean heuristic to solve the model as well as a GRASP meta-heuristic to find good solutions for the model.

3.6 Summary

The previous chapter presented the mathematical model for the Congestion and Economies Of Scale Uncapacitated Single Allocation p- Median Hub Location Problem (C-EOS-USA-p-MHLP). The derivation of the nonlinear cost function for the congestion effect was presented along with a linearization scheme inspired by the previous work done by Elhedhli and Wu [43] Elhedhli and Hu [42].Furthermore the derivation for a nonlinear cost function for a more realistic representation for the EOS was proposed in addition a linearization scheme was presented to tackle the difficulty of the nonlinear model. Finally, the integrated model was presented as a MIP model.

CHAPTER 4

METHODOLOGY: LAGRANGEAN RELAXATION

”Lagrangean relaxation is a tool that has been increasingly implemented in mathematical programming applications as it is proven an efficient method for solving large-scale integer programming problems. It is based upon the observation that a set of complicated constraints within a formulation can be removed and replaced with a penalty term in the objective function, involving the amount of violation of these constraints and their dual variables. The Lagrangean sub-problem is easier to solve and it provides a lower bound (for a minimization problem) on the optimal value of the original problem” Fisher [53].

For minimization problems, the objective achieved by the relaxation may have a lower value than the objective in the original problem, and this is why the objective obtained through relaxation constitutes a lower bound to the problem. When the solution generated by solving the relaxed version of the original problem we might get an infeasible solution; this is because the relaxed version of the problem does have constraints that are found in the original problem. In order to modify the solution, a feasible solution is generated at each iteration in during the Lagrangean heuristics via certain procedure to fix the infeasibility issue and the value of the solution generated is called the Upper Bound (UB). This heuristic is usually called a Lagrangean heuristic and the objective achieved through the fixed solution is called an upper bound for the objective value of the original problem, whereas the relaxed Lagrangean objective constitutes a lower bound, as previously mentioned earlier.

4.1 Lagrangean Relaxation for C-EOS-USA-p-MHLP

$$\min \quad \sum_k \sigma(\Omega_k)^\tau + \sum_{j\neq i}\sum_k\sum_m W_{ij}(c_{ik}+c_{mj})x_{ijkm} + \sum_k\sum_m c_{km}h_{km}(w_{km}) \tag{4.1}$$

$s.t.$

$$\sum_k z_{ik} = 1,\ \forall i \leftarrow \mu_i \tag{4.2}$$

$$z_{ik} \leq z_{kk},\ \forall i,k \tag{4.3}$$

$$\sum_k z_{kk} = p \tag{4.4}$$

$$\sum_m x_{ijkm} = z_{ik},\ \forall i,j,k : i \neq j \leftarrow \beta_{kij} \tag{4.5}$$

$$\sum_k x_{ijkm} = z_{jm},\ \forall i,j,m : i \neq j \leftarrow \gamma_{mij} \tag{4.6}$$

$$\sum_k\sum_m x_{ijkm} = 1,\ \forall i,j : i \neq j \tag{4.7}$$

$$w_{km} = \sum_l \omega^l \lambda^l_{km},\ \forall k,m \tag{4.8}$$

$$h_{km}(w_{km}) = \sum_l h_{km}(\omega^l)\lambda^l_{km},\ \forall k,m \tag{4.9}$$

$$\sum_l \lambda^l_{km} = 1,\ \forall k,m \tag{4.10}$$

$$w_{km} = \sum_i\sum_{j>i} W_{ij}x_{ijkm},\ \forall k,m \tag{4.11}$$

$$\Omega_k - \tau\left(\sum_i\sum_{j\neq i} w_{ij}z^h_{ik}\right)^{\tau-1}\left(\sum_i\sum_{j\neq i} w_{ij}z_{ik}\right) \geq (1-\tau)\left(\sum_i\sum_{j\neq i} w_{ij}z_{ik}\right),\ \forall h \in H_k \tag{4.12}$$

$$\lambda^l_{km} \in \quad SOS2,\ \forall k,m,l \tag{4.13}$$

$$x_{ijkm}, z_{ik} \in \{0,1\}, \tag{4.14}$$

The mathematical model represents the single allocation HLP with economies of scale and congestion. The first term in the objective function minimizes the cost of flow associated with congestion due to the allocation of non-hub node (i) to the hub node (k), second term minimizes the total transportation cost from the origin (i) to the first hub node (k) and from the second hub node (m) to the destination (j) and the third term in the objective function minimizes the total flow on the interhub link (k,m) where the parameters $a\&b$ represent different economies of scale functions with $a > 0$ and $0 < b < 1.0$.

Besides being nonlinear, C-EOS-USA-p-MHLP is typically large in size, as we have $n^4 + n^2$ binary variables, $n^2 * L$ SOS Type II variables, and n continues variables and $2n^3 + 6n^2 + n$ constraints in addition to constraint set 4.12 which is infinite, where n is the number of nodes in the network. Hence, we consider decomposing it into a number of smaller problems. For the USA-p-MHLP, Pirkul and Schilling [89] developed an efficient Lagrangean heuristic which manages to decompose USA-p-MHLP into two smaller subproblems by targeting the appropriate constraints to be relaxed. Since the structure of C-EOS-USA-p-MHLP is similar to the USA-p-MHLP, we start by relaxing the same constraints as in Pirkul and Schilling. Specifically, constraints 4.2, 4.5, and 4.6 with μ, β, and γ, respectively as a Lagrangean multipliers, respectively. This results in three new terms to be added to the objective function of C-EOS-USA-p-MHLP:

1 $\sum_i [\mu_i(\sum_k z_{ik} - 1)]$;

2 $\sum_i \sum_{j:j\neq i} \sum_k \left[\beta_{ijk}(\sum_m x_{ijkm} - z_{ik})\right]$;

3 $\sum_i \sum_{j:j\neq i} \sum_m \left[\gamma_{ijm}(\sum_k x_{ijkm} - z_{jm})\right]$.

All the three relaxed constraints are equality, so their Lagrangean multipliers are unrestricted in sign (can be either positive or negative). Adding the three terms to the objective of C-EOS-USA-p-MHLP, we get:

$$\begin{aligned} Z_R(\mu,\beta,\gamma) = \min \quad & \sigma \sum_k \left(\sum_i \sum_{j\neq i} W_{ij} z_{ik} \right)^{\tau} + \sum_i \sum_k \overline{C}_{ik} Z_{ik} + \\ & \sum_i \sum_j \sum_k \sum_m \overline{F}_{ijkm} x_{ijkm} - \sum_i \mu_i + \sum_k \sum_m c_{km} h_{km}(w_{km}) \end{aligned}$$

$s.t.$

$$z_{ik} \leq z_{kk},\ \forall i,k$$

$$\sum_k z_{kk} = p$$

$$\sum_k \sum_m x_{ijkm} = 1,\ \forall i,j : i \neq j$$

$$w_{km} = \sum_l \omega^l \lambda^l_{km},\ \forall k,m$$

$$h_{km}(w_{km}) = \sum_l h_{km}(\omega^l)\lambda^l_{km},\ \forall k,m$$

$$\sum_l \lambda^l_{km} = 1,\ \forall k,m$$

$$w_{km} = \sum_i \sum_{j>i} W_{ij} x_{ijkm},\ \forall k,m$$

$$\Omega_k - \tau\left(\sum_i \sum_{j\neq i} w_{ij} z^h_{ik}\right)^{\tau-1}\left(\sum_i \sum_{j\neq i} w_{ij} z_{ik}\right) \geq (1-\tau)\left(\sum_i \sum_{j\neq i} w_{ij} z_{ik}\right),\ \forall h \in H_k$$

$$\lambda^l_{km} \in \quad SOS2,\ \forall k,m,l$$

$$x_{ijkm}, z_{ik} \in \{0,1\},$$

Where $\overline{C}_{ik} = \mu_i - \sum_j \beta_{kij} - \sum_j \gamma_{kji}$ and $\overline{F}_{ijkm} = W_{ij}(c_{ik} + c_{mj}) + \beta_{kij} + \gamma_{mij}$.

This problem can be separated into two sub-problems without any loss of integrity as SUB1 with the domain of z variables and SUB2 with the domain of x variables.

SUB1:

$$\min \quad \sum_i \sum_k \overline{C}_{ik} z_{ik} + \sigma \sum_k \left(\sum_i \sum_{j\neq i} W_{ij} z_{ik}\right)^{\tau}$$

$s.t.$

$$z_{ik} \leq z_{kk},\ \forall i,k$$

$$\sum_k z_{kk} = p$$

$$z_{ik} \in \{0,1\},$$

Where in SUB1 the locations and allocations of hubs and nonhub nodes to hubs are found, respectively.

And SUB2:

$$\min \quad \sum_i \sum_j \sum_k \sum_m \overline{F}_{ijkm} x_{ijkm} \sum_i \sum_j \sum_k \sum_m W_{ij}(c_{ik} + c_{mj}) x_{ijkm} + \sum_k \sum_m c_{km} h_{km}(w_{km}) \quad (4.15)$$

$s.t.$

$$\sum_k \sum_m x_{ijkm} = 1, \; \forall i, j : i \neq j \quad (4.16)$$

$$w_{km} = \sum_l \omega^l \lambda^l_{km}, \; \forall k, m \quad (4.17)$$

$$h_{km}(w_{km}) = \sum_l h_{km}(\omega^l) \lambda^l_{km}, \; \forall k, m \quad (4.18)$$

$$\sum_l \lambda^l_{km} = 1, \; \forall k, m \quad (4.19)$$

$$w_{km} = \sum_i \sum_{j>i} W_{ij} x_{ijkm}, \; \forall k, m \quad (4.20)$$

$$\lambda^l_{km} \in \quad SOS2, \; \forall k, m, l \quad (4.21)$$

$$x_{ijkm} \in \{0, 1\}, \quad (4.22)$$

4.1.1 Solving SUB1

SUB1:

$$\min \quad \sum_i \sum_k \bar{C}_{ik} z_{ik} + \sigma \Omega_k$$

$s.t.$

$$\Omega_k - \tau \left(\sum_i \sum_{j:j \neq i} W_{ij} z^h_{ik} \right)^{\tau - 1} \left(\sum_i \sum_{j:j \neq i} W_{ij} z_{ik} \right) \geq (1 - \tau) \left(\sum_i \sum_{j:j \neq i} W_{ij} z^h_{ik} \right)^{\tau}, \; \forall h \in H_k \quad (4.23)$$

$$z_{ik} \leq z_{kk}, \; \forall i, k \quad (4.24)$$

$$\sum_k z_{kk} = p, \quad (4.25)$$

$$z_{ik} \in \{0, 1\}, \quad (4.26)$$

With the absence of constraint 4.25, SUB1 separates into a set of smaller problems one for each potential hub k.

SUB1-k:

$$\min \quad \sum_i \bar{C}_{ik} z_{ik} + \sigma \Omega_k$$

$$s.t.$$

$$\Omega_k - \tau \left(\sum_i \sum_{j:j \neq i} W_{ij} z_{ik}^h \right)^{\tau - 1} \left(\sum_i \sum_{j:j \neq i} W_{ij} z_{ik} \right) \geq (1-\tau) \left(\sum_i \sum_{j:j \neq i} W_{ij} z_{ik}^h \right)^{\tau}, \ \forall h \in H_k$$

$$z_{ik} \leq z_{kk}, \ \forall i, k$$

$$z_{ik} \in \{0, 1\},$$

As SUB1-k has an infinite number of constraints, its solution is done iteratively using a cutting plane approach, similar to the work done by Elhedhli Hu [42] and where an initial set of constraints is started with and the rest is added as needed. This is summarized in Algorithm 1, phase 1.

”Note that the procedure sets z_{kk} to 1 and finds the corresponding optimal solution. If the obtained solution has a negative objective, then it is the optimal solution to [SUB1-k], otherwise the optimal to [SUB1-k] is zero. Having solved problems [SUB1-k] for all k, the solution to [SUB1] amounts to picking the solutions corresponding to the problems with the p smallest objective values”. This is summarized in Algorithm 1, phase 2.

