Resource management in transportation networks: addressing challenges and optimizing efficiency

by

Mohammad Maleki

B.S., Golpayegan College of Engineering, 2008
M.S., Tarbiat Modares University, 2011

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Industrial and Manufacturing Systems Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

The American economy heavily relies on transportation, which contributed 8% to the GDP of the United States in 2020. The transportation sector, being the fourth-largest contributor to the overall GDP, encompasses various industries such as airlines, trucking, railroads, shipping, logistics firms, and transportation infrastructure providers. The efficiency of transport systems plays a vital role in facilitating better access to markets, employment opportunities, and additional investments. However, inefficiencies within the transportation sector, including poor resource management, pose challenges that need to be addressed for enhanced sector-wide efficiency and economic benefits.

This dissertation endeavors to develop effective optimization techniques and algorithms to tackle diverse transportation challenges, aiming to optimize resource allocation and enhance resource utilization within the transportation sector. This abstract highlights the key research contributions and findings of the dissertation:

Firstly, the dissertation focuses on addressing hub location problems in transportation networks. Existing hub covering models often lead to congestion and inefficiency, compromising network coverage. To overcome this, two hub covering location models are proposed, taking into account the busy probability of hub nodes. The first model assumes the known number of servers in each hub, while the second model considers the number of servers in each hub as a decision variable. Metaheuristics based on the Genetic algorithm and Tabu Search are developed to solve the models, demonstrating their efficiency through experimental results on American Airlines domestic flights in 2019.

Next, the complexity of the multiple allocation hub maximal covering problem (MAHMCP) is explored. A branch and cut approach is developed to solve this problem, providing

a stronger theoretical model compared to previous approaches. The effectiveness of the proposed model is demonstrated through an empirical study using the Australia Post (AP) dataset, showcasing optimal solutions and fast run times for instances up to 100 nodes.

Furthermore, the dissertation analyzes the chassis inventory management problem in an intermodal transportation system. A chassis connects shipping containers to trucks. Having a chassis available at the right time and place is essential for efficient loading, unloading and subsequent transportation of goods. Among the most significant challenges that intermodal transportation companies face is chassis shortages at terminals. To address this, A multi-time period mathematical model with n terminals is developed to determine the daily moves of chassis and empty containers between terminals, ensuring efficient loading, unloading, and transportation of goods while minimizing costs, and meeting demand requirements.

Lastly, the dissertation addresses one of the main challenges faced by the US truckload industry - high turnover rates among truckers due to demanding working conditions. To mitigate this challenge, a relay network approach is proposed, consisting of a depot and relay location problem as well as a routing problem. This comprehensive approach aims to reduce drivers' away-from-home times.

In conclusion, this dissertation emphasizes the importance of addressing sector-specific challenges and optimizing resource allocation and utilization within the transportation sector. The developed optimization techniques and algorithms contribute to enhancing resource management and system efficiency.

Resource management in transportation networks: addressing challenges and optimizing efficiency

by

Mohammad Maleki

B.S., Golpayegan College of Engineering, 2008
M.S., Tarbiat Modares University, 2011

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Industrial and Manufacturing Systems Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Ashesh Kumar Sinha

Copyright

Abstract

The American economy heavily relies on transportation, which contributed 8% to the GDP of the United States in 2020. The transportation sector, being the fourth-largest contributor to the overall GDP, encompasses various industries such as airlines, trucking, railroads, shipping, logistics firms, and transportation infrastructure providers. The efficiency of transport systems plays a vital role in facilitating better access to markets, employment opportunities, and additional investments. However, inefficiencies within the transportation sector, including poor resource management, pose challenges that need to be addressed for enhanced sector-wide efficiency and economic benefits.

This dissertation endeavors to develop effective optimization techniques and algorithms to tackle diverse transportation challenges, aiming to optimize resource allocation and enhance resource utilization within the transportation sector. This abstract highlights the key research contributions and findings of the dissertation:

Firstly, the dissertation focuses on addressing hub location problems in transportation networks. Existing hub covering models often lead to congestion and inefficiency, compromising network coverage. To overcome this, two hub covering location models are proposed, taking into account the busy probability of hub nodes. The first model assumes the known number of servers in each hub, while the second model considers the number of servers in each hub as a decision variable. Metaheuristics based on the Genetic algorithm and Tabu Search are developed to solve the models, demonstrating their efficiency through experimental results on American Airlines domestic flights in 2019.

Next, the complexity of the multiple allocation hub maximal covering problem (MAHMCP) is explored. A branch and cut approach is developed to solve this problem, providing

a stronger theoretical model compared to previous approaches. The effectiveness of the proposed model is demonstrated through an empirical study using the Australia Post (AP) dataset, showcasing optimal solutions and fast run times for instances up to 100 nodes.

Furthermore, the dissertation analyzes the chassis inventory management problem in an intermodal transportation system. A chassis connects shipping containers to trucks. Having a chassis available at the right time and place is essential for efficient loading, unloading and subsequent transportation of goods. Among the most significant challenges that intermodal transportation companies face is chassis shortages at terminals. To address this, A multi-time period mathematical model with n terminals is developed to determine the daily moves of chassis and empty containers between terminals, ensuring efficient loading, unloading, and transportation of goods while minimizing costs, and meeting demand requirements.

Lastly, the dissertation addresses one of the main challenges faced by the US truckload industry - high turnover rates among truckers due to demanding working conditions. To mitigate this challenge, a relay network approach is proposed, consisting of a depot and relay location problem as well as a routing problem. This comprehensive approach aims to reduce drivers' away-from-home times.

In conclusion, this dissertation emphasizes the importance of addressing sector-specific challenges and optimizing resource allocation and utilization within the transportation sector. The developed optimization techniques and algorithms contribute to enhancing resource management and system efficiency.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am grateful for the support and guidance I received during my PhD journey in Industrial & Manufacturing Systems Engineering at Kansas State University.

Special thanks to my advisor, Dr. Ashesh Kumar Sinha, for his exceptional mentorship and unwavering support. I am truly grateful for his expertise and commitment to my academic and personal growth, which have been invaluable. I am grateful for the countless hours spent discussing ideas, refining my research methodology, and pushing me to reach my full potential. His passion for the subject matter and his continuous encouragement have made a significant impact on my development as a researcher.

I extend my heartfelt appreciation to my committee members, Dr. Jessica Heier Stamm, Dr. Chih-Hang Wu, and Dr. Hongyu Wu, for their insightful feedback and valuable contributions to my research. I am particularly thankful to Dr. Heier Stamm for the opportunity to serve as her graduate teaching assistant for IMSE 560 - Operations Research I during my first semester, as it provided me with invaluable experience. Her guidance and support have had a significant impact on my academic and personal development. I extend my sincere gratitude to Dr. Chih-Hang Wu for teaching the foundational courses, IMSE 780 - Methods of Operations Research and IMSE 881 - Linear Programming. These courses laid the groundwork for my research and provided me with essential tools and knowledge. I am deeply thankful to Dr. Hongyu Wu for his invaluable expertise and guidance, which have played a vital role in shaping my research journey.

I am grateful to Dr. Todd Easton for the opportunity to serve as his graduate teaching assistant for multiple courses. His mentorship and belief in my abilities have been transformative. The recognition I received with the Outstanding Graduate Teaching Assistant award is truly

appreciated. Working with Dr. Easton has not only enhanced my teaching skills but also deepened my understanding of the subject matter.

I express my sincere appreciation to Dr. David Ben-Arieh for allowing me to serve as his graduate teaching assistant for IMSE 633 - Production Planning and Inventory Control. His expertise and support have been truly invaluable, and I am grateful for the opportunity to contribute to student education.

My heartfelt gratitude goes to Dr. Bradley Kramer, our esteemed Department Head, for his unwavering encouragement and invaluable support throughout my academic journey. I also extend my deep appreciation to the department staff, including Ms. Vicky Geyer, Ms. Danielle Brooks, and Ms. Debbie Harper, for their assistance, support, and administrative help.

I am grateful for the unwavering support of my peers throughout this journey, particularly Dr. Mohammad Bisheh and Dr. Ali Tolooie, who shared their invaluable experiences with me.

To my friends in Manhattan, who have made Kansas feel like home, they have become my family, and the memories we have shared will always hold a special place in my heart.

Furthermore, I want to express my deepest gratitude to Dr. Nahid Majlesinasab, my wife, for her unwavering support, understanding, and encouragement during the challenging times of my doctoral journey. Her presence and support have been a constant source of strength and motivation.

Finally, I want to extend my heartfelt appreciation to my parents, my two brothers, and my sister. Their unwavering love, support, and encouragement have been instrumental in my accomplishments. Despite the physical distance, their presence in my life remains a constant source of inspiration.

Thank you all for being an integral part of my academic and personal journey.

Dedication

To Shahanshah Aryamehr and Shahbanou Farah

Chapter 1 - Introduction

Transportation plays a vital role in a nation's economic and social fabric, facilitating the movement of people, goods, and services. Efficient transport systems are essential for unlocking economic potential, improving market access, generating job opportunities, and attracting additional investment. Conversely, a deficient transport system can hinder economic growth, limit opportunities, and adversely impact the quality of life. Inefficiencies within transportation systems arise due to various factors, including poor resource management. Addressing the sector-specific challenges and improving overall efficiency is of paramount importance to ensure economic prosperity and social well-being.

Efficient transport systems offer a plethora of economic and social benefits. Firstly, they enable better access to markets, allowing businesses to reach a wider customer base and facilitating trade. Improved connectivity reduces transportation costs, enhancing market competitiveness and enabling businesses to capitalize on emerging opportunities. Secondly, efficient transport systems play a pivotal role in job creation. They support the transportation industry itself, providing employment opportunities for individuals involved in airlines, trucking, railroads, shipping, logistics, and transportation infrastructure. Moreover, by improving access to markets, transport systems stimulate economic activities in various sectors, leading to additional job opportunities and economic growth.

Despite their importance, transportation systems face various challenges that impede their efficiency. One significant challenge lies in poor resource management. Inadequate utilization of resources, such as vehicles, infrastructure, and human capital, can result in suboptimal system performance, congestion, delays, and increased costs. Inefficient resource allocation and planning

can lead to underutilized capacity in some areas and overburdened infrastructure in others, creating bottlenecks and hindering smooth operations.

Additionally, different transportation sectors face unique challenges that require sector-specific solutions. For instance, the air transportation sector confronts congestion issues at hub nodes during peak times, leading to reduced network coverage and operational inefficiencies. The truckload industry grapples with high driver turnover due to prolonged periods away from home, affecting productivity and service quality. Similarly, intermodal transportation systems face challenges in managing chassis inventory efficiently to ensure seamless loading, unloading, and transportation of goods.

Through the development of advanced optimization techniques, this research aims to improve the allocation and utilization of resources, reduce congestion, enhance operational efficiency, and minimize costs. By addressing the specific challenges faced by each transportation sector, the study seeks to enhance the overall efficiency of the transportation system and unlock its full economic potential. Ultimately, the improved resource management and optimization approaches are expected to yield economic benefits, improve the quality of life, and contribute to the overall development of the nation.

In the subsequent chapters, we will delve into the specific transportation challenges, propose novel optimization models and algorithms, and present empirical studies to validate their effectiveness. The findings of this research have the potential to guide policymakers, transportation planners, and industry stakeholders in making informed decisions to optimize resource management and improve the efficiency of transportation networks. It is important to address the challenges specific to each sector of transportation so the whole sector becomes more efficient and the economy benefits.

1.1 Research motives and objectives

The main objective of this research is to develop new optimization techniques and algorithms for modeling a variety of transportation challenges in order to improve resource management and make transportation networks more efficient. By optimizing resource allocation and decision-making processes, transportation networks can become more efficient, resilient, and sustainable. The focus is on modeling and optimizing various aspects, including hub location and allocation problems, chassis inventory management problem, and depot and relay location problem. This research includes four main research objectives as follows:

- **Research objective 1: developing efficient hub maximal covering models considering service availability**

We consider hub maximal covering problem (HMCP), where the flow between the origin and the destination is transferred via multiple hubs such that the total flow covered by located hubs is maximized. However, the existing literature on HMCP, specifically in relation to uncertainties and congestion effects, is limited. For example, during rush-hour traffic at a major airport, the runways can become unavailable, causing long waiting times for planes. Motivated by this gap, the first objective is to develop hub maximal covering location problems while considering unavailability times for hub servers (e.g., runways in an airport hub). Additionally, it is important to note that this problem is NP-hard, necessitating efficient solutions for large-scale problems. The following tasks outline how we can achieve this research objective.

Task 1A: To the best of our knowledge, the busy probability has not received much attention in the hub covering location problem. A busy probability is integrated into the model and the importance of this is demonstrated. We analyze two variants of hub covering location models:

(*i*) fixed number of servers, and (*ii*) variable number of servers. The latter case results in a mix integer program with a non-linear objective function. To solve the model using commercial solvers such as GUROBI, we transform the objective function to a quadratic version. We also show how the change in the number of runways might impact the hub network configuration and the flow coverage.

Task 1B: We develop two metaheuristics based on Genetic algorithm (GA) and Tabu Search (TS) to solve large-scale NP-hard models. Furthermore, we propose a Lemma that makes use of the problem structure and finds the optimal allocation of demand points to the hubs in one simple step. Both algorithms take advantage of this lemma which eliminates the need for solving an allocation optimization problem whenever the location of hubs changes due to having neighborhood moves or crossover operators. We show that the proposed methods outperform traditional commercial software for small and medium size problems and can efficiently solve larger problems that commercial software cannot solve.

- **Research objective 2: developing a new model for the MAHMCP**

After conducting an extensive literature review on hub location problems, it has come to our attention that there has been limited focus on the multiple allocation hub maximal covering problem (MAHMCP). In particular, the model proposed by Qu and Weng (Qu & Weng, 2009) stands out as one of the most recent MAHMCP models; however, there is considerable room for improvement. Notably, Campbell (Campbell J. F., 1994) did not provide any numerical results for this problem, while Qu and Weng's path relinking (PR) algorithm only yields approximate solutions for $n > 25$, where *n* represents the number of nodes.

In their study, Qu and Weng employed the AP dataset and utilized the LINGO solver to solve their model. Furthermore, they evaluated their algorithm using a specific case involving the

hub airport locations for Chinese aerial freight flows between 82 cities in 2002. Their findings indicated that, for 82 nodes, LINGO would require more than 200 hours to obtain the optimal solution.

Given the potential applications of this problem in transportation, telecommunications, and delivery systems, where hub nodes play a vital role in receiving, collecting, and delivering commodities, any enhancements in solution quality and computational runtime would have a significant impact. Therefore, our objective is to develop a novel MAHMCP model that surpasses previous approaches in strength while minimizing the number of binary variables required. The subsequent task outlines the detailed steps involved in achieving this goal.

Task 2A: In this study, we introduce an $O(n^2)$ mixed-integer programming formulation for the MAHMCP and establish its superiority over one of the most recent MAHMCP models. When referring to the model's complexity as $O(n^2)$, it is important to explicitly emphasize that this measure quantifies the growth rate of computational resources with respect to the number of variables involved in the system. To assess the effectiveness of our new model, we conduct an empirical study using the Australia Post (AP) dataset as a benchmark.

By utilizing this formulation, we demonstrate exceptional performance in terms of both runtime and solution quality. Our results indicate that the proposed model achieves significant improvements compared to existing approaches.

- **Research objective 3**: **resource management and cost minimization: A multi-period model for intermodal terminals**

When it comes to shipping and logistics, intermodal transportation refers to the movement of a container using multiple modes of transportation, including roads, rails, air, and oceans. By combining these modes effectively, intermodal transportation offers a more efficient and cost-

effective approach to shipping compared to relying solely on long-haul trucks. A crucial aspect of intermodal transportation is the availability of chassis, which are used to connect shipping containers to trucks. Having an adequate supply of chassis at the right time and place is essential for smooth loading, unloading, and transportation of goods.

One of the major challenges faced by intermodal transportation companies is the shortage of chassis at terminals. This shortage leads to long queues, congestion at terminals, and delays in the supply chain. Consequently, it increases costs and contributes to higher driver turnover rates due to extended waiting times. Despite the existing literature on chassis inventory management, the focus has predominantly been on the maritime industry. Studies that examine truck-terminals often overlook crucial factors such as chassis management, demand for loaded containers, multiple time periods, multiple terminals, and delivery times.

Considering these gaps in the literature, this study aims to address the issue of chassis shortage at multiple terminals over an extended period, while accounting for both deterministic and uncertain events. To achieve this goal, the following tasks outline a detailed approach.

Task 3A: We formulate a mathematical model that operates across multiple time periods and *n* terminals. This model determines various aspects such as the total number of chassis and empty containers within the system, the inventory of empty containers and chassis at each terminal at the end of each day, the number of empty containers relocated between terminals, and the mode of relocation. Additionally, it considers the number of chassis that are transferred between terminals using trucks. The primary objective of this model is to fulfill the total demand for empty and loaded containers, while simultaneously minimizing the overall cost.

Task 3B: A scenario-based mathematical model in which disruptions in terminals are taken into account is developed. In light of the results of each scenario, we recommend that some terminals serve as intermodal hub locations.

Task 3C: In order to make the model more realistic and accurate, a more detailed analysis of the problem is being proposed. The original model assumed that each train could haul 300 containers per day, and that two trains would be available at the beginning of each day to transport containers. However, this simplified approach may not capture the complexities and constraints of the real-world scenario. To enhance the model, new decision variables and constraints will be defined to accurately determine train movements and container transportation. These additions will allow for a more precise and realistic representation of the problem.

- **Research objective 4: developing a depot and relay location model to mitigate high turnover rates in the trucking industry**

American's economy relies heavily on truck driving, but the industry has one of the highest turnover rates. Truckers have quit at alarming rates for decades, leading to a chronic shortage. The working conditions require drivers to be away from home, alone for days and sometimes weeks. Additionally, transit delays, and lifestyle-related health pressures, are the primary causes of a high driver turnover rate. The high turnover rate makes what consumers buy at the store more expensive. Companies spend thousands of dollars on sign-on bonuses, training, and recruiting each new driver. Also, as a result of the driver shortage, the entire supply chain is delayed, leading to higher costs. In turn, those companies pass along the additional costs to consumers. The last objective of this project, therefore, is to develop a mathematical model that specifies some locations as relay points so that drivers can switch loads and trucks. This could reduce away-from-home times for drivers, thereby reducing high turnover rates. In fact, having less time away from

home allows them to rest at home rather than at rest stops along the way. Besides trucking, other industries such as airlines could benefit from this, making it a significant problem. Typically, pilots spend around 3 nights away from home a week, and in extreme cases, many cargo pilots can spend several weeks away from home at once[1]. The turnover rate for the US pilots was around 46% in 2020. Sometimes airlines must drop their routes due to a lack of pilots. This happened in July 2022, when American Airlines, the only carrier providing regular service to Dubuque in Iowa, dropped its routes due to a lack of pilots. In fact, staffing shortage is one of the reasons for long delays or cancellations[2]. It is therefore possible to modify this problem and use it for managing flight resources. Following are the tasks that will help us achieve this research objective.

Task 4A: In previous models, depot and relay locations were either assumed to be known or a subset of potential locations. However, in our research, we introduce a novel concept called "link type," which treats depot and relay locations as decision variables. We formulate a depot-relay-point location and truck routing problem with pickup and delivery, aiming to determine the best depot locations, optimal relay locations, and the sequence of nodes for truck routing. The primary objective of the model is to minimize the total distance traveled by all trucks.

Task 4B: We compare the proposed model with a well-established existing model in the literature where either relay points or depots are pre-designated. In every scenario we examine, we show the new model consistently outperforms the existing model in terms of objective value and solution time.

---

[1] You're always at work! How often are pilots REALLY at home? | flyingbynumbers.com

[2] Airlines forced to drop service at these US airports due to the pilot shortage | CNN Business

These research findings can be used to design more efficient transportation networks for a variety of transport systems. To be more specific, this research helps 1) airline carriers design hub networks in order to reduce long queues of planes waiting for take-off or landing, 2) trucking industry develop efficient relay networks to reduce drivers' away from home times, thereby lowering their turnover rates, and 3) intermodal railroad companies specify certain terminals as hubs, which are key in resolving chassis shortages. Additionally, the algorithms presented in this research are highly efficient and effective at solving large-scale problems. It is possible to modify the models and algorithms so that they can be applied to a variety of problems with similar structures.

1.2 Dissertation outline

This dissertation consists of seven chapters. Chapter 2 delves into the related literature, providing a comprehensive review. Chapters 3 to 6 concentrate on specific problems within transportation and logistics systems, offering in-depth analyses and insights.

In Chapter 3, the focus is on hub covering location problems with an application in air transportation. It introduces two hub covering models that consider the busy probability of hub nodes, accounting for congestion and inefficiency during peak times. Metaheuristics based on the Genetic algorithm and Tabu Search are developed to solve these NP-hard models, demonstrating their efficiency and the importance of considering busy fractions of servers.

Chapter 4 explores the complexity of the multiple allocation hub maximal covering problem (MAHMCP). A branch and cut approach is proposed to solve this problem, offering a theoretically stronger formulation. Empirical results using the Australia Post dataset showcase the

model's effectiveness, providing high-quality solutions and fast run times for instances up to 100 nodes.

Chapter 5 analyzes the chassis inventory management problem in intermodal transportation systems. A mathematical model is developed to determine the optimal daily moves of chassis between terminals, considering the total demand for loaded and empty containers while minimizing costs. This approach has the potential to enhance the efficiency of loading, unloading, and subsequent transportation of goods.

Chapter 6 addresses a significant challenge in the US truckload industry – high driver turnover rates due to long periods away from home. This research presents a relay network approach, encompassing a depot and relay location problem, and a routing problem. The goal is to reduce drivers' away-from-home times and enhance overall working conditions.

Finally, Chapter 7 concludes the dissertation by summarizing the research contributions, emphasizing the development of optimization techniques and algorithms to enhance resource management in transportation. The findings underscore the importance of addressing sector-specific challenges and optimizing resource allocation and utilization. The chapter also outlines future research directions to further enhance transportation system efficiency and sustainability.

Chapter 2 – Literature review

The field of transportation faces various challenges that require effective solutions to optimize resource allocation and enhance overall system efficiency. This literature review examines several prominent challenges within the transportation sector and proposes innovative approaches to address them. The first challenge in the transportation sector revolves around the efficient location of hubs, where existing hub covering models frequently encounter congestion and inefficiency. Furthermore, the complexity associated with hub covering problems poses an additional challenge that requires careful consideration and effective solutions. The next challenge focuses on chassis inventory management in intermodal transportation, aiming to ensure the availability of chassis at the right time and place for efficient loading, unloading, and transportation of goods. The last challenge pertains to the high turnover rates among truckers in the trucking industry, requiring the development of a depot and relay location model that reduces drivers' away-from-home times. In this chapter, by delving into the existing literature and exploring relevant studies in each of these areas, this literature review aims to provide a comprehensive understanding of the challenges and advancements in optimizing transportation systems and propose innovative approaches to enhance resource management.

2.1 Hub location problem

In recent decades, numerous novel mathematical models have emerged for the purpose of facility location, aiming to minimize costs or maximize coverage of demand. These models have found applications in diverse fields and industries (Maleki, Majlesinasab, & Sepehri, 2014; Sepehri, Maleki, & Majlesinasab, 2013; Maleki, Majlesinasab, & Sinha, 2023). Before diving into the problem at hand, it is important to understand the concept of a hub and its significance. A hub

is typically defined as an airport, station, or similar facility that serves as a central location for numerous activities and operations. In the context of transportation or logistics, hubs play a crucial role in efficiently connecting origins and destinations. With that understanding, let's explore the hub location problem, which is an intriguing topic in location theory. Let $G = (V, E)$ be a complete undirected graph with the node set $V = \{1, 2, \ldots, n\}$. The node set (V) includes origins, destinations, and potential hub locations. Each pair of nodes is connected by an arc (route) (i, j) with cost c_{ij} (distance, time, etc.). W is the flow (demand) matrix and $w_{ij} \in W$ is defined as the demand from origin i to destination j. Hub facilities capitalize on economies of scale by consolidating flows, instead of directly serving each origin-destination pair. This consolidation occurs along the route from the origin to the hub, from the hub to the destination, and even between multiple hubs. Implementing hubs effectively reduces the overall number of transportation links required between origins and destinations. Suppose that a fully connected network has k nodes and no hub node, then it has $k \times (k - 1)$ origin–destination links. In contrast, if only one node is selected to connect all other nodes, there will only be $2 \times (k - 1)$ connections to serve all origin–destination pairs. Figure 2-1 provides an illustrative example of the advantage of hubs. A fully connected network with three nodes without hubs has six links (panel *a*), whereas a network with three nodes where one of them is hub has only four links (panel *b*). It is possible to extend this idea to networks with more than one hub node. Thus, hub networks can better serve demand pairs than fully connected structures by utilizing fewer resources.

Hub location problem spans its implementation to transportation systems, telecommunication systems, and delivery systems where more than one site delivers commodities such as data, people, or postal packages. These systems can be improved using hubs that route

flow between origin-destination pairs and gain economic profits from dense transportation (Daskin, 1997).

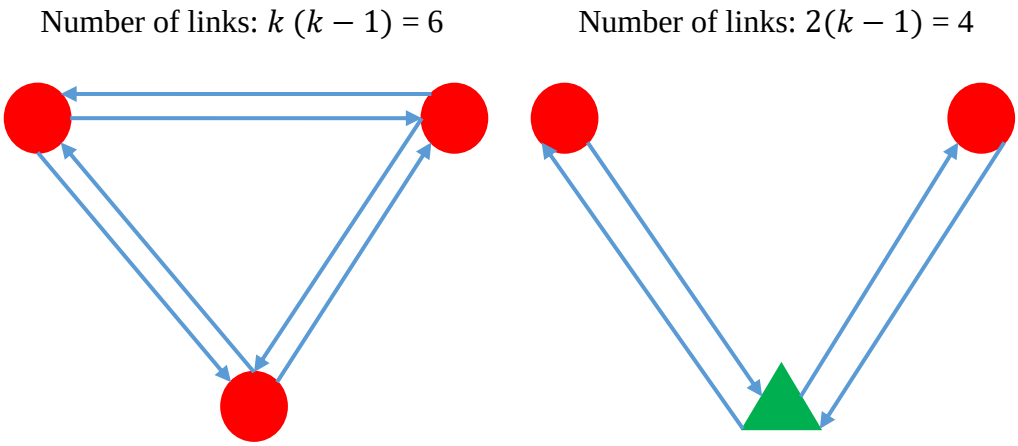

Figure 2-1 Reduction of transportation links using hub network

2.1.1 Categories of hub location problems

The most prevalent hub location problems can be categorized into three main groups: *p*-hub median problems, *p*-hub center problems, and hub covering problems. Each of these groups further consists of two subgroups: single allocation hub problem, where the flow between an origin-destination pair is routed through a single route, and multiple allocation hub problem, where the flow can be distributed across multiple routes. The hubs themselves may have capacity restrictions, either on the total flow entering the hubs (Drezner & Hamacher, 2004) or on the flow leaving non-hub locations and entering the hubs (Andreas T. Ernst & Krishnamoorthy, 1999).

2.1.1.1 The *p*-hub median problem

The *p*-hub median problem locates *p* hubs among a given set of nodes with the goal of minimizing the total transportation cost. O'Kelly was the first who proposed a quadratic integer programming formulation for a general hub location problem (E.O'kelly, 1987). This model is recognized as the single allocation *p*-hub median problem (SApHMP). In the SApHMP, each non-hub node is connected to a single hub. He developed two heuristics to solve the model on the CAB dataset. Later, O'Kelly presented a quadratic integer programming for the single allocation hub location problem with fixed costs and determined the optimal number of hubs (O'Kelly, 1992). Campbell (Campbell J. F., 1994), Skorin-Kapov et al. (Skorin-Kapov, Skorin-Kapov, & O'Kelly, 1996) and Ernst and Krishnamoorthy (T.Ernst & Mohan, 1996) formulated various linearizations of the quadratic models. Campbell also added a constraint to the model which put a capacity limit on hubs (Campbell J. F., 1994). Ebery presented two different types of the problem including the uncapacitated single allocation *p*-hub median problem and the *p*-hub allocation problem. Compared to previous formulations, the proposed models used fewer variables (Ebery, 2001).

The multiple allocation hub location problem is another type of problem in which each origin–destination pair can be transferred via more than one route. Campbell (Campbell J. F., 1992) proposed the first multiple allocation *p*-hub median problem. Campbell also introduced two models for the uncapacitated multiple allocation hub location problem (UMAHLP) and the capacitated multiple allocation hub location problem (CMAHLP) (Campbell J. F., 1994). Later, Ernst and Krishnamoorthy (Andreas T. Ernst & Krishnamoorthy, 1999), Ebery et al. (Ebery, Krishnamoorthy, Ernst, & Boland, 2000), Marin et al. (Marín, Cánovas, & Landete, 2006) and Wagner (Wanger, 2007) improved Campbell's models. Marianov and Serra (Marianov & Serra, 2003) concentrated on congestion effects on airline networks and modeled hub airports as M/D/c

queuing systems. They added a probabilistic constraint to the uncapacitated multiple allocation hub location model that limits the probability of having more than a given number of airplanes waiting in a queue. The objective of the model was to minimize the flow assignment cost and fixed cost of locating the hubs. They proposed two models: The first model assumed that the number of servers in each hub is known. In the second model, they incorporated the number of servers in each hub as a decision variable. While the first model was successfully solved using a Tabu Search algorithm, the second model remained unsolved. Köksalan and Soylu (Köksalan & Soylu., 2010) presented a bi-criteria uncapacitated, multiple allocation *p*-hub median location problem. The first objective function is the minimization of the total transportation cost considering a positive local transfer cost. The second objective function is the minimization of the total traveling cost between hub locations and other nodes. A general capacitated p-hub median model was introduced by Lin et al. (Lin, Lin, & Chen, 2012), which considered economies of scale and integral constraints on paths. The objective was to minimize the total cost in a pure capacitated hub-and-spoke network by selecting a specific *p* from a group of potential hubs.

2.1.1.2 The *p*-hub center problem

The primary concern of the p-hub center problem is the service time, and its objective is to determine the location of *p* hubs and assign non-hubs to these hubs in a way that minimizes the maximum distance between any pair of source and destination points (Campbell, Lowe, & Zhang, 2007). The *p*-hub center model, as introduced by Campbell (Campbell J. F., 1994), holds significant value due to its practical implications in various real-world scenarios. One such application is the identification of optimal locations for emergency service facilities and vehicles. In addition, the *p*-hub center model also provides valuable insights into worst-case scenarios, specifically regarding the maximum cost or travel time between locations. Other researchers also

studied and developed the *p*-hub center problem (e.g., Kara and Tansel (Kara & Tansel., 2000), Campbell et al. (Campbell, Lowe, & Zhang, 2007), Yaman et al. (Yaman, Y.Kara, & Ç.Tansel, 2007) and Ernst et al. (T.Ernst, Hamacher, Jiang, Krishnamoorthy, & Woeginger, 2009)). Bashiri et al. (Bashiri, Mirzaei, & Randall., 2013) presented a heuristic based on a genetic algorithm to solve the capacitated *p*-hub center problem. Yang et al. (Yang, Liu, & Yang, 2013) developed the problem considering uncertain travel times characterized by normal fuzzy vectors. The authors presented an approximation method to discretize the fuzzy travel times and formulated a mixed-integer programming model. They also developed a metaheuristic based on particle swarm optimization to solve the problem. Gao and Qin (Gao & Qin, 2016) presented a chance-constrained programming approach for an uncertain type of *p*-hub center location problem.

2.1.1.3 The hub covering problem

The hub covering problems include hub set covering and hub maximal covering problems. In these problems, every origin-destination pair is covered if hubs are close enough to serve the demand within a given threshold. The hub set covering problem locates hubs to cover all demand such that the cost is minimized. In case, the cost of covering all origin–destination pairs is greater than the budget, then some pairs are not covered. One alternative is formulating a covering problem that maximizes the flow covered by a certain number of hub nodes within a given travel time/distance or for a given cost. Campbell (Campbell J. F., 1994) formulated the single allocation hub covering problem (SAHCP) and the multiple allocation hub maximal covering problem (MAHMCP) with $O(n^4)$ variables and constraints. Marianov et al. (Marianov, Serra, & ReVelle, 1999) designed a new model for the single allocation hub maximal covering problem that locates hubs in a competitive environment. The primary goal of this model is to safeguard customers from being drawn towards competitor hubs. One scenario where this is applicable is when an airline's

management contemplates relocating hubs to reduce costs and improve overall performance. Another instance is when an airline endeavors to attract passengers who are not adequately served by their competitors.