Algorithm 1 SUB1-Solution

Phase 1: Solution of SUB1-k

Start with an initial set of $\bar{H}_k \subseteq H_k$;
Solve SUB1-k and get its optimal solution $\bar{z}_{ik}$ and its optimal objective value $\bar{v}_k$;
Set the Lower Bound $LB_k = \bar{v}_k$;
Plug $\bar{z}_{ik}$ into the objective function $\sum_i \bar{C}_{ik} z_{ik} + \sigma \left(\sum_i \sum_{j:j \neq i} W_{ij} z_{ik} \right)^{\tau}$ to obtain an upper bound UB_k;
if $LB_k = UB_k$ **then**
  if $\bar{v}_k > 0$ **then**
    $\bar{v}_k = 0$;
    $z_{ik} = 0 \quad \forall \quad i$;
    Stop.
  end if
else
  Add one more constraint corresponding to $\bar{z}_{ik}$ to SUB1-k:
  $\Omega_k - \tau \left(\sum_i \sum_{j:j \neq i} W_{ij} z_{ik}^h \right)^{\tau-1} \left(\sum_i \sum_{j:j \neq i} W_{ij} z_{ik} \right) \geq (1-\tau) \left(\sum_i \sum_{j:j \neq i} W_{ij} z_{ik}^h \right)^{\tau}.$
end if

Phase 2: Solution of SUB1

Sort $\bar{v}_k$ in increasing order;
For the first p hubs with the smallest $\bar{v}_k$ set $\bar{z}_{kk} = 1$ and keep the values of the corresponding $\bar{z}_{ik} = 1 \quad \forall \quad i$;
For the rest, set $\bar{z}_{kk} = 0$ and the corresponding $\bar{z}_{ik} = 0 \quad \forall \quad i$

4.1.2 Solving SUB2

SUB2:

$$\min \quad \sum_i \sum_j \sum_k \sum_m \overline{F}_{ijkm} x_{ijkm} + \sum_k \sum_m c_{km} h_{km}(w_{km}) \tag{4.27}$$

$s.t.$

$$\sum_k \sum_m x_{ijkm} = 1, \ \forall i, j : i \neq j \tag{4.28}$$

$$w_{km} = \sum_l \omega^l \lambda_{km}^l, \ \forall k, m \tag{4.29}$$

$$h_{km}(w_{km}) = \sum_l h_{km}(\omega^l) \lambda_{km}^l, \ \forall k, m \tag{4.30}$$

$$\sum_l \lambda_{km}^l = 1, \ \forall k, m \tag{4.31}$$

$$w_{km} = \sum_i \sum_{j>i} W_{ij} x_{ijkm}, \ \forall k, m \tag{4.32}$$

$$\lambda_{km}^l \in \quad SOS2, \ \forall k, m, l \tag{4.33}$$

$$x_{ijkm} \in \{0, 1\}, \tag{4.34}$$

The solution for SUB2 selects the optimal routes, the optimal hub connections x_{kmkm}, and the optimal amount of flow to use the selected interhub (k,m). This can be done using a commercial solver CPLEX.

Unfortunately, the initial results with this relaxation are disappointing and discouraging, we obtained poor quality Lagrangean bounds with gaps ranging from 15% up to 30%. Of course the low quality solutions in the Lagrangean relaxation context results from either low quality lower bounds or upper bounds. We found that SUB2 solution can be improved by the addition of good cuts as valid equalities.

After observing the solution of the HLP problem for many instances we found that for fully connected hub network the number of connections between hubs will always be $p*(p-1)$ connections. This desirable property can be utilized to provide a valid cut for the HLP formulation, and this cut can be written as:

$$\sum_{k:k\neq m}\sum_{m:m\neq k} x_{kmkm} = p*(p-1) \tag{4.35}$$

This new cut is a redundant for the USA-p-MHLP formulation, however, it does improve the solution for the decomposed formulation. After the addition of this cut to SUB2, the quality bounds improved. Also, knowing that the connection between the hubs is two way connection we add further add the following cut:

$$x_{kmkm} = x_{mkmk} \quad \forall \quad k,m : k\neq m \tag{4.36}$$

The addition of those two cuts improved the solution quality for the Lagrangean relaxation and thus they will be included in the solution of SUB2. Moreover, the addition of those two cuts is not expensive when it comes to the computational time of SUB2 and does not complicate the structure of SUB2.

4.2 Generating a feasible solution For C-EOS-USA-p-MHLP

The solution we get from solving those two subproblems $\bar{x}_{ijkm},\bar{z}_{ik}$ might not be feasible to the original problem C-EOS-USA-p-MHLP. As mentioned earlier constraints 4.2, 4.5, and 4.6 are relaxed, constraints 4.2 assure that each nonhub node is assigned to exactly 1 hub, having this constraint relaxed we might get a solution from SUB1 where a node is not assigned to any hub or assigned to more than one hub. Moreover, due to relaxing 4.5 and 4.6, a flow from (i, j) might be routed through nonhub nodes.

After solving SUB1 and SUB2 and obtaining $\bar{x}_{ijkm}, \bar{z}_{ik}$, we check whether each non-hub node i is assigned to one and only one hub. If yes, we leave its hub assignment $\{z_{ik}\}$ as it is; otherwise, we set all $\{z_{ik}\}$ to zero, and then select a hub n with lowest unit transportation cost among all hubs and set z_{ik} to 1, based on the $\bar{C}_{ik}$ cost. Any violation of constraint 4.2 can be eliminated with this strategy. After that, we simply set $x_{ijkm} = z_{ik} * z_{jm}$ for any pair of origin-destination pair $(i, j \neq i)$, thus constraints 4.5, and 4.6 can be satisfied, too.

Since x_{ijkm} is obtained from the values of $z_{ik} \quad \& \quad z_{jm}$ we need to find the values of h^l_{km}, therefore we implement a simple procedure to find the value of flow on each interhub link. Once we know the value of x_{ijkm} we define the set of nodes who are hubs and the links among them, then for each interhub link we calculate the total amount of flow used that link by identifying the interhub link used by each flow commodity W_{ij}. After finding a feasible solution for $z_{ik}, x_{ijkm} \quad \& \quad h^l_{km}$, we calculate the objective function of C-EOS-USA-p-MHLP model and obtain an UB.
The procedure is shown in algorithm2 and finds the upper bound as well.

Algorithm 2 Generating a feasible solution

Phase 1: Find feasible z_{ik}, x_{ijkm}

1: **for all** z_{ik} **do**
2: **if** $\sum_k \bar{z}_{ik} > 1$ **then**
3: Set $\bar{z}_{ik} = 0 \quad \forall k$
4: Find $C_{in} = \min_k\{C_{ik} \mid z_{kk} = 1\}$ (i.e. assign the node to the cheapest hub found in SUB1).
5: Set $z_{in} = 1$
6: **end if**
7: **if** $\sum_k \bar{z}_{ik} = 0$ **then**
8: Find $C_{in} = \min_k\{C_{ik} \mid z_{kk} = 1\}$ (i.e. assign the node to the cheapest hub found in SUB1).
9: Set $z_{in} = 1$.
10: **end if**
11: **end for**
12: For all $k, m, i, j : i \neq j$ set $x_{ijkm} = z_{ik} \times z_{jm}$

Phase 2: Calculate $h^l_{km}(w^l_{km})$

13: Given the feasible x_{ijkm} solution as found in Line **12**, find the k, m (hub nodes) associated with each origin-destination pair.
14: For each unique hub pairs k, m found in Line **13** sum the total flow flowing on this pair from all origin-destination pairs utilizing this hub pair.
15: According to total flow for each hub pair k, m find the correct slop and interception of the function (i.e. α^l_{km}, b^l_{km}).

Phase 3: Calculate the Upper Bound *(UB)*

16: Given the $x_{ijkm}, z_{ik}, \alpha^l_{km}, b^l_{km}, \& h^l_{km}$ found above.
17: $UB = \sum_k \sigma \left(\sum_i \sum_{j\neq i} W_{ij} z_{ik}\right)^{\tau} + \sum_i \sum_{j\neq i} \sum_k \sum_m W_{ij}(c_{ik} + c_{jm}) x_{ijkm} + \sum_l \sum_k \sum_m c_{km}(\alpha^l_{km} w^l_{km} + b^l_{km} z^l_{km})$

4.3 Updating the Lagrangean multipliers

Being a Lagrangean relaxation, the objective value of C-EOS-USA-p-MHLP-LR, denoted as $Z_R(\mu^*, \beta^*, \gamma^*)$, provides a lower bound to C-EOS-USA-p-MHLP. The best lower bound is denoted as:

$$Z_R(\mu^*, \beta^*, \gamma^*) = \max_{\mu,\beta,\gamma} Z_R(\mu, \beta, \gamma)$$

We used the subgradient algorithm to find a good set of Lagrangean multipliers $(\mu^*, \beta^*, \gamma^*)$. With an upper bound Z_b obtained from a feasible solution of C-EOS-USA-p-MHLP and a lower bound Z_L from C-EOS-USA-p-MHLP-LR, the subgradient algorithm adjusts the multipliers (μ, β, γ) according to the gap between Z_b and Z_L. Several ways can be used to update the

Lagrangean multipliers at each iteration, such as Pure Sub-gradient, Deflected Sub-gradient, Conditional Sub-gradient, Bundle Method, or a hybrid technique. We first applied the Pure Sub-gradient to adjust the Lagrangean multipliers but we found that the behavior of the Lagrangean relaxation exhibit ***zigzaging of kind I***, which occurs if at any two (or more) consecutive iterate points the Lagrangean multipliers, say u^{iter}, u^{iter+1}, the angle between corresponding step directions d^{iter} and d^{iter+1} is obtuse; i.e. $d^{iter}d^{iter+1} < 0$.

"Such zigzagging phenomena that might manifest itself at any stage of the subgradient procedure slow down the search process. In order to avoid such an unpleasant behavior one may need to deflect the subgradient direction whenever it forms an obtuse angle with the previous stepping direction". To this end, in order to form a smaller angle between the current stepping direction and the preceding direction than the traditional (pure) subgradient direction does, and hence to enhance the speed of convergence, Camerini et. al, [14] proposed a modification of the pure subgradient method in which the subgradient direction s^{iter} at an iterate u^{iter} is replaced by a deflected subgradient direction d^{iter}, given by:

$$d^{iter} = s^{iter} + \delta_{iter} d^{iter-1}$$

where δ_{iter} is a suitable scalar called deflection parameter and $d^{iter-1} = 0$ for iter=1.

$$\delta_{iter} = \begin{cases} -\tau_{iter} \frac{s^{iter} d^{iter-1}}{\|d^{iter-1}\|^2} & \text{if } s^{iter} d^{iter-1} < 0 \\ 0 & otherwise \end{cases};$$

with $0 \leq \tau_{iter} < 2$. Note that the choice of $\tau_{iter} = 1$ would amount to using a direction orthogonal to d^{iter1} . In Camerini et. al. [14], "the use of $\tau_{iter} = 1.5$ is recommended and its intuitive justification together with computational results are also given, which is indicating that also in practice the performance of deflected subgradient algorithm is superior to that of the pure subgradient algorithm". There are, in fact, various forms of the choices of the deflection parameter in literature (for instance, Sherali and Ulular [96] and Brannlund [11]) other than the one proposed by Camerini et. al. [14] which is discussed above. However, the scope of this work is to apply a good technique to eliminate zigzag type *I* rather than a discussion about techniques to eliminate this effect, and for more information the interested reader is referred to

[73].

The step-size parameter Δ_{iter} controls the step size along the subgradient direction s^{iter}. A first approach used by Held and Karp [59] and also recommended by Fisher [52] is to determine Δ_{iter} by setting $\Delta_0 = 2$, and halving Δ_{iter} whenever the lower bound has failed to increase in some fixed number of iterations. However, Caprara et. al. [24] observed that "in some particular instance of problems the classical approach halves the step-size parameter after so many iterations, although in these iterations the growth of the value of the dual function is far from regular and can cause a slow convergence. In order to obtain a faster convergence, they proposed the following strategy: Start with $\Delta_{iter} = 0.1$. For every L subgradient iterations compare the biggest and lowest values of Z_R computed on the last L iterations. If these two values differ by more than 1%, the current value of Δ is halved. If, on the other hand, the two values are with in 0.1% from each other, multiply the current value of Δ by 1.5".

In our work we used the deflected Subgradient optimization technique to eliminate the effect of zigzag type I. Moreover, we adapt the technique developed by Caprara et. al. [24] to found good values for Δ_{iter} to accelerate the convergence. Algorithm 3 describes in detail the procedure of the subgradient method.