Kara and Tansel (Kara & Tansel, 2003) proved that the SAHCP is NP-hard. In their study, the authors formulated and evaluated three linear versions of the original quadratic model. Additionally, they introduced a novel linear model and demonstrated its superior performance compared to all previous linear models for solving the SAHCP. Wagner (Wagner, 2008) proposed the improved formulation for the SAHCP. He also designed a model for the multiple allocation hub covering problem (MAHCP). Qu and Weng (Qu & Weng, 2009) developed a new model for the MAHCP with $O(n^2)$ variables and constraints. Calik et al. (Hatice Calık, 2009) proposed an integer programming formulation for the incomplete hub network case of the SAHCP. This model locates hub facilities, finds the hub links that connect hubs and allocates non-hubs to the hubs such that the travel time between any origin-destination pair is within a given time threshold. Ernst et al. (Ernst, Jiang, Krishanmoorthy, & Baatar, 2018) introduced a new variable called hub radius and presented a new version of SAHCP. Also, the authors added some cuts to the formulation and proved that they were valid. Hwang and Lee (Hwang & Lee, 2012) proposed a new uncapacitated allocation *p*-hub maximal covering problem by defining three coverage criteria. Since the formulation is Np-hard, the authors proposed two heuristics to solve the model. Peker and Kara (Peker & Kara, 2015) used partial coverage to remodel single allocation *p*-hub maximal covering problem (SApHMCP) and multiple allocation *p*-hub maximal covering problem (MApHMCP). They also strengthened the formulations by adding some valid inequalities to them. The authors used branch and bound to solve SApHMCP and MApHMCP on the Turkish network dataset and the CAB dataset. They compared their models to the models presented in (Campbell J. F., 1994)

and (Hwang & Lee, 2012) and showed the new formulations perform better. Silva and Cunha (Silva & Cunha, 2017) developed a Tabu Search to solve the uncapacitated single allocation version of the problem. Janković and Stanimirović (Janković & Stanimirović, 2017) developed Hwang and Lee's work (Hwang & Lee, 2012) to a new integer linear programming model. In this model, the authors assumed that each non-hub node can be assigned to at most r hubs such that $r \leq p$. They used a general variable neighbourhood search (GVNS) algorithm to solve the model on both CAB and AP datasets. A comparison of their heuristic with branch and bound indicated that GVNS finds optimal solutions in a shorter time. They also ran the algorithm on the AP dataset and showed that GVNS finds a solution in a reasonable time. In a similar work, Brimberg et al. (Brimberg, Mišković, Todosijević, & Uroševic, 2022) used the binary coverage criterion and proposed a new uncapacitated *r*-allocation p-hub maximal covering problem (UrApHMCP) which can be applied to single as well as multiple allocation problems. The authors proposed two GVNS heuristics due to the complex nature of the formulation.

2.1.2 Research gaps

We have observed a significant gap in the existing literature on the coverage problem concerning hub locations. Insufficient attention has been paid to addressing uncertainties and congestion effects within hub covering networks. For instance, during peak periods, major airport hubs often encounter a substantial volume of take-offs and landings, leading to runway unavailability and extensive delays for aircraft. Motivated by this gap, chapter 3 of this study focuses on developing hub maximal covering location problems considering this fact that hub servers (e.g., runways in an airport hub) are not always available. In other words, they become unavailable when they serve a flow. We formulate two models to address such an uncertainty by

maximizing the expected coverage of flow. Our models are extensions of the classical covering location problems. In the first model, each hub is assumed to have a fixed number of servers, while in the second model, this number is a variable to be decided. The proposed problems try to locate *p* hub facilities in a way that the origin–destination pair of two non-hub nodes is covered by a pair of hub nodes. In fact, origin–destination pairs are covered only if there are hub facilities in pre-specified distances (coverage threshold) from their links. There is a great deal of importance attached to this coverage threshold, especially for airline networks or wireless systems. Existing models, such as those proposed by Marianov and Serra (Marianov & Serra, 2003), do not take into account the possibility that certain origin-destination pairs may fall outside the coverage threshold. Furthermore, they did not provide any solution to their second model. However, in this research we solve both proposed models using exact solutions as well as metaheuristics based on GA and TS.

Campbell (Campbell J. F., 1994) did not provide any numerical results for the MAHMCP and Qu and Weng (Qu & Weng, 2009) proposed a path relinking (PR) algorithm which does not provide exact solutions for $n > 25$, where *n* is the number of nodes. Qu and Weng chose the AP dataset and used LINGO to solve the model by branch and bound. They also evaluated their algorithm using a special case on hub airports location of Chinese aerial freight flows between 82 cities in 2002. They argued that for 82 nodes, LINGO may work more than 200 hours to get the optimal solution. Chapter 4 of this dissertation addresses some ideas for developing a new model for the MAHMCP using integer programming techniques and shows that the new formulation is stronger than the past models.

2.2 Chassis inventory management in intermodal transportation

When it comes to shipping and logistics, intermodal transportation refers to moving a container using more than one mode of transportation. By using roads, rails, air, and oceans in conjunction with each other, it is possible to meet the time and cost demands of modern-day shipping more efficiently and affordably than long-haul, over-the-road trucking can. To move raw materials, parts, and products across the supply chain, it is imperative to manage assets and logistics efficiently. Using intermodal transportation resources appropriately is crucial when shipping goods across the globe. One of these resources is chassis. When shipping containers are moved over land, a chassis is essential. A chassis connects shipping containers to trucks. Furthermore, chassis play an essential role in intermodal transportation. Efficiently loading, unloading, and transporting goods relies heavily on the timely and strategic availability of chassis. Chassis play a crucial role in facilitating the transportation of containers, making them indispensable. Intermodal transportation companies encounter a notable hurdle in the form of chassis shortages at terminals, posing one of their biggest challenges. Shipping container backlogs at rail ramps have experienced exponential growth throughout the United States, leading to a widespread scarcity of chassis in nearly every major city. This scarcity has emerged as a significant impediment to the import and export of goods[1].

To prevent shortages, one solution is to keep an inventory of empty containers and chassis in terminals and reposition them between terminals in order to satisfy the demand for loaded and empty containers. Researchers have attempted to solve this transportation challenge over the past decades. Justice recognized and discussed the significant challenges involved in chassis logistics

[1] Chassis Shortage across America - The Scarbrough Group

within the realm of intermodal freight transport. In response to these challenges, he developed a comprehensive planning model aimed at effectively tackling the issue. The model aimed to answer crucial questions such as when, where, how many, and how to efficiently relocate chassis. By analyzing various factors and employing optimization techniques, Justice aimed to optimize the entire process of chassis relocation within the intermodal freight transport system. Using mathematical formulations, the problem was modeled as a bi-directional time-based network transport problem (Justice, 1995). Lopez (Lopez, 2003) emphasized the significance of routing empty containers in an intermodal chain. The author examined how ocean carriers repositioned their empty containers in the USA. Bin and Zhongchen (Bin & Zhongchen, 2007) proposed an integer programming model specifically designed for relocating intermodal empty containers in land-carriage, disregarding considerations such as chassis and delivery time. The primary objective of the model was to minimize transportation costs associated with empty containers while ensuring the satisfaction of container demands and supplies. Corry and Kozan (Corry & Kozan, 2006) proposed a multi-period assignment model to match containers to slots. They ignored the chassis in their model. Shintani et al. (Shintani, Imai, Nishimura, & Papadimitriou, 2007) addressed the problem of designing container shipping service networks and optimizing containership routing, while also considering the repositioning of empty containers. The approach involved developing a model that took into account various factors such as operating and capital costs, as well as the penalty cost associated with unmet customer demand. By utilizing this model, he determined the most profitable voyage route for a liner shipping company. The transportation of empty and loaded containers was primarily conducted by ships, and the research did not account for factors such as time. Alvarez (Alvarez, 2009) proposed a heuristic approach for the integrated distribution of empty and full containers in an intermodal network. Le-Griffin et al. (Hanh D.Le-Griffin, Lam

Mai, Mark Griffin, 2011) used simulation to evaluate the current management chassis systems. They concluded that the lack of a cooperative chassis pool negatively impacts the performance of container terminals in terms of effective capacity, system operation times, and air emissions. The chassis pool is a place where intermodal chassis are stored near or at terminals (ocean terminals, terminals, etc.) to support motor carrier's daily use of them. Song and Dong (Song & Dong, 2012) examined the issue of combining cargo routing and repositioning of empty containers in the operational phase within a shipping network that had multiple service routes, multiple vessels in operation, and multiple scheduled voyages. The goal was to minimize the overall costs associated with various factors throughout the planning period. These factors included the expenses related to loading and unloading containers at ports, costs incurred due to customer demand backlog, charges for storing loaded containers temporarily at transshipment ports, expenses related to maintaining an inventory of empty containers at ports, and the costs of transporting empty containers.

Hartman and Clott (Bruce C.Hartman and Christopher Clott, 2015) proposed a model based on a Bayesian game to manage the number of chassis in the system. Ng and Tally (Ng & Talley, 2017) developed a mathematical model that could assist in chassis pool planning. They used the model as a solution for the chassis shortage in the maritime industry. Using the Los Angeles/Long Beach port area as an example of interest, Chassiakos et al. (Chassiakos, Jula, VanderBeek, Shellhammer, & Dona, 2017) proposed an analytical framework for modeling and optimizing the concept of centralized processing of chassis around marine container terminals. Ng and Tally (Nga & Talley, 2020) developed an integer formulation to maximize the utilization of the space on a train. Ng (Ng, 2021) proposed a strategy for chassis pool operators to use when making repositioning decisions.

The chassis shortage at intermodal transportation networks, including trucking and terminals, has not been thoroughly examined. Using a stochastic inventory management model, Couchman (Couchman, 2020) showed how to optimally allocate and reposition chassis and empty shipping containers across terminals while minimizing costs. The demand at terminals was divided into two categories: empty containers and loaded containers. A multi-period time was examined, but only two terminals were considered. With *n* terminals, Collins (Collins, 2021) formulated a deterministic model to optimally relocate empty containers within an intermodal truck-rail network. However, the author ignored loaded containers in her model and did not take the time into consideration.

2.2.1 Research gaps

Examining the current literature on intermodal transportation indicates that the majority of research has primarily concentrated on the maritime sector, disregarding the distinct challenges encountered in intermodal transportation at truck-rail terminals. Moreover, the existing literature frequently overlooks significant factors such as shortages of chassis, the presence of multiple terminals, the demand for loaded containers, multiple time periods, or lead time. These gaps in the literature emphasize the necessity for a comprehensive strategy to tackle the hurdles associated with managing chassis inventory in intermodal transportation.

Driven by these identified gaps, Chapter 5 introduces a comprehensive mathematical framework that leverages a multi-time period and *n* terminals. This innovative approach effectively manages the inventory of chassis and empty containers, strategically reallocating them between terminals as required to meet the overall demand for empty and loaded containers. The objective is to minimize the total cost incurred throughout the process.

2.3 Relay networks

The trucking industry has long faced the challenge of high turnover rates among drivers, which can significantly impact operational efficiency and overall productivity. One potential solution to address this issue is the implementation of relay points. Relay points serve as strategic locations along trucking routes where drivers can exchange trailers or transfer cargo to another driver. By utilizing relay points, trucking companies can optimize their operations by minimizing the need for drivers to travel long distances without rest, ultimately reducing driver fatigue and improving safety. This approach not only enhances operational efficiency and reduces transportation costs but also holds the potential to address the issue of driver turnover. By minimizing the time drivers spend away from home and allowing them to return more frequently, relay points offer a way to alleviate the strain on truck drivers and potentially reduce turnover rates. This literature review aims to provide a comprehensive analysis of the existing research, identify gaps in current knowledge, and propose directions for advancing the field of relay point optimization in the trucking industry.

The concept of utilizing relay networks as an alternative dispatching method for Truckload transportation emerged from the research conducted by Taha and Taylor (Taha & Taylor, 1994). They suggested a simulation-based model that utilized rules to create a network of relay points, with specific nodes assigned to these points, along with a designated service area for each relay point. This approach has gained recognition as a potential solution to mitigate the prevalent issue of driver retention in various studies. Üster and Maheshwari conducted a study on the strategic design of a network to facilitate truckload shipments in various zones. They created a model to efficiently determine the optimal positioning of relay points, the coverage area for each relay point,

and the allocation of non-relay points to their corresponding relay points. The authors proposed using Tabu Search as a solution approach since the problem they defined could achieve optimal solutions only for small cases (Üster & Maheshwari, 2007). Later on, Üster and Kewcharoenwong introduced an algorithm based on Bender decomposition to efficiently tackle larger cases (Üster & Kewcharoenwong, 2011). Melton and Ingalls introduced a model that provided recommendations for locating relay points with the aim of minimizing overall costs. Their mixed integer program took into account driver-related factors and incorporated the cost of driver turnover into the objective function (Melton & Ingalls, 2012). Rais et al. introduced a mixed integer-programming model to tackle the pickup-and-delivery problem involving transshipment locations. These transshipment points serve as locations where vehicles can stop to facilitate cargo transfers and load adjustments, enable drivers to switch vehicles, similar to relay points, or adhere to release times in accordance with policy requirements. Furthermore, they provide an opportunity for well-rested drivers to replace fatigued drivers, ensuring optimal driver performance throughout the transportation process (Rais, Alvelos, & Carvalho, 2014). Kewcharoenwong and Üster presented the incorporation of link capacity constraints and the concept of link imbalance into the strategic design of relay networks. The capacity of each link connecting relay points was defined as the maximum cumulative number of truckloads that could be transported in both directions. By defining the link imbalance, the authors were able to control the difference between the number of truckloads in opposing directions on a relay point to relay point link, thereby reducing the mileage traveled with empty loads (Kewcharoenwong & Üster, 2017).

2.3.1 Research gaps

Rais et al. introduced three sets of binary decision variables, two with three indices and another with four indices. They made the assumption that the depot locations are already known, and relays are selected from a subset of nodes (Rais, Alvelos, & Carvalho, 2014). In our research, we introduce a novel concept called "link type," which allows us to consider the depot locations and relay points as decision variables. Additionally, we simplify the formulation by defining only a single binary decision variable with four indices. Moreover, our approach guarantees that two trucks converge at a relay point solely when there is a load exchange, effectively minimizing the distance traveled with empty loads.

Chapter 6 of this dissertation formulates a depot-relay-point location and truck routing problem. The problem aims to determine the best locations for depots, identify optimal locations for relays, and establish the sequence of nodes that a truck will visit after it leaves a depot

Chapter 3 - Two maximal hub covering location models considering service availability

This chapter is derived from our manuscript titled "Two Maximal Hub Covering Location Models Considering Service Availability," which we have diligently prepared and intend to submit to a reputable scholarly journal specializing in the field of transportation.

3.1 Introduction

Hub location problem is one of the novel topics in location problems. It spans its implementation to transportation systems, telecommunication systems, and delivery systems where more than one site delivers commodities such as data, people, or postal packages. These systems can be improved using hubs that route flow between origin-destination pairs and gain economic profits from dense transportation (Daskin, 1997).

This research focuses on the hub maximal covering problem (HMCP) also known as the multiple allocation hub maximal covering problem (MAHMCP), first introduced by Campbell (Campbell J. F., 1994) and then developed by (Bo Qu, Kerui Weng, 2009) that provides optimal decisions for the location of the hubs and allocation of non-hub nodes to the hubs. In hub systems, the flow between an origin-destination pair must pass through one or two hubs resulting in an increase in travel distance. To address this issue, the hub maximal covering problems limit the travel path within an acceptable range while locating p hubs (Bo Qu, Kerui Weng, 2009). Hub covering location problems are classified into: (*i*) the hub set covering problem (HSCP) that locates hubs to cover all the demand such that the cost is minimized, and (*ii*) the hub maximal covering problem (HMCP) that is used if the cost of coverage of all demand is greater than the available budget.

The main motivation of this study arises from the fact that the HMCP problem assumes that every hub facility is always available (or not congested) and there is an adequate number of servers in hubs. However, hubs may get congested and become unavailable in certain cases. Take cargo transportation as an example. In the second half of 2021, port congestion on the West Coast of North America was severe, forcing many shippers to transfer their goods to the East Coast. With 479,700 TEUs (Twenty-foot equivalent units) in January 2022, the Port of Savannah, Georgia, set a new monthly record and has done so for the past 18 months[1]. This can severely impact on various businesses including trucking industries, maritime transportation, and railways, causing service times to increase and delivery dates to be missed. A similar situation often happens in air transportation. For three months in a row, the Dallas-Fort Worth International Airport hub had the largest number of takeoffs and landings around the globe (22,831 takeoffs and landings since May 2020)[2]. The significant increase in the number of takeoffs and landings leads to server (runway) unavailability, resulting in lengthy queues of airplanes both on the runways and in the sky. Therefore, when the airlines design their networks, they need to consider the probability that sometimes hubs might be unavailable. Moreover, the airports and their runway capacities are expanding in response to the growing annual number of passengers. For instance, the annual number of passengers on U.S. airlines was 647 million in 2003, which increased to 927 million in 2019 (Bureau of Transportation Statistics, 2021). To enhance customer service and accommodate the growth in demand, airlines make substantial investments in infrastructure at their chosen hub

[1] Long Beach container imports rebound in January as terminal congestion eases | S&P Global Commodity Insights (spglobal.com)

[2] How Dallas-Fort Worth Airport Became the Busiest in the World | Condé Nast Traveler (cntraveler.com)

airports, amounting to hundreds of millions of dollars. As an example, in 2020, Southwest Airlines announced a $1.5 billion growth investment directed towards Denver International Airport. (Murray, 2020). In 2017, American Airlines and several other commercial air carriers operating at Philadelphia International Airport reached an agreement on an extensive plan worth nearly $900 million. The plan aimed to sustain essential improvements and develop new infrastructure at both PHL and Philadelphia Northeast Airport (PNE). They allocated $26 million of that budget towards the rehabilitation of taxiways and runways, as well as upgrades to various facilities (Philadelphia International Airport, 2017). Hence, when investing in the development of new infrastructure, having a clear understanding of the optimal number of runways necessary for a hub to meet the airline's demand is immensely advantageous.

In this study, we analyze an HMCP problem where each hub consists of several servers (e.g., runways in an airport) that operate as an M/M/c queuing system. In the proposed formulations, we incorporate the concept of busy probability to model congestion within hubs. In relation to busy fraction parameters, there are a few points to consider: An airport is used by many airlines throughout the world. Dubai International Airport, for example, hosts more than 140 airlines that transport passengers and cargo. Hence, the arrival rate of a given airport depends on the flow coming from various airlines. In this research, we study the optimal location of hubs and allocation of demand points to the hubs from a single airline point of view, namely the American Airlines Group, as opposed to most previous studies in this field. That being said, the traffic that is allocated to the hubs in the proposed models is only a small fraction of the total traffic flowing to an airport. Therefore, when calculating busy fraction parameters, and in particular the arrival rate λ, we need to look at the total flow coming from all airlines which is an exogenous parameter.

Leaving this fact out and making traffic allocation dependent on hubs leads to the following constraint proposed by (Marianov & Serra, 2003):

$$\sum_i \sum_j \sum_k W_{ij} Y_{ij}^{km} + \sum_i \sum_j \sum_k W_{ij} Y_{ij}^{mk} \leq \lambda_m \tag{3.1}$$

In the above inequality, W_{ij} is the average hourly flow American Airlines made between origin *i* and destination *j*. $Y_{ij}^{km}$is the fraction of the flow from origin i to destination j that is routed via hubs at locations *k* and *m* in that specific order. The left-hand side is the hourly flow transported by the American Airlines through hub *m* which should be less than or equal to λ_m. In contrast, λ_m is the total rate of demand (the average hourly flow) carried by all airlines through hub m. Therefore, this constraint is always met, making it redundant.

In this study, we use airport tower operations, the sum of the number of airport arrivals and departures, provided by the Federal Aviation Administration which is a very good approximation for the arrival rate λ. Furthermore, we derive the airport's service (call) rate and the number of runways from the Federal Aviation Administration data.

The main contributions of this research are as follows: Firstly, to the best of our knowledge, the busy probability has not received much attention in the hub covering location problems. We plan to integrate busy probability in the model. We also analyze two variants of hub covering location models: (*i*) fixed number of servers, and (*ii*) variable number of servers. The latter case results in a mix-integer program with a non-linear objective function. To solve the model using commercial solvers such as GUROBI, we transform the objective function to a quadratic version. Secondly, we develop two metaheuristics based on the Genetic algorithm and Tabu Search to solve these NP-hard models. We propose Lemma 1 which makes use of the problem structure and finds the optimal allocation of demand points to the hubs in one simple step. Both algorithms take

advantage of this lemma which eliminates the need for solving an allocation optimization problem whenever the location of hubs changes due to having neighborhood moves or crossover operators. We show that the proposed methods outperform traditional commercial software for small and medium size problems and can efficiently solve larger problems that commercial software cannot solve. Thirdly, we show how the change in the number of runways might impact the hub network configuration and the flow coverage.

The remainder of this chapter is organized as follows: Section 3.2 introduces two proposed models. In Section 3.3, we delve into the complexity analysis of these models and present two metaheuristics, namely the Genetic algorithm and Tabu Search algorithm, designed to effectively solve them. Computational analysis is then addressed in Section 3.4. Lastly, the chapter concludes with a summary of findings and concluding remarks in the final section.

3.2 Mathematical formulation

Campbell (Campbell J. F., 1994) formulated the HMCP in order to locate hubs and allocate non-hubs to hubs such that the coverage of demand is maximized. This model assumes that every hub facility is always available (or not congested) and there is an adequate number of servers in hubs. However, hubs may get congested and become unavailable in certain cases. Note that including the unavailability of hubs may impact the location of hubs as well as the allocation of non-hubs to the hubs. Our study tackles this uncertainty by formulating two models specifically designed to address it. The first model assumes that the number of servers is given but may be different from hub to hub. This model is called the hub covering location problem with different busy probabilities and different number of servers (HCLP-(q_h,s_h)). It uses the current number of servers in the hubs as a parameter that might not be optimal to serve the network. In the second

model, called HCLP-(q_h ,x_h), we determine the optimal number of servers where hubs can have different busy probabilities. Both models in this study make the assumption that the service time remains independent of the state of the system. This implies that the rate of service is not influenced by the number of customers present in the system (e.g., airplanes in queues waiting to land or take off). The service is individual, meaning that each server (runway) can serve at most one customer (airplane) at a time. The servers are also independent. In these models, each hub has one queue for all its servers and the order of the queue is first in, first out (FIFO).

3.2.1 Hub covering location problem-(q_h ,s_h)

Let $G = (V, E)$ be a complete undirected graph with the node set $V = \{1, 2, \ldots, n\}$. The node set (V) includes origins, destinations, and potential hub locations. Each pair of nodes is connected by an arc (i, j) with cost c_{ij} (distance, time, etc.). W is the flow matrix and $W_{ij} \in W$ is defined as the demand from origin *i* to destination *j*. It is assumed to be advantageous to pass by hubs on the route $i-j$, with a discount factor α, where $0 \leq \alpha \leq 1$. The main goal is to select *p* hubs. We assume that at most two hubs can be selected on a route from one non-hub node to another. The hubs have no capacity restrictions to collect, transport, and distribute demands. Let $C_{ij}^{km} = c_{ik} + \alpha c_{km} + c_{mj}$ be the transportation cost from node *i* to node *j* passing through hubs *k* and *m*. The parameter β is the maximum acceptable transportation cost from node *i* to node *j*. V_{ij}^{km} is a binary parameter, which indicates whether the origin–destination pair (i, j) can be covered by the candidate hub nodes *k* and *m* and is defined as follows:

$$V_{ij}^{km} = \begin{cases} 1, & if\, C_{ij}^{km} \leq \beta \\ 0, & otherwise \end{cases} \tag{3.2}$$

This study considers each individual hub node as a queuing system. Therefore, the busy probability of each server in hub *m* can be calculated using the equation $q_m = \frac{\lambda_m}{s_m \mu_m}$, where λ_m is the demand arrival rate for hub *m*, μ_m is the average service rate of the servers existing in hub *m*, and s_m is the number of servers existing in hub *m*. s_m enforces a range of zero to one for q_m. It is assumed that servers in hub *m* serve customers independently and each is busy with the same probability of q_m. Therefore, the probability of having *j* free servers in hub *m* can be calculated using the binomial distribution which is given by $\binom{s_m}{j}(1-q_m)^j(q_m)^{s_m-j}$. To ensure adequate coverage for the demand between each origin-destination pair, it is required that either one hub or a pair of hubs handles the demand. The probability of coverage by hub *m* is determined by the probability that at least one server is available in hub *m*, which can be calculated as $1\text{-}q_m{}^{s_m}$. The probability of achieving coverage by a pair of hubs, denoted as (*k*, *m*), is contingent upon the availability of at least one server in both the first hub (*k*) and the second hub (*m*). As the hubs operate independently, the probability of being covered by the pair of hubs (*k*, *m*) can be calculated as $(1\text{-}q_m{}^{s_m})(1\text{-}q_k{}^{s_k})$.

The deterministic coverage of the demand is equal to $\sum_i \sum_j \sum_k \sum_m W_{ij} Y_{ij}^{km} V_{ij}^{km}$. Therefore, the expected coverage of demand by hubs is calculated as follows:

Expected coverage of demand = $\sum$ (demand × probability of coverage)

$$\sum_i \sum_j \sum_k \sum_{m \neq k} W_{ij} Y_{ij}^{km} V_{ij}^{km} (1 - {q_m}^{s_m})(1 - {q_k}^{s_k}) + \sum_i \sum_j \sum_k \sum_{m=k} W_{ij} Y_{ij}^{km} V_{ij}^{km} (1 - {q_m}^{s_m}) \quad (3.3)$$

The decision variables are as follows:

H_m= 1, if node *m* is determined as a hub; otherwise, it takes 0.

Y_{ij}^{km}: The fraction of the flow from origin *i* to destination *j* that is routed via hubs at locations *k* and *m* in that specific order.

The HCLP-(q_h ,s_h) model is presented as a mixed integer programming formulation as follows:

$$Max \sum_i \sum_j \sum_k \sum_{m \neq k} W_{ij} Y_{ij}^{km} V_{ij}^{km} (1 - {q_m}^{s_m})(1 - {q_k}^{s_k}) + \sum_i \sum_j \sum_k \sum_{m=k} W_{ij} Y_{ij}^{km} V_{ij}^{km} (1 - {q_m}^{s_m}) \quad (3.3)$$

S.t.

$$\sum_m H_m = p \quad (3.4)$$

$$\sum_k \sum_m Y_{ij}^{km} = 1 \quad \forall i, j \quad (3.5)$$

$$Y_{ij}^{km} \leq H_m \quad \forall i, j, k, m \quad (3.6)$$

$$Y_{ij}^{km} \leq H_k \quad \forall i, j, k, m \quad (3.7)$$

$$Y_{ij}^{km} \geq 0 \quad \forall i, j, k, m \quad (3.8)$$

$$H_m \in \{0,1\} \quad \forall m \quad (3.9)$$

The objective function (3.3) maximizes the excepted coverage of flow. Constraint (3.4) guarantees that exactly *p* hubs are chosen. Constraints set (3.5) ensures that the flow for every origin-destination pair is transported via some hub pair. It also rules out any possibility of calculating the probability of a particular hub being busy twice in (3.3)**.** Constraint sets (3.6) and

(3.7) ensure that flows are routed via locations that are hubs. Constraint sets (3.8) and (3.9) show the range of variables.

3.2.2 Hub covering location problem-(q_h ,x_h)

This model assumes that the number of servers in each hub is a variable to be determined. The goal is to locate the hubs, determine the optimal number of servers, and allocate non-hubs to the hubs based on the busy probability of the servers, thereby maximizing coverage. The HCLP-(q_h ,x_h) is presented as a mixed integer programming formulation by adding constraints (3.11)-(3.14) to the HCLP-(q_h ,s_h) and modifying (3.3) as follows:

$$Max \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\x_k\neq 0}}^{n}\sum_{\substack{m=1\\m\neq k\\x_m\neq 0}}^{n} W_{ij}\, Y_{ij}^{km} V_{ij}^{km}\left[1-\left(\frac{\lambda_m}{x_m\mu_m}\right)^{x_m}\right]\left[1-\left(\frac{\lambda_k}{x_k\mu_k}\right)^{x_k}\right]$$

$$+\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\x_k\neq 0}}^{n}\sum_{m=k} W_{ij}\, Y_{ij}^{km} V_{ij}^{km}\left[1-\left(\frac{\lambda_k}{x_k\mu_k}\right)^{x_k}\right] \qquad (3.10)$$

S.t.

$$\sum_{m} H_m = p \qquad (3.4)$$

$$\sum_{k}\sum_{m} Y_{ij}^{km} = 1 \qquad \forall i,j \qquad (3.5)$$

$$Y_{ij}^{km} \le H_m \qquad \forall i,j,k,m \qquad (3.6)$$

$$Y_{ij}^{km} \le H_k \qquad \forall i,j,k,m \qquad (3.7)$$

$$\sum_{m} x_m \le R \qquad (3.11)$$

$$x_m \geq H_m \frac{\lambda_m}{\mu_m} \qquad \forall m \tag{3.12}$$

$$x_m \leq (R - p + 1)H_m \qquad \forall m \tag{3.13}$$

$$Y_{ij}^{km} \geq 0 \qquad \forall i, j, k, m \tag{3.8}$$

$$H_m \in \{0,1\} \qquad \forall m \tag{3.9}$$

$$x_m \in Z_0^+ \qquad \forall m \tag{3.14}$$

x_m is an integer decision variable that indicates the number of servers in hub *m*, while *R* represents the maximum number of servers that can be opened. The value of *R* will be determined based on the available budget.

The objective function (3.10) consists of two parts: (*i*) the expected coverage of demand by the hub pairs; and (*ii*) the expected coverage of demand by the single hubs. The goal is to maximize the expected coverage of demand. Busy probabilities ($0 \leq q_m \leq 1$) are raised to the power of x_m variables. Since the objective is maximization, x_m variables can be as large as possible. Hence, we add constraint (3.11) that limits the total servers that can be established. The constraint sets (3.12) and (3.13) establish the upper and lower bounds, respectively, for the $x_m$variables. In conjunction with constraint (3.4), they ensure the opening of a total of *p* hubs. The $x_m$variables are constrained such that their values cannot exceed $R - p + 1$. This restriction ensures that at least one server is allocated to each of the other $p - 1$ hubs, in addition to the primary hub, resulting in a total of *p* hubs being operational. According to the rules of the Probability Theory, the probability value ($\frac{\lambda_m}{x_m \mu_m}$) should be less than or equal to 1. Hence, x_m variables should be greater than λ_m/μ_m. Constraint (3.14) shows that x_m variables are non-negative integers.

The objective function (3.10) is non-linear (polynomial and rational) and cannot be solved using commercial solvers such as GUROBI. Given that integer variable x_m takes discrete values, we use linearization techniques to reformulate the objective function as a quadratic one, making the model solvable by GUROBI. In this regard, first, we define a new binary variable, called F_{ms}, which takes 1 if *s* servers are allocated to hub *m*, and 0 otherwise. The new objective function can be written as follows:

$$Max \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\substack{m=1 \\ m\neq k}}^{n} W_{ij}\, Y_{ij}^{km} V_{ij}^{km} \left[1 - \sum_{s}\left(\frac{\lambda_m}{s \times \mu_m}\right)^{s} \times F_{ms}\right]\left[1 - \sum_{t}\left(\frac{\lambda_k}{t \times \mu_k}\right)^{t} \times F_{kt}\right]$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{m=k}^{n} W_{ij}\, Y_{ij}^{km} V_{ij}^{km} \left[1 - \sum_{t}\left(\frac{\lambda_k}{t \times \mu_k}\right)^{t} \times F_{kt}\right] \qquad (3.15)$$

After multiplying every term inside the first set of parentheses by every term in the second set, the objective function is converted to (3.16).

$$Max \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\substack{m=1 \\ m\neq k}}^{n} W_{ij}\, Y_{ij}^{km} V_{ij}^{km} \left(1 - \sum_{s}\left(\frac{\lambda_m}{s\mu_m}\right)^{s} \times F_{ms} - \right.$$

$$\left. \sum_{t}\left(\frac{\lambda_k}{t\mu_k}\right)^{t} \times F_{kt} + \sum_{s}\sum_{t}\left(\frac{\lambda_m}{s\mu_m}\right)^{s} \times \left(\frac{\lambda_k}{t\mu_k}\right)^{t} \times F_{ms} \times F_{kt}\right) +$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{m=k} W_{ij}\, Y_{ij}^{km} V_{ij}^{km} \left(1 - \sum_{t}\left(\frac{\lambda_k}{t\mu_k}\right)^{t} \times F_{kt}\right) \qquad (3.16)$$

The objective function (3.16) has the product $F_{ms} \times F_{kt}$ where F_{ms} and F_{kt} are binary variables. This product can be linearized by introducing a new binary variable Z_{mk}^{st} which represents the value of $F_{ms} \times F_{kt}$. By adding the three inequalities below, we make that product linear:

$Z_{mk}^{st} \leq F_{ms},\ Z_{mk}^{st} \leq F_{kt}, Z_{mk}^{st} \geq F_{ms} + F_{kt} - 1$

Since, we still have some products of Y_{ij}^{km} and Z_{mk}^{st} or Y_{ij}^{km} and F_{ms}, the new objective function (3.17) is a quadratic function. However, commercial software such as GUROBI can solve the problem in this form. In the new formulation, we do not need x_m variables. Also, constraints sets (3.12) and (3.13) are removed. Instead, constraints sets (3.19) -(3.22) are added to the model, and constraint (3.11) is changed to constraint (3.18).

We call the quadratic formulation QHCLP-(q_h ,f_{hs}) and present it as follows:

$$Max \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\substack{m=1 \\ m\neq k}}^{n} W_{ij}\, Y_{ij}^{km} V_{ij}^{km} \left(1 - \sum_{s}(\frac{\lambda_m}{s\mu_m})^s \times F_{ms} - \sum_{t}(\frac{\lambda_k}{t\mu_k})^t \times F_{kt} + \sum_{s}\sum_{t}\left(\frac{\lambda_m}{s\mu_m}\right)^s \times \left(\frac{\lambda_k}{t\mu_k}\right)^t \times Z_{mk}^{st}\right) + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{m=k} W_{ij}\, Y_{ij}^{km} V_{ij}^{km} \left(1 - \sum_{t}(\frac{\lambda_k}{t\mu_k})^t \times F_{kt}\right) \quad (3.17)$$

S.t.

$$\sum_{m} H_m = p \quad (3.18)$$

$$\sum_{k}\sum_{m} Y_{ij}^{km} = 1 \qquad \forall i,j \quad (3.5)$$

$$Y_{ij}^{km} \leq H_m \qquad \forall i,j,k,m \quad (3.6)$$

$$Y_{ij}^{km} \leq H_k \qquad \forall i,j,k,m \quad (3.7)$$

$$\sum_{m}\sum_{s} s \times F_{ms} \leq R \quad (3.18)$$

$$\sum_{s} F_{ms} = H_m \qquad \forall m \quad (3.19)$$

$$Z_{mk}^{st} \leq F_{ms} \qquad \forall m,k,s,t \tag{3.20}$$

$$Z_{mk}^{st} \leq F_{kt} \qquad \forall m,k,s,t \tag{3.21}$$

$$Z_{mk}^{st} \geq F_{ms} + F_{kt} - 1 \qquad \forall m,k,s,t \tag{3.22}$$

$$Y_{ij}^{km} \geq 0 \qquad \forall i,j,k,m \tag{3.8}$$

$$H_m \in \{0,1\} \qquad \forall m \tag{3.9}$$

$$F_{ms} \in \{0,1\} \qquad \forall m,s \tag{3.23}$$

$$Z_{km}^{st} \in \{0,1\} \qquad \forall m,k,s,t \tag{3.24}$$

According to Campbell et al., the determination of hub locations in the hub location problem is considered NP-hard for both single and multiple allocations (J. F. Campbell, A. T. Ernst, M. Krishnamoorthy, 2005). If every origin-destination pair is viewed as a single demand point, then the HCLP-(q_h ,s_h) becomes the MEXCLP formulated by Daskin (Mark S. Daskin, 1983). This means that the MEXCLP can be considered a special case of HCLP-(q_h ,s_h), indicating that HCLP-(q_h ,s_h) is NP-hard. Similarly, the HCLP-(q_h ,x_h) and QHCLP-(q_h ,F_{hs}) are extensions of the HCLP-(q_h ,s_h) and therefore also NP-hard.

The HCLP-(q_h ,s_h) model involves $n^4 + n$ variables including *n* binary variables. It also requires $2n^4 + n^2 + 1$ constraints. The HCLP-(q_h ,x_h) model involves $n^4 + 2n$ variables with $2n$ integer variables and it requires $2n^4 + n^2 + 2n + 2$ constraints. Finally, the QHCLP-(q_h ,F_{hs}) involves $2n^4 + n^2 + n$ variables including $n^4 + n^2 + n$ binary variables. It also requires $5n^4 + n^2 + n + 2$ constraints. As the size of the problem increases with the growth of *n*, the proposed models become increasingly complex, with a large number of variables and constraints. Consequently, solving such problems becomes exceptionally challenging. In the following section, we address

this issue by presenting two algorithms specifically designed to tackle large-scale problems effectively.

3.3 Heuristics to solve the models

In this section, we propose two heuristic methods and incorporate Lemma 1 to solve the proposed NP-hard problems.

Lemma 1. The HCLP-(q_h ,s_h) has an optimal solution in which all of the Y_{ij}^{km} variables are either zero or one. In other words, there exists an optimal solution where the flow between any origin-destination pair of nodes is transferred along a single path.

Proof. Given that $W_{ij} \geq 0$, hub graph is a complete graph, with $p \geq 1$, and edges that have no capacity restrictions, it can be concluded that there will always exist a feasible solution for the HCLP-(q_h ,s_h).

If Y_{ij}^{km} is fractional for some *i*, *j*, *k*, and *m*, then there exists at least one other fractional Y_{ij}^{tz} for some *i*, *j* since Eq. (3.5) holds. In this case, the coverage parameters V_{ij}^{km} and V_{ij}^{tz} should be equal to 1 for the flow to be transported via hub pair (*k*, *m*) as well as hub pair (*t*, *z*). Now if $(1 - {q_t}^{s_t})(1 - {q_z}^{s_z}) > (1 - {q_k}^{s_k})(1 - {q_m}^{s_m})$, then the solution is not optimal since a solution with a greater expected coverage could be found by setting

$Y_{ij}^{tz} \leftarrow Y_{ij}^{tz} + Y_{ij}^{km} \quad and \quad Y_{ij}^{km} \leftarrow 0.$

A similar argument holds if $(1 - {q_t}^{s_t})(1 - {q_z}^{s_z}) < (1 - {q_k}^{s_k})(1 - {q_m}^{s_m})$. Finally, if $(1 - {q_t}^{s_t})(1 - {q_z}^{s_z}) = (1 - {q_k}^{s_k})(1 - {q_m}^{s_m})$, the flow could be routed via either the hub set *k*, *m* or the hub set *t*, *z*. A similar argument holds true if we seek to cover the origin-destination (*i*, *j*) by single hubs. This shows that for a given set of hubs (*H*), the optimal solution of the allocation

problem can be obtained by picking some $k, m \in H$ for each $i, j \in V$ that corresponds to $Max\left\{\underset{t,z\in H}{Max}(1-{q_t}^{s_t})(1-{q_z}^{s_z})V_{ij}^{tz},\ \underset{t\in H}{Max}(1-{q_t}^{s_t})V_{ij}^{tt}\right\}$ while Y_{ij}^{km} is set equal to one. ■

This lemma also holds true for the HCLP-(q_h ,x_h). There will be a similar argument to show that for a given set of hubs (*H*) and a given number of servers in them, the optimal solution of the allocation problem can be obtained by picking some $k, m \in H$ for each $i, j \in V$ that corresponds to $Max\left\{\underset{t,z\in H}{Max}\left\{\left(1-\left(\frac{\lambda_t}{\mu_t x_t}\right)^{x_t}\right)\left(1-\left(\frac{\lambda_z}{\mu_z x_z}\right)^{x_z}\right)V_{ij}^{tz}\right\}, \underset{t\in H}{Max}\left\{\left(1-\left(\frac{\lambda_t}{\mu_t x_t}\right)^{x_t}\right)V_{ij}^{tt}\right\}\right\}$ while Y_{ij}^{km} is set equal to one.

3.3.1 A Tabu Search algorithm to solve the HCLP-(q_h ,x_h)

Tabu Search (TS) is a technique for solving combinatorial optimization problems. The general idea was introduced by Glover (Glover, 1989; Glover, 1990). TS may be viewed as an interactive technique that explores a set of problem solutions by repeatedly making moves from one solution to another in its neighborhood. TS is based on an evaluation function that chooses the best move in terms of objective function and tabu restrictions. This function selects a solution that produces the best improvement or the least non-improvement in the objective function in each iteration.

Solution: The solution to this problem consists of two parts. The first part includes the variables that specify the location of hubs, H_m, and the number of servers in each hub, x_m. The second part includes those variables that determine the allocation of non-hubs to the hubs, Y_{ij}^{km}.

Solution representation: The first part of the solution is represented by a numeric array with *n* elements (the number of nodes in the network). According to constraints (3.12), (3.13), and (3.14), the value for each element may range from 0 to $R - p + 1$ servers such that constraint

(3.11) holds. If an element is equal to 0, this node has no server and is not selected as a hub. In the initial solution, some *p* elements will take non-zero values while others will take a zero value to ensure that constraints (3.4) and (3.9) are satisfied. Figure 3-1 illustrates a solution with 10 nodes in which nodes 3, 6, and 10 are hubs. Node 3 has two servers, node 6 has one, node 10 has three, and the rest are non-hubs.

Element Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Element Value	**0**	**0**	**2**	**0**	**0**	**1**	**0**	**0**	**0**	**3**

Figure 3-1. Solution representation in TS algorithm

Once hub nodes are determined, the non-hubs will be allocated to the hubs represented by Y_{ij}^{km} variables to form the second part of the solution.

Initial solution: This stage of the algorithm includes two phases. Phase 1 locates the hubs and determines the number of servers. Phase 2 allocates non-hubs to hubs.

Location phase: In this phase, three scenarios are considered. The first scenario selects *p* first nodes in decreasing order of λ_m/μ_m as hubs, where λ_m is the demand arrival rate for hub *m* and μ_m is the average service rate of the servers existing in hub *m.* In the second scenario, the hubs are selected randomly. The third scenario selects *p* first nodes in non-increasing order $Sum_i = \sum_j \sum_k \sum_m V_{ij}^{km}$ as hubs. After determining the location of hubs, all three scenarios generate random numbers to determine the number of servers in each hub, while adhering to constraints (3.11) to (3.13). Non-hub nodes will have a value of zero. Experimental results indicate that the first scenario performs the best in terms of both the quality of the final solution and the solution time.

Allocation phase: In this phase, each pair of origin-destination (O-D) is allocated to hubs by Lemma 1. In this regard, the value of $\left(1-\frac{\lambda_t}{\mu_t x_t}^{x_t}\right)\left(1-\frac{\lambda_z}{\mu_z x_z}^{x_z}\right)V_{i'j'}^{tz}$ is calculated for each pair of O-D (i',j') and each pair of hubs (t, z). The maximum value determines the hub pair (*k*,*m*) which transports the demand of O-D (i',j'); in other words, $Y_{i'j'}^{km}$variable will be equal to 1. Allocation by Lemma 1 guarantees that constraints (3.4), (3.6), (3.7), and (3.8) are met.

Once Y_{ij}^{km}variables are determined, the value of the objective function is calculated for the initial solution. To improve the solution, the TS algorithm enters the improvement stage.

Improvement stage: In this stage, the new solution is created by generating two random numbers within the range of $\{1,2,\dots,n\}$. These random numbers are used to determine the hub nodes in the new solution. If any of them are already existing hubs from the previous solution or tabu nodes, the process will be repeated. Next, the algorithm makes 2×*p* neighborhoods of the previous solution. In this regard, the existing hubs of the current solution (solution A) are numbered from 1 to *p*. At first, hub 1 is replaced by the first randomly selected node to form a new set of hubs and the new hub will take the same number of servers as hub 1. This ensures constraint (3.11) is always met. Then, the non-hubs are allocated to this set of hubs by Lemma 1, and the objective function (3.10) is calculated. This solution is saved as a feasible solution (solution B). Then, hub 2 of solution A is replaced by the first randomly selected node, the non-hubs are allocated to this set of hubs by Lemma 1, and the objective function (3.10) is calculated. This solution is saved as a feasible solution (solution C). The process continues until hub *p* is replaced by the first randomly selected node. The same process is repeated using solution A and the second randomly selected node to make another set of *p* neighborhoods. Finally, the solution with the best

objective function value among these 2×p neighborhoods is selected as the new solution, and the algorithm proceeds to the next iteration.

Tabu list: Every hub node that is replaced by a non-hub in an iteration will be added to the tabu list. This node will be prohibited from being selected as a hub for a certain number of iterations. The number of these iterations is determined by the tabu list-hub parameter, which is set to the ceiling of the square root of the total number of nodes, denoted as to $\lceil\sqrt{n}\rceil$. For example, in a network with $n = 25$ nodes, if node 10 was selected as a hub in iteration 1 but changed to a non-hub in iteration 6, it will be saved in the tabu list. As a result, it cannot be chosen as a hub again until $\lceil\sqrt{25}\rceil = 5$ new iterations have passed.

Stopping criterion: If the best solution does not improve over a specified number of iterations, the Tabu Search algorithm will terminate. This number is determined by the *MaxIter* parameter, which is set to 80 iterations.

Algorithm 1 depicts the pseudo code for the proposed Tabu Search heuristic.

Algorithm 1 Tabu Search

Step 1: {Initialization}

- Set *Iter* = 0, *MaxIter* = θ, and *Tabu List* = φ
- Construct the initial solution by selecting p nodes as hubs in descending order of the ratio λ_m/μ_m.
- Generate two random numbers as the number of servers for each hub.
- Allocate each pair of O-D to hubs by Lemma 1 and calculate the objective function.
- Consider this as the current solution X_c as well as the best solution X_{best}. Go to the improvement stage.

Step 2: {Improvement}

- Select two nodes from those that are not currently existing hubs or tabu nodes.
- Make 2×*p* neighborhoods of the current solution.
- Perform the best exchange and generate a new solution. Update the tabu list.
- If the new solution improves the best solution, then:

  Record it as the best new solution X_{best} and Set *Iter* = 0.

  Else, let *Iter* ← *Iter*+1.

Step 3: {Restart}

While *Iter* ≤ θ*:*

Proceed with the improvement stage.

End while

For the best solution, X_{best}, optimally distribute *R* servers among the *p* hubs and save the resulting configuration.

Step 4: {Output}

- Report the best solution, X_{best}, and its objective function value.

3.3.2 A Genetic algorithm to solve the HCLP-(q_h ,x_h)

The Genetic algorithm (GA), introduced by Holland in the 1970s, is an optimization technique and stochastic search method based on a set of genetic-inspired operations. Solutions are represented in the form of chromosomes, which collectively form a population. These

operations encompass selection, crossover, mutation, improvement, and replacement, all of which are applied to the population. The objective of the GA is to ensure the survival and propagation of the fittest individuals or structures, ultimately driving the optimization process (Holland, 1992).

Chromosome representation: In the proposed GA-based approach, the representation of each chromosome is a $1 \times n$ vector, where *n* represents the number of nodes in the network. The values assigned to each gene within the chromosome adhere to the constraints (3.12), (3.13), and (3.14). These constraints specify that the gene values must be integers ranging from 0 to $R - p + 1$, where the upper bound denotes the number of servers available. By satisfying this range, constraint (3.11) is ensured.

If a gene within the chromosome is assigned a value of 0, it signifies that the corresponding node does not have a server and is classified as a non-hub node. To guarantee that the initial solution meets constraints (3.4) and (3.9), p genes are assigned non-zero values, while the remaining genes are set to zero.

Fitness function: In GA approach, chromosomes (solutions) are evaluated based on their fitness values. The fitness value of each chromosome shows the quality of the solution. In our proposed GA algorithm, we define the fitness function based on the value of the objective function (3.10). Since we want to maximize the coverage, higher fitness values are favorable. To do this, we need to allocate the non-hubs to hubs by Lemma 1.

Population size: Let *R* be the maximum number of servers. We can select *p* hubs out of *n* nodes in $\binom{n}{p}$ ways, and the number of possibilities *R* servers can be distributed among *p* hubs is equal to the number of positive, integer-valued vectors $(x_1, x_2, \dots, x_p)$ such that $x_1 + x_2 + \cdots + x_p = R$. In order to determine this number, imagine that we have *R* indistinguishable items lined

up and that we want to divide them into p nonempty groups. To accomplish this, we can select $p - 1$ of the $R - 1$ spaces between adjacent items as our dividing points. Therefore, according to the Fundamental Counting Principle, in total, there are $\binom{n}{p} \times \binom{R-1}{p-1}$ potential solutions for the problem of finding p hubs out of n nodes and distributing R servers among the p hubs. We performed several experiments on $n/2$, n, and $2\times n$ problem sizes to determine the population size. The experiments showed that a population with size n results in a better final solution. We discuss the results of our experiments in Section 5.

Initial population: The initial population is generated randomly.

Selecting the parents: The parents are selected randomly from the population. We experimented with biased selection mechanisms where fitter parents were more likely to be selected for reproduction, but no significant impact was observed on the performance of the algorithm. While both versions of the algorithm produced excellent results, we obtained marginally better results with randomly selected parents. Figure 3-2 illustrates the situation in which there are two parents in the current population.

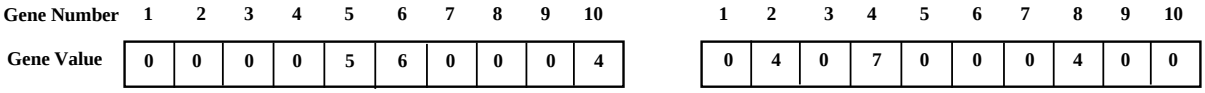

Figure 3-2 Parents representation in GA

Crossover: We present a novel crossover operation where four offspring are reproduced. Initially, all genes are set to zero. In order to produce offspring 1, the first gene of parent 1 is considered. In the case that this gene is a hub node, offspring 1's first gene inherits the value of

that gene; otherwise, parent 2's n^{th} gene determines the value of offspring 1's n^{th} gene. Next, the second gene of parent 1 is considered. If this gene is a hub node, offspring 1's second gene inherits the value of that gene; else, the operator will consider the $(n\text{-}1)^{th}$ gene of parent 2. This continues until the number of hubs in offspring 1 reaches p. Offspring 2 is reproduced in a similar way, only the order is swapped. The first gene of parent 2 is considered to determine the first gene of offspring 2 and if it is a hub node, offspring 2's first gene inherits the value of that gene; otherwise, parent 1's n^{th} gene determines the value of offspring 2's n^{th} gene. This continues until the number of hubs in offspring 2 reaches p. In order to produce offspring 3, the first gene of parent 1 is considered. If this gene is a hub node, offspring 3's first gene inherits the value of that gene; else, the first gene of parent 2 is considered. This continues until the number of hubs in offspring 3 reaches p. Finally, in order to generate offspring 4, the n^{th} gene of parent 1 is considered. If this gene is a hub node, offspring 4's n^{th} gene inherits the value of that gene; else, the n^{th} gene of parent 2 is considered. Again, this continues until the number of hubs in offspring 4 reaches p. Finally, the offspring with the best fitness value is selected as the candidate member. The crossover mechanism is shown in Figure 3-3 for an example of $n = 10$ nodes and $p = 3$ hubs.

Replacement: We admit a candidate member into the population if it is different from existing members, and if its fitness value is higher than the lowest fitness value in the population. In this situation we admit the candidate member and discard the worst one. Such a policy gradually improves the average fitness value of the population while maintaining genetic diversity. We keep track of the worst and best members of the population after every population update. The steps of the replacement operator are as follows:

Input: One candidate member.

Step 1. If the candidate offspring is not different from all existing chromosomes in the population, discard the candidate and stop the replacement operator. Else, go to step 2.

Step 2. If the fitness value of the candidate is less than the minimum fitness value, discard the candidate and stop the replacement operator. Else, go to step 3.

Step 3. Remove the worst member of the population and accept the candidate offspring.

Output: Update the population.

Mutation: The mutation operator plays a crucial role in enabling the algorithm to avoid getting stuck in local optima. In our approach, we utilize two mutation operators. The first operator is triggered with a low probability and is applied to a randomly chosen chromosome within the population. This operator then proceeds to select both a hub gene and a non-hub gene at random. The values of these selected genes are swapped, resulting in the non-hub gene becoming a hub node, while the hub gene is assigned a value of zero. This exchange of values helps introduce novel configurations and promotes exploration of different solutions, potentially allowing the algorithm to escape local optima and discover more favorable solutions. The second mutation operator is applied when the algorithm observes the stopping criterion. First, the best solution is introduced to the mutation operator. Then, the operator selects each hub gene and replaces it with each non-hub gene. Finally, the hub gene is swapped for the one non-hub gene that makes the best improvement in the fitness value and a new member is produced. Lastly, all possible combinations of distributing R servers across p hubs are considered and the optimal combination is chosen. The number of possibilities R servers can be distributed among p hubs is equal to the number of

positive, integer-valued vectors $(x_1, x_2, \dots, x_p)$ such that $x_1 + x_2 + \cdots + x_p = R$ and $x_i \geq k$ for $i \in \{1, \dots, p\}$ where k is the minimum number of servers in each hub. In order to determine the number of combinations, imagine that we have R indistinguishable items lined up and that we want to divide them into p nonempty groups and each group has at least k items. It can be shown the number of solutions to the equation is equal to $\binom{R-(k-1)\times p-1}{p-1}$. In this study, we assume each hub contains at least 4 servers (k = 4). In fact, $x_i \geq k = 4$ enforces a range of zero to one for q_m. Our largest problem size has $R = 30$ and $p = 5$. Therefore, the number of solutions (combinations) to the equation $x_1 + x_2 + x_3 + x_4 + x_5 = 30$ is equal to $\binom{30-(4-1)\times 5-1}{5-1}$ = 1001. The solution time for the current part is negligible due to the small number of combinations.

By utilizing this mutation operator when the stopping criterion is met, the algorithm introduces additional diversity and explores different potential solutions, aiming to refine and improve the overall quality of the best solution found thus far.

Termination: If the best solution is not improved in several successive iterations, GA will stop. The number of these iterations is shown by the *MaxIter* parameter which is set to 80.

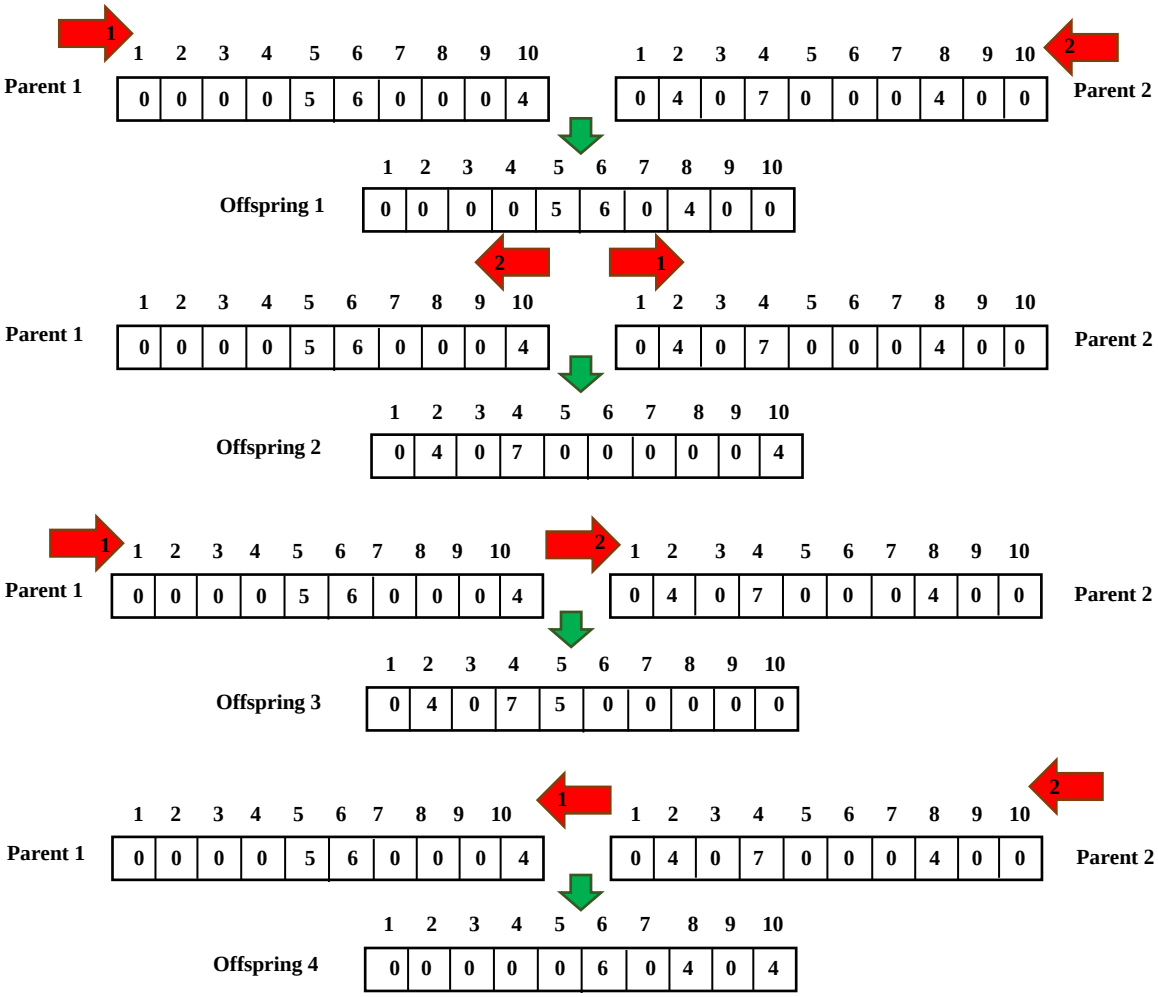

Figure 3-3 Offspring representation

Algorithm 2 depicts the pseudo code for the proposed Genetic algorithm.

Algorithm 2 Genetic algorithm

Step 1: {Initialization}

- Generate an initial population of size *n* as described.
- Initialize *Iter* = 0 to keep track of successive iterations where the best solution found does not change.
- Set *MaxIter* = θ
- Set Prob = ρ

Step 2: {Improvement}

While *Iter* < θ*:*

- **Parents selection:** Select two members from the current population randomly.
- **Crossover:** Input these two members and produce 4 candidate members.
- **Lemma 1**: Allocate the non-hubs to hubs by Lemma 1 and calculate its fitness value.
- **Replacement**: Input each candidate member if possible.
- **Mutation Type 1**: Generate a random value. If it is greater than ρ, apply the Mutation Type 1 as described to replace a randomly chosen chromosome with a new one.
- If the best solution found so far has not changed, then increment *Iter*.

End While

Step 3: {Mutation Type 2}

- Run the Mutation Operator Type 2 and improve the best solution if possible.

Step 4: {Output}

- Report the best member as the final solution of the algorithm.

These two algorithms were also implemented to solve the HCLP-(q_h ,s_h) model, with some modifications.

3.4 Experimental results

To evaluate the proposed models, our focus is on the queues formed by airplanes waiting to land on or take off from airport runways, which are considered as servers. An airplane must pass through runways, ramps, and gates. Considering that 50% to 70% of takeoff delays are related to

runways (Idris, et al., 2016), the focus of this study is only on runways. The airplanes enter the airport for landing or takeoff operation. If there are free runways, they get serviced immediately; else, they are queued up until one becomes available. We collected data from the US department of Transportation's Bureau of Transportation Statistics (BTS). The collected data set contains all domestic passenger flights' origins and destinations for the year 2019. At each airport, we aggregated the total number of incoming and outgoing flights. Subsequently, we sorted the airports based on the sum of their landings and departures. Table 3-1 presents the names and three-letter codes of the 50 busiest airports, while Figure 3-4 visually represents their locations on the map.

Table 3-1 Top 50 busiest airports

1- Chicago O'Hare International Airport (ORD)	14- Boston Logan International Airport (BOS)	27- Chicago Midway International Airport (MDW)	40- Cincinnati/Northern Kentucky International Airport (CVG)
2- Hartsfield-Jackson Atlanta International Airport (ATL)	15- Miami International Airport (MIA)	28- San Diego International Airport (SAN)	41- Indianapolis International Airport (IND)
3- Dallas/Fort Worth International Airport (DFW)	16- Minneapolis−Saint Paul International Airport (MSP)	29- Memphis International Airport (MEM)	42- Pittsburgh International Airport (PIT)
4- Los Angeles International Airport (LAX)	17- Detroit Metropolitan Wayne County Airport (DTW)	30- Tampa International Airport (TPA)	43- Louis Armstrong New Orleans International Airport (MSY)
5- Denver International Airport (DEN)	18- Philadelphia International Airport (PHL)	31- Portland International Airport (PDX)	44- William P. Hobby Airport (HOU)
6- Charlotte Douglas International Airport (CLT)	19- LaGuardia Airport (LGA)	32- Nashville International Airport (BNA)	45- Kansas City International Airport (MCI)
7- Harry Reid International Airport (LAS)	20- Orlando International Airport (MCO)	33- St. Louis Lambert International Airport (STL)	46- Cleveland Hopkins International Airport (CLE)
8- George Bush Intercontinental Airport (IAH)	21- Salt Lake City International Airport (SLC)	34- Dallas Love Field Airport (DAL)	47- San Antonio International Airport (SAT)
9- John F. Kennedy International Airport (JFK)	22- Fort Lauderdale-Hollywood International Airport (FLL)	35- Raleigh-Durham International Airport (RDU)	48- Ontario International Airport (ONT)
10- San Francisco International Airport (SFO)	23- Daniel K. Inouye International Airport (HNL)	36- Norman Y. Mineta San Jose International Airport (SJC)	49- General Mitchell International Airport (MKE)
11- Newark Liberty International Airport (EWR)	24- Dulles International Airport (IAD)	37- Louisville International Airport (SDF)	50- Jacksonville International Airport (JAX)
12- Seattle-Tacoma International Airport (SEA)	25- Ronald Reagan Washington National Airport (DCA)	38- Austin-Bergstrom International Airport (AUS)	-
13- Phoenix Sky Harbor International Airport (PHX)	26- Baltimore/Washington International Thurgood Marshall Airport (BWI)	39- Oakland International Airport (OAK)	-

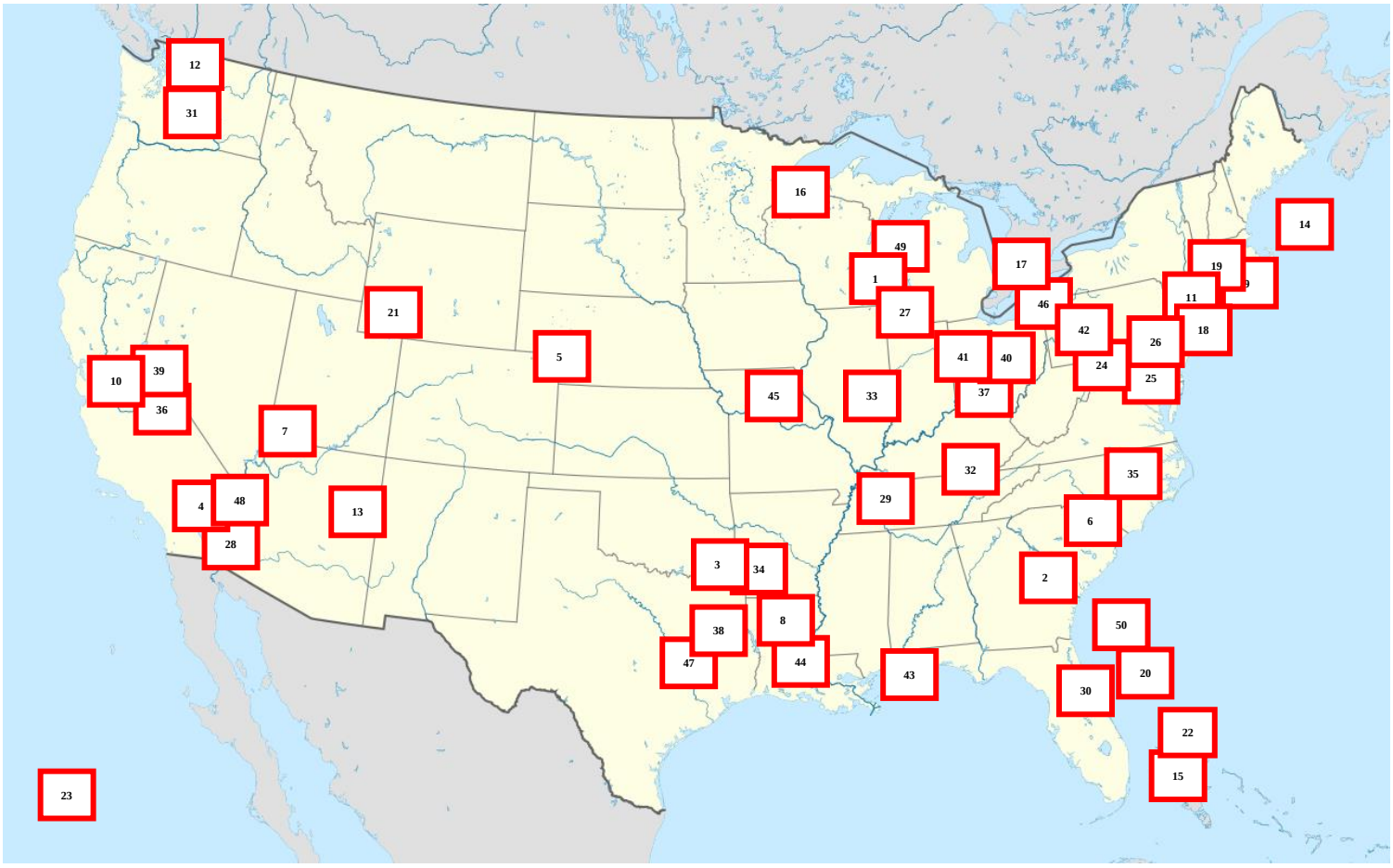

Figure 3-4 50 busiest airports on map

This study aims to design a network of hubs for American Airlines, one of the largest domestic carriers in 2019, by routing the traffic between each pair of the top 50 airports. The arrival rates of airplanes known as airport tower operations, the peak hour service rates known as call rate, and the number of existing runways of each airport for the year 2019 have been published (Federal Aviation Administration, 2020). The arrival rates for airports are calculated based on the combined average number of peak takeoffs and landings, rather than being computed as separate average values for each metric. This approach considers both takeoffs and landings together to determine the overall arrival rate at an airport. By combining these two metrics, a comprehensive

understanding of the airport's traffic flow can be obtained, enabling more accurrate calculations and planning. Therefore, we assume that the existing runways can be used for both landing and takeoff. Arrival rates, service rates, number of runways, q values used in the HCLP-(q_h ,s_h) model, and λ/μ values used in the HCLP-(q_h ,x_h) model are shown in Table 3-2. These parameters are calculated using real data.