For any choice of the Lagrangean multipliers (μ, β, γ) a Lagrangean lower bound is given by$Z_R(\mu, \beta, \gamma)$. The best such lower bound is the solution of the dual Lagrangean problem:

$$Z_R(\mu^*, \beta^*, \gamma^*) = \max_{\mu,\beta,\gamma} Z_R(\mu, \beta, \gamma)$$

Algorithm 3 SubgradientLR

Step 1 Initialization

step size (Δ);

The Upper Bound $Zb = Z0(x,z,h)$;

The Lower Bound $Zl = ZR(\mu,\beta,\gamma,\bar{x},\bar{z})$;

Step 2 Adjust multipliers

$s1^{iter} = \sum_k z_{ik} - 1 \quad \forall \quad i$

$d1^{iter} = s1^{iter} + \delta_1 d1^{iter-1}$

$\mu_i^{iter+1} = \mu_i^{iter} + t_{iter}\left(d1^{iter}\right)$;

$s2^{iter} = \sum_m x_{ijkm} - z_{ik} \quad \forall \quad k,i,j : i \neq j$

$d2^{iter} = s2^{iter} + \delta_2 d2^{iter-1}$

$\beta_{kij}^{iter+1} = \beta_{kij}^{iter} + t_{iter}\left(d2^{iter}\right)$;

$s3^{iter} = \sum_k x_{ijkm} - z_{jm} \quad \forall \quad m,i,j : i \neq j$

$d3^{iter} = s3^{iter} + \delta_3 d3^{iter-1}$

$\gamma_{mij}^{iter+1} = \gamma_{mij}^{iter} + t_{iter}\left(d3^{iter}\right)$;

$t_{iter} = \Delta_{iter}\left(\frac{z^*_{feas} - z^*_L(\mu,\beta,\gamma)}{\sum_i(\sum_k(z_{ik}-1)^2 + \sum_k\sum_i\sum_j(\sum_m x_{ijkm} - z_{ik})^2 + \sum_k\sum_i\sum_j(\sum_k x_{ijkm} - z_{jm})^2}\right)$

Step 3 Adjust Step Size

$$\Delta_{iter} = \begin{cases} 0.5\Delta_{iter-1} & \text{if } \bar{\phi} - \underline{\phi} > 0.01\underline{\phi} \\ 1.5\Delta_{iter-1} & \text{if } \bar{\phi} - \underline{\phi} < 0.001\underline{\phi} \\ \Delta_{iter-1} & otherwise \end{cases};$$

Where $\bar{\phi} = \max\{\phi(u^t) : t = k-p+1, k-p+2, ..., k\}$ and $\underline{\phi} = \min\{\phi(u^t) : t = k-p+1, k-p+2, ..., k\}$

4.4 Summary

This chapter presented the development of Lagrangean heuristic to solve the C-EOS-USA-p-MHLP as it was shown that C-EOS-USA-p-MHLP is mixed integer program that is difficult to be solved directly and large-scale, as it has $n^4 + n^2$ binary variables, $n^2 * L$ SOS Type II variables, and n continues variables and $2n^3 + 6n^2 + n$ constraints in addition to constraint set 4.12 which is infinite, where n is the number of nodes in the network. Hence, we considered developing a Lagrangean relaxation to decompose the problem into a number of smaller problems that are easier to be solved. Furthermore, a valid cuts were added to SUB2 to enhance the solution provided by the Lagrangean relaxation. The addition of the valid cuts was essential for the success of the Lagrangean relaxation heuristic as found experimentally that the proposed Lagrangean relaxation would yield poor quality gaps without the addition of the valid cuts.

A heuristic is developed to find a feasible solution for the relaxed solution obtained by solving SUB1 and SUB2. As shown in section 4.3 we utilized the deflected sub-gradient technique to find good values for the Lagrangean multipliers at each iteration of the sub-gradient algorithm. The use of the deflected sub-gradient algorithm was found to provide a faster convergence over the pure sub-gradient techniques and thus we adapted the deflected sub-gradient algorithm.

CHAPTER 5

METHODOLOGY: GRASP

”Today, a variety of heuristic approaches are available to the operations research practitioner. One methodology that has a strong intuitive appeal, a prominent empirical track record, and is trivial to efficiently implement on parallel processors is GRASP (Greedy Randomized Adaptive Search Procedures)” [51]. GRASP is a multi-start or iterative procedure where each GRASP iteration consists of a construction phase, where a feasible solution is constructed, followed by a local search procedure that finds a locally optimal solution. The construction phase of GRASP is essentially a randomized greedy algorithm [60]. GRASP is a recently exploited method combining the power of greedy, randomness, and local search. In this thesis I designed GRASP algorithm for the C-EOS-USA-p-MHLP formulation. In this chapter I will illustrate step-by-step various components of GRASP and how to develop this heuristic for the C-EOS-USA-p-MHLP model.

5.1 Overview of GRASP

The following sections will introduce and explain the components pertaining to the construction of a GRASP.

The GRASP (Greedy Randomized Adaptive Search Procedure) metaheuristic is a multi-start or iterative process, in which each iteration consists of two phases: construction and local search. The construction phase builds a feasible solution, whose neighborhood is investigated until a local minimum is found during the local search phase. The best overall solution is kept as the result. The pseudo-code in Algorithm 4 shows the main parts of a GRASP algorithm for minimization problem, where *Max Iterations* iterations are performed and *Seed* is used as the initial seed for the *pseudorandom* number generator. These include the construction of initial solution and local search techniques to improve the built solution.

Algorithm 4 A generic GRASP pseudo-code

```
Read_ Input();
for k = 1 → Max _ iterations do
    Solution ← Greedy_ Randomized_ Construction(Seed);
    Solution ← Local _ Search (Solution);
    Update_ Solution ← Greedy_ Randomized_ Construction(Seed);
end for
return Best_ Solution
```

5.1.1 Construction Phase in GRASP

The construction phase is done iteratively, by the addition of one element at a time. For each iteration in the construction phase, the selection of the next element to be added to the constructed initial solution is determined by ordering all the elements in a candidate list with respect to a greedy function depending on the design of the algorithm. The greedy function measures the (myopic) benefit of selecting each element. The value associated with every element is updated at each iteration of the construction phase to reflect the changes brought on by the selection of the previous element and therefore the algorithm is adaptive. The randomness in choosing one of the best candidate elements in the list characterizing the probabilistic component of a GRASP, but not necessarily the top candidate. The list of best candidates is called the *restricted* candidate list (RCL). This selection technique allows for different solutions to be obtained at each GRASP iteration, but does not necessarily compromise the power of the adaptive greedy component of the method. Algorithm 5 displays pseudo-code for the construction phase of GRASP. The solution to be constructed is initialized in line 1 of the pseudo-code. The loop from line 3 to 8 is responsible for the construction of the solution. In line 4, the restricted candidate list is built. A candidate from the list is selected, at random, in line 5 and is added to the solution in line 6. The effect of the selected solution element s on the benefits associated with every element is taken into consideration in line 7, where the greedy function is adapted.

Algorithm 5 GRASP construction phase pseudo-code

```
Solution ← ∅ ;
Evaluate the incremental costs of the candidate elements;
while Solution is not a complete solution do
    Build the restricted candidate list (RCL);
    Select an element s from the RCL at random;
    Solution ← Solution∪{s} ;
    Reevaluate the incremental costs;
end while
return Solution
```

The solutions generated by a greedy randomized construction are not necessarily optimal, even with respect to simple neighborhoods. The local search phase usually improves the constructed solution. A local search algorithm works in an iterative fashion by successively replacing the current solution by a better solution in the neighborhood of the current solution. It terminates when no better solution is found in the neighborhood. The pseudo-code of a basic local search algorithm starting from the solution constructed in the first phase and using a neighborhood N is given in Algorithm 5.

5.1.2 Local Search in GRASP

Even though with respect to simple local search procedure applied to obtain a feasible solution in the construction phase of the GRASP, it is not guaranteed that the feasible solution obtained by the local search is local optimal. Therefore, it is always beneficial to apply local search to the feasible solution obtained by the construction phase. By iteratively replacing the current solution by a better solution through the local search algorithm, the algorithm terminates when no better solution is found in the neighborhood of the current solution. For the problem P we have the neighborhood structure N, which relates a solution s of the problem P to the solutions $N(s)$. s is the solution that is claimed to be locally optimal if there is no better solution in $N(s)$. For a given neighborhood structure N, the general form of local search algorithm is shown in Algorithm 6. The key to success for a local search algorithm consists of the suitable choice of a neighborhood structure, efficient neighborhood search techniques, and the starting solution.

Algorithm 6 GRASP local search phase pseudo-code

```
for s not locally optimal do
    Find a better solution t ∈ N(s);
    Let s = t;
end for
return s as local optimal for P
```

Empirically, the efficiency of the local search optimization procedures significantly improves as the initial solution improves, this is because it might take exponential time from an arbitrary starting point to construct a feasible solution and implement a local search to find local optimal with respect to the obtained feasible solution. One way to improve the efficiency of the local search is by the use of customized data structures. The result is that often many GRASP solutions are generated in the same amount of time required for the local optimization procedure to converge from a single random start.

As stated by Feo et. al. [51]: "it is difficult to formally analyze the quality of solution values found by using the GRASP methodology. However, there is an intuitive justification that views GRASP as a repetitive sampling technique. Each GRASP iteration produces a sample solution from an unknown distribution of all obtainable results. The mean and variance of the distribution are functions of the restrictive nature of the candidate list. For example, if the cardinality of the restricted candidate list is limited to one, then only one solution will be produced and the variance of the distribution will be zero. Given an effective greedy function, the mean solution value in this case should be good, but probably suboptimal. If a less restrictive cardinality limit is imposed, many different solutions will be produced implying a larger variance. Since the greedy function is more compromised in this case, the mean solution value should degrade. Intuitively, however, by order statistics and the fact that the samples are randomly produced, the best value found should outperform the mean value. Indeed, often the best solutions sampled are optimal".

5.2 GRASP Applied to the C-EOS-USA-p-MHLP

Section 5.1 presented brief introduction about the GRASP heuristic, and the following section will provide the details for designing a GRASP heuristic that is able to find good solutions for the C-EOS-USA-p-MHLP model.

5.2.1 Construction Phase for EOS-C-HLP

To obtain a solution for the C-EOS-USA-p-MHLP we first select the p nodes to be located as hubs *(step 1)* and then the allocation of non-hub nodes to the hubs is performed *(step 2)*.

Several ways can be considered to create the RCL to solve the HLP. In our work we will use the method used by Klincewicz [70] to create RCL based on the total flow originating and terminating from the node (i.e., by decreasing values of $\sum_j (w_{ij} + w_{ji})$. So the RCL contains the top predefined number of nodes with the highest originating and terminating flow.

The first step in the construction of feasible solution is a random selection of p nodes to be located as hubs. In step 2 of the construction phase, 1 of the p hubs is chosen to be allocated to each node. In the construction of the solution phase each node will be allocated to the nearest hub based on the distance criterion.

5.2.2 Solution presentation

I present my representation for the solutions of the phub problem and its utilization within the GRASP approach. Generally speaking, the matrix for the USA-p-MHLP formulations is a natural representation of a solution. Nevertheless, this representation requires huge amount of memory for large instances. Thus, I will use the following representation, inspired by the solution presentation developed by Pérez et al. [88]

Each solution is represented by an array of size n that indicates the hubs and the assignment of the rest of nodes to these hubs. Let me assume that the set of non-hub nodes is indexed by the set $L = \{1, ..., n-p\}$, the idea of this representation is the following: the first p positions of the array, called locationarray, are the ordered indexes of the hubs; and the last $(n-p)$ positions, called allocationarray, designate the allocations of the non hub nodes, assuming that the hubs are allocated to themselves. Therefore, I will need one array for solution representation which is: $s = \{H|A\}$, where H is the set of hub nodes and A is the allocated array.

For example, suppose that we have 5 nodes and we want to locate 2 hubs, $p = 2, L = (1\,2\,3\,4\,5)$, let us assume that nodes 4 and 5 are the set of hubs and nodes 1 and 3 are allocated to hub 4 and node 2 is allocated to hub 5, is presented in the matrix X,

$$X = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

is represented by the following array

$$s = \begin{pmatrix} 4 & 5 & | & 1 & 2 & 1 \end{pmatrix}$$

The array s has the following meaning: As p is equal 2, the first two positions within s, represent the hubs (4 and 5). Then, these two values are eliminated from L and $L = \{1,2,3\}$. The last three positions on the array s are the assignments of the remaining nodes in L. That is, nodes 1 and 3 are assigned to hub 4 and the node 2 is assigned to hub 5. Note that this representation always produce feasible solutions for the single allocation uncapacitated HLP.

5.2.3 GRASP Local search for C-EOS-USA-p-MHLP

Since the proposed method consists of two steps, I can apply two types of local searches to the results obtained at step 1 and step 2: changing the hubs selected, and changing the assignments of hubs to nodes, specifically the arrays $H\&A$. Given a solution s I consider two neighborhoods, N_H and N_A, to improve s. N_H implements a classical exchange in which a hub node is replaced by a non-hub node $h_{j^o} \in N \backslash H$. In other words, the local search moves the hub from node h_j to node h_{j0}, and therefore obtaining $H^o = \{h_1, h_2, ..., h_{j^o}, ..., h_p\}$. The element h_j to be removed from the solution and replacing it with h_{j^o} by scanning the non-hub nodes and the first non-hub found to minimize the total cost of the network is used to replace h_j (instead of scanning the entire neighborhood to determine the best one starting from a random element). This step is costly when it comes to computational time because each time a new hub set H is formed we first assign all non-hub nodes to the nearest hub, as in the construction of feasible solution, then local search for the assignment solution A is performed. However, this method is found to be better as found by Piero et. al. [86]. On the other hand, neighborhood N_A does not change the hub selection and it only considers the node assignments. In particular, for a node i assigned to hub h_j it considers the exchanging of the hub to be h_{j^o}, $A^o = \{h_1, h_2, ..., h_{i^o}, ..., h_{n-p}\}$.