Table 3-2 The q and λ/μ values

Potential Hub number	Airport Code	$\frac{\lambda}{\mu}$	$q = \frac{\lambda}{s\mu}$	Potential Hub number	Airport Code	$\frac{\lambda}{\mu}$	$q = \frac{\lambda}{s\mu}$
1	ORD	3.959	0.495	26	BWI	1.297	0.432
2	ATL	2.261	0.452	27	MDW	1.881	0.376
3	DFW	3.054	0.436	28	SAN	0.547	0.547
4	LAX	2.300	0.575	29	MEM	0.709	0.177
5	DEN	2.177	0.363	30	TPA	0.627	0.209
6	CLT	1.738	0.434	31	PDX	0.788	0.263
7	LAS	2.507	0.627	32	BNA	0.693	0.173
8	IAH	1.713	0.343	33	STL	0.806	0.202
9	JFK	2.441	0.610	34	DAL	0.391	0.195
10	SFO	2.262	0.566	35	RDU	0.681	0.340
11	EWR	1.945	0.649	36	SJC	0.320	0.160
12	SEA	1.733	0.578	37	SDF	0.660	0.220
13	PHX	1.130	0.377	38	AUS	0.541	0.270
14	BOS	3.407	0.568	39	OAK	0.895	0.224
15	MIA	1.479	0.370	40	CVG	1.424	0.356
16	MSP	1.292	0.323	41	IND	0.709	0.355
17	DTW	1.791	0.299	42	PIT	1.370	0.342
18	PHL	1.774	0.444	43	MSY	0.482	0.241
19	LGA	1.171	0.585	44	HOU	0.625	0.156
20	MCO	1.177	0.294	45	MCI	0.281	0.094
21	SLC	1.168	0.292	46	CLE	0.479	0.160
22	FLL	0.879	0.440	47	SAT	0.569	0.190
23	HNL	1.347	0.337	48	ONT	0.318	0.159
24	IAD	1.054	0.263	49	MKE	0.543	0.109
25	DCA	1.522	0.507	50	JAX	0.297	0.148

There were 102,424,429 domestic passengers carried by American Airlines in 2019 within the 50 busiest airports. To form the flow matrix, W, we calculated the exact number of flow

American Airlines made between each of the airport pairs for the year 2019. We divided the number of flows by 8760 hours thus obtaining the average hourly flow for each origin-destination pair.

In this section, computational experiments are performed to solve the HCLP-(q_h ,s_h) and HCLP-(q_h ,x_h) models and to evaluate the proposed metaheuristics. We conduct a series of experiments on different problem sizes, including a selection of small (25 airports), medium (40 airports), and large (50 airports) numbers of airports. The experiments conducted involve testing different values of the discount factor α, selected from the set {0.2, 0.5, 0.8}, and the parameter p, chosen from the set {3, 4, 5}. The maximum number of servers (R) is set to $6\times p$.

The total flow of domestic flights operated by American Airlines during the peak hour is as follows: 8213 for the top 25 busiest airports, 9193 for the top 30 busiest airports, 10501 for the top 40 busiest airports, and 11692 for the top 50 busiest airports. The HCLP-(q_h ,s_h) model, QHCLP-(q_h ,f_{hs}) model, the GA, and TS are coded in C++ programming language and GUROBI 9.1.2 is used as the solver. Computational experiments were performed using the same PC running on an Intel Core i72.8 GHz machine with 32 GB RAM.

To find the desired level for population size, we performed several experiments on $n/2$, n, and $2\times n$ population sizes where n is the problem size. We set n = 40, p = 5, and α = 0.5 for all these three problems. Then we ran the algorithm 50 times for each population size. We collected 50 fitness values as well as 50 solution times for each problem. We performed one-way ANOVA analysis for both fitness values and solutions times.

One-way ANOVA analysis for solution times

This experiment involves a single factor, population size, at three levels ($n/2$, n, and $2\times n$). Thus, we have a completely randomized design with $k = 3$ treatments. Let μ_1, μ_2, and μ_3 represent the means of solution times for populations of sizes $n/2$, n, and $2\times n$, respectively. Then we want to test

$H_0: \mu_1 = \mu_2 = \mu_3$

against

H_a: At least two of the three treatment means differ.

The SAS representation analysis for the hypothesis is shown in Table 3-3. The F statistic for testing the hypothesis is F =43.04, where the distribution of F is based on $v_1 = (k - 1) = 3 - 1 = 2$ and $v_2 = (N - k) = 150 - 3 = 147$ degree of freedom. For $\alpha = 0.05$, the critical value is $F_{0.05} = 3.058$. Since the calculated F, 43.04, is greater than the critical value, $F_{0.05} = 3.058$, We reject H_0 and conclude that the mean of solution times for a problem with $n = 40$, $p = 5$, and $\alpha = 0.05$ differs in at least two of the three population sizes.

Table 3-3 One-way ANOVA Analysis for solution times

Source	DF	Sum Of Squares	Mean Square	F Value	Pr > F
Model	2	11314.36032	5657.18016	43.04	<.0001
Error	147	19319.56020	131.42558		
Corrected Total	149	30633.92053			

Pairwise *t* tests for solution times

Since ANOVA doesn't tell which treatments are significantly different from each other, we perform multiple pairwise comparison analysis using one-tailed *t*-test. Referring to Table 3-4, the results suggest that we fail to reject the null hypothesis for the pairwise comparison between population of sizes $n/2$ and n, at significance level $\alpha = 0.05$. However, for the pairwise comparison between population of sizes $n/2$ and $2\times n$ and the comparison between population of sizes n and $2\times n$, since *p-value* < 0.05, we reject the null hypothesis and conclude that the means of solution times for population of sizes $n/2$ and n are significantly less than the mean of solution times for population of size $2\times n$. Figure 3-5 shows the distributions of solution times of three populations. We see that the mean of solution times for population of size $2\times n$ is much greater than the mean of solution times for population of sizes $n/2$ and n.

Table 3-4 Pairwise *t* tests between solution times for population of sizes $n/2$, n, and $2\times n$

One-Tailed Test	t statistic	DF	p-value
$H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 < 0$	-0.5346	98	0.2971
$H_0: \mu_1 - \mu_3 = 0$ $H_a: \mu_1 - \mu_3 < 0$	-7.4095	98	<0.00001
$H_0: \mu_2 - \mu_3 = 0$ $H_a: \mu_3 - \mu_3 < 0$	-7.6759	98	<0.00001

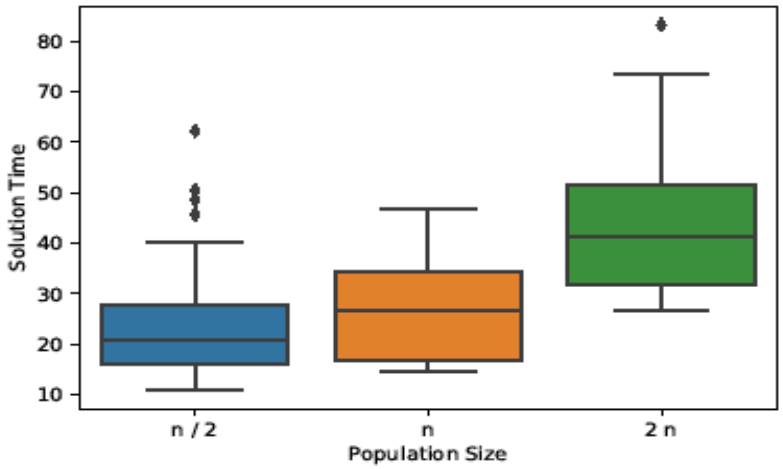

Figure 3-5 The distribution of solution times for population of sizes *n*/2, *n*, and 2×*n*

One-way ANOVA analysis for fitness values

This experiment involves a single factor, population size, at three levels (*n*/2, *n*, and 2×*n*). Thus, we have a completely randomized design with $k = 3$ treatments. Let μ_1, μ_2, and μ_3 represent the means of fitness values for populations of sizes *n*/2, *n*, and 2×*n*, respectively. Then we want to test

$H_0: \mu_1 = \mu_2 = \mu_3$

against

H_a: At least two of the three treatment means differ.

Table 3-5 shows the SAS representation analysis for the hypothesis. The *F* statistic for testing the hypothesis is $F = 0.31$, where the distribution of *F* is based on $v_1 = (k - 1) = 3 - 1 = 2$ and $v_2 = (N - k) = 150 - 3 = 147$ degree of freedom. For $\alpha = 0.05$, the critical value is $F_{0.05} = 3.058$. Since the calculated *F*, 0.31, is less than the critical value, $F_{0.05} = 3.058$, we fail to reject H_0. There is insufficient evidence (at $\alpha = 0.05$) of a difference among the means of fitness values for population of sizes *n*/2, *n*, and 2×*n*. Three population sizes are shown in Figure 3-6 along with

their fitness distributions. Based on the analysis, the variation observed between the sample means is relatively small compared to the variation within each sample. As a result, there is insufficient evidence to suggest significant differences between the means.

Table 3-5 One-way ANOVA Analysis for fitness values

Source	DF	Sum Of Squares	Mean Square	F Value	Pr > F
Model	2	56444.19	28222.09	0.31	0.7314
Error	147	13236709.90	90045.65		
Corrected Total	149	13293154.09			

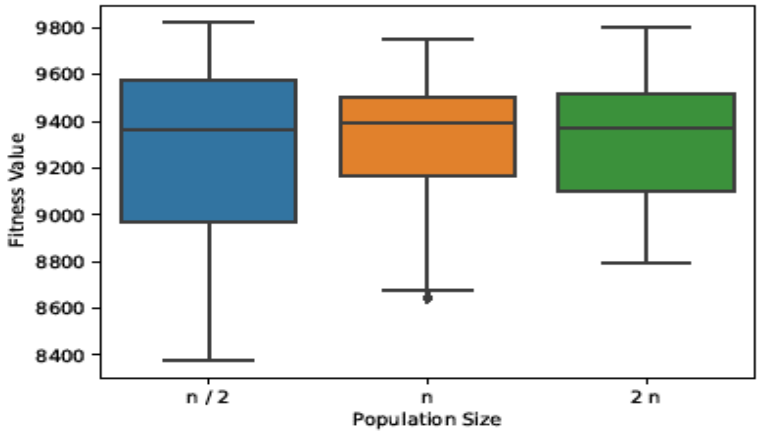

Figure 3-6 The distribution of fitness values for population of sizes $n/2$, n, and $2\times n$

Regarding the mean of solution times, we observed that the algorithm with population of sizes n and $n/2$ outperform the algorithm when population size is set to $2\times n$. Also, we observed that there is insufficient evidence (at $\alpha = 0.05$) of a difference among the means of fitness values

for three population sizes. Referring to Figure 3-6, since the within-sample variation for population of size $n/2$ is higher than that for population of size n, we choose n as the desired population size for our GA.

Table 3-6, Table 3-7, and Table 3-8 present the results of the HCLP-(q_h ,s_h) model solved by GUROBI, TS, and GA, respectively, for n = 30, 40, and 50. In order to assess the performance of TS and GA, each test instance mentioned in Table 3-6, Table 3-7, and Table 3-8 was repeated 10 times. The meaning of the column headings are as follows: The first two columns specify the test problem where p is the required number of hubs, and α is the discount factor. The optimal objective value, solution time for solving the model by GUROBI, and the optimal hub locations are given in columns 3, 4, and 5, respectively. The columns under the headings "TS" and "GA" represent the average percentage gap (*agap%*), the average run-time of the algorithm over 10 runs in seconds (t(sec)), and the number of times the algorithm obtained the optimal solution out of 10 runs (No. of optimal). The average percentage gap is defined as the gap between the 10 solutions by the algorithm and Opt_{sol}; i.e., $Gap = \left(\frac{\sum_{i=1}^{10} gap_i}{10}\right)$, where $gap_i = \frac{100(Opt_{sol} - sol_i)}{Opt_{sol}}$ and sol_i represents the solution obtained by the algorithm in its i^{th} execution.

The results presented in these three tables provide clear evidence that both algorithms are effective in solving the HCLP-(q_h ,s_h) model and achieving optimal solutions across small, medium, and large instances within a short timeframe. In contrast, GUROBI solver requires several hours to obtain optimal solutions for the majority of medium instances and all large instances. This highlights the superior efficiency and effectiveness of the algorithms in solving the problem compared to GUROBI in these scenarios. For example, GUROBI's solution time is

101,688 seconds when n = 50, p = 4, and α = 0.2, whereas TS and GA both find the optimal solution in less than 50 seconds (Table 3-8).

As shown in Table 3-6 and Table 3-7, both algorithms perform similarly across small and medium instances. GA also appears to be faster than TS for large-scale instances (Table 3-8). As an example, when n = 50, p = 5, and α = 0.2, the GA's solution time is 33.98 seconds, compared to 53.90 seconds for TS. We also observe that the run-time of GA is not dependent upon the number of hubs, but that of TS increases with the number of hubs.

Table 3-6 Computational results of HCLP-(q_h ,s_h) for small instances (n = 30)

Test bed		GUROBI			TS			GA		
p	α	Opt_{sol}	t(sec)	Hub locations	agap (%)	t(sec)	No. of optimal	agap (%)	t(sec)	No. of optimal
3	0.2	8017.19	3795.50	1,3,6	0.000	10.61	10	0.000	10.89	10
3	0.5	7955.80	3641.63	3,6,27	0.653	10.43	9	0.026	8.94	9
3	0.8	7942.61	1291.77	1,3,6	0.000	10.89	10	0.152	9.44	8
4	0.2	8390.84	3830.89	1,3,6,13	0.204	17.01	9	0.418	11.66	6
4	0.5	8363.92	2669.89	3,6,13,27	0.781	16.76	6	0.152	14.72	9
4	0.8	8350.71	512.14	1,3,6,13	0.245	16.82	9	0.000	17.08	10
5	0.2	8644.41	1950.41	1,3,6,9,13	0.176	18.77	7	0.510	16.30	4
5	0.5	8611.56	1303.45	3,6,9,13,27	0.112	21.09	8	0.236	13.02	9
5	0.8	8598.35	615.03	1,3,6,9,13	0.124	17.41	7	0.867	16.62	5

Table 3-7 Computational results of HCLP-(q_h ,s_h) for medium instances (n = 40)

Test bed		GUROBI			TS			GA		
p	α	Opt_{sol}	t(sec)	Hub locations	agap (%)	t(sec)	No. of optimal	agap (%)	t(sec)	No. of optimal
3	0.2	9073.60	12196.73	1,3,6	0.000	14.48	10	0.000	15.36	10
3	0.5	8996.68	2320.08	3,6,27	0.000	14.87	10	0.000	13.50	10
3	0.8	8983.91	2717.13	1,3,6	0.000	14.22	10	0.073	13.24	9
4	0.2	9579.71	38703.13	1,3,6,13	0.393	23.83	9	0.000	14.28	10
4	0.5	9537.33	2704.80	3,6,13,27	0.057	23.36	9	0.167	20.49	7
4	0.8	9524.52	2268.70	1,3,6,13	0.000	24.75	10	0.000	29.53	10
5	0.2	9890.26	91773.31	1,3,6,9,13	0.000	27.08	10	0.096	24.98	8
5	0.5	9833.83	3341.94	3,6,13,18,27	0.067	29.89	9	0.034	32.80	9
5	0.8	9820.90	4311.20	1,3,6,13,18	0.000	25.12	10	0.000	19.96	10

Table 3-8 Computational results of HCLP-(q_h ,s_h) for large instances (n = 50)

Test bed		GUROBI			TS			GA		
p	α	Opt_{sol}	t(sec)	Hub locations	agap (%)	t(sec)	No. of optimal	agap (%)	t(sec)	No. of optimal
3	0.2	10089.44	87543.18	1,3,6	0.000	29.49	10	0.125	23.13	8
3	0.5	10002.85	58771.19	3,6,27	0.024	28.62	8	0.062	19.88	9
3	0.8	9990.74	57108.05	1,3,6	0.000	30.16	10	0.243	25.74	8
4	0.2	10683.66	101688.73	1,3,6,13	0.205	47.69	8	0.000	34.72	10
4	0.5	10631.61	61925.18	3,6,13,27	0.895	47.17	8	0.217	26.08	8
4	0.8	10619.47	44087.09	1,3,6,13	0.006	48.91	9	0.000	23.46	10
5	0.2	11016.87	112771.15	1,3,6,13,18	0.015	53.90	6	0.276	33.98	8
5	0.5	10963.06	62820.25	3,6,13,18,27	0.000	59.08	10	0.000	31.80	10
5	0.8	10950.79	11962.75	1,3,6,13,18	0.216	51.79	9	0.024	21.96	9

In order to compare the HCLP-(q_h ,s_h) and the HMCP presented by Campbell (Campbell J. F., 1994), we solved the latter using GUROBI for n = 40, p = 5, and α = 0.2, 0.5, 0.8. As shown

in Table 3-9, this problem exhibits alternate optimal solutions across all instances, possibly due to HMCP's disregard for the server busy probability and its sole focus on the coverage radius. The next part of our discussion will be about how the server busy probability impacts our solution. In this regard, to compare the HCLP-(q_h ,s_h) with the HMCP, the objective value of the HCLP-(q_h ,s_h) was calculated at the optimal solution of HMCP for n = 40 and p = 5. The meaning of the column headings in Table 3-9 are as follows: The first column represents the discount factor α. The optimal hub locations selected by the HMCP and the HCLP-(q_h ,s_h) are given in columns 2 and 3, respectively. Column 4 shows the objective value of the HCLP-(q_h ,s_h) at the optimal solution of the HMCP. The optimal objective value of the HCLP-(q_h ,s_h) is given in column 5. The last column gives the difference between the objective value of the HCLP-(q_h ,s_h) and that of the HMCP. The findings derived from Table 3-9 demonstrate that, in the majority of instances, the objective value of the HCLP-(q_h ,s_h) model at the HMCP solution differs from the optimal expected coverage. This discrepancy indicates that incorporating the busy probability factor results in a modification of the hub network configuration. By accounting for this uncertainty, the model captures real-world characteristics more accurately. It acknowledges the dynamic nature of network environments, where varying factors, such as busy probabilities, can influence the optimal configuration of hub locations. Consequently, incorporating such uncertainty into the model provides a more realistic representation of the problem and improves the applicability of the results to real-world scenarios.

Table 3-9 Comparison between the HCLP-(q_h ,s_h) and the HMCP for n = 40 & p = 5

α	Hub locations in HMCP model	Hub location in the presented model	Obj value of the proposed model at the HMCP solution	Opt.sol of the presented model	gap
0.2	1,3,6,9,13	1,3,6,9,13	9890.26	9890.26	0.00
0.2	3,6,9,13,27		9883.39	9890.26	6.87
0.2	1,3,6,13,19		9816.64	9890.26	73.62
0.2	1,6,9,13,34		9771.71	9890.26	118.55
0.5	3,6,13,19,27	3,6,13,18,27	9747.72	9833.83	86.11
0.5	3,6,9,13,27		9813.09	9833.83	20.74
0.5	6,9,13,27,34		9695.27	9833.83	138.56
0.5	6,13,19,27,34		9629.90	9833.83	203.93
0.8	6,13,18,27,34	1,3,6,13,18	9599.46	9820.90	221.44
0.8	3,6,13,18,27		9815.19	9820.90	5.71
0.8	1,3,6,13,18		9820.90	9820.90	0.00
0.8	1,6,9,13,34		9670.68	9820.90	150.22
0.8	3,6,9,13,27		9778.85	9820.90	42.05
0.8	1,3,6,9,13		9784.55	9820.90	36.35

The existing number of runways in the airports ranges from 1 to 8. Therefore, we require the decision variable x_h in the HCLP-(q_h ,x_h) to take values from the same range. In the quadratic model, Q-HCLP-(q_h ,f_{hs}), f_{hs} takes 1 if s servers are allocated to hub h; otherwise, it is equal to zero. Equivalent to x_h, s can range from 1 to 8. We solve Q-HCLP-(q_h ,f_{hs}) for n = 25 using GUROBI and HCLP-(q_h ,x_h) using the TS as well as the GA. Since these two models are equivalent, their optimal solutions are the same. The results are shown in Table 3-10.

In order to assess the performance of the TS and GA, each test instance repeated 10 times. The column headings in Table 3-10 have the following meanings: As shown in the first three columns, p is the number of hubs, α is the discount factor, and R is the maximum number of servers that can be opened. The optimal objective, run-time for solving the model by GUROBI, the optimal hub locations, and the optimal number of servers in each hub (No. of servers) are given in columns 4-6, respectively. The columns under the headings “TS” and “GA” are similar to those in Table 3-6, Table 3-7 and Table 3-8.

Taking a look at Table 3-10, both algorithms demonstrated high efficiency in solving the HCLP-(q_h ,x_h) model and achieving optimal solutions across all instances within a remarkably short time frame. As an example, for p = 5 and α = 0.2, GUROBI requires 41,944 seconds to find the optimal solution. However, both algorithms find the optimal solution within 5 seconds. We also observe that the largest average gaps for TS and GA are equal to 0.35% and 0.36%, respectively. This indicates that algorithms' final solutions are optimal or very close to the optimal solutions.

GUROBI was not able to obtain the optimal solution for Q-HCLP-(q_h ,f_{sh}) when n = 50. Instead, we solve the equivalent problem, HCLP-(q_h ,x_h), using the proposed metaheuristics for instances with n = 50.

Table 3-11 displays the best solutions obtained using the GA and TS algorithms. For both algorithms, *MAXITER* was set to 400 to investigate whether longer runs would lead to significantly improved results. The table reveals that although the best solutions remain unchanged, the number of instances achieving the best solution considerably increases.

While we lack the optimal solutions for the HCLP-(q_h ,x_h) when n = 50, the exceptional outcomes produced by both algorithms across all instances of the HCLP-(q_h ,s_h) and the majority of instances of the HCLP-(q_h ,x_h), coupled with the fact that increasing *MAXITER* does not alter the best solutions found, allow us to confidently infer that the results presented in Table 3-11 from both algorithms are also of exceptional quality.

Table 3-10 Computational results of QHCLP-(q_h ,f_{hs}) and HCLP-(q_h ,x_h) for small instances (n=25)

Test bed			GUROBI			TS			GA		
p	R	α	Opt_{sol}	t(sec)	Hub locations: No. of Servers	agap (%)	t(sec)	No. of optimal	agap (%)	t(sec)	No. of optimal
3	18	0.2	7172.54	13940.48	1:7 - 3:6 - 6:5	0.35	1.52	2	0.07	1.51	7
3	18	0.5	7099.98	1633.20	1:7 - 3:6 - 6:5	0.23	1.46	3	0.22	1.94	4
3	18	0.8	7099.98	1435.09	1:7 - 3:6 - 6:5	0.08	1.44	7	0.07	1.18	7
4	24	0.2	7558.90	14220.12	1:7 – 3:7 – 6:6 -13:4	0.01	2.08	9	0.19	2.62	4
4	24	0.5	7520.39	3008.05	1:7 – 3:7 – 6:6 -13:4	0.12	2.66	6	0.05	2.40	7
4	24	0.8	7520.35	1770.34	1:7 – 3:7 – 6:6 -13:4	0.26	2.19	4	0.36	2.64	3
5	30	0.2	7867.47	41944.06	1:8 – 3:8 – 6:6 – 13:4 – 19:4	0.03	4.43	7	0.05	4.14	8
5	30	0.5	7821.99	7641.16	1:8 – 3:8 – 6:6 – 13:4 – 19:4	0.09	5.15	7	0.02	5.94	9
5	30	0.8	7821.95	2153.73	1:8 – 3:8 – 6:6 – 13:4 – 19:4	0.31	4.96	4	0.03	5.03	8

Table 3-11 Computational results of HCLP-(q_h ,x_h) for large instance (n = 50)

Test bed			Hub locations: No. of Servers	TS						GA					
				MAXITER= 80			MAXITER=400			MAXITER= 80			MAXITER=400		
p	*R*	*α*		Best Obj	No. of Best (Out of 10)	t(sec)	Best Obj	No. of Best (Out of 10)	t(sec)	Best Obj	No. of Best (Out of 10)	t(sec)	Best Obj	No. of Best (Out of 10)	t(s)
3	18	0.2	3:7 - 6:6 - 27:5	10181.71	8	33.42	10181.71	10	244.63	10181.71	8	31.06	10181.71	10	229.84
3	18	0.5	3:7 - 6:6 - 27:5	10101.29	6	25.53	10101.29	10	205.16	10101.29	5	41.36	10101.29	9	392.08
3	18	0.8	3:7 - 6:6 - 27:5	10082.05	9	10.71	10082.05	10	98.23	10082.05	9	22.02	10082.05	10	162.92
4	24	0.2	6:7 -13:6 - 27:7 - 34:4	10832.05	2	33.67	10832.05	9	281.54	10832.05	1	57.32	10832.05	6	543.44
4	24	0.5	6:7 -13:6 - 27:7 - 34:4	10788.18	9	34.12	10788.18	9	293.17	10788.18	10	27.75	10788.18	10	263.04
4	24	0.8	6:7 -13:6 - 27:7 - 34:4	10768.80	8	32.39	10768.80	10	274.93	10768.80	6	53.53	10768.80	10	396.08
5	30	0.2	6:8-13:7-19:6-27:5-34:4	11201.01	1	70.46	11201.01	8	576.81	11201.01	1	69.15	11201.01	5	511.72
5	30	0.5	6:7-13:6-18:6-27:7-34:4	11132.75	2	57.24	11132.75	9	441.23	11132.75	4	51.84	11132.75	10	383.64
5	30	0.8	6:7-13:6-18:6-27:7-34:4	11113.24	8	56.26	11113.24	10	545.66	11113.24	6	61.36	11113.24	10	473.24

A random selection of two, three, and four non-hub nodes was tested in the improvement stage of the Tabu Search algorithm in order to find the right number of randomly selected nodes. According to the Table 3-12, for $n = 50$, except for one case, it has been observed that the solution times decrease when two non-hub nodes are selected. Additionally, the algorithm can find the best solution more frequently despite the fact that the objective solution remains the same. Increasing the number of randomly selected non-hubs to 3 or 4 results in a longer solution time, but we do not necessarily get the best solution more often. Thus, the improvement step includes two randomly selected non-hubs.

Table 3-12 Tabu Search computational results of HCLP-(q_h ,x_h) for 1, 2, 3, and 4 randomly selected non-hubs (n = 50)

Test bed			Hub locations: No. of Servers	TS											
				No. of randomly selected non-hubs = 1			No. of randomly selected non-hubs = 2			No. of randomly selected non-hubs = 3			No. of randomly selected non-hubs = 4		
p	*R*	*α*		Best Obj	No. of Best (Out of 10)	t(sec)	Best Obj	No. of Best (Out of 10)	t(sec)	Best Obj	No. of Best (Out of 10)	t(sec)	Best Obj	No. of Best (Out of 10)	t(s)
3	18	0.2	3:7 - 6:6 - 27:5	10181.71	8	35.70	10181.71	9	33.42	10181.71	7	39.97	10181.71	9	51.60
3	18	0.5	3:7 - 6:6 - 27:5	10101.29	6	29.18	10101.29	8	25.53	10101.29	6	33.95	10101.29	7	49.58
3	18	0.8	3:7 - 6:6 - 27:5	10082.05	9	15.11	10082.05	9	10.71	10082.05	9	20.43	10082.05	8	28.16
4	24	0.2	6:7 -13:6 - 27:7 - 34:4	10832.05	2	40.54	10832.05	4	33.67	10832.05	3	43.85	10832.05	1	45.83
4	24	0.5	6:7 -13:6 - 27:7 - 34:4	10788.18	9	32.64	10788.18	9	34.12	10788.18	8	38.57	10788.18	9	43.90
4	24	0.8	6:7 -13:6 - 27:7 - 34:4	10768.80	8	37.76	10768.80	9	32.39	10768.80	7	40.17	10768.80	9	47.32
5	30	0.2	6:8-13:7-19:6-27:5-34:4	11201.01	1	80.98	11201.01	5	70.46	11201.01	2	86.67	11201.01	3	89.06
5	30	0.5	6:7-13:6-18:6-27:7-34:4	11132.75	2	60.43	11132.75	5	57.24	11132.75	2	63.74	11132.75	2	66.39
5	30	0.8	6:7-13:6-18:6-27:7-34:4	11113.24	8	58.36	11113.24	9	56.26	11113.24	9	68.07	11113.24	6	71.27

In addition, we conduct statistical tests to compare the objective values and solution times of the two algorithms. These tests aim to determine whether there are any statistically significant differences between the performance of the algorithms in terms of the objectives achieved and the time required to obtain solutions. The HCLP-(q_h ,x_h) is executed on instances with n = 50, α = 0.5, and p = 3 using the proposed metaheuristics. Each algorithm is run 30 times, and the objective values and solution times are recorded for analysis. The first step in our comparison involves examining the mean of the objective values obtained from these runs. Let μ_{tso} represents the mean of 30 objective values for the TS and μ_{gao} represents the mean of 30 objective values for the GA. Then we want to test

$H_0: \mu_{tso} - \mu_{gao} = 0$

against

$H_a: \mu_{tso} - \mu_{gao} \neq 0.$

We also test if there is a significant difference between the mean of solution times. Let μ_{tss} represents the mean of 30 solution times for the TS and μ_{gas} represents the mean of 30 solution times for the GA. We want to test

$H_0: \mu_{tss} - \mu_{gas} = 0$

against

$H_a: \mu_{tss} - \mu_{gas} \neq 0.$

Referring to Table 3-13, the results suggest that we fail to reject the null hypothesis of the first test. Since *p-value* > 0.05, there is insufficient evidence (at α = 0.05) of a difference between the mean of objective values. Similarly, we fail to reject the null hypothesis of the second test. Again, with *p-value* > 0.05, there is insufficient evidence (at α = 0.05) of a difference between the

mean of solution times. Therefore, there is insufficient evidence (at α = 0.05) to conclude any differences between the two metaheuristics are statistically significant. Figure 3-7 and Figure 3-8 show the distribution of objective values and solutions times, respectively. It is evident that the variation between the sample means is small relative to the within-sample variation, and therefore, there is little evidence to imply that the means are significantly different.