The local search LS_H performs moves in N_H as long as the objective function improves. At each iteration, it selects a hub at random and randomly selects a node to replace with. Two different strategies might be implemented the so-called first find strategy, in which I perform the first improving move in the neighborhood(instead of scanning the entire neighborhood to determine the best one) or extensive search where we explores the entire neighbor then we select the neighbor with best value according to the objective function. LS_H terminates when all the elements in set L are explored to produce the new set of hubs L^o.

The second neighborhood N_A only considers moves on A without changing H. In other words, N_A explores the possibility of exchanging a hub hia to which node i is assigned, with one of the hub nodes h_{ja^o} to which i is not assigned. This local search terminates when all the hubs in set H explored then the node is allocated to the hub that minimizes the total network cost.

The overall GRASP FOR HLP is shown in Algorithm 7.

Algorithm 7 GRASP Algorithm for HLP

for $i = 1 \rightarrow n$ **do**
  $T(i) = \sum_j (w_{ij} + w_{ji})$
end for
$T \leftarrow$ sort in decreasing order T
$RCL \leftarrow$ highest Q nodes in T
$H \leftarrow$ Randomly pick p elements from RCL
for $i = 1 \rightarrow N$ **do**
  if $i \notin H$ **then**
    Assign i to the nearest $h_j \in H$ ([$-$,Index]=min(Distance(i, H)))
    $a_{Index} \leftarrow a_i$
  end if
end for
Incumbent=Network Cost
Counter=0
while counter$< C_{max}$ **do**
  $R-Vector \leftarrow$ random integer numbers from 1:p
  $N-Vector \leftarrow$ random integer numbers from 1:n
  for $j = 1 \rightarrow p$ **do**
    $h_j = H(R-Vector(j))$ (Pick a hub at random from the hub set H)
    for $i = 1 \rightarrow n-p$ **do**
      $h_j \leftarrow i$
      for $i = 1 \rightarrow N$ **do**
        if $i \notin H$ **then**
          Assign i to the nearest $h_j \in H$ ([$-$,Index]=min(Distance(i, H)))
          $a_{Index} \leftarrow a_i$
        end if
      end for
      for $ii = 1 \rightarrow nc-p$ **do**
        for $jj = 1 \rightarrow p$ **do**
          $a_{ii}^{jj} \leftarrow a_{ii}^{j}$
          Network Cost(jj)=Cost
          counter $\leftarrow$ 0
        end for
        [Cost-N, Index]=min(Network Cost)
        $a_{ii}^{Index} \leftarrow a_{ii}^{j}$ (Changes the assignment of the selected node to the hub that minimizes the total network cost)
      end for
      if Cost-N$<$ Incumbent **then**
        Incumbent $\leftarrow$ Cost-N
      else
        counter $\leftarrow$ counter+1
      end if
    end for
  end for
end while

5.3 Summary

This chapter started by introducing the GRASP solution methodology and its components. After introducing the GRASP solution methodology, the GRASP heuristic to solve the C-EOS-p-MHLP was presented. The solution representation of the C-EOS-p-MHLP model using efficient data structure was presented, we used the array representation for the solution obtained by the GRASP as well as using this structure for the local search procedure.

To solve the C-EOS-p-MHLP model using GRASP two local search procedures were developed into the GRASP algorithm, namely the search for the optimal set of nodes to be located as hubs and the local search for the allocation decisions for non-hub nodes to hub nodes (i.e. which hub should be allocated to each node).

CHAPTER 6

Implementation and Testing

For our computational experiments, we utilize what has become a standard testbed for hub location problems, i.e., a set of problems based on 1970 airline passenger travel, as evaluated by the Civil Aeronautics Board. This set of problems was first used by OKelly [81]. we utilize the 25-node problem (based on the 25 largest U.S. cities), as well as the 10-node problem, 15-node problem and the 20-node problem that are obtained by considering the first 10 nodes, the first 15 nodes or the first 20 nodes, respectively, from the list of 25 nodes in alphabetical order. The cities used in the data set are shown on the US map in Figure 6.1 and displayed in Table 6.1. The data includes values for the pairwise traffic W_{ij}, as well as longitude and latitude coordinates with which to compute distances between the nodes. Costs c_{ij} on links are determined to be proportional to the distance between the end nodes. We will test values of $p = 3$ and 4.

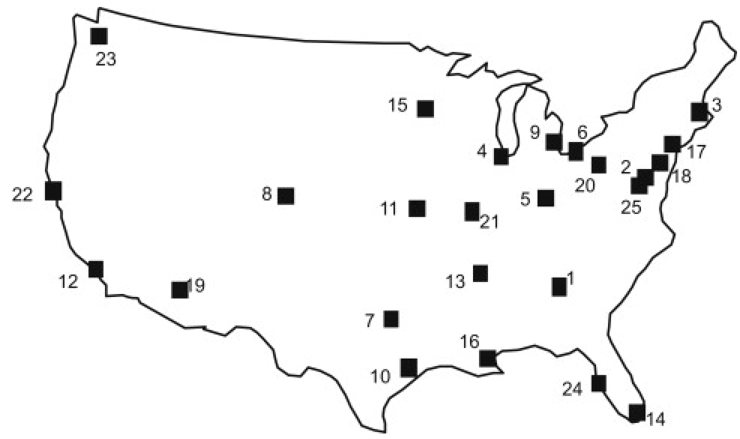

Figure 6.1: US cities in the CAB data set.

1	Atlanta	6	Cleveland	11	Kansas City	16	New Orleans	21	St. Louis
2	Baltimore	7	Dallas-Fort Worth	12	Los Angeles	17	New York	22	San Francisco
3	Boston	8	Denver	13	Memphis	18	Philadelphia	23	Seattle
4	Chicago	9	Detroit	14	Miami	19	Phoenix	24	Tampa
5	Cincinnati	10	Houston	15	Minneapolis	20	Pittsburgh	25	Washington DC

Table 6.1: Cities in the CAB data set and their assigned numbers.

The piecewise-linear concave cost functions used have four pieces each. Three different concave cost functions are included in my computational experiments, as proposed by Klincewicz [70], having slopes as defined in Table 6.2. These three functions represent different degrees of economy-of-scale. Function 1 provides for a very modest discount as the amount of traffic increases, with the minimum slope equal to 0.7. Function 2 provides a greater degree of discount, with slopes as low as 0.4. Function 3 provides a very aggressive discount scheme, with a minimum slope of 0.2. We used 8 break points to calculate w_{km}, which are $[0, 50, 100, 200, 800, 2300, 4000, 8600] * 10^3$, the selection of those points assure that the total network flow is always within the specified range and proper selection of the values for λ^l_{km}.

Total flow on *km* (*1,000)	Function 1 slopes	Function 2 slopes	Function 3 slopes
$0 \leq w^1_{km} < 50$	1.0	1.0	0.8
$50 \leq w^2_{km} < 100$	0.9	0.8	0.6
$100 \leq w^3_{km} < 200$	0.8	0.6	0.4
$200 \leq w^4_{km}$	0.7	0.4	0.2

Table 6.2: Piecewise-Linear Concave Functions

The parametric settings for the Lagrangean heuristic (C-EOS-USA-p-MHLP-LR) are presented in Algorithm 3. After an initial investigation, Lagrangean multipliers $\alpha, \beta, \& \gamma$ are initialized at $(-10^6, 10^6, 10^6)$.

As mentioned in Section 4.3, we adapted the method developed by Caprara et. al. [24] to found good values for Δ_{iter} to accelerate the convergence of the Lagrangean heuristic (C-EOS-USA-p-MHLP-LR). For our work we found experimentally that the value of $L = 5$ and $\Delta_0 = 0.2$ are good values.

The Lagrangean algorithm is stopped when the relative gap between Zb and Zl drops below 1%, or the number of iteration reaches the maximum *Max_ Iter=1500*.

For the GRASP design, the value of Q which controls the length of RCL was set to be $ceil(0.7n)$ and $C_{max} = ceil(1.5p)$.

The algorithm was coded in MATLAB R2012b and run on a Dell workstation with Intel(R)Core(TM)i5 2540M CPU @2.60GHz processors with Windows 7 operating system. In addition, because Matlab is not very efficient in running loops, all the loops are coded in C and are called from MATLAB. GAMS version 24.0.2 was used to solve some instances and the solver was CONOPT 3 version 3.15I. In addition, we used IBM ILOG CPLEX 12.5 solver to solve SUB1 and SUB2 through calling the appropriate solvers from the MATLAB environment.

6.1 Numerical Results

Computational results with the 10-node problem using the Lagrangean heuristic are summarized in Table 6.3. Each row corresponds to a problem instance. The first three columns indicate the number of hubs p, the index of the cost function used (from Table 6.2) and the value of τ for the convex function of congestion cost function, the fourth column shows the gap between the upper and lower limits found in the Lagrangean Heuristic ($\frac{UB-LB}{LB}$), the fifth, sixth, seventh, and eighth columns show the computational time in seconds spent in the heuristic, solving sub-problem 1, solving sub-problem 2, and generating a feasible solution, the ninth column shows the cost of the network, and the last column displays the set of selected nodes as optimal hubs.

No. of hubs	Cost Function used	τ	Gap (%)	Heur.	SUB1	SUB2	Feas.	Cost ($*10^6$)	Optimal Hubs
3	f1	1	0.66	0.87	1.50	3.26	0.648	759.315	4,7,9
3	f1	1.3	0.42	1.20	53.65	4.62	0.889	804.746	4,7,9
3	f1	1.5	0.86	0.58	62.11	2.23	0.428	1353.696	4,7,9
3	f2	1	0.61	0.85	1.22	2.45	0.636	744.081	4,7,9
3	f2	1.3	0.77	1.17	42.18	3.86	0.872	787.118	4,7,9
3	f2	1.5	0.93	0.82	44.66	2.44	0.617	1335.182	4,7,9
3	f3	1	0.62	0.88	1.31	2.47	0.648	677.219	4,6,7
3	f3	1.3	0.76	0.71	25.97	1.83	0.528	722.686	4,6,7
3	f3	1.5	0.88	2.04	157.30	6.80	1.457	1263.911	4,6,7
4	f1	1	0.42	1.28	2.31	4.70	0.941	730.279	4,7,8,9
4	f1	1.3	0.55	1.33	55.23	4.84	0.973	768.797	4,7,8,9
4	f1	1.5	0.97	3.39	314.39	13.06	2.487	1247.356	4,5,7,8
4	f2	1	0.57	2.21	3.61	7.84	1.629	713.335	4,7,8,9
4	f2	1.3	0.72	0.85	28.14	2.76	0.619	756.294	4,7,8,9
4	f2	1.5	0.93	7.89	524.89	30.40	5.932	1239.414	1,4,6,7
4	f3	1	0.64	1.34	2.08	4.31	0.979	632.731	4,6,7,8
4	f3	1.3	0.60	1.05	35.48	3.21	0.769	676.576	4,6,7,8
4	f3	1.5	0.82	3.83	258.23	13.57	2.806	1171.398	4,5,7,9

Table 6.3: Computational results for 10-node models with economies of scale and congestion.

Computational results with the 10-node problem using the GRASP heuristic are summarized in Table 6.4. Each row corresponds to a problem instance. The first two columns indicate the number of hubs p and the index of the cost function used (from Table 6.2), the third column shows the value of τ, the fourth column shows the objective function value, the fifth column shows the computational time in seconds, the sixth column displays the set of selected nodes as optimal hubs, the seventh column shows the number of iterations in the GRASP Algorithm, the eighth column shows the gap (%) between the solution value found by GRASP compared to GAMS solution, and the last column shows the computational time solving the model exactly using the commercial solver GAMS. We included the last column to illustrate the superior performance of my solution methods developed in this work. As clearly shown, the computational time grows exponentially as the non-linearity in the congestion function and/or in the economies of scale function increase. Therefore, the use of the Lagrangean heuristic or GRASP heuristic is justified when the judgment is based on computational time and solution quality.

No. of hubs	Cost Function used	τ	Cost	Time (sec)	Optimal Hubs	Iterations	Gap (%)	GAMS Time (sec)
3	f1	1	759.315	0.12	4,7,9	6	0	110
3	f1	1	804.746	0.11	4,7,9	6	0	201
3	f1	1.5	1353.696	0.11	4,7,9	7	0	9261
3	f2	1	744.081	0.12	4,7,9	6	0	270
3	f2	1.3	787.118	0.12	4,7,9	6	0	985
3	f2	1.5	1335.182	0.12	4,7,9	6	0	28940
3	f3	1	677.219	0.11	4,6,7	6	0	416
3	f3	1.3	722.686	0.12	4,6,7	6	0	1431
3	f3	1.5	1263.911	0.12	4,6,7	6	0	$**$
4	f1	1	730.279	0.15	4,7,8,9	8	0	238
4	f1	1.3	768.797	0.14	4,7,8,9	7	0	578
4	f1	1.5	1247.556	0.13	4,5,7,9	7	0	36844
4	f2	1	713.335	0.13	4,7,8,9	7	0	332
4	f2	1.3	756.294	0.19	4,7,8,9	8	0	3231
4	f2	1.5	1239.414	0.13	4,5,7,9	7	0	48657
4	f3	1	632.731	0.16	4,6,7,8	8	0	478
4	f3	1.3	676.576	0.16	4,6,7,8	8	0	$**$
4	f3	1.5	1171.398	0.17	4,5,7,9	9	0	70124

Table 6.4: Computational results for 10-node models using the GRASP heuristic.