Table 3-13 One-tailed *t* tests for TS and GA

Test	Two-Tailed Test	t statistic	DF	p-value
Test 1: Difference between objective values means of two algorithms	H_0: $\mu_{tso} - \mu_{gao} = 0$ H_a: $\mu_{tso} - \mu_{gao} \neq 0$	-0.8244	58	0.4130
Test 2: Difference between solution times means of two algorithms	H_0: $\mu_{tss} - \mu_{gas} = 0$ H_a: $\mu_{tss} - \mu_{gas} \neq 0$	-0.4052	58	0.6868

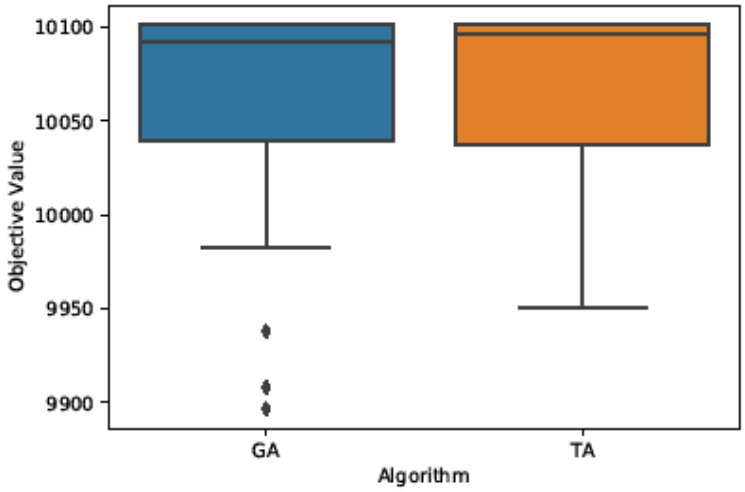

Figure 3-7 The distribution of objective values for GA and TS

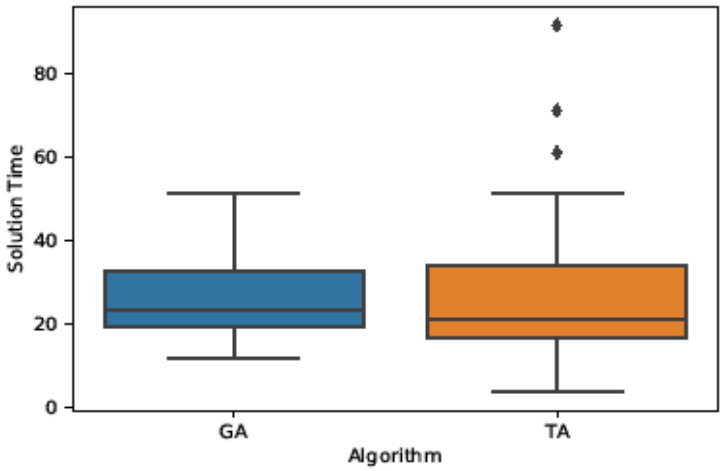

Figure 3-8 The distribution of solution times for GA and TS

As a final step, we analyze how the change in the number of runways impacts the hub network configuration and flow coverage. To show this impact, we solve the QHCLP-(q_h ,f_{hs}) and HCLP-(q_h ,s_h) using GUROBI for n = 25 and α = 0.2, 0.5, 0.8. Table 3-14 compares the two models based on the selected hubs, number of servers in each hub, and coverage. As presented, we observe that for p = 3, the same hubs are selected by both models. However, QHCLP-(q_h ,f_{hs}) allocates different runways to the hubs than are actually available in the hubs. Despite the fact that the total number of existing runways in hubs located by the HCLP-(q_h ,s_h) is greater than those allocated to the hubs by the QHCLP-(q_h ,f_{hs}), the objective function value of the latter is slightly greater. It is due to the fact that hub 6 is one of the busiest hubs and has fewer operational runways than what is allocated by QHCLP-(q_h ,f_{hs}). Moreover, we find that for p = 4 and 5, going beyond fixed number of runways results in different hub network configurations, and consequently, greater gaps between two objective function values.

Table 3-14 Comparison between QHCLP-(q_h ,f_{hs}) and HCLP-(q_h ,s_h) for n = 25

Test bed			HCLP-(q_h ,s_h)		QHCLP-(q_h ,f_{hs})		Gap
p	*R*	*α*	Opt_{sol}	Hubs: Servers	Opt_{sol}	Hubs: Servers	
3	18	0.2	7171.45	1:8 – 3:7 – 6:4	7172.54	1:7 - 3:6 - 6:5	1.09
3	18	0.5	7097.03	1:8 – 3:7 – 6:4	7099.98	1:7 - 3:6 - 6:5	2.95
3	18	0.8	7096.88	1:8 – 3:7 – 6:4	7099.98	1:7 - 3:6 - 6:5	3.10
4	24	0.2	7317.64	3:7 – 6:4 – 13:3 – 24:4	7558.90	1:7 – 3:7 – 6:6 -13:4	241.26
4	24	0.5	7394.57	1:8 – 3:7 – 4:4 – 6:4	7520.39	1:7 – 3:7 – 6:6 -13:4	125.82
4	24	0.8	7314.98	1:8 – 3:7 – 20:4 – 24:4	7520.35	1:7 – 3:7 – 6:6 -13:4	205.37
5	30	0.2	7715.34	1:8 - 3:7 - 13:3 - 20:4 - 24:4	7867.47	1:8 – 3:8 – 6:6 – 13:4 – 19:4	152.13
5	30	0.5	7608.75	1:8 – 3:7 – 4:4 – 6:4 – 13:3	7821.99	1:8 – 3:8 – 6:6 – 13:4 – 19:4	213.24
5	30	0.8	7664.92	1:8 – 3:7 – 13:3 – 20:4 – 24:4	7821.95	1:8 – 3:8 – 6:6 – 13:4 – 19:4	157.03

3.5 Conclusion

Common hub maximal covering models assume that hub nodes are always available. They do not take into account the probability that a particular hub will be busy at a given time. The problem is that hub nodes can become clogged during peak hours, causing hub-and-spoke networks to run inefficiently. To deal with this problem, we considered each hub node as a queueing system. We developed two hub covering location models in which the hub nodes' busy probability was considered. The first model assumed that the number of servers in each hub is known. In the second model, we considered the number of servers in each hub as a decision variable. Since the proposed models were NP-hard, two metaheuristics were developed based on the Genetic algorithm and Tabu Search. Experimental results on American Airlines domestic flights in 2019 showed the importance of considering the busy fractions of servers. Furthermore, the results demonstrated that the algorithms are efficient for solving large-scale problems.

Chapter 4 – An efficient model for the multiple allocation hub maximal covering problem

Chapter 4 is based on the manuscript "An Efficient Model for the Multiple Allocation Hub Maximal Covering Problem" Published in Optimization Methods and Software (Maleki, Majlesinasab, & Sinha, 2023).

4.1 Introduction

The hub location problem is one of the interesting subjects in location theory. This problem can benefit transportation, telecommunication, or delivery systems, in which hub nodes are responsible for receiving, collecting, and delivering commodities. The hub node routes traffic between the origin-destination pair and makes a profit when the traffic is dense (Daskin, 1997). There are two major decisions that need to be made in hub location problems: the location of hub nodes and the allocation of non-hubs to hub nodes. The hub facilities should be located such that all demand is covered and the cost for the hubs is minimized. Sometimes the limited budget does not allow to cover all origin-destination pairs. As a result, some origin-destination pairs are not covered. One option is to solve a hub maximal covering problem which maximizes the demand covered with a predetermined number of hubs nodes (Bo Qu, Kerui Weng, 2009).

This study analyzes multiple allocation hub maximal covering problem (MAHMCP) first introduced by Campbell (Campbell J. F., 1994) and then developed by Qu and Weng (Bo Qu, Kerui Weng, 2009). In this problem, the flow of an origin-destination pair is transferred via multiple hubs such that the total flow covered by located hubs is maximized. The model proposed by Qu and Weng (Bo Qu, Kerui Weng, 2009), referred to as P1 in this chapter, has less constraints and variables than Campbell's model (Campbell J. F., 1994). However, it has more binary

variables which makes the model hard to solve. Although the authors provided an evolutionary algorithm to solve the model, they could not get optimal solutions for $n > 25$, where n denotes the number of nodes.

In this research, we develop a new model for the MAHMCP. We improve the root relaxation value by tightening the constraints of the proposed model such that the linear constraints form a smaller polyhedron. We theoretically show that the new model is stronger than the past models and uses a significantly smaller number of binary variables. We can solve the new model for larger sizes in a very short time compared to P1. For example, the branch and bound running time for solving our model with $n = 70$ nodes and 5 hubs is improved by 99.15 percent compared to the branch and bound running time for P1. Also, the branch and bound is not able to solve P1 in 3 hours when $n \geq 90$. However, it solved our proposed model with 100 nodes and 5 hubs in 174.51 seconds.

The rest of this chapter is structured as follows: In section 4.2, we provide a summary of the MAHMCP by Qu and Weng (Bo Qu, Kerui Weng, 2009) and propose theories for developing the new model. We consider the complexity of the proposed model. Section 4.3 provides numerical studies to investigate the performance of the proposed approach. Finally, the chapter is concluded in section 4.4 with discussion and future work.

4.2 Mathematical models

In this section, first, we present one of the most recent MAHMCP models, and then, we develop a new formulation.

4.2.1 Initial MIP formulation of the problem

Let $G = (V, E)$ be a complete undirected graph with the node set $V = \{1, 2, \ldots, n\}$. The node set V includes origins, destinations, and potential hub locations. Each pair of nodes is connected by an arc $(i, j) \in E$ with distance d_{ij}, $d_{ij} = d_{ji}$. Each arc from node *i* to node *j* has a demand $W_{ij} \geq 0$. Let $C_{ij}^{km} = \chi c_{ik} + \alpha c_{km} + \delta c_{mj}$ be the transportation cost from node *i* to node *j* passing through hubs *k* and *m* where χ and δ are the cost coefficients for transporting from a non-hub to a hub and from a hub to a non-hub, respectively. α is the discount factor for inter-hub transportation costs and $0 \leq \alpha \leq 1$. The parameter β_{ij} is the maximum acceptable transportation cost from node *i* to node *j*. Let V_{ij}^{km} denotes a binary parameter which indicates whether the origin-destination pair (i, j) can be covered by the candidate hub nodes *k* and *m* and is defined as follows:

$$V_{ij}^{km} = \begin{cases} 1, & if\, C_{ij}^{km} \leq \beta \\ 0, & otherwise \end{cases} \tag{4.1}$$

Then the formulation for MAHMCP is as follows ($P1$):

$$max \sum_i \sum_j W_{ij} R_{ij} \tag{4.2}$$

S.t.

$$\sum_m Z_m = p \tag{4.3}$$

$$R_{ij} \leq \sum_k \sum_m V_{ij}^{km} x_{km} \qquad \forall i, j \tag{4.4}$$

$$z_m + z_k \geq 2x_{km} \qquad \forall k, m \tag{4.5}$$

$$z_m \in \{0,1\} \qquad \forall m \tag{4.6}$$

$$x_{km} \in \{0,1\} \qquad \forall k, m \tag{4.7}$$

$R_{ij} \in \{0,1\}$ $\forall i, j$ (4.8)

where,

R_{ij}=1, if located hubs cover the flow between origin *i* and destination *j* and 0 otherwise,

x_{km}=1, if both nodes *k* and *m* are selected as hubs and 0 otherwise,

z_m= 1, if node *m* is determined as a hub; otherwise, it is equal to 0.

The objection function (4.2) maximizes the total flows covered by hubs. Constraint (4.3) guarantees that exactly *p* hubs are chosen. Constraints set (4.4) ensures that the flow of pair (*i*, *j*) is covered only if there is at least one hub to cover it. Constraints set (4.5) guarantees that x_{km}takes 1 if only nodes *k* and *m* are chosen to serve as hubs (Bo Qu, Kerui Weng, 2009).

We can see that formulation P1 has $2n^2 + n$ variables and $2n^2 + 1$ constraints. Megiddo et al. (Megiddo, Zemel, & Hakimi, 1983) proved that the maximal covering problem (MCP) is NP-hard. Qu and Weng (Bo Qu, Kerui Weng, 2009) argued that if we consider every origin-destination pair as a single node which is equivalent to $E = \{(i,i) : W_{ii} > 0, i \in V\}$, then P1 reduces to a general maximal covering problem (MCP). It means that the MCP is a special case of the MAHMCP and since the MCP is NP-hard, the MAHMCP is also NP-hard. In the next subsections, we provide details for improving formulation P1.

4.2.2 Improved MAHMCP model

The main idea to improve the formulation P1 is to improve the root relaxation value by tightening the linear programming formulation such that the linear constraints form a smaller polyhedron. This provides a better upper bound. We use the Boolean quadric polytope (Padberg, 1989) to provide a stronger formulation compared to P1 by tightening the constraint (4.5) and substituting it by the following equations.

$$z_m \geq x_{km} \qquad \forall k, m \qquad (4.9)$$

$$z_k \geq x_{km} \qquad \forall k, m \qquad (4.10)$$

We call the new formulation P2. Now, we prove that LP relaxation for P2 is stronger than P1.

Proposition 4.1. Let LP1 denote the initial LP relaxation feasible region for P1, i.e, LP1 is the set of $(z, x, R) \in [0,1]^n \times [0,1]^{n\times n} \times [0,1]^{n\times n}$ satisfying (4.3), (4.4), (4.5). Similarly, let LP2 be the set of all $(z, x, R) \in [0,1]^n \times [0,1]^{n\times n} \times [0,1]^{n\times n}$ satisfying (4.3), (4.4), (4.9), (5.10). LP2 is stronger than LP1. In other words, LP2 ⊆ LP1, but LP1⊈ LP2.

Proof. Let $(z^*, x^*, R^*) \in$ LP2. It satisfies (5.3) and (5.4). Besides $\forall m, k \in [n] \times [n]$, we have $z_m^* \geq x_{km}^*$ and $z_k^* \geq x_{km}^*$, therefore $z_m^* + z_k^* \geq 2x_{km}^*$. Thus $(z^*, x^*, R^*) \in$ LP1 and LP2 ⊆ LP1. Now we want to find $(z', x', R') \in$ LP1 such that $(z', x', R') \notin$ LP2. Let $z'_m = 1$ for $m = 1, ..., p$ and $z'_m = 0$ for $m = p + 1, ..., n$, $x'_{1(p+1)} = 0.5$ and $x'_{km} = 0$ for $(k, m) \neq (1, p + 1)$, and finally $R'_{ij} = 0$, $\forall i, j$. The solution belongs to LP1; however, it fails to satisfy constraint (4.10), indicating that it does not belong to LP2. Consequently, we can deduce that LP1⊈LP2. In conclusion, LP2 proves to be stronger than LP1. ■

Next, we determine valid inequalities for the above formulation. As stated by Campbell (Campbell J. F., 1994) and O'Kelly (O'kelly, 1987), discrete hub location problems involve locating a set of fully interconnected hubs. Hence, the subgraph induced by hub nodes is complete and we can consider the subset of hub vertices as a clique. We know that there is an edge between all the nodes in a clique. Here z_m variables correspond to vertices in the clique and x_{km} variables correspond to edges. Therefore, if $z_m = z_k = 1$ then $x_{km} = 1$. Moreover, in formulation P2, it is obvious that if $z_m = z_k = 1$ for an arbitrary pair of hubs (k, m), then x_{km}=1 because it increases the LHS of the constraint (4.4) and so increases the value of objective function. On the other hand, if at least

one of z_m or z_k is equal to zero, then according to constraints (4.9) and (4.10) x_{km}= 0. Hence, we add the valid inequality (4.11) to make the formulation stronger and we observed that adding this constraint provides a better upper bound.

$$x_{km} \geq z_m + z_k - 1 \qquad \forall k, m \tag{4.11}$$

In the following lemma, we prove that z_m is the same as x_{mm}, therefore, we can replace z_m by x_{mm} and reduce the number of variables to $2n^2$ variables.

Lemma 4.1. Variable z_m is the same as variable x_{mm}.

Proof. We want to prove if z_m = 1, then x_{mm} = 1. Let z_m = 1. By constraint (4.9), we have $x_{mm} \leq 1$. By constraint (4.11), we have $x_{mm} \geq 1$. Therefore, $x_{mm} = 1$.

Now, we show that if z_m = 0, then x_{mm} = 0. Let z_m = 0. By constraint (4.9), we have $x_{mm} \leq 0$ and by constraint (4.11), we have $x_{mm} \geq -1$. Therefore, x_{mm} = 0.

Similarly, we can show that if x_{mm} = 1, then z_m = 1, and if x_{mm} = 0, then z_m = 0. ■

Now, we substitute variables z_m and z_k with x_{mm} and x_{kk}, respectively and rewrite constraints (4.3), (4.9)-(4.11) as follows:

$$\sum_{m=1}^{n} x_{mm} = p \tag{4.12}$$

$$x_{mm} \geq x_{km} \qquad \forall k, m \tag{4.13}$$

$$x_{kk} \geq x_{km} \qquad \forall k, m \tag{4.14}$$

$$x_{km} \geq x_{mm} + x_{kk} - 1 \qquad \forall k, m \tag{4.15}$$

For the next improvement, we calculate $max\{V_{ij}^{km}, V_{ij}^{m,k}\}$ for any $i, j, k, m < k$. Then replace V_{ij}^{km} by $max\{V_{ij}^{km}, V_{ij}^{m,k}\}$. After that we need to solve the new formulation P3 for $m < k$.

Since the hub vertices induces a clique, we also add the following inequalities to the model which is called star inequalities (Macambira & Souza, 2000).

$$\sum_{m=1}^{k-1} x_{km} + \sum_{m=k+1}^{n} x_{mk} \leq (p-1)x_{kk} \qquad \forall k \tag{4.16}$$

This inequality means that the number of incident edges to a hub node is $p-1$ and to a non-hub node is 0 because the non-hub node does not belong to the clique. We call the new formulation P4. In Proposition 2, we show that the inequality presented in (4.16) is valid for the model P3.

Proposition 4.2. Let P3 denote the set of $(x, R) \in [0,1]^{n\times n} \times [0,1]^{n\times n}$ satisfying (4.4), (4.12) - (4.15). The inequality (4.16) is valid for P3.

Proof. At first, we add x_{kk} to both sides of the inequality (4.16), therefore, we have

$$\sum_{m=1}^{k} x_{km} + \sum_{m=k+1}^{n} x_{mk} \leq p x_{kk} \qquad \forall k \tag{4.17}$$

Let $(x', R') \in P3$. According to constraint (4.13), for $1 \leq m \leq k$

$$x'_{km} \leq x'_{mm} \qquad \forall k$$

and similarly, by constraint (4.14) for $k+1 \leq m \leq n$ we have

$$x'_{mk} \leq x'_{mm} \qquad \forall k$$

Summing over $1 \leq m \leq n$, we get

$$\sum_{m=1}^{k} x'_{km} + \sum_{m=k+1}^{n} x'_{mk} \leq \sum_{m=1}^{n} x'_{mm} = p \qquad \forall k \tag{4.18}$$

where the last equality is obtained by constraint (4.12).

On the other hand, according to constraint (4.14), for $1 \leq m \leq k$

$$x'_{km} \leq x'_{kk} \qquad \forall k$$

and similarly, by constraint (4.13) for $k + 1 \leq m \leq n$, we have

$$x'_{mk} \leq x'_{kk} \qquad \forall k$$

Summing over $1 \leq m \leq n$, we get

$$\sum_{m=1}^{k} x'_{km} + \sum_{m=k+1}^{n} x'_{mk} \leq n x'_{kk} \qquad \forall k \tag{4.19}$$

Now we know that for any hub vertex *k*, $x'_{kk} = 1$. By inequalities (4.18) and (4.19) (and since $p \leq n$), we get

$$\sum_{m=1}^{k} x'_{km} + \sum_{m=k+1}^{n} x'_{mk} \leq p = p x'_{kk} \qquad \forall k$$

And for any non-hub vertex *k*, $x'_{kk} = 0$. By inequalities (4.18) and (4.19), we get

$$\sum_{m=1}^{k} x'_{km} + \sum_{m=k+1}^{n} x'_{mk} \leq 0 = p x'_{kk} \qquad \forall k$$

Therefore, for both hub nodes and non-hub nodes the inequality (4.17) and so (4.16) holds. In Proposition 3, we show that LP4 is stronger than LP3.

Proposition 4.3. Let LP3 denote the initial LP relaxation feasible region for P3, i.e, LP3 is the set of $(x, R) \in [0,1]^{n \times n} \times [0,1]^{n \times n}$ satisfying (4.4), (4.12) -(4.15). Similarly, let LP4 be the set of all $(x, R) \in [0,1]^{n \times n} \times [0,1]^{n \times n}$ satisfying (4.4), (4.12) -(4.16). LP4 is stronger than LP3. In other words, LP4 $\subseteq$ LP3, LP3 $\nsubseteq$ LP4.

Proof. Let $(x^*, R^*) \in$ LP4. Obviously, it satisfies (4.4), (4.12) -(4.15). Thus $(x^*, R^*) \in$ LP3 and LP4 $\subseteq$ LP3.

Now, we want to find $(x', R') \in$ LP3 such that $(x', R') \notin$ LP4. Let x'_{mm}= 0.5 for $m = 1, \ldots, 2p$ and x_{mm} = 0 for $m = 2p + 1, \ldots, n$, $x'_{j,2p}$ = 0.5 for $j = 1, \ldots, 2p - 1$ and x'_{km}= 0 for the rest of (k, m), and finally R'_{ij} = 0, $\forall i, j$. This solution belongs to LP3, but obviously does not satisfy constraint (4.16) for $k = 2p$ and so does not belong to LP4. Consequently LP3 $\nsubseteq$ LP4. We conclude that LP4 is stronger than LP3.

There is another valid inequality that we add to the model which is related to the number of edges in the clique induced by hubs as follows:

$$\sum_{k=1}^{n} \sum_{m=1}^{k-1} x_{km} = \frac{p(p-1)}{2} \tag{4.20}$$

Let us call the new formulation P5. In Proposition 4, we show that the inequality presented in (4.20) is valid for the model P4.

Proposition 4.4. Let P4 denote the set of $(x, R) \in [0,1]^{n \times n} \times [0,1]^{n \times n}$ satisfying (4.4), (4.12) - (4.16). The inequality (4.20) is valid for P4.

Proof. Let $(x', R') \in$ P5. We need to show that

$$\sum_{k=1}^{n} \sum_{m=1}^{k-1} x_{km} \leq \frac{p(p-1)}{2} \tag{4.21}$$

According to constraint (4.16),

$$\sum_{m=1}^{k-1} x'_{km} + \sum_{m=k+1}^{n} x'_{mk} \leq (p-1) x'_{kk} \qquad \forall k$$

Summing over all *k*, we get

$$\sum_{k=1}^{n}\sum_{m=1}^{k-1} x'_{km} + \sum_{k=1}^{n}\sum_{m=k+1}^{n} x'_{mk} \leq (p-1)\sum_{k=1}^{n} x'_{kk} = (p-1)p$$

where the last inequality is obtained by constraint (4.12). It is trivial that

$$\sum_{k=1}^{n}\sum_{m=1}^{k-1} x'_{km} = \sum_{k=1}^{n}\sum_{m=k+1}^{n} x'_{mk}$$

and hence,

$$\sum_{k=1}^{n}\sum_{m=1}^{k-1} x_{km} \leq \frac{p(p-1)}{2}.$$

Now, we are to show that

$$\sum_{k=1}^{n}\sum_{m=1}^{k-1} x_{km} \geq \frac{p(p-1)}{2} \tag{4.22}$$

By definition we know that $E = \{(i,j) \mid i \in V,\ j \in V\}$. Let H define the set of all hub nodes, $H \subseteq V$. We split set E into two subsets as follows:

$$H \times H = \{(i,j) \mid i \in H, j \in H\},$$

$$H' = E - (H \times H),$$

Where $H \times H$ denotes all arcs connecting hub nodes, and H' denotes all existing arcs in graph G which do not belong to $H \times H$. Then, for any $(m,k) \in H'$, we know that $x'_{mm} = 0$ or $x'_{kk} = 0$. Therefore, using constraints (7), (13), and (14), we have $x'_{km} = 0$. On the other hand, when $(m,k) \in H \times H$, we know that $x'_{mm} = 1$ and $x'_{kk} = 1$. Therefore, by constraint (15), we have $x'_{km} \geq 1$. Hence,

$$\sum_{k=1}^{n}\sum_{m=1}^{k-1} x'_{km} \geq \sum_{k=1}^{n}\sum_{\substack{m=1 \\ (m,k)\in H'}}^{k-1} x'_{km} + \sum_{k=1}^{n}\sum_{\substack{m=1 \\ (m,k)\in H\times H}}^{k-1} x'_{km}$$

$$\geq 0 + \sum_{k=1}^{n} \sum_{\substack{m=1 \\ (m,k)\in H\times H}}^{k-1} 1 = \sum_{k=1}^{p} \sum_{m=1}^{k-1} 1 = \frac{p(p-1)}{2}$$

Constraint (4.20) is obtained by combining inequalities (4.21) and (4.22). ■

Due to the fact that the subgraph induced by hub nodes is complete and there are *p* hubs, each hub is connected to $p-1$ other hubs. Therefore, the following inequalities should hold,

$$x_{21} \leq (p-1),$$

$$x_{31} + x_{32} \leq (p-1),$$

$$x_{41} + x_{42} + x_{43} \leq (p-1),$$

.

.

.

$$x_{n1} + x_{n2} + \cdots + x_{n,n-1} \leq (p-1),$$

Or in summary,

$$\sum_{m=1}^{k-1} x_{km} \leq (p-1) \quad \forall k. \tag{4.23}$$

Now, we prove that inequality (4.23) is valid for the model P5.

Proposition 4.5. Let P5 denote the set of $(x, R) \in \{0,1\}^{n\times n} \times \{0,1\}^{n\times n}$ satisfying (4.4), (4.12) - (4.16), and (4.20). The inequality (4.23) is valid for *P*5.

Proof. From constraint (4.16), we have

$$\sum_{m=1}^{k-1} x_{km} + \sum_{m=k+1}^{n} x_{mk} + \leq (p-1)x_{kk} \leq (p-1) \quad \forall k.$$

Since $\sum_{m=k+1}^{n} x_{mk} \geq 0$, inequality (4.23) holds. ■

In the final step, we relax the assumption that $R_{ij} = \{0,1\}$ and replace it by $0 \leq R_{ij} \leq 1 \ \forall i,j$. We expect that it remarkably reduces the solution time.

Proposition 4.6. Let *P* denote the feasible region for P5 with a difference that $0 \leq R_{ij} \leq 1 \ \forall i,j$. *P* is a perfect formulation for *P*5, i.e., we can solve the integer programming problem by solving a mixed integer programming problem.

Proof. We need to show that, for any $W \in \mathbb{R}_{+}^{n\times n}$, the $max\{\sum_i \sum_j W_{ij} R_{ij} | (x,R) \in P\}$. Since *P* is bounded, the $max\{\sum_i \sum_j W_{ij} R_{ij} | (x,R) \in P\}$ is itself finite.

Let $W \in \mathbb{R}_{+}^{n\times n}$ and consider an optimal solution (x^*, R^*) to the *P*. According to constraint (4.12), $x_{mm}^* = 1$ for *p* number of variables x_{mm} and the rest are equal to 0. Based on constraints (4.7), (4.13) -(4.15), for any *k*, $m \leq k$

$$x_{km}^* = \begin{cases} 1 \ if \ x_{kk}^* = x_{mm}^* = 1, \\ 0 \qquad otherwise \end{cases}$$

Since $W_{ij} \geq 0$ and $0 \leq R_{ij}^* \leq 1$, the objective function to *P* is $\sum_{W_{ij}>0} W_{ij} R_{ij}^* + \sum_{W_{ij}=0} 0$.

According to constraint (4.4) and maximum objective function, we have

$$R_{ij}^* = \begin{cases} 1, & if \ W_{ij} > 0 \ and \ \exists (k,m) \ni V_{ij}^{km} = 1 \ \vee \ x_{km}^* = 1 \\ 0, & if \ W_{ij} > 0 \ and \ \forall (k,m) \ x_{km}^* = 0 \\ \in [0,1], & if \ W_{ij} = 0 \end{cases}$$

(Here notice that based on constraint (4.12) and $p \geq 1$, there exist (k,m) such that $x_{km}^* = 1$). Therefore, there is always an optimal solution to *P* where $R_{ij} = \{0,1\}$. ■

After all, we end up with a formulation with $n^2 + n$ binary variables, n^2 continuous variables and $5 \times \frac{n^2}{2} + 7 \times \frac{n}{2} + 1$ constraints. Let us call the final model P6 which is as follows:

$$max \sum_i \sum_j W_{ij} R_{ij} \quad (4.2)$$

S.t.

$$\sum_{m=1}^{n} x_{mm} = p \quad (4.12)$$

$$R_{ij} \leq \sum_k \sum_m V_{ij}^{km} x_{km} \qquad \forall i,j \quad (4.4)$$

$$x_{mm} \geq x_{km} \qquad \forall k,m \quad (4.13)$$

$$x_{kk} \geq x_{km} \qquad \forall k,m \quad (4.14)$$

$$x_{km} \geq x_{mm} + x_{kk} - 1 \qquad \forall k,m \quad (4.15)$$

$$\sum_{m=1}^{k-1} x_{km} + \sum_{m=k+1}^{n} x_{mk} \leq (p-1)x_{kk} \qquad \forall k \quad (4.16)$$

$$\sum_{k=1}^{n} \sum_{m=1}^{k-1} x_{km} = \frac{p(p-1)}{2} \quad (4.20)$$

$$\sum_{m=1}^{k-1} x_{km} \leq (p-1) \quad \forall k. \quad (4.23)$$

$$x_{km} \in \{0,1\} \qquad \forall k,m \quad (4.7)$$

$$0 \leq R_{ij} \leq 1 \qquad \forall i,j \quad (4.24)$$

4.3 Numerical experiments

We adopt the Australia Post (AP) dataset [1] (T.Ernst & Mohan, 1996) that is derived from the mail flows in an Australian cityas shown in Figure 4-1. The dataset contains 200 nodes each representing a postal district. It includes the coordinates of nodes as well as flow between nodes.

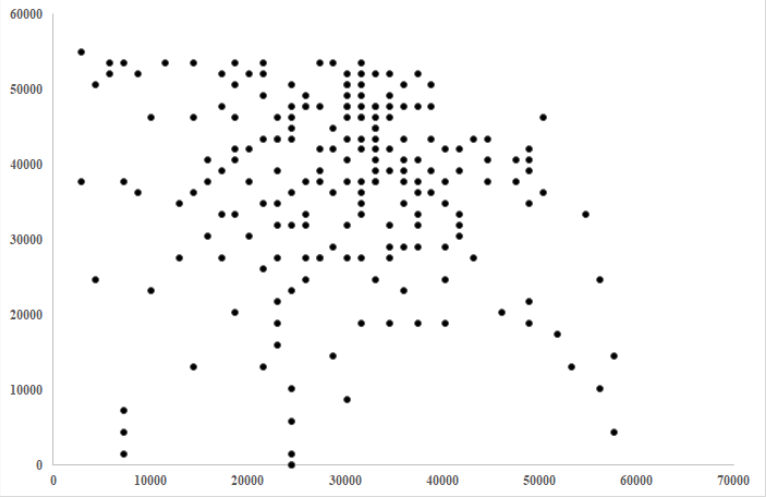

Figure 4-1 AP dataset nodes

Using the provided dataset, we utilize the coordinates of the nodes to calculate the distances between them. For AP dataset, χ and δ are set equal to 3 and 2, respectively. At each node *i*, we added up the total flow coming in and going out. Then, we sorted the nodes based on their flow. The top 20 nodes are shown in Table 4-1.