$**$ GAMS was not able to find feasible integer solution at the completion.

Several observations can be made from the data in Table 6.3. For the number of hubs equals 3 as well as for the case of 4 hubs, there is variation in the set of hubs selected. Thus, the degree of nonlinearity in the concave cost function can affect the selected hubs in the network under the congestion effect. Also, this observation was noted by Klincewicz [70] for the multiple allocation version of the HLP as well. Moreover, the set of hub for when $p = 4$ changes within the same EOS but under different congestion values. Also, the heuristics performed well. GRASP obtained the optimal solution in all cases.

Table 6.5 shows the optimal hubs for the single allocation with constant discount factor and without congestion. In the rows staring from the second till the last row, the incremental values of the discount factor 0.1, for instance in the second row we tested the discount factor values (0.2,0.3,0.4,0.5,0.6 & 0.7).

No. of hubs	Discount factor	Optimal Hubs
3	0.1	3,4,7
3	0.2-0.7	4,6,7
3	0.8-1.0	4,7,9
4	0.1-0.3	3,4,6,7
4	0.4-0.6	4,6,7,8
4	0.8-1.0	4,7,8,9

Table 6.5: Computational results for 10-node models with constant discount factor.

Computational results with the 15 nodes problem is summarized in tables 6.6 and 6.7, for the Lagrangean heuristic and GRASP, respectively, having the same format as table 6.3. Once again, we observe that the set of optimal hubs does change for different cost functions and under different congestion costs. Also the performance of the GRASP heuristic is excellent with only one case where the optimal of GRASP is not same as Lagrangean heuristic, the case of network with three hubs, function 2 and $\tau = 1.3$, however, the GRASP selected the optimal set of hubs but I found that the assignment is different. The gap in Table 6.7 is calculated based on the difference between objective value found by GRASP and the lower bound found in the Lagrangean heuristic.

No. of hubs	Cost Function used	τ	Gap (%)	Heur.	SUB1	SUB2	Feas.	Cost ($*10^6$)	Optimal Hubs
3	1	1	0.53	2.06	2.29	21.13	1.51	2655.946	4,7,8
3	1	1.3	0.67	1.36	11.86	19.87	0.97	2809.156	4,7,8
3	1	1.5	0.83	2.76	43.17	31.20	2.02	4820.148	4,5,11
3	2	1	0.65	1.64	1.85	23.46	1.20	2542.092	4,8,13
3	2	1.3	0.77	1.53	16.50	21.77	1.12	2684.901	4,8,13
3	2	1.5	0.81	3.62	46.55	52.85	2.55	4736.184	4,5,11
3	3	1	0.48	0.90	0.97	11.00	0.65	2343.830	4,12,13
3	3	1.3	0.86	4.60	43.63	70.23	3.42	2490.438	4,12,13
3	3	1.5	0.87	2.79	38.96	41.49	2.04	4575.254	4,5,7
4	1	1	0.55	4.49	5.11	49.27	3.30	2530.973	1,4,7,8
4	1	1.3	0.77	1.95	19.54	29.32	1.41	2668.828	1,4,7,8
4	1	1.5	0.78	5.91	101.00	99.66	4.36	4483.201	1,4,9,11
4	2	1	0.92	2.07	2.41	31.29	1.53	2415.171	1,4,7,8
4	2	1.3	0.58	3.36	34.14	53.88	2.44	2553.027	1,4,7,8
4	2	1.5	0.72	1.95	24.14	28.45	1.44	4388.809	1,4,6,7
4	3	1	0.62	1.07	1.14	13.83	0.78	2173.156	1,4,7,12
4	3	1.3	0.81	5.27	50.22	82.77	3.92	2309.599	1,4,7,12
4	3	1.5	0.90	2.27	32.80	32.88	1.65	4170.506	1,4,6,7

Table 6.6: Computational results for 15-node models with economies of scale and congestion.

No. of hubs	Cost Function used	τ	Cost	Time (sec)	Optimal Hubs	Iterations	Gap(%)
3	f1	1	2655.946	0.22	4,7,8	7	0.53
3	f1	1.3	2809.156	0.19	4,7,8	7	0.67
3	f1	1.5	4820.148	0.23	4,5,11	8	0.83
3	f2	1	2542.092	0.19	4,8,13	6	0.65
3	f2	1.3	2684.901	0.30	4,8,13	6	0.77
3	f2	1.5	4736.184	0.18	4,5,11	6	0.81
3	f3	1	2343.830	0.20	4,12,13	6	0.48
3	f3	1.3	2499.677	0.19	4,12,13	7	1.06
3	f3	1.5	4575.254	0.25	4,5,7	8	0.87
4	f1	1	2530.973	0.87	1,4,7,8	7	0.55
4	f1	1.3	2668.828	0.63	1,4,7,8	8	0.77
4	f1	1.5	4483.201	0.88	1,4,9,11	7	0.78
4	f2	1	2415.171	1.13	1,4,7,8	9	0.92
4	f2	1.3	2553.027	0.95	1,4,7,8	8	0.58
4	f2	1.5	4388.809	0.92	1,4,6,7	8	0.72
4	f3	1	2173.156	1.13	1,4,7,12	9	0.62
4	f3	1.3	2309.599	0.92	1,4,7,12	7	0.81
4	f3	1.5	4170.506	0.97	1,4,6,7	9	0.90

Table 6.7: Computational results for 15-node models using GRASP heuristic.

Computational results with the 20 nodes problem is summarized in tables 6.6 and 6.7, for the Lagrangean heuristic and GRASP, respectively, having the same format as table 6.3. Once again, we observe that the set of optimal hubs does change for different cost functions and under different congestion costs. In addition, the performance of the GRASP heuristic is excellent.

No. of hubs	Cost Function used	τ	Gap (%)	Heur.	SUB1	SUB2	Feas.	Cost ($*10^6$)	Optimal Hubs
3	1	1	0.61	9.34	6.14	291.26	5.45	6214.151	4,7,17
3	1	1.3	0.72	8.07	50.50	279.44	5.11	6656.634	4,7,17
3	1	1.5	0.81	11.89	121.76	402.01	7.44	14265.499	6,11,17
3	2	1	0.90	2.96	2.26	92.47	1.90	5517.943	4,7,17
3	2	1.3	0.93	2.47	10.61	77.46	1.60	5960.426	4,7,17
3	2	1.5	0.74	13.75	139.93	482.72	8.80	13657.265	4,13,17
3	3	1	0.56	1.94	1.45	57.59	1.26	4863.374	4,7,17
3	3	1.3	0.61	1.57	6.70	47.75	1.01	5323.822	4,12,17
3	3	1.5	0.86	1.83	9.47	52.59	1.17	13011.7715	4,7,17
4	1	1	0.61	11.55	8.46	405.26	7.30	5855.541	1,4,8,17
4	1	1.3	0.86	11.69	75.81	408.96	7.43	6265.614	1,4,8,17
4	1	1.5	0.92	13.05	139.80	448.93	8.22	12928.424	4,7,17,20
4	2	1	0.77	2.83	2.10	85.67	1.81	5205.408	4,7,14,17
4	2	1.3	0.83	1.71	7.17	51.05	1.09	5624.056	4,7,14,17
4	2	1.5	0.97	12.81	120.56	432.38	8.03	12373.475	4,7,17,20
4	3	1	0.88	4.60	3.31	139.35	2.91	4408.583	1,4,12,17
4	3	1.3	0.90	2.00	8.45	59.71	1.25	4826.063	1,4,12,17
4	3	1.5	0.92	10.77	99.97	339.34	6.58	11720.991	1,4,17,19

Table 6.8: Computational results for 20-node models with economies of scale and congestion.

No. of hubs	Cost Function used	τ	Cost$*10^6$	Time (sec)	Optimal Hubs	Iterations	Gap(%)
3	f1	1	6214.151	2.10	4,7,17	9	0.61
3	f1	1.3	6656.634	1.85	4,7,17	6	0.72
3	f1	1.5	14265.499	1.63	6,11,17	6	0.81
3	f2	1	5517.943	1.72	4,7,17	7	0.90
3	f2	1.3	5960.426	2.04	4,7,17	7	0.93
3	f2	1.5	13657.265	1.72	4,13,17	7	0.74
3	f3	1	4863.374	1.96	4,7,17	8	0.56
3	f3	1.3	5323.822	1.94	4,12,17	7	0.61
3	f3	1.5	13011.7715	1.93	4,7,17	7	0.86
4	f1	1	5855.541	3.77	1,4,8,17	10	0.61
4	f1	1.3	6265.614	2.66	1,4,8,17	7	0.86
4	f1	1.5	12928.424	5.11	4,7,17,20	9	0.92
4	f2	1	5205.408	3.18	4,7,14,17	9	0.77
4	f2	1.3	5624.056	3.04	4,7,14,17	9	0.83
4	f2	1.5	12373.475	3.33	4,7,17,20	9	0.97
4	f3	1	4408.583	3.24	1,4,12,17	9	0.88
4	f3	1.3	4826.063	2.99	1,4,12,17	9	0.90
4	f3	1.5	11720.991	3.42	1,4,17,19	10	0.92

Table 6.9: Computational results for 20-node models

Computational results with the 25 nodes problem is summarized in table 6.10 and 6.11, for the Lagrangean heuristic and GRASP, respectively, having the same format as table 6.3.

The stopping criterion for all instances in Table 6.10 is the number of iterations which was set previously to be 1500 iterations.

No. of hubs	Cost Function used	τ	Gap (%)	Heur.	SUB1	SUB2	Feas.	Cost ($*10^6$)	Optimal Hubs
3	1	1	2.34	429	6685	17834	194	9697.2	2,4,12
3	1	1.3	3.88	432	24818	21402	185	1045.25	2,4,12
3	1	1.5	10.3	442	27506	17692	201	24949	4,12,18
3	2	1	2.81	419	4351	16538	188	8383	4,12,18
3	2	1.3	2.43	370	23086	16244	188	9134	4,12,18
3	2	1.5	3.70	410	30522	18169	180	23481	4,12,18
3	3	1	1.85	405	6616	16090	184	7249.9	4,12,17
3	3	1.3	2.69	311	26517	25051	150	8000.6	4,12,17
3	3	1.5	0.90	305	33956	16313	178	21368	4,12,25
4	1	1	12.9	284	4301	15168	165	9253.6	4,7,12,17
4	1	1.3	0.61	415	20984	17107	177	9902.3	1,4,12,18
4	1	1.5	3.40	224	28896	17581	154	2216.2	2,12,21,25
4	2	1	4.52	332	7214	17496	172	7751.1	4,7,12,17
4	2	1.3	8.81	340	19509	16847	170	8438.7	4,7,12,17
4	2	1.5	2.58	357	23923	17339	183	20821	1,4,12,17
4	3	1	9.87	421	6595	15392	163	6355	4,12,17,21
4	3	1.3	8.60	293	18001	14986	162	7038	4,7,12,17
4	3	1.5	6.50	656	36377	16350	251	19295	4,7,12,17

Table 6.10: Computational results for 25-node models with economies of scale and congestion.

No. of hubs	Cost Function used	τ	Cost$*10^6$	Time (sec)	Optimal Hubs	Iterations
3	f1	1	9697.185	8.04	2,4,12	8
3	f1	1.3	10452.519	6.74	2,4,12	8
3	f1	1.5	24766.627	6.71	4,8,18	9
3	f2	1	8383.015	4.47	4,12,18	6
3	f2	1.3	9134.039	4.85	4,12,18	6
3	f2	1.5	23477.848	4.68	4,12,18	7
3	f3	1	7249.906	5.21	4,12,17	8
3	f3	1.3	8000.609	4.93	4,12,17	7
3	f3	1.5	22267.868	5.55	4,12,17	8
4	f1	1	9219.826	33.17	1,4,12,18	8
4	f1	1.3	9902.361	31.90	1,4,12,18	8
4	f1	1.5	22182.273	20.78	4,12,17,20	9
4	f2	1	7851.468	20.45	1,4,12,17	8
4	f2	1.3	8538.465	20.61	1,4,12,17	8
4	f2	1.5	20821.907	23.76	1,4,12,17	11
4	f3	1	6552.170	25.39	4,12,17,24	10
4	f3	1.3	7246.815	22.00	1,4,12,17	10
4	f3	1.5	19455.955	21.41	1,4,12,17	10

Table 6.11: Computational results for 25-node models

Since the design of the GRASP Algorithm is independent on the number of segments for nonlinear function of economies of scale, this will enable us to consider new instances with new parameters, for this purpose we proposed the use of b=0.6, 0.7, 0.8, and 0.9. Those functions present the actual use of economies of scale with concave function is tested on two set of data, namely AP and CAB.