[1] Hub location (brunel.ac.uk)

Table 4-1 Selected Nodes

Num	Node	Flow	Num	Node	Flow	Num	Node	Flow	Num	Node	Flow
1	158	1114.83	6	156	165.06	11	100	107.79	16	188	79.14
2	150	288.68	7	26	145.82	12	19	106.57	17	32	77.73
3	160	184.25	8	128	119.48	13	34	92.86	18	15	77.06
4	41	177.58	9	126	115.03	14	106	83.64	19	155	75.13
5	146	168.38	10	159	114.70	15	18	79.49	20	173	67.99

We selected small, moderate, and large number of nodes (e.g., 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, and 100 nodes) to evaluate the MIP formulation P6 and its bounds. We ran the model for the cases with p = 3, 4, 5. Qu and Weng (Bo Qu, Kerui Weng, 2009) set α equal to 0.6 and also required the flows passing through hubs to not exceed 20% of their direct path travel cost. To compare our model with P1, we chose the same parameters. We also set χ and δ equal to 3 and 2, respectively for both P1 and P6. We defined χ and δ as collection and distribution costs, respectively. In the context of postal delivery networks, these costs are especially significant. During collection, mail is transported from postcode districts to their designated sorting centers (the hub). Distribution is the process by which mail is moved from a terminal hub to a postal district at its destination. α is the discount factor for inter-hub transportation costs and $0 \leq \alpha \leq 1$ (T.Ernst & Mohan, 1996). In our experiments, we used a coefficient of 3 to calculate the cost of direct path d_{ij}, i.e., $3\times d_{ij}$ as well. Then, as explained we required that the flows passing through hubs should not exceed 20% of their direct path travel cost, i.e., $1.2 \times 3 \times d_{ij}$.

Therefore, we defined V_{ij}^{km} as:

$$V_{ij}^{km} = \begin{cases} 1, & if\ \chi d_{ik} + 0.6 d_{km} + \delta d_{mj} \leq \beta_{ij} \\ 0, & otherwise \end{cases} \tag{4.25}$$

The parameters are presented in Table 4-2.

Table 4-2 Problem parameters

Symbol	Description	Value
n	Number of nodes	15,20,25,40,50,60,70,80,90
p	Number of hubs	3,4,5
χ	Collection cost	3
α	Discount factor on inter-hub flows	0.6
δ	Distribution cost	2
β_{ij}	Cost coefficient	$1.2 \times 3 \times d_{ij}$

Next, we investigate the strength of LP relaxations of formulations P2-P6 as well as the solution times. The MIP formulations are coded in Python and GUROBI 9.1.1 is used as the solver. All experiments were performed on an Intel Core i7 2.80 GHz machine with 16.0 GB RAM.

4.3.1 Computational results

To compare the stronger formulation P2 with $P1$, at first, we carried out some experiments on the AP dataset and obtained the root relaxation values, optimal values and solution time. We set the time limit to 3 hours (10800 seconds). We observed that the solver provides a much better upper bound (root relaxation value) for P2 and solution time is significantly less. In order to see the strength of the proposed valid inequalities in practice, in the next step we added valid inequalities one by one to the model. By performing our proposed preprocessing and consequently reducing the number of variables and constraints, we observed that solution time of branch and bound reduces remarkably.

Table 4-3 shows the strength of the LP relaxation of formulation P6 compared to formulation P1. We observe that the solver provides a much better upper bound (root relaxation value) for formulation P6. We also noticed that for 26 out of 33 instances, the root relaxation value of formulation P6 is equal to the optimal objective value. Also, for other instances that the root

relaxation value of formulation P6 is not equal to the optimal objective value, the gap between them is insignificant. On the other hand, we observe that the root relaxation value of formulation P1 is different from the optimal objective value for all instances, and the gap between them is significantly greater than that of formulation P6. This is the reason why GUROBI spends much more time on finding the optimal solution for P1.

Table 4-4 presents the time it takes to get the optimal solution and time it takes to get the root relaxation for solving formulation P1 and P6 using branch and bound. It is observed that for a given *p*, as *n* increases, the run-time increases for branch and bound. However, for a given *n*, as *p* increases, the run-time may increase or decrease for branch and bound. For example, time to get the optimal solution for solving P1 increases from 807.98 seconds to 1643.95 seconds when n = 60 and *p* increases from 3 to 4, while we observe a decrease in run-time when *p* goes from 4 to 5.

It can be seen that branch and bound running time for solving P6 is significantly less than the running time for P1. For example, the branch and bound running time for solving P6 with n = 70 nodes and p = 5 hubs is improved by 99.15 percent compared to the branch and bound running time of P1 (28.82 seconds vs 3386.43 seconds). The GUROBI spent more than 2 hours to solve problem P1 when n = 80. Also, it was not able to solve Problem P1 in 3 hours when n = 90. However, it solved formulation *P*6 with 100 nodes and 5 hubs in 364.57 seconds. We also observe that for all instances, the number of branch and cut nodes explored (node counts) for formulation P6 is at most 1. However, the node counts significantly increase by the number of nodes, *n*, for formulation P1. In fact, our valid cuts tighten the LP relaxation of the mixed-integer problem P6 such that the root relaxation value is improved significantly. Therefore, the optimal solution is found before exploring many branch and cut nodes (Table 4-3 and Table 4-4).

Table 4-3 Comparing P1 and P6 in terms of root relaxation

				P6		P1	
n	*p*	Objective Value	Hubs	Root relaxation value	Gap (%)	Root relaxation value	Gap (%)
	3	537.193	1,4,5	537.193	0.00	542.485	0.99
15	4	548.608	1,3,4,5	548.608	0.00	554.840	1.14
	5	557.705	1,3,4,5,7	557.705	0.00	562.941	0.94
	3	676.766	1,5,17	676.766	0.00	693.095	2.41
20	4	695.134	1,3,5,17	695.134	0.00	706.422	1.62
	5	706.946	1,3,5,13,17	706.946	0.00	714.889	1.12
	3	797.555	1,5,17	797.555	0.00	823.546	3.26
25	4	816.903	1,4,5,7	817.426	0.06	840.537	2.89
	5	835.270	1,3,4,5,7	835.964	0.08	851.749	1.97
	3	919.642	1,4,5	919.642	0.00	961.572	4.56
30	4	947.519	1,5,14,17	947.519	0.00	979.935	3.42
	5	969.631	1,3,5,17,23	969.631	0.00	991.512	2.26
	3	1154.584	1,5,17	1154.584	0.00	1229.169	6.46
40	4	1197.890	1,5,14,17	1197.890	0.00	1256.197	4.87
	5	1228.740	1,2,5,17,23	1228.740	0.00	1272.644	3.57
	3	1397.436	1,5,17	1397.436	0.00	1510.821	8.11
50	4	1455.602	1,2,5,17	1455.602	0.00	1539.418	5.76
	5	1490.776	1,2,5,17,23	1490.776	0.00	1557.937	4.51
	3	1611.810	1,5,17	1611.810	0.00	1773.460	10.03
60	4	1688.851	1,2,5,17	1688.851	0.00	1806.414	6.96
	5	1737.337	1,2,5,17,23	1737.337	0.00	1851.543	6.57
	3	1834.354	2,5,6	1834.354	0.00	2072.388	12.98
70	4	1919.804	2,3,5,6	1919.804	0.00	2099.614	9.37
	5	1973.873	2,3,6,18,25	1973.873	0.00	2117.187	7.26
	3	2043.296	2,5,6	2043.296	0.00	2334.314	14.24
80	4	2139.258	2,3,5,6	2139.258	0.00	2362.460	10.43
	5	2197.150	2,3,6,8,65	2201.066	0.18	2380.626	8.35
	3	2213.857	2,5,6	2213.857	0.00	-	-
90	4	2333.078	2,3,5,6	2333.078	0.00	-	-
	5	2398.977	2,3,6,18,25	2406.048	0.29	-	-
	3	2373.418	2,5,6	2374.415	0.04	-	-
100	4	2505.396	2,3,5,6	2510.017	0.18	-	-
	5	2580.864	2,3,6,18,25	2595.747	0.58	-	-

Table 4-4 Comparing P1 and P6 in terms of solution times and number of branch and cut nodes explored

n	*p*	Objective Value	Hubs	P6			P1		
				Time to get the optimal solution (sec)	Time to get the root relaxation (sec)	Number of branch-and-cut nodes explored	Time to get the optimal solution (sec)	Time to get the root relaxation (sec)	Number of branch-and-cut nodes explored
15	3	537.193	1,4,5	0.03	0.00	0	0.05	0.00	1
	4	548.608	1,3,4,5	0.04	0.04	0	0.05	0.00	1
	5	557.705	1,3,4,5,7	0.04	0.00	0	0.07	0.00	1
20	3	676.766	1,5,17	0.08	0.00	0	0.13	0.00	1
	4	695.134	1,3,5,17	0.09	0.00	0	0.15	0.00	1
	5	706.946	1,3,5,13,17	0.07	0.00	0	0.17	0.00	1
25	3	797.555	1,5,17	0.21	0.01	0	0.40	0.01	1
	4	816.903	1,4,5,7	0.24	0.02	1	0.43	0.01	1
	5	835.270	1,3,4,5,7	0.24	0.00	1	0.43	0.00	1
30	3	919.642	1,4,5	0.49	0.01	0	1.32	0.02	1
	4	947.519	1,5,14,17	0.49	0.02	0	1.02	0.02	1
	5	969.631	1,3,5,17,23	0.62	0.01	0	1.06	0.02	1
40	3	1154.584	1,5,17	1.99	0.04	0	25.10	0.12	1
	4	1197.890	1,5,14,17	1.93	0.04	0	16.10	0.11	1
	5	1228.740	1,2,5,17,23	1.77	0.04	0	11.67	0.08	1
50	3	1397.436	1,5,17	8.72	0.23	0	93.89	0.35	214
	4	1455.602	1,2,5,17	8.19	0.40	0	134.95	0.23	507
	5	1490.776	1,2,5,17,23	9.11	0.34	1	34.42	0.31	1
60	3	1611.810	1,5,17	13.35	2.04	0	807.98	0.72	1442
	4	1688.851	1,2,5,17	11.72	1.06	0	1643.95	1.11	3586
	5	1737.337	1,2,5,17,23	11.95	1.08	0	656.89	0.33	1080
70	3	1834.354	2,5,6	28.00	4.12	0	3612.54	1.19	2859
	4	1919.804	2,3,5,6	26.81	2.62	0	3174.80	0.86	1086
	5	1973.873	2,3,6,18,25	28.82	4.34	0	3386.43	0.63	1133
80	3	2043.296	2,5,6	59.86	18.43	0	7518.12	38.36	1096
	4	2139.258	2,3,5,6	51.92	10.38	0	7822.19	9.19	1121
	5	2197.150	2,3,6,8,65	64.33	12.00	0	8794.31	6.28	1804
90	3	2213.857	2,5,6	242.78	179.94	0	-	-	-
	4	2333.078	2,3,5,6	185.10	124.20	0	-	-	-
	5	2398.977	2,3,6,18,25	199.88	113.27	1	-	-	-
100	3	2373.418	2,5,6	488.84	369.36	1	-	-	-
	4	2505.396	2,3,5,6	439.58	305.13	1	-	-	-
	5	2580.864	2,3,6,18,25	364.57	174.51	1	-	-	-

We also compared the dual and primal solvers for P6 provided in GUROBI. Table 4-5 shows that primal problems have a longer solution time than dual problems, which indicates that the difficulty is on the primal side.

Table 4-5 Comparing dual and primal in terms of their solution time for P6

		Solution time (sec)	
n	*p*	Dual	Primal
15	3	0.03	0.04
	4	0.04	0.04
	5	0.04	0.05
20	3	0.08	0.09
	4	0.09	0.09
	5	0.07	0.09
25	3	0.21	0.23
	4	0.24	0.25
	5	0.24	0.25
30	3	0.49	0.52
	4	0.49	0.52
	5	0.62	0.51
40	3	1.99	2.19
	4	1.93	2.09
	5	1.77	2.17
50	3	8.72	10.62
	4	8.19	11.44
	5	9.11	10.74
60	3	13.35	14.76
	4	11.72	16.16
	5	11.95	18.34
70	3	28.00	92.50
	4	26.81	58.76
	5	28.82	62.93
80	3	59.86	84.50
	4	51.92	87.11
	5	64.33	162.71
90	3	242.78	199.10
	4	185.10	219.07
	5	199.88	341.31
100	3	488.84	707.01
	4	439.58	566.42
	5	364.57	492.26

Figure 4-2 compares the run-time of branch and bound for solving P1 and P6 when p = 3, 4, 5. As the graphs show, the run-time of branch and bound grows exponentially when $n \geq 60$ for P1. However, the run-time of branch and bound increases almost linearly for $P6$.

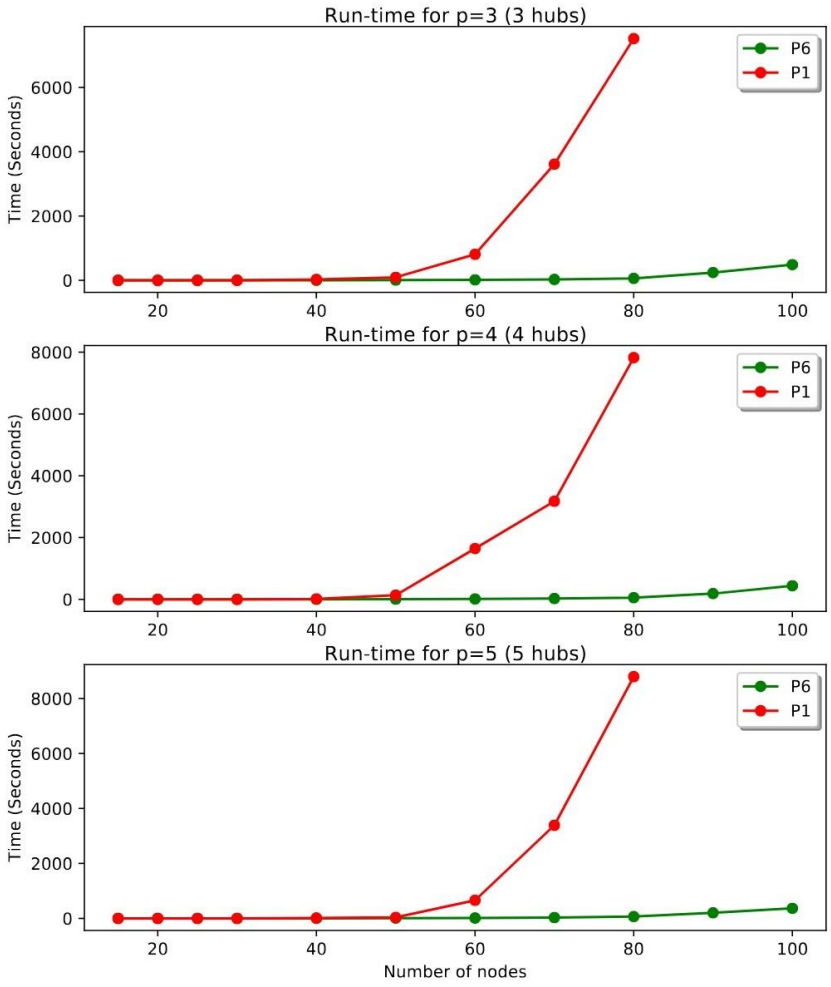

Figure 4-2 Comparison of P1 and P6 in terms of run-time

To compare how each formulation (P2, P3, P4, P5, and P6) adds to the reduction of the total run time, we carried out some experiments. We selected moderate and large number of nodes (e.g., 60 and 80 nodes) to evaluate each formulation. We ran the model for the cases with p = 3, 4,

5. The results are shown in Table 4-6. From P1 to P2, P2 to P3, P3 to P4, P4 to P5, and P5 to P6, we observe that the solution time improves for all cases. Improvements from P1 to P2, P2 to P3, and P5 to P6 are significant while improvements from P3 to P4 and P4 to P5 are smaller. Model P6 was also run without constraints (4.16) and (4.20) in order to see how these constraints affect the results. According to the last column, the solution time of P6 without constraints (4.16) and (4.20) increases significantly. We can conclude that the improvement is remarkably affected by all the valid cuts and changes combined even though it is not significant from P3 to P4 or between P4 and P5.

Table 4-6 Formulation effects on solution time and root relaxation value

n	*p*	Objective Value	P1		P2		P3		P4		P5		P6		P6 without 16 and 20	
			Root value	Sol. time (sec)	Root value	Sol. time (sec)	Root value	Sol. time (sec)	Root value	Sol. time (sec)	Root value	Sol. time (sec)	Root value	Sol. time (sec)	Root value	Sol. time (sec)
60	3	1611.810	1773.460	807.98	1679.203	126.17	1619.206	38.17	1611.810	31.97	1611.810	27.04	1611.810	13.35	1619.158	4504
	4	1688.851	1806.414	1643.95	1734.828	137.99	1698.053	37.42	1688.851	30.60	1688.851	26.71	1688.851	11.72	1698.035	40.52
	5	1737.337	1851.543	656.89	1773.900	115.87	1743.045	34.38	1737.337	28.76	1737.337	26.98	1737.337	11.95	1743.068	38.30
80	3	2043.296	2334.314	7518.12	2111.141	784.57	2098.526	330.96	2043.296	112.65	2043.296	105.62	2043.296	59.86	2098.261	336.84
	4	2139.258	2362.460	7822.19	2184.567	981.62	2177.191	277.05	2139.258	127.49	2139.258	98.19	2139.258	51.92	2177.083	326.76
	5	2197.150	2380.626	8794.31	2239.687	703.60	2234.715	226.89	2201.067	114.81	2201.066	112.60	2201.067	64.33	2234.592	239.05

4.4 Conclusion

In this chapter, we studied the multiple allocation hub maximal covering problem (MAHMCP). We presented an $O(n^2)$ mixed-integer programming formulation of the problem and proved it is stronger than one of the most recent MAHMCP models. We performed an empirical study to benchmark the effectiveness of the new model on the Australia Post (AP) dataset. The new formulation is more effective in terms of run-time and root relaxation value.

Chapter 5 - Chassis inventory management to optimally serve the demand at the intermodal terminals

This chapter is derived from our manuscript titled "Chassis Inventory Management to Optimally Serve the Demand at the Intermodal Terminals," which we have diligently prepared and intend to submit to a reputable scholarly journal specializing in the field of transportation.

5.1 Introduction

The field of shipping and logistics relies heavily on intermodal transportation, which involves the movement of containers using multiple modes of transportation. Intermodal transportation plays a vital role in efficiently transporting goods by seamlessly coordinating various modes such as road, rail, air, and ship. A critical resource in this process is the chassis, which serves as the connecting link between shipping containers and trucks. Chassis play an essential role in intermodal transportation by facilitating the loading, unloading, and subsequent transportation of goods across different transportation modes.

Efficient management of chassis availability is crucial for the smooth operation of intermodal transportation systems. However, a significant challenge faced by intermodal transportation companies is the shortage of chassis at terminals and ports, which are essential hubs for container handling. Figure 5-1 depicts a typical situation at an intermodal terminal, where containers are unloaded from rail cars and placed on chassis. However, during periods of chassis shortages, the available chassis may be insufficient to meet the demand. As a result, containers are stacked on top of each other in a practice known as ground stacking (Figure 5-2). This situation leads to extended waiting times for drivers who have containers at the bottom of the stack, causing

congestion, queues, and delays in the supply chain. The negative consequences of chassis shortages extend beyond increased costs and operational inefficiencies, impacting driver turnover rates and even contributing to inflation.

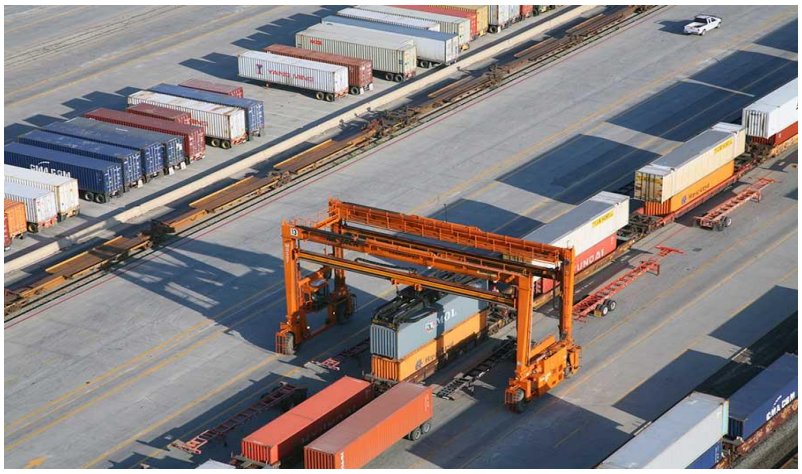

Figure 5-1 Intermodal terminal: unloading rail cars, loading chassis [1]

Figure 5-2 Chassis shortages: ground-stacked containers due to limited available chassis [2]

[1] BNSF Alliance IntermodalFacility Expansion - TranSystems

[2] Chassis Shortage across America - The Scarbrough Group

A review of the existing literature on intermodal transportation reveals that most studies have predominantly focused on the maritime industry, neglecting the unique challenges faced by intermodal transportation at truck-rail terminals. Furthermore, the literature often fails to consider crucial factors such as multiple terminals, the demand for loaded containers, multiple time periods, chassis shortages, and lead time. These gaps in the literature highlight the need for a comprehensive and innovative approach to address the challenges of chassis inventory management in intermodal transportation.

Motivated by these identified research gaps, Chapter 5 introduces a mathematical framework for the optimization of chassis inventory management in intermodal transportation. This innovative approach leverages a multi-time period and multiple terminal perspective, enabling effective management and relocation of chassis and empty containers between terminals based on demand. The primary objective of this research is to minimize the total cost incurred throughout the entire process. By developing a mathematical model, we aim to tackle the challenges associated with chassis shortages and provide practical solutions that fill the existing research gaps in intermodal transportation.

The rest of this chapter is organized as follows: In section 5.2, the mathematical formulations will be presented. Section 5.3 is dedicated to experimental results and recommendations. In section 5.4, we examine a tornado's potential to disrupt one of the nation's busiest terminals. In the last section, we present some concluding remarks.

5.2 Mathematical formulation

In this section, we present the mathematical formulation that underlies our approach for optimizing chassis inventory management in intermodal transportation. We outline the key

variables, constraints, and objectives of the model, providing a clear understanding of the proposed framework. The problem is based on certain assumptions:

• Loaded containers are considered heterogeneous and cannot be stored in inventory.

• On the other hand, empty containers are considered homogeneous and can be stored in the inventory of each terminal.

• Chassis are considered homogeneous and can be stored in the inventory of each terminal.

• Regardless of whether there is an empty or full container to transport, a chassis is always required for road transportation.

• The inventory of chassis at a terminal increases as trucks bring empty containers, loaded containers, or only chassis to the terminal.

• The inventory of chassis at a terminal decreases as trucks take empty containers, loaded containers, or only chassis away from the terminal.

• When loaded containers arrive at a terminal via train, chassis are required to transport them to their final destination. This process reduces the number of chassis held in inventory.

• When an empty container arrives at a terminal via train, the railroad will remove it from the train and place it on a chassis. In this case, the chassis' busy time is negligible, and it does not change the quantity of chassis stocked there.

• Chassis are needed to load and unload containers from the trains.

• Upon arriving at the destination rail facility, the railroad will remove a loaded container from the train and place it on a chassis.

• After the container has been loaded onto a chassis, the rail will send a notification to the draymen and the intermodal provider, informing them that the container is ready for delivery.

• The inventory of empty containers at a terminal increases as trains and trucks bring empty containers to meet the demand of that facility.

• The inventory of empty containers at a terminal decreases as trains and trucks take empty containers away from the terminal to meet the demand at other terminals.

• Each train can haul 300 containers per day, and two trains are available at the beginning of each day for container transportation. Thus, the maximum number of containers that can be hauled in one day is 600.

5.2.1 Chassis and empty container management model (CECM)

To develop the CECM model, we define some notations from graph theory. Let $G = (V, E)$ be an undirected graph with the terminal set $V = \{1,2, \dots, n\}$ and the arc set E. Each pair of terminals is connected by an arc (i, j) where cost depends on the mode of transportation, weight of the load, and the distance to be travelled. Let $K = \{0,1\}$ be the mode of transportation where 0 denotes truck and 1 denotes train. Let *T* denotes time periods indexed by $t = 0,1, \dots, T$.

The parameters used in the model are as follows:

h_{ij}^{k}: distance from *i* to *j* by transportation mode of type *k.*

a_{ij}^{stk}: 1, if it takes $(t - s)$ days to travel from terminal *i* to terminal *j* by transportation mode of type *k*; otherwise, it takes 0.

DE_j^t: the number of empty containers needed at terminal *j* on day *t*.

DL_j^t: the number of loaded containers needed at terminal *j* on day *t*.

w_{ij}^{t}: 1, if there is a demand for loaded container at terminal *j* on day *t* with a pickup at terminal *i*; otherwise, it takes 0.

C^e: price of one empty container.

C^c: price of one chassis.

p^k: The average price to ship by transportation mode of type *k* per ton-mile.

lb^f: Weight of component *f* where $f \in$ {chassis, empty container, loaded container}.

NT: Number of available trains at the beginning of every day to transport containers.

All decision variables in the CECM model are integer and as follows:

Transportation variables:

X_{ij}^{stk}: Number of empty containers picked up from terminal *i* on day *s* and delivered to terminal *j* on day *t* by transportation mode of type *k*.

Y_{ij}^{stk}: Number of loaded containers picked up from terminal *i* on day *s* and delivered to terminal *j* on day *t* by transportation mode of type *k*.

H_{ij}^{st}: Number of chassis picked up from terminal *i* on day *s* and delivered to terminal *j* on day *t* by trucks.

Inventory variables:

IE_j^t: Number of empty containers in inventory at terminal *j* at the end of day *t*.

IC_j^t: Number of chassis in inventory at terminal *j* at the end of day *t*.

Other variables:

NE: Total number of empty containers in the system.

NC: Total number of chassis in the system.

To incorporate the lead time into the model, we introduce two indices, *s* and *t*. These indices enable us to account for the time it takes for the truck/train to travel from one terminal to another. The CECM model is presented as an integer programming formulation as follows:

$$Min\ Z = \ C^e \times NE + \ C^c \times NC + \sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}[X_{ij}^{st0} \times h_{ij}^{0} \times (lb^c + lb^e) \times p^0] +$$

$$\sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[X_{ij}^{st1} \times h_{ij}^{1} \times (lb^e) \times p^1\right] + \sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[Y_{ij}^{st0} \times h_{ij}^{0} \times (lb^c + lb^l) \times p^0\right] +$$

$$\sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[Y_{ij}^{st1} \times h_{ij}^{1} \times (lb^l) \times p^1\right]$$

$$+ \sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[H_{ij}^{st} \times h_{ij}^{0} \times lb^c \times p^0\right] \qquad (5.1)$$

S.t.

$$IE_j^{t-1} + \sum_{k=0}^{1}\sum_{s=1}^{t}\sum_{i=1}^{n} X_{ij}^{stk} \times a_{ij}^{stk} -$$

$$\sum_{k=0}^{1}\sum_{s=t}^{T}\sum_{i=1}^{n} X_{ji}^{tsk} \times a_{ji}^{tsk} = IE_j^t \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.2)$$

$$IE_j^t \geq DE_j^t \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.3)$$

$$\sum_{k=0}^{1}\sum_{s=1}^{t}\sum_{i=1}^{n} Y_{ij}^{stk} \times a_{ij}^{stk} \times w_{ij}^{t} \geq DL_j^t \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.4)$$

$$IC_j^{t-1} + \sum_{i=1}^{n}\sum_{s=1}^{t}[(X_{ij}^{st0} + Y_{ij}^{st0} + H_{ij}^{st}) \times a_{ij}^{st0} - Y_{ij}^{st1} \times a_{ij}^{st1}]$$

$$- \sum_{i=1}^{n}\sum_{s=t}^{T}[(X_{ji}^{ts0} + Y_{ji}^{ts0} + H_{ji}^{ts}) \times a_{ji}^{ts0} = IC_j^t \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.5)$$

$$\sum_{i=1}^{n}\sum_{t=s}^{T}\sum_{k=0}^{1}(X_{ji}^{stk} \times a_{ji}^{stk}) \leq IE_j^{s-1} \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, s \in \{1, \dots, T\} \qquad (5.6)$$

$$\sum_{i=1}^{n}\sum_{s=1}^{t} Y_{ij}^{st1} \times a_{ij}^{st1} +$$

$$\sum_{i=1}^{n}\sum_{s=t}^{T}[(X_{ji}^{ts0} + Y_{ji}^{ts0} + H_{ji}^{ts}) \times a_{ji}^{ts0} \leq IC_j^{t-1} \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.7)$$

$$\sum_{j=1}^{n} IE_j^0 = NE \qquad (5.8)$$

$$\sum_{j=1}^{n} IE_j^t \leq NE \qquad \forall\, t \in \{1, \dots, T\} \tag{5.9}$$

$$\sum_{j=1}^{n} IC_j^0 = NC \tag{5.10}$$

$$\sum_{j=1}^{n} IC_j^t \leq NC \qquad \forall\, t \in \{1, \dots, T\} \tag{5.11}$$

$$NE \geq \sum_{j=1}^{n} DE_j^t \qquad \forall\, t \in \{1, \dots, T\} \tag{5.12}$$

$$\sum_{j=1}^{n} \sum_{i=1}^{n} \sum_{t=s}^{T} [(X_{ij}^{st1} + Y_{ji}^{st1}) \times a_{ij}^{st1} \leq 300 \times NT \qquad \forall\, s \in \{1, \dots, T\} \tag{5.13}$$

$$NE, NC, IE_j^t, IC_j^t, X_{ij}^{stk}, Y_{ij}^{stk}, H_{ij}^{st} \in Z^+ \cup \{0\} \tag{5.14}$$

The objective function (5.1) minimizes the total cost. This function includes 7 terms: *I*) the price of total number of empty containers in the system, *II*) the price of total number of chassis in the system, *III*) the total road transportation cost of empty containers, *IV*) the total rail transportation cost of empty containers, *V*) the total road transportation cost of loaded containers, *VI*) the total rail transportation cost of loaded containers, and *VII*) the total road transportation cost of chassis. Constraints (5.2) are the empty containers inventory balance at each terminal. Constraints (5.3) states that demand for empty containers must be satisfied. Similarly, constraints (5.4) ensure that demand for loaded containers is met. Constraints (5.5) are the chassis inventory balance at each terminal. Following constraints set (5.6), the inventory level of empty containers at terminal *j* at the end of day $s-1$ should be enough to cover the number of empty containers leaving there at time *s*. According to the constraints set (5.7), the number of chassis in inventory at terminal *j* at the end of day $t-1$ must be more than or equal to the number of chassis required to unload the loaded containers arriving by train at terminal *j* on day *t* plus the number of chassis

leaving there on day *t*. Constraint (5.8) states that the inventory of empty containers at all terminals on day 0 equals the number of empty containers in the system. In constraints (5.9), the inventory of empty containers at all terminals at the end of day *t* must be less than or equal to the total number of empty containers in the system. Constraints (5.10) and (5.11) are similar to constraints (5.8) and (5.9), except they are for chassis instead of empty containers. According to constraints (5.12) all empty containers in the system must be greater than or equal to the demand for empty containers at all terminals on day *t*. Constraints (5.13) state that the maximum number of containers which can be hauled in one day is $300 \times NT$. Constraints (5.14) require all decision variables to be non-negative integer.

5.2.2 Chassis and empty container management with tracking trains model (CECMTT)

The CECM model assumes each train could haul 300 containers per day and that two trains would be available at the beginning of every day to transport containers.

The CECM model assumes that each train could haul 300 containers per day and that two trains would be available at the beginning of every day to transport containers. However, in the real world, when a train is transporting containers, it will not be immediately available for further transportation until it arrives at a designated terminal. That said, in a real situation, we may need more trains to handle the exact number of containers that can already be handled by the CECM model with this simplified assumption. As part of our effort to make the model more realistic, we will track trains each day to determine the terminals where they start and end. To do this, we need to relax the constraints (5.13). We also define the following binary decision variable:

Z_{ij}^{stk}: if train *k* departs from terminal *i* on day *s* and arrives to terminal *j* on day *t*; otherwise, it is set to 0.

After defining this new decision variable, the index *k* in variables X_{ij}^{stk} and Y_{ij}^{stk} takes values 1, 2, ..., *K* to denote the train ID number. For road transportation, the index *k* continues to take a value of 0.