No. of hubs	b	τ	Cost$*10^6$	Time (sec)	Optimal Hubs	Iterations
3	0.6	1	5400.449	3.17	4,12,17	6
3	0.6	1.3	6145.448	5.82	4,12,17	7
3	0.6	1.5	20391.061	4.39	4,12,17	9
3	0.7	1	5481.079	4.11	4,12,17	6
3	0.7	1.3	6226.078	4.11	4,12,17	6
3	0.7	1.5	20471.691	5.48	4,12,17	7
3	0.8	1	5788.503	3.22	4,12,17	8
3	0.8	1.3	6533.501	3.05	4,12,17	8
3	0.8	1.5	20779.115	3.02	4,12,17	7
3	0.9	1	6921.620	3.15	4,12,17	7
3	0.9	1.3	7672.323	3.17	4,12,17	8
3	0.9	1.5	21953.365	3.08	4,12,17	8
4	0.6	1	3987.472	13.97	4,12,17,24	8
4	0.6	1.3	4672.786	12.40	4,12,17,24	8
4	0.6	1.5	16876.020	13.43	4,12,16,17	8
4	0.7	1	4094.884	14.31	4,12,17,24	8
4	0.7	1.3	4780.199	12.85	4,12,17,24	8
4	0.7	1.5	16983.365	14.10	4,12,16,17	8
4	0.8	1	4247.643	14.98	4,7,12,17	8
4	0.8	1.3	4924.077	14.90	4,7,12,17	8
4	0.8	1.5	17022.046	17.40	4,7,12,17	9
4	0.9	1	5895.424	16.04	4,12,17,24	8
4	0.9	1.3	6584.660	16.75	4,12,17,24	10
4	0.9	1.5	18801.873	12.46	1,4,12,17	8

Table 6.12: Computational results for 25-node models using GRASP.

Also we have tested the algorithm on the AP (Australian Post) data set. It is based on real data from the Australian postal service and was presented by Ernst and Krishnamoorthy in 1996 [45] with 50 nodes.

No. of hubs	b	τ	Cost	Time (sec)	Optimal Hubs	Iterations
3	0.6	1	141778.635	157.04	14,29,35	8
3	0.6	1.3	174611.6298	126.35	14,29,35	7
3	0.6	1.5	294818.1269	144.86	15,33,35	9
3	0.7	1	144428.5973	108.49	14,29,35	8
3	0.7	1.3	177232.7507	97.72	14,29,35	7
3	0.7	1.5	297019.3448	95.02	15,33,35	8
3	0.8	1	148906.7887	86.85	14,28,35	8
3	0.8	1.3	181684.6881	105.84	14,28,35	9
3	0.8	1.5	300904.5908	87.85	15,33,35	9
3	0.9	1	156623.2756	99.82	14,28,35	10
3	0.9	1.3	189353.6029	106.44	14,28,35	9
3	0.9	1.5	307890.9686	85.02	15,33,35	6
4	0.6	1	125220.3768	333.82	6,29,32,35	10
4	0.6	1.3	154199.5756	303.38	6,28,32,35	9
4	0.6	1.5	253970.844	351.88	6,28,33,35	10
4	0.7	1	128664.8008	375.88	14,29,32,35	10
4	0.7	1.3	157454.0148	362.27	6,28,32,35	9
4	0.8	1	133617.7165	397.97	14,28,32,35	11
4	0.8	1.3	162314.0656	384.63	14,28,33,35	11
4	0.8	1.5	261701.6845	358.17	14,28,33,38	11
4	0.9	1	141893.4629	339.14	14,28,33,35	10
4	0.9	1.3	170362.4674	287.03	14,28,33,35	8
4	0.9	1.5	269865.0802	325.44	14,28,33,35	10

Table 6.13: Computational results for 50-node models using GRASP.

6.2 Analysis

In the computational experiments conducted by OKelly and Bryan [83] and Klincewicz [70] they found that interhub traffic flow tended to be concentrated on a few interhub arcs. That is, some interhub arcs had carry relatively large traffic flows that were highly discounted, while others had very small amounts of traffic. As mentioned by Bryan [12] analysts should be aware of this potential imbalance in the interhub network (at least as witnessed in these CAB problems) when applying the model to real world networks. One way to mitigate the imbalance effect is to consider congestion effect to count for the incoming flow on the hubs, which is the solution proposed by Elhedhli and Hu [42]. In our work we integrated the congestion and economies of scale in the same model, therefore, we provided more insights about this issue. Table 6.14 provides some analysis for the C-EOS-USA-p-MHLP model the data is based on the AP dataset results in Table 6.13, the first three columns in Table 6.14 follows the same format as the first three columns in Table 6.13. As shown in the table the max/min ration increases as the nonlinearity factor b in the EOS function decreases (i.e. more discount is provided) which is the same imbalance effect mentioned by Bryan [12] in his study; however, for the same discount rate as the value if τ increases (i.e. the congestion cost) the max/min ratio decreases. For a network with three hubs is it reasonable to have a congestion cost of 1.5 and a discount rate of 0.7, as this provides more balanced network with good discounts on the flow between hubs.

No. of hubs	b	τ	Max flow (Value)	Min flow (Value)	max/min ratio
3	0.6	1	2419.669197	445.98463	5.43
3	0.6	1.3	2324.475167	445.98463	5.21
3	0.6	1.5	1739.611367	896.13833	1.94
3	0.7	1	2453.517797	445.98463	5.50
3	0.7	1.3	2295.703047	552.19868	4.16
3	0.7	1.5	1739.611367	896.13833	1.94
3	0.8	1	2424.745677	552.19868	4.39
3	0.8	1.3	2424.745677	552.19868	4.39
3	0.8	1.5	1786.984697	896.13833	1.99
3	0.9	1	2424.745677	552.19868	4.39
3	0.9	1.3	2424.745677	600.54075	4.04
3	0.9	1.5	1822.691447	896.13833	2.03

Table 6.14: Computational results for 50-node models

To compare the network design of hub-and-spoke without congestion and with constant discount factor for the EOS, with congestion and with constant discount factor for the EOS, and with congestion and nonlinear function for the EOS, we utilized the results for 20 nodes network to illustrate the difference between the three models. Table 6.15 shows the results for 20 node network with constant discount factor and without congestion, the first column shows the number of hubs, the second column shows the discount factor, the third column shows the optimal set of hubs, the fourth, fifth, sixth, and seventh columns provides the value of flow directed to each hub ordered in an descending order and the last column is the ratio between the max. and min. flow in the network. Clearly the constant discount model without congestion exploits economies of scale, but fails to account for the resulting congestion. Moreover, in this model the way the costs of economies of scale are modeled does not reflect the real-life aspects of the economies of scale.

No. of hubs	Discount factor	Optimal Hubs	FH1	FH2	FH3	FH4	Max./Min. ratio
3	0.2, 0.4, 0.6	4,12,17	2627432	2569068	558094		4.71
3	0.8	4,8,17	2569068	2462029	723497		3.55
4	0.2	4,12,16,17	2156360	1810993	1229147	558094	3.86
4	0.4,0.6	1,4,12,17	2156360	1810993	1229147	558094	3.86
4	0.8	1,4,8,17	2156360	1645590	1229147	723497	2.98

Table 6.15: Computational results for 20-node models without congestion and with constant discount factor.

For a clear comparison between the models with nonlieanr EOS versus nonlinear EOS with congestion. Table 6.16 shows the results for 20 node network with nonlinear calculation for EOS and without congestion, having the same format as Table 6.15 except for the second column which provides the index of EOS function. Table6.17 shows the results for 20 node network with nonlinear calculation for EOS and with congestion, having the same format as Table 6.16 with a third column providing the value of τ. As shown in Table 6.17 the model with nonlieanr EOS and congestion is always able to achieve a balance on the flow directed to the hubs, the effect of congestion cost is clear always for the case of $\tau = 1.5$ when compared to its counterpart for all the EOS functions, this balancing effect will yield more reality aspects to the design of hub-and-spoke network with nonlinear EOS and congestion effect to balance the utilization of EOS.

No. of hubs	Cost function	Optimal Hubs	FH1	FH2	FH3	FH4	Max./Min. ratio
3	f1	4,7,17	2569068	1926693	1258833		2.04
3	f2	4,7,17	2569068	1926693	1258833		2.04
3	f3	4,7,17	2569068	1844661	1340865		1.92
4	f1	1,4,8,17	2156360	1645590	1229147	723497	2.98
4	f2	4,7,14,17	2156360	1926693	1258833	412708	5.22
4	f3	1,4,12,17	2156360	2030907	1009233	558094	3.86

Table 6.16: Computational results for 20-node models without congestion and nonlinear EOS.

No. of hubs	Cost Function used	τ	Optimal hubs	FH1	FH2	FH3	FH4	Ratio
3	f1	1	4,7,17	2569068	1926693	1258833		2.04
3	f1	1.3	4,7,17	2569068	1926693	1258833		2.04
3	f1	1.5	6,11,17	2162524	1946154	1645916		1.31
3	f2	1	4,7,17	2569068	1926693	1258833		2.04
3	f2	1.3	4,7,17	2569068	1926693	1258833		2.04
3	f2	1.5	4,13,17	2156360	1810993	1787241		1.21
3	f3	1	4,7,17	2569068	1844661	1340865		1.94
3	f3	1.3	4,12,17	2627432	2569068	558094		4.71
3	f3	1.5	4,7,17	2156360	1844661	1753573		1.22
4	f1	1	1,4,8,17	2156360	1645590	1229147	723497	2.98
4	f1	1.3	1,4,8,17	2156360	1645590	1229147	723497	2.98
4	f1	1.5	4,7,17,20	1820455	1469346	1258833	1205960	1.51
4	f2	1	4,7,14,17	2156360	1926693	1258833	412708	5.22
4	f2	1.3	4,7,14,17	2156360	1926693	1258833	412708	5.22
4	f2	1.5	4,7,17,20	1820455	1469346	1258833	1205960	1.51
4	f3	1	1,4,12,17	2156360	2030907	1009233	558094	3.86
4	f3	1.3	1,4,12,17	2156360	2030907	1009233	558094	3.86
4	f3	1.5	1,4,17,19	2156360	1645590	1229147	723497	2.98

Table 6.17: Computational results for 20-node models with congestion and nonlinear EOS.

Finally Table 6.18 shows the results for the model with nonlinear congestion and constant discount factor. Several observations can be made on the results. As the discount factor value increases, the max./min ration decreases for the same value of τ. Also, max./min ration decreases as the value of τ increases. However, as mentioned by O'Kelly and Bryan [83] hub location models that utilize a constant discount factor represent something of an oversimplification. Therefore, if the model were used to study realistic communications or transportation networks, the model would apply a discount that would not be warranted.

No. of hubs	Discount factor	τ	Optimal hubs	FH1	FH2	FH3	FH4	Ratio
3	0.2, 0.4,, 0.6	1.0	4,12,17	2627432	2569068	558094		4.06
3	0.8	1.0	4,8,17	2569068	2462029	723497		3.55
3	0.2, 0.4, 0.6	1.3	4,12,17	2627432	2569068	558094		4.71
3	0.8	1.3	4,8,17	2569068	2462029	723497		3.55
3	0.2	1.5	4,7,17	2156360	1844661	1753573		1.23
3	0.4, 0.6	1.5	7,9,17	2358862	1921341	1474391		1.60
3	0.8	1.5	6,11,17	2162524	1946154	1645916		1.31
4	0.2	1.0	4,12,16,17	2156360	1810993	1229147	558094	3.86
4	0.4, 0.6	1.0	1,4,12,17	2156360	1810993	1229147	558094	3.86
4	0.8	1.0	1,4,8,17	2156360	1645590	1229147	723497	2.98
4	0.2	1.3	4,12,16,17	2156360	1810993	1229147	558094	3.86
4	0.4, 0.6	1.3	1,4,12,17	2156360	1810993	1229147	558094	3.86
4	0.8	1.3	1,4,8,17	2156360	1645590	1229147	723497	2.98
4	0.2	1.5	4,12,16,17	2156360	1645590	1229147	723497	2.98
4	0.4	1.5	1,4,12,17	2156360	1645590	1229147	723497	2.98
4	0.6	1.5	4,7,17,20	1820455	1469346	1258833	1205960	1.51
4	0.8	1.5	4,7,17,20	1728741	1561060	1258833	1205960	1.43

Table 6.18: Computational results for 20-node models with congestion and constant discount factor.

6.3 Summary

In this work we presented a Lagrangean relaxation heuristic to tackle the model of Uncapacitited Single Allocation p Median Hub Location Problem (USA-p-MHLP) incorporating the congestion effect and the economies of scale together. The use of Lagrangean relaxation enabled us to decompose the model into two sub-problems that are solved to optimality at each iteration, and a GRASP algorithm was developed to solve larger instances.

The two main contributions of this analysis provided can be summarized in terms of new insights and considerations for the design of an efficient hub-and-spoke network that is balanced and also captures real-life aspects which are the nonlinearity in the presentation of economies of scale.