Then, we add the following constraints:

$$\sum_{i=1}^{n} Z_{ii}^{01k} = 1 \qquad \forall k \in \{1,2,\dots,K\} \tag{5.15}$$

$$\sum_{j=1}^{n}\sum_{t=1}^{7} Z_{ij}^{1tk} = Z_{ii}^{01k} \qquad \forall\, i \in \{1,\dots,n\}, \forall k \in \{1,2,\dots,K\} \tag{5.16}$$

$$\sum_{j=1}^{n}\sum_{t=s}^{7} Z_{ij}^{stk} = \sum_{l=1}^{n}\sum_{h=1}^{s} Z_{li}^{hs-1k} \qquad \forall\, i \in \{1,\dots,n\}, \forall\, s \in \{2,\dots,7\}, \forall k \in \{1,2,\dots,K\} \tag{5.17}$$

$$X_{ij}^{stk} + Y_{ij}^{stk} \le 300 \times Z_{ij}^{stk} \qquad \forall\, i \in \{1,\dots,n\}, \forall\, j \in \{1,\dots,n\}, \forall\, s \in \{1,\dots,t+1\}, \forall\, t \in \{1,\dots,7\}, \forall k \in \{1,2,\dots,K\} \tag{5.18}$$

$$Z_{ij}^{stk} \in \{0,1\} \quad \forall\, i \in \{1,\dots,n\}, \forall\, j \in \{1,\dots,n\}, \forall\, s \in \{1,\dots,t+1\}, \forall\, t \in \{1,\dots,7\}, \forall k \in \{1,2,\dots,K\} \tag{5.19}$$

Constraints set (5.15) specifies that each train originates from a single terminal. Constraints set (5.16) states that train *k* can either stay at terminal *i* or depart from terminal *i* on day 1, provided that the train originates from the same terminal *i*. Constraint set (5.17) represents the conservation flow constraints at terminal *i* on day *s* for train *k*. It specifies that train *k* can either stay at terminal *i* or depart from terminal *i* on day *s*, under the condition that the train was at the same terminal *i* on the previous day, $s-1$. Constraints set (5.18) ensures that the variables X_{ij}^{stk} and Y_{ij}^{stk} can take

positive values only if the corresponding variable Z_{ij}^{stk} is equal to 1. This implies that if a train exists at terminal *i* on day *s*, it is possible to transport both empty and loaded containers from terminal *i* to other terminals starting from day *s*. Moreover, this constraint set includes a capacity restriction on the total number of containers that can be transported by the train, which is set to a maximum of 300 containers. Finally, Constraints (5.19) require Z_{ij}^{stk} variables to be binary.

There are instances when a train undergoes relocation without carrying any containers, and this is commonly referred to as deadhead travel:

a) Let's consider a scenario where a train is situated at terminal *i* without any surplus of empty containers in its inventory. Furthermore, let's assume that terminal *j* requires empty containers. In such a case, the train can be relocated to a different terminal, say terminal *m*, which possesses a surplus of empty container inventory. The train can then pick up those containers from terminal *m* and transport them to terminal *j*.

b) Another situation occurs when a train is located at terminal *i* and there is a demand for loaded containers to be picked up at terminal *m* and delivered to terminal *j*. In this scenario, the train can be relocated to terminal *m* in order to collect the loaded containers and transport them to terminal *j*. During the journey from terminal *i* to terminal *j*, the train may travel without any cargo.

To accommodate these cases, it is necessary to allow the left-hand side of constraints set (5.18) to take a value of 0, even when the corresponding Z variable takes a value of 1. This flexibility in the constraint allows for the consideration of scenarios where the train undergoes deadhead relocation, as discussed earlier.

In order to obtain a more accurate objective value, it is crucial to include the train deadhead relocation cost in the objective function. The following part, therefore, is added to the objective function (5.1):

$$\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{j=1}^{n}\sum_{i=1}^{n} max(Z_{ij}^{stk} - X_{ij}^{stk} - Y_{ij}^{stk}, 0) \times h_{ij}^{1} \times p^{1} \tag{5.20}$$

When train *k* travels from terminal *i* to terminal *j* while carrying empty or loaded containers, the expression $Z_{ij}^{stk} - X_{ij}^{stk} - Y_{ij}^{stk} \leq 0$ holds. As a result, $max\left(Z_{ij}^{stk} - X_{ij}^{stk} - Y_{ij}^{stk}, 0\right)$ equals 0 indicating no additional cost is added. However, if train *k* travels without any cargo, $X_{ij}^{stk} = 0$, $Y_{ij}^{stk} = 0$, and $Z_{ij}^{stk} = 1$. Therefore, $max\left(Z_{ij}^{stk} - X_{ij}^{stk} - Y_{ij}^{stk}, 0\right)$ equals 1. Note that in the given cost term, we do not consider the weight of the empty container or the weight of the loaded container. This term focuses solely on the deadhead transportation cost, represented by $h_{ij}^{1} \times p^{1}$, without taking into account the container weights.

The expression $max\left(Z_{ij}^{stk} - X_{ij}^{stk} - Y_{ij}^{stk}, 0\right)$ makes the model nonlinear. To linearize it, we can introduce an additional binary variable, let's call it V_{ij}^{stk}. We add the following linear constraints:

$$V_{ij}^{stk} \geq Z_{ij}^{stk} - X_{ij}^{stk} - Y_{ij}^{stk} \quad \forall\, i \in \{1, \dots, n\}, \forall\, j \in \{1, \dots, n\}, \forall\, s \in \{1, \dots, t+1\}, \forall\, t \in \{1, \dots, 7\}, \forall k \in \{1,2, \dots, K\} \tag{5.21}$$

$$V_{ij}^{stk} \geq 0 \quad \forall\, i \in \{1, \dots, n\}, \forall\, j \in \{1, \dots, n\}, \forall\, s \in \{1, \dots, t+1\}, \forall\, t \in \{1, \dots, 7\}, \forall k \in \{1,2, \dots, K\} \tag{5.22}$$

Constraints set (5.22) holds since V_{ij}^{stk} is a binary variable. We also modify the expression in the objective function accordingly by replacing $max\left(Z_{ij}^{stk} - X_{ij}^{stk} - Y_{ij}^{stk}, 0\right)$ with V_{ij}^{stk}.

The new model is called "chassis and empty container management with tracking trains" (CECMTT) and it is structured as follows:

$$Min\ Z = \ C^e \times NE + \ C^c \times NC + \sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}[X_{ij}^{st0} \times h_{ij}^{0} \times (lb^c + lb^e) \times p^0] +$$
$$\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[X_{ij}^{stk} \times h_{ij}^{1} \times (lb^e) \times p^1\right] + \sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[Y_{ij}^{st0} \times h_{ij}^{0} \times (lb^c + lb^l) \times p^0\right] +$$
$$\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[Y_{ij}^{stk} \times h_{ij}^{1} \times (lb^l) \times p^1 + V_{ij}^{stk} \times h_{ij}^{1} \times p^1\right]$$
$$+ \sum_{t=1}^{T}\sum_{s=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[H_{ij}^{st} \times h_{ij}^{0} \times lb^c \times p^0\right] \qquad (5.23)$$

S.t.

$$IE_j^{t-1} + \sum_{k=0}^{K}\sum_{s=1}^{t}\sum_{i=1}^{n} X_{ij}^{stk} \times a_{ij}^{stk} -$$
$$\sum_{k=0}^{K}\sum_{s=t}^{T}\sum_{i=1}^{n} X_{ji}^{tsk} \times a_{ji}^{tsk} = IE_j^t \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.24)$$

$$IE_j^{t-1} + \sum_{k=0}^{K}\sum_{s=1}^{t}\sum_{i=1}^{n} X_{ij}^{stk} \times a_{ij}^{stk} -$$
$$\sum_{k=0}^{K}\sum_{s=t}^{T}\sum_{i=1}^{n} X_{ji}^{tsk} \times a_{ji}^{tsk} \geq DE_j^t \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.25)$$

$$\sum_{k=0}^{K}\sum_{s=1}^{t}\sum_{i=1}^{n} Y_{ij}^{stk} \times a_{ij}^{stk} \times w_{ij}^{t} \geq DL_j^t \qquad \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.26)$$

$$IC_j^{t-1} + \sum_{i=1}^{n}\sum_{s=1}^{t}[(X_{ij}^{st0} + Y_{ij}^{st0} + H_{ij}^{st}) \times a_{ij}^{st0} - Y_{ij}^{st1} \times a_{ij}^{st1}]$$
$$- \sum_{i=1}^{n}\sum_{s=t}^{T}[(X_{ji}^{ts0} + Y_{ji}^{ts0} + H_{ji}^{ts}) \times a_{ji}^{ts0} = IC_j^t \ \ \forall\, j \in \{1, \dots, n\}\ \forall\, t \in \{1, \dots, T\} \qquad (5.5)$$

$$\sum_{i=1}^{n}\sum_{t=s}^{T}\sum_{k=0}^{K}(X_{ji}^{stk}\times a_{ji}^{stk})\le IE_j^{s-1}\qquad \forall\, j\in\{1,\dots,n\}\,\forall\, s\in\{1,\dots,T\} \tag{5.27}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{s=1}^{t}Y_{ij}^{stk}\times a_{ij}^{stk}+$$

$$\sum_{i=1}^{n}\sum_{s=t}^{T}[(X_{ji}^{ts0}+Y_{ji}^{ts0}+H_{ji}^{ts})\times a_{ji}^{ts0}\le IC_j^{t-1}\qquad \forall\, j\in\{1,\dots,n\}\,\forall\, t\in\{1,\dots,T\} \tag{5.28}$$

$$\sum_{j=1}^{n}IE_j^0=NE \tag{5.8}$$

$$\sum_{j=1}^{n}IE_j^t\le NE\qquad \forall\, t\in\{1,\dots,T\} \tag{5.9}$$

$$\sum_{j=1}^{n}IC_j^0=NC \tag{5.10}$$

$$\sum_{j=1}^{n}IC_j^t\le NC\qquad \forall\, t\in\{1,\dots,T\} \tag{5.11}$$

$$NE\ge\sum_{j=1}^{n}DE_j^t\qquad \forall\, t\in\{1,\dots,T\} \tag{5.12}$$

$$\sum_{i=1}^{n}Z_{ii}^{01k}=1\qquad \forall k\in\{1,2,\dots,K\} \tag{5.15}$$

$$\sum_{j=1}^{n}\sum_{t=1}^{7}Z_{ij}^{1tk}=Z_{ii}^{01k}\qquad \forall\, i\in\{1,\dots,n\},\forall k\in\{1,2,\dots,K\} \tag{5.16}$$

$$\sum_{j=1}^{n}\sum_{t=s}^{7}Z_{ij}^{stk}=\sum_{l=1}^{n}\sum_{h=1}^{s}Z_{li}^{hs-1k}\qquad \forall\, i\in\{1,\dots,n\},\forall\, s\in\{2,\dots,7\},\forall k\in\{1,2,\dots,K\} \tag{5.17}$$

$$X_{ij}^{stk}+Y_{ij}^{stk}\le 300\times Z_{ij}^{stk}\qquad \forall\, i\in\{1,\dots,n\},\forall\, j\in\{1,\dots,n\},\forall\, s\in\{1,\dots,t+1\},\ \forall\, t\in\{1,\dots,7\},\forall k\in\{1,2,\dots,K\} \tag{5.18}$$

$$V_{ij}^{stk}\ge Z_{ij}^{stk}-X_{ij}^{stk}-Y_{ij}^{stk}\qquad \forall\, i\in\{1,\dots,n\},\forall\, j\in\{1,\dots,n\},\forall\, s\in\{1,\dots,t+1\},\ \forall\, t\in\{1,\dots,7\},\forall k\in\{1,2,\dots,K\} \tag{5.21}$$

$$Z_{ij}^{stk}, V_{ij}^{stk} \in \{0,1\} \quad \forall\, i \in \{1,\dots,n\}, \forall\, j \in \{1,\dots,n\}, \forall\, s \in \{1,\dots,t+1\}, \forall\, t \in \{1,\dots,7\}, \forall k \in \{1,2,\dots,K\} \qquad (5.22)$$

$$NE, NC, IE_j^t, IC_j^t, X_{ij}^{stk}, Y_{ij}^{stk}, H_{ij}^{st} \in Z^+ \cup \{0\} \qquad (5.14)$$

5.3 Experimental results

To evaluate the proposed model, we selected 10 terminals from the BNSF Railway network that are accessible online[1]. These terminals are located at the following cities (Table 5-1):

Table 5-1 Location of terminals

0- Houston, TX	1- Fort Worth, TX	2- Kansas City, KS	3- Memphis, TN	4- Chicago, IL
5- ST. Paul, MN	6- Oakland, CA	7- Long Beach, CA	8- Seattle, WA	9- Denver, CO

Figure 5-3 shows the aforementioned 10 terminals on a map of the United States. Using Google Maps, we calculated the rail and road distances (mileage and time) between each pair of terminals. The obtained time distances were used to create the binary matrix A = $[a_{ij}^{tsk}]$. As we mentioned earlier, a_{ij}^{tsk}=1 if it takes $(t-s)$ days to travel from terminal i to terminal j by transportation mode of type k; otherwise, it takes 0. Our time starts on day 0 and ends on day 7. On day 0, we have no demand.

[1] Totally intermodal: The big facilities that get freight where it needs to go | Rail Talk | BNSF

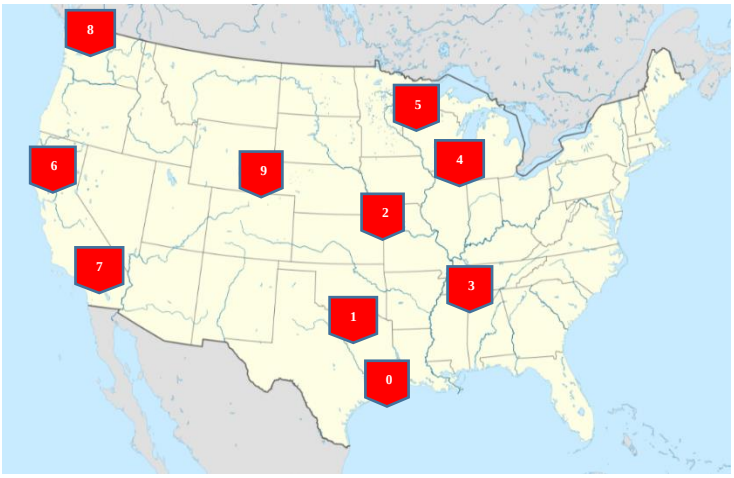

Figure 5-3 Locations of 10 terminals on the map

For each terminal, random numbers are generated to represent the demand for empty containers on days 1, 2, 3, ..., and 7. These numbers are presented in Table 5-2. The first column of the table indicates the delivery terminals. Similarly, random numbers are generated to represent the demand for loaded containers on days 1, 2, 3, ..., and 7 for each terminal. In connection with each demand for loaded containers, a pair of terminals is designated for pickups and deliveries (Table 5-3). In this table, the first column shows delivery terminals. Each tuple has two values: the demand for the loaded containers and the pickup terminal.

Table 5-2 Demand for empty containers

Day / Terminal	1	2	3	4	5	6	7
0- Houston, TX	0	153	0	0	0	156	160
1- Fort Worth, TX	0	0	0	139	139	0	82
2- Kansas City, KS	0	0	0	0	0	211	0
3- Memphis, TN	0	0	125	100	132	0	49
4- Chicago, IL	147	53	0	0	0	85	0
5- ST. Paul, MN	0	0	99	172	72	0	0
6- Oakland, CA	70	0	0	0	0	131	0
7- Long Beach, CA	62	146	111	53	121	0	0
8- Seattle, WA	85	0	0	0	67	0	0
9- Denver, CO	0	128	173	0	0	0	153
Total	364	480	508	464	531	583	444

Table 5-3 Demand for loaded containers

Terminal \ Day	1	2	3	4	5	6	7
0- Houston, TX	0	0	0	(122,3)	(65,3)	(135,7)	(86,9)
1- FT. Worth, TX	0	0	0	0	(103,8)	0	0
2- Kansas City, KS	(105,2)	(134,5)	(118,5)	(183,5)	0	0	0
3- Memphis, TN	(218,2)	(103,10)	0	0	(96,5)	(178,10)	(149,3)
4- Chicago, IL	0	0	0	(74,1)	0	0	0
5- ST. Paul, MN	0	0	(165,4)	(132,3)	0	0	0
6- Oakland, CA	(50,8)	(93,7)	0	0	0	(146,1)	(87,3)
7- Long Beach, CA	0	0	(93,1)	0	0	0	0
8- Seattle, WA	(112,7)	(201,10)	0	0	(167,4)	(67,7)	0
9- Denver, CO	0	0	0	(146,8)	0	0	(170,6)
Total	485	531	376	657	431	526	492

All chassis are homogeneous and have a uniform weight of 3.3 tons each. A 40-foot container is used for both empty and loaded containers. Empty 40-foot containers have a weight of 4 tons each, while loaded containers weigh 21 tons each. A chassis costs $15,000 and a container costs $5,000. Over all markets and types of freight, the average cost to ship by truck was 15.6 cents per ton mile, compared with 5.1 cents for rail in 2015 (Lin A. , 2015). As of the time of writing this chapter in 2022, there has been an 18% decline in the value of the U.S. dollar since 2015. Consequently, we will make necessary adjustments to the prices and rates for the year 2022. Both models (CECM and CECMTT) were coded in Python 3.9 programming language and GUROBI 9.1.2 is used as the solver. Computational experiments were performed using the same PC running on an Intel Core i7 2.8 GHz machine with 16 GB RAM. GUROBI finds the optimal solution in less than a second for 10 terminals and one week.

5.3.3 Observations and recommendations: normal case

In our study, we begin by presenting our observations on the results derived from the CECM model. Based on these findings, we provide insightful recommendations that have the capability to improve the overall efficiency and effectiveness of the system. Additionally, we analyze the results obtained from the CECMTT model and compare its solution with that of CECM.

5.3.3.1 CECM results

The optimal objective value is $35,114,239. On day six, the total demand for empty containers reaches its peak (see Table 5-2). *NE*, the optimal number of empty containers, is 654, slightly higher than the total demand on day six. The optimal number of chassis in the system is 1058. In terms of loaded containers, trucks handle 70.64% of the demand (2471 vs 1027), which explains why there are so many chassis (Table 5-4). The main reason for primarily handling loaded containers through trucks is that every demand for loaded containers has a specified pickup node. The pickup node may be far from the delivery node. That said, it may take several days for a train to move some loaded containers from a specified pickup terminal to the delivery terminal. In this case, we may not be able to make the delivery date. However, truck transportation which is faster makes it possible at the cost of an increase in the objective function.

Table 5-5 shows whether the demand for empty containers at each terminal and day is satisfied by inventory, transportation, or a combination of both. Additionally, it indicates the mode of transportation used to meet the demand. The table includes the following columns: "Day" indicates the start day of transportation. "DE_{jt}" represents the demand quantity at a specific terminal on a given day. "Inventory on day $t-1$" displays the inventory levels at the end of the

day $t-1$. "# Met by Inventory" shows the number of demand quantities fulfilled by existing inventory. "Pickup Info" includes details on terminal and pickup day. "Delivery Info" provides information on terminal and delivery day. "Mode" specifies transportation mode (rail or truck). "No. of empty containers moved" counts empty container transfers.

Our observation is that most of the demand for the empty containers is satisfied either from the previous day's inventory, or via rail. The reason is that we have no specified pickup nodes for empty containers. Consequently, in this scenario, rail transportation is more cost-effective, and we have an inventory of empty containers, so we can fulfill their needs by relocating them from nearby railyards.

Table 5-6 ranks the terminals based on their total number of pickups and deliveries as well as their demands for empty and loaded containers. The Kansas City terminal holds the sixth position with 617 empty and loaded container demands. In terms of total pickups and deliveries, however, this terminal ranks first. Because of this, Kansas City is well positioned to be an intermodal hub for the BNSF Railway network. Additionally, Kansas City's proximity to the geographical center of the country will prove that it is suitable as a hub for intermodal US transportation. Among the top three cities with the most truck pickups and deliveries are Kansas City, Chicago, and Memphis with 725, 708, and 687, respectively. The intermodal network of the BNSF Railway may benefit from having depots and enough facilities for trucks near these three terminals. Details of these statistics can be found in Table 5-6.

Table 5-4 Number of empty and loaded containers transported by truck or train

Empty containers		Loaded containers	
Truck	Train	Truck	Train
186	1370	2471	1027

Table 5-5 Empty containers: inventory and transportation

Day	DE_j^t	Demand Quantity	Inventory on day t-1	# Met by Inventory	Pickup Info	Delivery Info	Mode	No. of empty containers moved
1	DE_4^1	147	147	147	N/A	N/A	N/A	None
	DE_6^1	70	70	70	N/A	N/A	N/A	None
	DE_7^1	62	76	62	N/A	N/A	N/A	None
	DE_8^1	85	85	85	N/A	N/A	N/A	None
2	DE_0^2	153	153	153	N/A	N/A	N/A	None
	DE_4^2	53	147	53	Terminal: Chicago Day: 2	Terminal: Denver Day: 2	Rail	5
	DE_7^2	146	76	76	Terminal: Oakland Day: 2	Terminal: Long Beach Day: 2	Rail	70
	DE_9^2	128	123	123	Terminal: Chicago Day: 2	Terminal: Denver Day: 2	Rail	5
3	DE_3^3	125	0	0	Terminal: Houston Day: 3	Terminal: Memphis Day: 3	Rail	125
	DE_5^3	99	0	0	Terminal: Chicago Day: 3	Terminal: St. Paul Day: 3	Rail	142
	DE_7^3	111	146	111	N/A	N/A	N/A	None
	DE_9^3	173	128	128	Terminal: Seattle Day: 2	Terminal: Denver Day: 3	Rail	45
4	DE_1^4	139	0	0	Terminal: Memphis Day: 4	Terminal: FT. Worth Day: 4	Rail	25
					Terminal: Denver Day: 4	Terminal: FT. Worth Day: 4	Rail	114
	DE_3^4	100	125	100	Terminal: Memphis Day: 4	Terminal: FT. Worth Day: 4	Rail	25
	DE_5^4	172	147	147	Terminal: Seattle Day: 3	Terminal: ST. Paul Day: 4	Rail	30
	DE_7^4	53	146	53	Terminal: Long Beach Day: 4	Terminal: Seattle Day: 5	Rail	8
5	DE_1^5	139	139	71	Terminal: FT. Worth Day: 5	Terminal: Chicago Day: 6	Rail	68
					Terminal: ST. Paul Day: 5	Terminal: FT. Worth Day: 5	Truck	68
	DE_3^5	132	100	100	Terminal: ST. Paul Day: 5	Terminal: Memphis Day: 5	Truck	32
	DE_5^5	72	172	72	Terminal: ST. Paul Day: 5	Terminal: FT. Worth Day: 5	Truck	68
					Terminal: ST. Paul Day: 5	Terminal: Memphis Day: 5	Truck	32
	DE_7^5	121	138	121	Terminal: Long Beach Day: 5	Terminal: Kansas City Day: 6	Rail	17
	DE_8^5	67	10	10	Terminal: Long Beach Day: 4	Terminal: Seattle Day: 5	Rail	8
					Terminal: Denver Day: 4	Terminal: Seattle Day: 5	Rail	49

Day	DE_j^t	Demand Quantity	Inventory on day t-1	# Met by Inventory	Pickup Info	Delivery Info	Mode	No. of empty containers moved
6	DE_0^6	156	28	28	Terminal: FT, Worth Day: 6	Terminal: Houston Day: 6	Truck	86
					Terminal: FT, Worth Day: 6	Terminal: Houston Day: 6	Rail	42
	DE_2^6	211	0	0	Terminal: FT. Worth Day: 6	Terminal: Kansas City Day: 6	Rail	7
					Terminal: Memphis Day: 6	Terminal: Kansas City Day: 6	Rail	132
					Terminal: ST. Paul Day: 6	Terminal: Kansas City Day: 6	Rail	55
					Terminal: Long Beach Day: 5	Terminal: Kansas City Day: 6	Rail	17
	DE_4^6	85	0	0	Terminal: FT. Worth Day: 5	Terminal: Chicago Day: 6	Rail	68
					Terminal: ST. Paul Day: 6	Terminal: Chicago Day: 6	Rail	17
	DE_6^6	131	0	0	Rail Long Beach Day: 6	Terminal: Oakland Day: 6	Rail	121
					Terminal: Denver Day: 4	Terminal: Oakland Day: 6	Rail	10
7	DE_0^7	160	156	156	Terminal: FT. Worth Day: 7	Terminal: Houston Day: 7	Rail	4
	DE_1^7	82	4	0	Terminal: FT. Worth Day: 7	Terminal: Houston Day: 7	Rail	4
					Terminal: Kansas City Day: 7	Terminal: FT. Worth Day: 7	Rail	82
	DE_3^7	49	0	0	Terminal: Chicago Day: 7	Terminal: Memphis Day: 7	Rail	49
	DE_9^7	153	0	153	Terminal: Kansas City Day: 7	Terminal: Denver Day: 7	Rail	129
					Terminal: Chicago Day: 7	Terminal: Denver Day: 7	Rail	24

Table 5-6 Terminals ranking in terms of total pickups & deliveries.

Terminal	Road Pickups	Road Deliveries	Rail Pickups	Rail Deliveries	Pickups & Deliveries (Total)	Demand	Ranking (Demand)	Ranking (Pickups & Deliveries)
Houston, TX	203	273	235	267	978	877	2	6
Ft. Worth, TX	409	220	121	306	1056	597	7	4
Kansas City, KS	319	406	447	211	1383	617	6	1
Memphis, TN	192	495	297	352	1336	1047	1	2
Chicago, IL	531	177	220	85	1013	462	10	5
ST. Paul, MN	131	390	211	172	904	733	5	7
Oakland, CA	205	137	272	277	891	484	9	8
Long Beach, CA	214	243	231	121	809	787	3	9
Seattle, WA	0	139	161	264	564	498	8	10
Denver, CO	453	177	202	342	1174	770	4	3

Table 5-7 illustrates the inventory level of empty containers and chassis for each terminal throughout the reporting period. Houston terminal exhibits the highest average number of empty containers, with 107 units, whereas Kansas City terminal has the lowest inventory, with an average of 26 empty containers. Additionally, we observe that Denver terminal maintains the highest average number of chassis in its inventory, with 139 units, while Oakland terminal has the lowest average inventory, with 50 chassis. Given these observations, it becomes apparent that BNSF Railway and its main customers can benefit significantly by increasing their investment in parking lots and chassis pools strategically located in Houston, TX, and Denver, CO. This would enable them to effectively store and manage the inventory of both empty containers and chassis, ensuring smooth operations and efficient supply chain management.

Table 5-7 Inventory of empty containers and chassis

Day	Houston		Ft. Worth		Kansas City		Memphis		Chicago		St. Paul		Oakland		Long Beach		Seattle		Denver	
	IE	IC	IE	IC	IE	IC	IE	IC	IE	IC	IE	IC	IE	IC	IE	IC	IE	IC	IE	IC
0	153	42	0	323	0	0	0	0	147	134	0	0	70	205	76	50	85	0	123	304
1	153	42	0	0	0	105	0	218	147	134	0	0	70	50	76	0	85	112	123	304
2	153	0	0	134	0	86	0	192	142	148	0	93	0	0	146	201	40	112	128	0
3	28	74	0	0	0	254	125	27	0	183	142	165	0	0	146	164	10	112	173	0
4	28	122	139	85	0	183	100	0	0	96	172	132	0	0	138	18	10	140	0	146
5	28	135	139	86	0	0	132	128	0	0	72	31	0	59	121	0	67	67	0	189
6	156	86	4	0	211	0	0	149	85	0	0	31	131	87	0	0	67	0	0	139
7	160	0	82	0	0	0	49	0	12	0	0	0	131	0	0	0	67	0	153	31
Average	107	63	46	79	26	79	51	89	67	87	48	57	50	50	88	54	54	68	88	139

5.3.3.2 CECMTT results

Table 5-8 displays the results of solving the CECMTT for one train, gradually increasing the number of trains until we achieve an objective value comparable to that of the CECM model,

which is 35,114,239. When we use 12 trains, we achieve a smaller objective value (34,983,252). This implies that having 12 trains in the system is nearly equivalent to assuming that two trains are available at the start of each day for container transportation. The difference can be attributed to the assumptions made in the CECM model. In the CECM model, it is assumed that each train has the capacity to haul 300 containers per day, and two trains are available at the beginning of every day for transporting containers. However, in real-world scenarios, when a train is actively transporting containers, it cannot be immediately available for further transportation until it arrives at a designated terminal and completes its current task. This downtime in transit leads to the need for more trains to efficiently handle container transportation.

It is apparent that there is an inverse relationship between the number of trains and the number of containers transported by trucks. As the number of trains increases, the quantity of containers conveyed by trucks diminishes. This trend can be attributed to the higher cost associated with truck transportation. When a train option is available, it is generally preferred due to its cost-effectiveness. As a result, the increased availability of trains leads to a decrease in the number of containers transported by trucks and a corresponding rise in train utilization for container transportation.

With an increasing number of trains, the total number of chassis in the system remains relatively stable, showing minor fluctuations. The figures consistently fall within the range of 1050 to 1137, without a discernible upward or downward trend. While there are slight variations, no consistent pattern emerges. One possible explanation is that as the number of trains rises, the quantity of containers transported by trucks declines, leading to a reduced demand for chassis on the road. However, the volume of loaded containers transported by trains increases, and these containers utilize chassis upon their arrival at the terminal. Consequently, the decrease in chassis

usage for road transportation is balanced by the increase in chassis usage on the rail side, resulting in a roughly even distribution (Figure 5-4). On the other hand, as the number of trains increases, the trend on the total number of empty containers in the system appears to be a decreasing trend. The values consistently decrease from 1005 to 657. The trend is quite clear and indicates a downward movement in the data points (see Figure 5-5). Transporting containers by truck is significantly more expensive than using trains. Therefore, in situations where train availability is limited, maintaining a larger inventory of empty containers becomes advantageous as a strategic measure to mitigate higher trucking costs. By ensuring an ample supply of empty containers, business can adapt to the circumstances and ensure streamlined logistics operations despite the scarcity of trains and the associated higher expenses. Furthermore, it becomes more cost-effective to relocate empty containers rather than purchasing additional ones. This means that instead of buying new containers, it is more economical to redistribute the existing empty containers to where they are needed. This approach offers a more efficient use of resources and helps to save costs in comparison.