Analysis provided based on the results found by the integrated model, as shown in the previous section on OKelly and Bryan [83] and Klincewicz [70] they found that interhub traffic flow tended to be concentrated on a few interhub arcs. That is, some interhub arcs had to carry relatively large traffic flows that were highly discounted, while others had very small amounts of traffic. As mentioned by Bryan [12] analysts should be aware of this potential imbalance in the interhub network (at least as witnessed in these CAB problems) when applying the model to real world networks. One way to mitigate the imbalance effect is to consider congestion effect to count for the incoming flow on the hubs, which is the solution proposed by Elhedhli and Hu [42]. In our work we integrated the congestion and economies of scale in the same model, therefore, we provided more insights about this issue. As shown in the numerical results the imbalance effect is clear as the nonlinearity factor in the EOS function decreases (i.e. more discount is provided) which is the imbalance effect mentioned by Bryan [12] in his study; however, for the same discount rate considering the congestion effect eliminates the imbalance effect found in the network. Thus, the use of the intergated model is justified when it comes to considering real-life aspects of the design of hub-and-spoke network.

The use of the proposed two solution methodologies was justified in the previous section based on the solution quality and computational time required to find the solution for the proposed integrated model. Specifically, the leading edge best commercial solver GAMS required several days to solve instances that took less than 1 hour to be solved using the Lagrangean relaxation heuristic and less than 1 second for the GRASP, and it should be noted that this is for small size instances. For larger instances it was prohibited for GAMS to find a solution even given more that 15 days of running time, while the developed solution methodologies were able to find near-optimal solutions effectively in less than one hour.

CHAPTER 7

Conclusion

Although the main motivation of installing hub-and-spoke networks is to exploit the economies of scale, the way the economies of scale is modeled in the literature is an oversimplification, as there are few publications which considered the use of the nonlinear term to model the economies of scale. Furthermore, while hub-and-spoke networks are an efficient way to exploit economies of scale, they inevitably cause congestion at hubs. Thus, it is reasonable to seek models that achieve a balance between benefits of economies of scale and the drawbacks of congestion, taking into consideration more realistic aspects in the design of hub-and-spoke networks.

Previous studies which considered the congestion effect or the economies of scale separately provided valuable insights for the design of hub-and-spoke networks. However, up to now there is no work in the literature, to the best of our knowledge, to tackle the economies of scale for the single allocation case nor the integration of congestion and economies of scale effects on the design of hub-and-spoke networks. Previous work considered the congestion effect or the economies of scale separately. This is because of the great complication resulting from the incorporation of those two effects in one single model, moreover for the single allocation case the allocation decision variables are binary taking either 0 or 1, unlike the multiple allocation version of HLP where the allocation decision variables are linear taking values from 0 to 1, which presents a more challenging task to be solved.

In addition of having a complicated model that considers the effect of nonlinear representation of the Economies Of Scale and congestion together, the previous work carried out by OKelly and Bryan [83] and Klincewicz [70] they found that interhub traffic flow tended to be concentrated on a few interhub arcs. That is, some interhub arcs had carry relatively large traffic flows that were highly discounted, while others had very small amounts of traffic. As mentioned by Bryan [12] analysts should be aware of this potential imbalance in the interhub network (at least as witnessed in these CAB problems) when applying the model to real world networks. One way to mitigate the imbalance effect is to consider congestion effect to count for the incoming flow on the hubs, which is the solution proposed by Elhedhli and Hu [42]. In our work we integrated the congestion and economies of scale in the same model, therefore, we provided more insights and justifications for the use of integrated model in terms of impact of the structure of the network in addition the routing of flows in the network.

In this thesis we have considered the ”p-hub median problem with single allocations integrating the effect of congestion and economies of scale”, in which the congestion was modeled as a nonlinear convex function and the economies of scale was modeled as concave function”. Two solution methodologies were proposed, in particular, the Lagrangean relaxation heuristic and GRASP. The Lagrangean relaxation heuristic is able to solve efficiently instances up to 25 nodes and GRASP is able to solve larger problems.

The use of the proposed two solution methodologies was justified in Chapter based on the solution quality and computational time required to find the solution for the proposed integrated model. Specifically, the leading edge best commercial solver GAMS required several days to solve instances that took less than 1 hour to be solved using the Lagrangean relaxation heuristic and less than 1 second for the GRASP. It should be also noted that this is for small size instances, and that for larger instances it was prohibited for GAMS to find a solution even given more that 15 days of running time, while the developed solution methodologies were able to find near-optimal solutions effectively in less than one hour.

Another benefit of this study is the insights and the detailed analysis provided by analyzing the integrated model when compared to prevoiusly developed models. As shown in Chapter ,

the integrated model was able to achieve a balance in the network design in terms of flow for the hub nodes as well as a more realistic presentation for the economies of scale that is the main motivation for installing hub-and-spoke networks.
In the computational experiments conducted by OKelly and Bryan [83] and Klincewicz [70] they found that interhub traffic flow tended to be concentrated on a few interhub arcs. That is, some interhub arcs had carry relatively large traffic flows that were highly discounted, while others had very small amounts of traffic. As mentioned by Bryan [12] analysts should be aware of this potential imbalance in the interhub network (at least as witnessed in these CAB problems) when applying the model to real world networks. One way to mitigate the imbalance effect is to consider congestion effect to count for the incoming flow on the hubs, which is the solution proposed by Elhedhli and Hu [42]. In our work we integrated the congestion and economies of scale in the same model, therefore, we provided more insights about this issue. As shown in the numerical results the imbalance effect is clear as the nonlinearity factor in the EOS function decreases (i.e. more discount is provided) which is the imbalance effect mentioned by Bryan [12] in his study; however, for the same discount rate considering the congestion effect eliminates the imbalance effect found in the network. Thus, the use of intergated model is justified when it comes to considering real-life aspects of the design of hub-and-spoke network.

Further research on the ”p-hub median problem with congestion and economies of scale” may consider ways to solve the Mixed Integer Nonlinear Programming formulation on larger instances using special Benders decomposition techniques or hybrid Lagrangean relaxation technique, such as Lagrangean relaxation with Brach-and-Bound, variable fixing, or reduction tests, such as the Lagrangean relxation that is enhanced by reduction tests that limits the set of potential hub locations similar to the work developed by Contreras et. al. [28] which was able to solve large scale problems. Such an approach would reduce the number of decision variables in the integer programming formulation, thus allowing for faster solution times. In addition, as pointed by Klincewicz [70] this may include the development of alternate formulations, and use of valid inequalities and other mathematical programming techniques.

The focus of this work has been on new insights in the design of hub-and-spoke networks that has not been considered before for the single allocation case of HLP. Future work could also consider the multiple allocation case of HLP or the so-called r-allocation model, which was developed recently by Yaman [105] where the non-hub node is allowed to be located to r number of hubs rather than one or all the hubs in the network.

Bibliography

[1] Sue Abdinnour-Helm. A hybrid heuristic for the uncapacitated hub location problem. *European Journal of Operational Research*, 106(2):489–499, 1998.

[2] Sue Abdinnour-Helm and MA Venkataramanan. Solution approaches to hub location problems. *Annals of Operations Research*, 78:31–50, 1998.

[3] Sibel Alumur and Bahar Y Kara. Network hub location problems: The state of the art. *European Journal of Operational Research*, 190(1):1–21, 2008.

[4] Sibel A Alumur, Hande Yaman, and Bahar Y Kara. Hierarchical multimodal hub location problem with time-definite deliveries. *Transportation Research Part E: Logistics and Transportation Review*, 48(6):1107–1120, 2012.

[5] Yu An, Yu Zhang, and Bo Zeng. The reliable hub-and-spoke design problem: Models and algorithms, 2011.

[6] Turgut Aykin. Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *European Journal of Operational Research*, 79(3):501–523, 1994.

[7] Turgut Aykin. Networking policies for hub-and-spoke systems with application to the air transportation system. *Transportation Science*, 29(3):201–221, 1995.

[8] Philippe Barla and Christos Constantatos. Airline network structure under demand uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 36(3): 173–180, 2000.

[9] Silke Baumgartner. *Polyhedral analysis of hub center problems*. PhD thesis, Diploma Thesis, Universität Kaiserslautern, Germany, 2003.

[10] Natashia Boland, Mohan Krishnamoorthy, Andreas T Ernst, and Jamie Ebery. Preprocessing and cutting for multiple allocation hub location problems. *European Journal of Operational Research*, 155(3):638–653, 2004.

[11] Ulf Brännlund. A generalized subgradient method with relaxation step. *Mathematical Programming*, 71(2):207–219, 1995.

[12] Deborah Bryan. Extensions to the hub location problem: Formulations and numerical examples. *Geographical Analysis*, 30(4):315–330, 1998.

[13] Hatice Calık, Sibel A Alumur, Bahar Y Kara, and Oya E Karasan. A tabu-search based heuristic for the hub covering problem over incomplete hub networks. *Computers & Operations Research*, 36(12):3088–3096, 2009.

[14] Paolo M Camerini, Luigi Fratta, and Francesco Maffioli. On improving relaxation methods by modified gradient techniques. In *Nondifferentiable Optimization*, pages 26–34. Springer, 1975.

[15] Ann Melissa Campbell, Timothy J Lowe, and Li Zhang. The¡ i¿ p¡/i¿-hub center allocation problem. *European Journal of Operational Research*, 176(2):819–835, 2007.

[16] James F Campbell. Location and allocation for distribution systems with transshipments and transportion economies of scale. *Annals of Operations Research*, 40(1):77–99, 1992.

[17] James F Campbell. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, 72(2):387–405, 1994.

[18] James F Campbell. Hub location and the p-hub median problem. *Operations Research*, 44(6):923–935, 1996.

[19] James F Campbell. Hub location for time definite transportation. *Computers & Operations Research*, 36(12):3107–3116, 2009.

[20] James F Campbell and Morton E O'Kelly. Twenty-five years of hub location research. *Transportation Science*, 46(2):153–169, 2012.

[21] James F Campbell, Andreas T Ernst, and Mohan Krishnamoorthy. Hub location problems. *Facility location: applications and theory*, 1:373–407, 2002.

[22] James F Campbell, Andreas T Ernst, and Mohan Krishnamoorthy. Hub arc location problems: part iintroduction and results. *Management Science*, 51(10):1540–1555, 2005.

[23] Lázaro Cánovas, Sergio Garcia, and Alfredo Marín. Solving the uncapacitated multiple allocation hub location problem by means of a dual-ascent technique. *European Journal of Operational Research*, 179(3):990–1007, 2007.

[24] Alberto Caprara, Matteo Fischetti, and Paolo Toth. A heuristic method for the set covering problem. *Operations research*, 47(5):730–743, 1999.

[25] Giuliana Carello, F Della Croce, Marco Ghirardi, and Roberto Tadei. Solving the hub location problem in telecommunication network design: a local search approach. *Networks*, 44(2):94–105, 2004.

[26] Selim Çetiner, Canan Sepil, and Haldun Süral. Hubbing and routing in postal delivery systems. *Annals of Operations Research*, 181(1):109–124, 2010.

[27] Jeng-Fung Chen. A hybrid heuristic for the uncapacitated single allocation hub location problem. *Omega*, 35(2):211–220, 2007.

[28] Ivan Contreras, Juan A Díaz, and Elena Fernández. Lagrangean relaxation for the capacitated hub location problem with single assignment. *OR spectrum*, 31(3):483–505, 2009.

[29] Ivan Contreras, Jean-François Cordeau, and Gilbert Laporte. Benders decomposition for large-scale uncapacitated hub location. *Operations research*, 59(6):1477–1490, 2011.

[30] Ivan Contreras, Jean-François Cordeau, and Gilbert Laporte. Stochastic uncapacitated hub location. *European Journal of Operational Research*, 212(3):518–528, 2011.

[31] Ivan Contreras, Juan A Díaz, and Elena Fernández. Branch and price for large-scale capacitated hub location problems with single assignment. *INFORMS Journal on Computing*, 23(1):41–55, 2011.

[32] Claudio B Cunha and Marcos Roberto Silva. A genetic algorithm for the problem of configuring a hub-and-spoke network for a ltl trucking company in brazil. *European Journal of Operational Research*, 179(3):747–758, 2007.

[33] Maria da Graça Costa, Maria Eugénia Captivo, and João Clímaco. Capacitated single allocation hub location problema bi-criteria approach. *Computers & Operations Research*, 35(11):3671–3695, 2008.

[34] Ricardo Saraiva de Camargo, G Miranda Jr, and HP Luna. Benders decomposition for the uncapacitated multiple allocation hub location problem. *Computers & Operations Research*, 35(4):1047–1064, 2008.

[35] Ricardo Saraiva de Camargo, Gilberto de Miranda, and Henrique Pacca L Luna. Benders decomposition for hub location problems with economies of scale. *Transportation Science*, 43(1):86–97, 2009.

[36] Ricardo Saraiva de Camargo, G Miranda Jr, Ricardo Poley Martins Ferreira, and HP Luna. Multiple allocation hub-and-spoke network design under hub congestion. *Computers & Operations Research*, 36(12):3097–3106, 2009.

[37] Ricardo Saraiva De Camargo, Gilberto De Miranda, and Ricardo PM Ferreira. A hybrid outer-approximation/benders decomposition algorithm for the single allocation hub location problem under congestion. *Operations Research Letters*, 39(5):329–337, 2011.

[38] RS De Camargo and G Miranda. Single allocation hub location problem under congestion: Network owner and user perspectives. *Expert Systems with Applications*, 39(3): 3385–3391, 2012.

[39] Jamie Ebery. Solving large single allocation¡ i¿ p¡/i¿-hub problems with two or three hubs. *European Journal of Operational Research*, 128(2):447–458, 2001.

[40] Jamie Ebery, Mohan Krishnamoorthy, Andreas Ernst, and Natashia Boland. The capacitated multiple allocation hub location problem: Formulations and algorithms. *European Journal of Operational Research*, 120(3):614–631, 2000.

[41] Horst A Eiselt and Vladimir Marianov. A conditional p-hub location problem with attraction functions. *Computers & Operations Research*, 36(12):3128–3135, 2009.

[42] Samir Elhedhli and Frank Xiaolong Hu. Hub-and-spoke network design with congestion. *Computers & Operations Research*, 32(6):1615–1632, 2005.

[43] Samir Elhedhli and Huyu Wu. A lagrangean heuristic for hub-and-spoke system design with capacity selection and congestion. *INFORMS Journal on Computing*, 22(2):282–296, 2010.

[44] Andreas Ernst, Houyuan Jiang, and Mohan Krishnamoorthy. Reformulations and computational results for uncapacitated single and multiple allocation hub covering problems. *Unpublished reportt, CSIRO Mathematical and Information Sciences, Australia*, 2005.

[45] Andreas T Ernst and Mohan Krishnamoorthy. Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location science*, 4(3):139–154, 1996.

[46] Andreas T Ernst and Mohan Krishnamoorthy. An exact solution approach based on shortest-paths for p-hub median problems. *INFORMS Journal on Computing*, 10(2): 149–162, 1998.

[47] Andreas T Ernst and Mohan Krishnamoorthy. Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research*, 86:141–159, 1999.

[48] Andreas T Ernst, Horst Hamacher, Houyuan Jiang, Mohan Krishnamoorthy, and Gerhard Woeginger. Uncapacitated single and multiple allocation p-hub center problems. Technical report, CSIRO Mathematical and Information Sciences, Australia.

[49] Andreas T Ernst, Horst Hamacher, Houyuan Jiang, Mohan Krishnamoorthy, and Gerhard Woeginger. Uncapacitated single and multiple allocation p-hub center problems. *Computers & Operations Research*, 36(7):2230–2241, 2009.

[50] Reza Zanjirani Farahani, Masoud Hekmatfar, Alireza Boloori Arabani, and Ehsan Nikbakhsh. Hub location problems: A review of models, classification, techniques and application. *Computers & Industrial Engineering*, 2013.

[51] Thomas A Feo and Mauricio GC Resende. Greedy randomized adaptive search procedures. *Journal of global optimization*, 6(2):109–133, 1995.

[52] Marshall L Fisher. The lagrangian relaxation method for solving integer programming problems. *Management science*, 27(1):1–18, 1981.

[53] Marshall L Fisher. An applications oriented guide to lagrangian relaxation. *Interfaces*, 15(2):10–21, 1985.

[54] Elena O Gavriliouk. Aggregation in hub location problems. *Computers & Operations Research*, 36(12):3136–3142, 2009.

[55] Shahin Gelareh, Stefan Nickel, and David Pisinger. Liner shipping hub network design in a competitive environment. *Transportation Research Part E: Logistics and Transportation Review*, 46(6):991–1004, 2010.

[56] Fred Glover and Gary A Kochenberger. *Handbook of metaheuristics*. Springer, 2003.

[57] Peter G Grove and Morton E O'Kelly. Hub networks and simulated schedule delay. *Papers in Regional Science*, 59(1):103–119, 1986.

[58] Horst W Hamacher and Tanja Meyer. Hub cover and hub center problems. *Departament of Mathematics, University of Kaiserslautern, Germany*, 2006.

[59] Michael Held and Richard M Karp. The traveling-salesman problem and minimum spanning trees: Part ii. *Mathematical programming*, 1(1):6–25, 1971.

[60] Reiner Horst, Panos M Pardalos, and H Edwin Romeijn. *Handbook of global optimization*, volume 2. Springer, 2002.

[61] E Hult, H Jiang, and D Ralph. Exact computational approaches to a stochastic uncapacitated single allocation p-hub center problems. *Judge Business School, University of Cambridge, Trumpington Street, Cambridge, United Kingdom*, 2011.

[62] Aleksandar Ilić, Dragan Urošević, Jack Brimberg, and Nenad Mladenović. A general variable neighborhood search for solving the uncapacitated single allocation p-hub median problem. *European Journal of Operational Research*, 206(2):289–300, 2010.

[63] Rafay Ishfaq and Charles R Sox. Design of intermodal logistics networks with hub delays. *European Journal of Operational Research*, 220(3):629–641, 2012.

[64] Bahar Y Kara and Barbaros C Tansel. On the single-assignment¡ i¿ p¡/i¿-hub center problem. *European Journal of Operational Research*, 125(3):648–655, 2000.

[65] Bahar Y Kara and Barbaros Ç Tansel. The latest arrival hub location problem. *Management Science*, 47(10):1408–1420, 2001.

[66] BY Kara and BC Tansel. The single-assignment hub covering problem: Models and linearizations. *Journal of the Operational Research Society*, 54(1):59–64, 2003.

[67] Hyun Kim and Morton E O'Kelly. Reliable p-hub location problems in telecommunication networks. *Geographical Analysis*, 41(3):283–306, 2009.

[68] John G Klincewicz. Heuristics for the p-hub location problem. *European Journal of Operational Research*, 53(1):25–37, 1991.

[69] John G Klincewicz. Avoiding local optima in thep-hub location problem using tabu search and grasp. *Annals of Operations Research*, 40(1):283–302, 1992.

[70] John G Klincewicz. Enumeration and search procedures for a hub location problem with economies of scale. *Annals of Operations Research*, 110(1-4):107–122, 2002.

[71] Martine Labbé, Hande Yaman, and Eric Gourdin. A branch and cut algorithm for hub location problems with single assignment. *Mathematical programming*, 102(2):371–405, 2005.

[72] Cheng-Chang Lin, Jr-Yung Lin, and Yin-Chieh Chen. The capacitated p-hub median problem with integral constraints: An application to a chinese air cargo network. *Applied Mathematical Modelling*, 36(6):2777–2787, 2012.

[73] Francesco Maffioli. Subgradient optimization methods in integer programming with an application to a radiation therapy problem. 2003.

[74] Vladimir Marianov and Daniel Serra. Location models for airline hubs behaving as m/d/*c* queues. *Computers & Operations Research*, 30(7):983–1003, 2003.

[75] Vladimir Marianov, Daniel Serra, and Charles ReVelle. Location of hubs in a competitive environment. *European Journal of Operational Research*, 114(2):363–371, 1999.

[76] Alfredo Marın, Lázaro Cánovas, and Mercedes Landete. New formulations for the uncapacitated multiple allocation hub location problem. *European Journal of Operational Research*, 172(1):274–292, 2006.

[77] Christopher Mayer and Todd Sinai. Network effects, congestion externalities, and air traffic delays: Or why all delays are not evil. Technical report, National Bureau of Economic Research, 2002.

[78] Gabriela Mayer and Bernd Wagner. Hublocator: an exact solution method for the multiple allocation hub location problem. *Computers & Operations Research*, 29(6):715–739, 2002.

[79] T Meyer, AT Ernst, and M Krishnamoorthy. A 2-phase algorithm for solving the single allocation¡ i¿ p¡/i¿-hub center problem. *Computers & Operations Research*, 36(12): 3143–3151, 2009.

[80] Morton E O'Kelly. Activity levels at hub facilities in interacting networks. *Geographical Analysis*, 18(4):343–356, 1986.

[81] Morton E O'kelly. A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research*, 32(3):393–404, 1987.

[82] Morton E O'Kelly. Hub facility location with fixed costs. *Papers in Regional Science*, 71 (3):293–306, 1992.

[83] Morton E OKelly and DL Bryan. Hub location with flow economies of scale. *Transportation Research Part B: Methodological*, 32(8):605–616, 1998.

[84] Morton E O'Kelly, Deborah Bryan, Darko Skorin-Kapov, and Jadranka Skorin-Kapov. Hub network design with single and multiple allocation: A computational study. *Location Science*, 4(3):125–138, 1996.

[85] F Selcen Pamuk and Canan Sepil. A solution to the hub center problem via a single-relocation algorithm with tabu search. *IIE Transactions*, 33(5):399–411, 2001.

[86] Juanjo Peiró, Ángel Corberán, and Rafael Martí. Grasp for the uncapacitated r-allocation p-hub median problem. *Computers & Operations Research*, 43:50–60, 2014.

[87] Eric Pels and Erik T Verhoef. The economics of airport congestion pricing. *Journal of Urban Economics*, 55(2):257–277, 2004.

[88] Melquíades Pérez, Francisco Almeida, and J Marcos Moreno-Vega. A hybrid grasp-path relinking algorithm for the capacitated p–hub median problem. In *Hybrid Metaheuristics*, pages 142–153. Springer, 2005.

[89] Hasan Pirkul and David A Schilling. An efficient procedure for designing single allocation hub and spoke systems. *Management Science*, 44(12-Part-2):S235–S242, 1998.

[90] Hrvoje Podnar, Jadranka Skorin-Kapov, and Darko Skorin-Kapov. Network cost minimization using threshold-based discounting. *European Journal of Operational Research*, 137(2):371–386, 2002.

[91] Bo Qu and Kerui Weng. Path relinking approach for multiple allocation hub maximal covering problem. *Computers & Mathematics with Applications*, 57(11):1890–1894, 2009.

[92] Illia Racunica and Laura Wynter. Optimal location of intermodal freight hubs. *Transportation Research Part B: Methodological*, 39(5):453–477, 2005.

[93] Marianne Raffarin. Congestion in european airspace. *Journal of Transport Economics and Policy (JTEP)*, 38(1):109–125, 2004.

[94] Marcus Randall. Solution approaches for the capacitated single allocation hub location problem using ant colony optimisation. *Computational Optimization and Applications*, 39(2):239–261, 2008.

[95] Mihiro Sasaki and Masao Fukushima. On the hub-and-spoke model with arc capacity constraints. *Journal of the Operations Research Society of Japan-Keiei Kagaku*, 46(4): 409–428, 2003.

[96] Hanif D Sherali and Osman Ulular. A primal-dual conjugate subgradient algorithm for specially structured linear and convex programming problems. *Applied Mathematics and Optimization*, 20(1):193–221, 1989.

[97] Thaddeus Sim, Timothy J Lowe, and Barrett W Thomas. The stochastic p-hub center problem with service-level constraints. *Computers & Operations Research*, 36(12):3166–3177, 2009.

[98] Darko Skorin-Kapov and Jadranka Skorin-Kapov. On tabu search for the location of interacting hub facilities. *European Journal of Operational Research*, 73(3):502–509, 1994.

[99] Darko Skorin-Kapov, Jadranka Skorin-Kapov, and Morton O'Kelly. Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research*, 94(3):582–593, 1996.

[100] Haluk Topcuoglu, Fatma Corut, Murat Ermis, and Gülsah Yilmaz. Solving the uncapacitated hub location problem using genetic algorithms. *Computers & Operations Research*, 32(4):967–984, 2005.

[101] Adriano D Vasconcelos, Carlos D Nassi, and Luiz AS Lopes. The uncapacitated hub location problem in networks under decentralized management. *Computers & Operations Research*, 38(12):1656–1666, 2011.

[102] Bernd Wagner. Model formulations for hub covering problems. *Journal of the Operational Research Society*, 59(7):932–938, 2007.

[103] Huyu Wu. Capacitated hub-and-spoke system design with congestion. Master's thesis, University of Waterloo.

[104] Huyu Wu. *Capacitated Hub-and-spoke System Design with Congestion*. PhD thesis, University of Waterloo, 2005.

[105] Hande Yaman. Allocation strategies in hub networks. *European Journal of Operational Research*, 211(3):442–451, 2011.

[106] Hande Yaman and Sourour Elloumi. Star p-hub center problem and star p-hub median problem with bounded path lengths. *Computers & Operations Research*, 39(11):2725–2732, 2012.

[107] Hande Yaman, Oya Ekin Karasan, and Bahar Y Kara. Release time scheduling and hub location for next-day delivery. *Operations Research*, 60(4):906–917, 2012.

[108] Ta-Hui Yang. Stochastic air freight hub location and flight routes planning. *Applied Mathematical Modelling*, 33(12):4424–4430, 2009.