Table 5-8 Performance metrics for varying train numbers

# of trains	Objective value	Run time	# of chassis	# of empty containers	Truck			Train		
					# of empty containers handled by trucks	# of loaded containers handled by trucks	# of empty and loaded	# of empty containers handled by trains	# of loaded containers handled by trains	# of empty and loaded
1	42,174,726	0.56	1137	1005	403	3246	3649	335	252	587
2	40,751,096	1.73	1109	862	433	3143	3576	437	355	792
3	39,496,590	3.09	1063	965	349	2950	3299	491	548	1039
4	38,651,107	20.37	1051	1040	129	2865	2994	645	633	1278
5	37,581,134	28.18	1065	973	190	2785	2975	742	740	1482
6	36,879,765	39.75	1065	825	228	2758	2986	942	740	1682
7	36,301,259	65.49	1069	767	226	2601	2827	1053	897	1950
8	35,915,230	131.19	1065	727	156	2585	2741	1198	909	2107
9	35,671,874	234.31	1065	680	66	2588	2654	1396	910	2306
10	35,428,338	143.59	1055	680	66	2531	2597	1389	967	2356
11	35,227,159	685.94	1058	722	79	2440	2519	1302	1058	2360
12	34,983,253	843.52	1063	657	119	2435	2554	1328	1063	2391

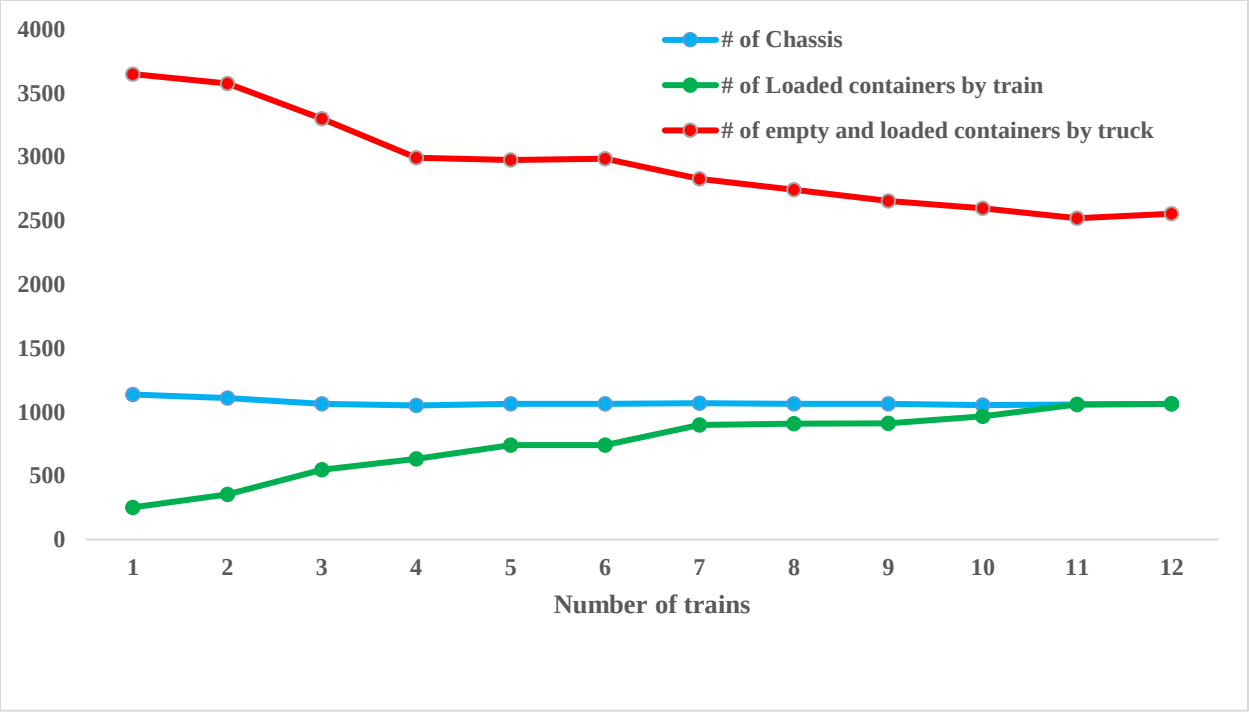

Figure 5-4 Number of chassis with increasing trains showing stable trend

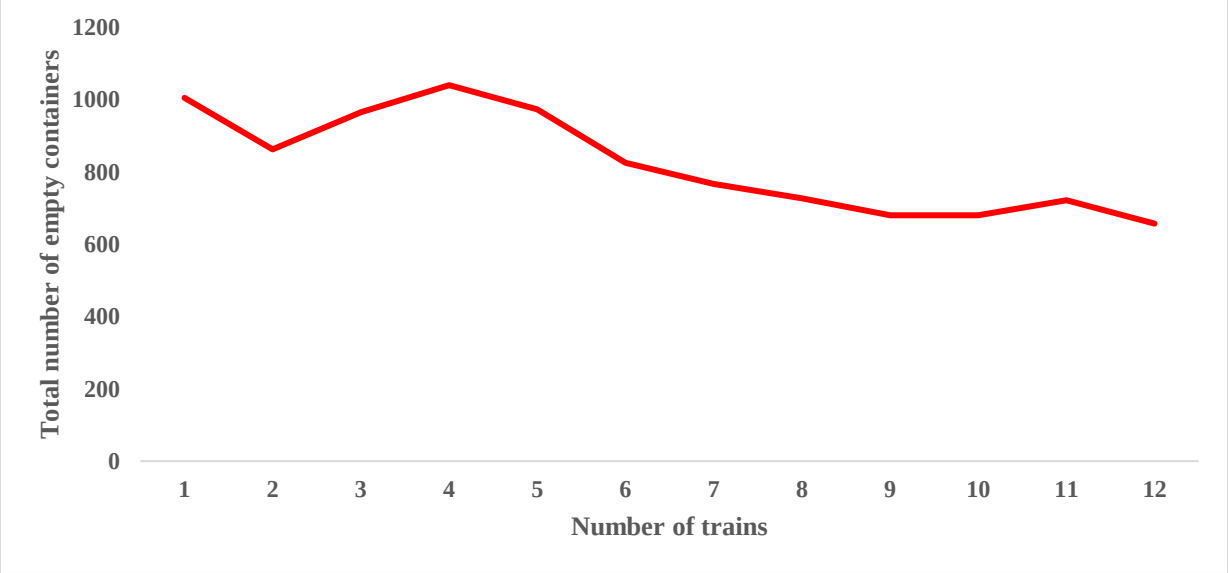

Figure 5-5 Decreasing trend of empty containers as train numbers increase

5.4 Disruption in terminals

Tornadoes are most active in the United States during April, May, and June, according to long-term weather records (Donegan, 2022). They can deliver disruption to US supply chain. Tornadoes are a major cause of massive destruction along key supply chain hubs every year. These storms have a significant impact on major transportation companies in the US, resulting in delays for trains, trucks, planes, and other modes of transportation. It would be interesting to consider how such an uncertainty will affect our transportation network. This section examines the possibility of disruption caused by a tornado in one of the busiest terminals in the country, Denver. We assume that Denver is susceptible to tornadoes any day of the week. We consider three scenarios to account for the potential impact of a tornado on our intermodal network. Please note that both the assigned probabilities and the increased hours in each scenario are selected arbitrarily and are used solely for the purpose of analysis and scenario planning. The scenarios are as follows:

- First scenario: In this scenario, the tornado is assumed to have no effect on our intermodal network, and as a result, travel times remain unaffected. A probability of 0.5 is associated with this scenario.
- Second scenario: Under this scenario, the tornado only affects trains traveling to and from the Denver terminal. As a consequence of delays at the terminal, all travel times from and to Denver increase by 20 hours. This scenario is assigned a probability of 0.20.
- Third scenario: The third scenario is based on the assumption that only trucks traveling to and from the Denver terminal will be impacted by the tornado. Delays at the Denver terminal result in an extension of truck travel times from and to Denver by 10 hours. The probability assigned to this scenario is 0.30.

We solved an extensive formulation of the problem using GUROBI. While the model is similar to the CECM, we have added another dimension to all decision variables to capture uncertainty. Additionally, the objective function of the new model is the expected value of the objectives of all scenarios (5.23).

$$
\begin{aligned}
Min\ Z = \sum_{s=1}^{|S|} Prob^s \Bigg[& C^e \times NE^s + C^c \times NC^s + \sum_{t=1}^{T}\sum_{r=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}[X_{ij}^{rt0s} \times h_{ij}^0 \times (lb^c + lb^e) \times p^0 \\
& + \sum_{t=1}^{T}\sum_{r=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[X_{ij}^{rt1s} \times h_{ij}^1 \times (lb^e) \times p^1\right] \\
& + \sum_{t=1}^{T}\sum_{r=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[Y_{ij}^{rt0s} \times h_{ij}^0 \times (lb^c + lb^e + lb^l) \times p^0\right] \\
& + \sum_{t=1}^{T}\sum_{r=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[Y_{ij}^{rt1s} \times h_{ij}^1 \times (lb^l) \times p^1\right] \\
& + \sum_{t=1}^{T}\sum_{r=1}^{t}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[H_{ij}^{rts} \times h_{ij}^0 \times lb^c \times p^0\right] \Bigg]
\end{aligned}
\tag{5.23}
$$

5.4.1 Observations and recommendations: disruption case

It is expected that the objective function value will be $35,228,527, which is only $114,288 higher than the value in a normal situation. This indicates that the solution provided by the model is a great response to the disruption, considering the marginal increase in cost. In all scenarios, the optimal number of chassis remains constant at 1058. However, the total number of empty containers is highest when tornadoes impact train travel times (NE^2=807). There are several reasons contributing to this observation. Under this scenario, trains experience longer travel times, which may result in them being unable to fulfill certain demand requirements. Despite the fact that tornadoes have not affected truck travel times, it is more expensive to transport goods via trucks.

Therefore, it becomes advantageous to maintain a larger inventory of empty containers as a strategic approach to mitigate the impact of increased train travel times and higher transportation costs associated with trucks.

Our observations indicate that trains compensate for the increased truck travel times in scenario 3 by accommodating a greater number of shipments than they would in normal circumstances. This increase in road travel times would also lead to a notable decline in chassis-only shipments.

Another noteworthy observation is that when train travel times are impacted (scenario 2), the solution differs significantly from the others. However, the solution remains robust to the sudden change in truck travel times (scenario 3). As a recommendation, we suggest that BNSF Railway consider investing in additional railroads connecting to and from its intermodal hub terminals. This strategic measure would help mitigate the adverse effects of disruptions and enhance the overall resilience of the network. Table 5-9 provides more details on these statistics. While recommending additional rail connections to enhance the network's resilience and mitigate disruption costs, it is essential to conduct a comprehensive cost-benefit analysis that evaluates the financial feasibility and potential long-term benefits of this strategic measure, considering factors such as disruption frequency, investment costs, operational savings, and the overall enhancement of the network's reliability.

Table 5-9 Solutions under different scenarios

Scenario	Total number of empty containers in the system	Total number of chassis in the system	Empty and loaded containers transported to and from Denver		Number of Chassis transported to and from Denver
			Trucks	Trains	
Scenario 1	654	1058	630	544	224
Scenario 2	807	1058	599	219	255
Scenario 3	654	1058	609	565	121

5.5 Conclusion

The effective utilization of intermodal transportation resources is of paramount importance when it comes to shipping goods on a global scale. Among these essential resources, the chassis stands out as a crucial component that facilitates the smooth and secure loading and unloading of containers. Throughout this chapter, we delved into the intricacies of the chassis inventory management problem within an intermodal transportation system.

Our primary objective was to minimize the costs involved in transporting loaded and empty containers, while efficiently managing inventory of chassis and empty containers at terminals. To achieve this, we meticulously developed a comprehensive mathematical model that focuses on the daily movement of chassis and containers between terminals over multiple time periods. By incorporating various factors and variables into our model, we aimed to optimize the efficiency and effectiveness of the chassis management process.

We also explored a potential scenario that could disrupt the operations of one of the busiest terminals in the country—the occurrence of a tornado. By considering this extreme event, we aimed to emphasize the need for robust contingency plans. Adequate preparation and proactive measures can help minimize the impact of such disruptions, ensuring the continuity of operations.

Based on our examination, we discovered a number of important observations and discoveries. Primarily, we emphasized the importance of effectively distributing and moving chassis and empty containers. By making appropriate decisions in this area, the utilization of resources can be greatly improved, resulting in reduced expenses and optimal transportation. Expanding on these findings, we propose the designation of particular terminals as intermodal hubs, which would serve as central points for operations and continue functioning even during

disruptions. Implementing the insights and recommendations from this study will undoubtedly contribute to cost reduction, enhanced operational efficiency, and improved overall performance within the global supply chain.

Chapter 6 Depot-relay-point location and truck routing problem

This chapter is derived from our manuscript titled "Depot-relay-point location and truck routing problem," which we have diligently prepared and intend to submit to a reputable scholarly journal specializing in the field of transportation.

6.1 Introduction

Employee turnover is a critical concern for organizations across various industries. In the context of trucking, high turnover rates have resulted in a chronic shortage of drivers, leading to significant operational challenges. For instance, in 2019, the trucking industry experienced a staggering turnover rate of 91 percent, implying that out of every 100 individuals who joined as drivers, 91 ultimately left their positions[1]. This situation can be attributed to a combination of factors, including the demanding nature of working conditions, prolonged periods away from home, transit delays, and lifestyle-related health pressures.

Truck drivers often endure extended periods away from home, spending days and sometimes weeks on the road. Their demanding job involves prolonged periods of solitude, combined with transit delays and lifestyle factors such as sleep deprivation and unhealthy eating habits. These factors significantly contribute to the high turnover rate within the industry. As an example, let's consider a scenario where there is a container at the Port of Savannah, Georgia destined for a main distribution center in Chicago, Illinois, and another container in Chicago, Illinois that needs to be delivered to the Port of Savannah, Georgia for export. In this scenario, a truck is designated to collect the container from the Port of Savannah, while another truck is

[1] Opinion | How Life as a Trucker Devolved Into a Dystopian Nightmare - The New York Times (nytimes.com)

assigned to retrieve the container in Chicago. After loading their respective containers, the drivers commence their journeys, deliver the containers, and return home, traversing a combined distance of approximately 3700 miles. However, due to the long distances and necessary stops for rest and compliance with regulations, drivers can find themselves away from home for multiple days (Figure 6-1).

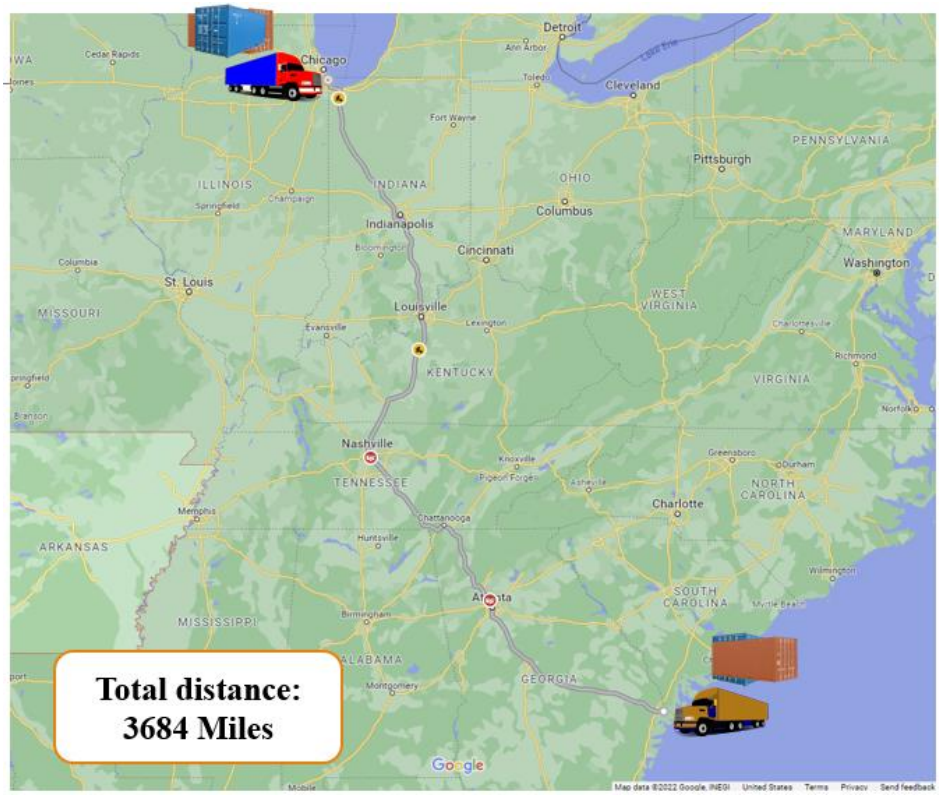

Figure 6-1 Truck drivers: long distances, high turnover

To effectively address this issue, we can implement an improved strategy. This approach involves establishing a centralized meeting point, such as Nashville, Tennessee, where drivers can gather and exchange their loads. For example, the driver who initially collected the container from Savannah, Georgia, would exchange it with the driver responsible for the container from Chicago,

Illinois. Consequently, the first driver would return to Savannah to deliver the second load, while the second driver would head back to Chicago to deliver the first load. By implementing this load exchange solution, the total distance traveled can be significantly reduced to approximately 1,940 miles, resulting in an impressive decrease of around 50% compared to the previous scenario (Figure 6-2). The implementation of a relay network, where truckloads switch drivers or loads during transportation, has the potential to reduce the amount of time drivers spend away from home. This allows drivers to rest at home instead of having to rely solely on rest stops along the way. These locations where load exchanges occur are referred to as relay-point locations.

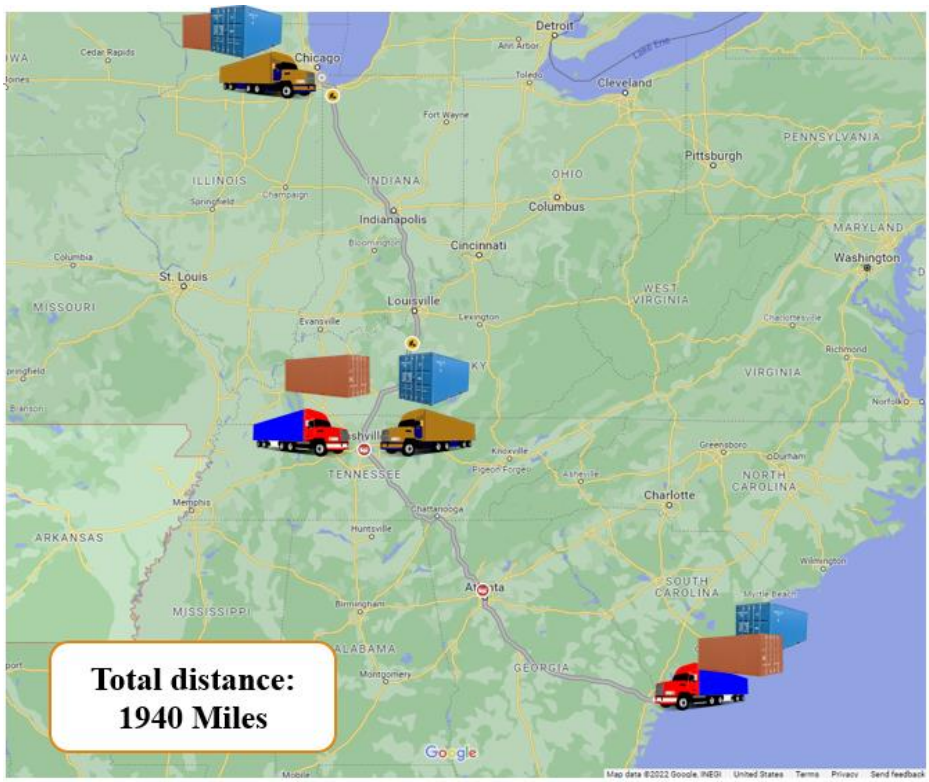

Figure 6-2 Relay network: shorter distance, more driver rest

Drawing parallels to the airline industry, pilots also face similar challenges in terms of time away from home. Cargo pilots, for instance, can spend several weeks away from home continuously. In 2020, the turnover rate for U.S. pilots was approximately 46%, highlighting the importance of addressing staffing shortages (American Airlines, 2020). Furthermore, the shortage of staff can result in the cancellation of routes and significant delays, as evidenced by American Airlines' decision to discontinue its routes to Dubuque, Iowa, in July 2022 due to a lack of pilots. The airline is discontinuing service in Islip, New York, located on eastern Long Island, as well as in Ithaca, New York, the upstate home of Cornell University, and Toledo, Ohio. The decision to cease operations in these locations is driven by the same underlying reason. However, it is important to note that Dubuque and the affected airports are not the only ones experiencing this problem. According to experts, an increasing number of cities are being impacted by this issue, and it is likely to become a growing concern in the years ahead. Other major airlines like United and Delta, which rely on a hub-and-spoke network using regional jets, are also reducing their services to some of the smaller airports they serve[1]. Currently, airlines are compelled to expedite the student training process to compensate for these shortages while also increasing pilots' compensation to retain them.

In addition to relay points, it is crucial to identify depot locations that are close to pickup and delivery points. Placing depots near pickup and delivery points plays a vital role in reducing the time drivers spend away from home. When depots are strategically located, drivers can minimize the distance they need to travel from depots to pick up points or from delivery points to depots. This reduces overall time spent on the road and allows drivers to return home more

[1] Airlines forced to drop service at these US airports due to the pilot shortage | CNN Business

frequently, contributing to higher job satisfaction, improved driver well-being, and ultimately helping to reduce turnover rates within the industry.

This chapter aims to address key questions related to efficient resource management in the context of depot and relay location optimization. Specifically, it seeks to determine the optimal depot locations, identify suitable relay locations, and define the sequence of nodes that a truck should visit after leaving a depot. By addressing these questions, this research aims to improve operational efficiency and minimize the time away from home for truck drivers, thereby mitigating turnover rates. Furthermore, the lessons learned from this problem can be adapted and applied to flight resource management within the airline industry, aiding in the reduction of staffing shortages and ensuring smoother operations.

The remaining portion of this chapter is structured as follows.: Section 6.2 provides an overview of the mathematical model. Section 6.3 focuses on the computational analysis conducted. Finally, Section 6.4 offers concluding remarks on the subject matter.

6.2 Mathematical formulation

In this section, we begin by presenting the mixed integer programming model developed by Rais et al. (Rais, Alvelos, & Carvalho, 2014) for addressing the pickup-and-delivery problem with transshipment. Transshipment points serve as locations where vehicles can make stops to facilitate transfers and adjustments of their cargo loads. These points enable drivers to switch vehicles, receive release times to comply with policy-related matters, allow fresh or well-rested drivers to replace fatigued ones, and perform similar functions as relay points. Then, we introduce our relay point model with pickup and delivery, which forms the foundation of our approach aimed at minimizing overall time spent on the road and enabling drivers to return home more frequently.

We outline the essential decision variables, constraints, and objectives of the model, offering a comprehensive understanding of the proposed framework. It is important to note that the new model discussed in this section is based on specific assumptions:

- There are predetermined pickup and delivery nodes with specified demand.
- A fixed number of depots exist, serving as the base for trucks and drivers.
- A specific number of relay points are available.
- The locations of both depots and relay points are unknown.
- Each truck/driver is capable of accommodating a maximum of 2 requests, with only one request being served at a time.
- The trucks are considered homogeneous, meaning that all vehicles possess the same capacity.

6.2.1 The mixed integer programming model for the pickup-and-delivery problem with transshipment (PDPT)

Consider an undirected graph denoted as $G = (V, E)$, where *V* represents the set of nodes and *E* represents the set of arcs. The node set *V* is defined as $V = \{1,2,\dots,n\}$, where *n* represents the total number of nodes in the graph. The arcs in *E* are represented as *ij*, indicating an arc between node *i* and node *j*. Let *K* be a set of vehicles, denoted as $k = 1,\dots,|K|$, where $|K|$ represents the total number of vehicles. The initial depot of vehicle *k*, denoted as $o(k)$, and the final depot, denoted as $o'(k)$, both belong to the node set *V*. The set *Q* represents the customer pickup and delivery requests, where $q = 1,\dots,|Q|$. Each request q is associated with a pickup node, denoted as $p(q)$, and a delivery node, denoted as $d(q)$, both of which belong to the node set *V*. The set *T* represents the transshipment nodes, which are also part of the node set *V*. The unit cost of

transportation from node $i \in V$ to node $j \in V$ using vehicle $k \in K$ is represented as c_{ij}^k. The decision variables in the model are binary and as follows:

x_{ij}^k: 1, if truck *k* takes the arc *ij*, otherwise, it is set to 0.

y_{ij}^{kq}: 1, if truck *k* carries the request *q* on the arc *ij*, otherwise, it is set to 0.

z_{ij}^k: 1, if node *i* precedes (not necessarily immediately) node *j* in the route of the vehicle *k*, otherwise, it is set to 0.

The model is presented as an integer programming formulation as follows:

$$Min \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}^k c_{ij}^k \tag{6.1}$$

S.t.

$$\sum_{j=1}^{n} x_{ij}^k \leq 1 \qquad \forall k \in K, \forall i = o(k) \tag{6.2}$$

$$\sum_{j=1}^{n} x_{ij}^k = \sum_{j=1}^{n} x_{jl}^k, \qquad \forall k \in K, \forall i = o(k), \forall l = o'(k) \tag{6.3}$$

$$\sum_{j=1}^{n} x_{ij}^k - \sum_{j=1}^{n} x_{ji}^k = 0, \qquad \forall k \in K, \forall i \in V \setminus \{o(k), o'(k)\} \tag{6.4}$$

$$\sum_{k=1}^{K} \sum_{j=1}^{n} y_{ij}^{kq} = 1, \qquad \forall q \in Q, \forall i \in p(q) \tag{6.5}$$

$$\sum_{k=1}^{K} \sum_{j=1}^{n} y_{ji}^{kq} = 1, \qquad \forall q \in Q, \forall i \in d(q) \tag{6.6}$$

$$\sum_{k=1}^{K} \sum_{j=1}^{n} y_{ij}^{kq} - \sum_{k=1}^{K} \sum_{j=1}^{n} y_{ji}^{kq} = 0, \quad \forall q \in Q, \forall i \in T \tag{6.7}$$

$$\sum_{j=1}^{n} y_{ij}^{kq} - \sum_{j=1}^{n} y_{ji}^{kq} = 0, \qquad \forall k \in K, \qquad \forall q \in Q, \forall i \in V \setminus T \tag{6.8}$$

$$y_{ij}^{kq} \leq x_{ji}^{k}, \qquad \forall ij \in E, \forall k \in K, \forall q \in Q \tag{6.9}$$

$$\sum_{q=1}^{Q} y_{ij}^{kq} \leq 1 \qquad \forall ij \in E, \forall k \in K \tag{6.10}$$

$$x_{ij}^{k} \leq z_{ij}^{k}, \qquad \forall i,j \in V, \forall k \in K, o(k) \neq i, o'(k) \neq j \tag{6.11}$$

$$z_{ij}^{k} + z_{ji}^{k} \leq 1, \qquad \forall i,j \in V, \forall k \in K, o(k) \neq i, o'(k) \neq j \tag{6.12}$$

$$z_{ij}^{k} + z_{jl}^{k} + z_{li}^{k} \leq 2, \qquad \forall i,j,l \in V, \forall k \in K, o(k) \neq i,j, o'^{(k)} \neq l \tag{6.13}$$

$$x_{ij}^{k} \in \{0,1\} \qquad \forall ij \in E, \forall k \in K \tag{6.14}$$

$$y_{ij}^{kq} \in \{0,1\} \qquad \forall ij \in E, \forall k \in K, \forall q \in \tag{6.15}$$

$$z_{ij}^{k} \in \{0,1\} \qquad \forall ij \in E, \forall k \in K \tag{6.16}$$

The objective function (6.1) identifies a set of vehicle routes that incur the least cost while fulfilling all customer requests. Constraints (6.2) ensure that each vehicle is allowed to start only one route from its original depot. Constraints (6.3) ensure that the route of a vehicle must conclude at its designated final depot. Constraints (6.4) ensure that the flow of vehicles through the nodes in the network is conserved. Constraints (6.5) and (6.6), respectively, are responsible for ensuring that all customer pickups and deliveries are carried out as required. Constraints (6.7) ensure that the flow of requests is conserved at the transshipment nodes, allowing requests to transfer from one vehicle to another. On the other hand, constraints (6.8) maintain the flow conservation of requests at the non-transshipment nodes, requiring that a vehicle that brings a request must also depart with the same request. Constraints (6.9) ensure that there is a flow of vehicles on an arc if there is a corresponding flow of requests in the same vehicle on that arc. Constraints (6.10)

guarantee that a vehicle can carry, at most, one request on any given arc. Constraints (6.11) -(6.13) are introduced to address the issue of potential subtours within the solution. Constraints (6.14) – (6.16) require the variables to be binary.

6.2.2 Depot-relay-point location and truck routing model (DRPLTR)

To develop the DRPLTR model, we define some notations from graph theory. Similar to the PDPT, we have an undirected graph $G = (V, E)$ consisting of a set of nodes *V* and a set of arcs *E*. The node set *V* is defined as $V = \{1,2,\dots,n\}$, where *n* represents the total number of nodes in the graph. The arcs in *E* are denoted by ij, indicating an arc between node *i* and node *j*. We also have a set of vehicles, *K*, represented by $k = 1,\dots,|K|$, where $|K|$ denotes the total number of vehicles. The set Q represents the customer pickup-and-delivery requests, with $q = 1,\dots,|Q|$. Each request q is associated with a pickup node $p(q)$ and a delivery node $d(q)$, both belonging to the node set *V*.

Additionally, we have a set of link types, *T*, represented by $t = 0,\dots,|T|$. The link type is a new concept that we introduce to aid us in formulating the problem. Using different link types, the model is able to treat each node as a potential candidate for either a depot or a relay. The distance between nodes i and *j*, denoted by d_{ij}, Finally, we have two parameters, N_d and N_r, representing the maximum number of depots and relays, respectively.

All decision variables in the DRPLTR model are binary and as follows:

x_{ij}^{kt}: 1, if truck *k* takes arc type *t* from node *i* to node *j*, otherwise, it is set to 0.

D_i: 1, if we open a depot at node i, otherwise it is set to 0.

R_j: 1, if we open a relay at node *j*, otherwise it is set to 0.

In this problem, we encounter five different types of links. The link types are defined as follows:

- Link type 0: When the start of an arc is a depot and the end point is a pickup, the link type is 0. In this case, the corresponding variable for the truck is set to 1.
- Link type 1: When the truck moves from a pickup to a delivery, the link type is 1. The corresponding variable for the truck is set to 1.
- Link type 2: For a pickup-relay pair, the link type is 2. The corresponding variable for the truck is set to 1.
- Link type 3: When a truck moves from a relay to a delivery, the link type is 3. The corresponding variable for the truck is set to 1.
- Link type 4: If a truck moves from a delivery to a depot, the link type is 4. The corresponding variable for the truck is set to 1.

These link types and their associated variables provide a framework for addressing the different movements and actions of the trucks in the problem. Instead of solely exchanging loads, drivers also have the option to exchange their trucks.

Drawing a connection between this problem and the airline industry, we can view it through the lens of pilots, planes, and airports. Pilots can be assigned to various flights at hubs or relays, allowing them to spend more nights at home. Additionally, it is important to note that the link types mentioned earlier are subject to potential changes based on the regulations set by the Federal Aviation Administration (FAA) and individual airlines.

The DRPLTR model is presented as an integer programming formulation as follows:

$$Min \sum_{t=1}^{T} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}^{kt} d_{ij} \tag{6.17}$$

S.t.

$$\sum_{k=1}^{K}\sum_{i=1}^{n} x_{ip(q)}^{k0} = 1 \qquad \forall q \in Q \qquad (6.18)$$

$$\sum_{k=1}^{K} x_{ip(q)}^{k0} \leq D_i \qquad \forall i \in V, \forall q \in Q \qquad (6.19)$$

$$\sum_{q=1}^{Q}\sum_{i=1}^{n} x_{ip(q)}^{k0} = 1 \qquad \forall k \in K \qquad (6.20)$$

$$\sum_{k=1}^{K} x_{p(q)j}^{k2} \leq R_j \qquad \forall j \in V, \forall q \in Q \qquad (6.21)$$

$$\sum_{i=1}^{n} D_i \leq N_d \qquad (6.22)$$

$$\sum_{j=1}^{n} R_j \leq N_r \qquad (6.23)$$

$$\sum_{q=1}^{Q} x_{ip(q)}^{k0} = \sum_{q=1}^{Q} x_{d(q)i}^{k4} \qquad \forall i \in V\ , \forall k \in K \qquad (6.24)$$

$$x_{p(q)d(q)}^{k1} + \sum_{j=1}^{n} x_{p(q)j}^{k2} - \sum_{i=1}^{n} x_{ip(q)}^{k0} = 0 \qquad \forall q \in Q \qquad (6.25)$$

$$\sum_{q=1}^{Q} x_{jd(q)}^{k3} - \sum_{q=1}^{Q} x_{p(q)j}^{k2} = 0 \qquad \forall j \in V, \forall k \in K \qquad (6.26)$$

$$\sum_{j=1}^{n} x_{d(q)j}^{k4} - x_{p(q)d(q)}^{k1} - \sum_{j=1}^{n} x_{jd(q)}^{k3} = 0 \qquad \forall q \in Q, \forall k \in K \qquad (6.27)$$

$$x_{p(q)j}^{k2} + x_{jd(q)}^{k3} \leq 1 \qquad \forall j \in V, \forall q \in Q, \forall k \in K \qquad (6.28)$$

$$x_{ij}^{kt} \in \{0,1\} \qquad \forall i, j \in V, \forall k \in K, \forall t \in T \qquad (6.29)$$

$$D_i \in \{0,1) \qquad \forall i \in V \qquad (6.30)$$

$$R_j \in \{0,1) \qquad \forall j \in V \qquad (6.31)$$

In this formulation, the objective function (6.17) minimizes the total distance traveled by all trucks. Constraints (6.18) – (6.20) are depot-pickup related constraints. Constraints (6.18) ensure that each pickup node is covered by only one truck from one depot. Constraints (6.19) ensure that a truck can go from node *i* to a pickup node, if a depot is opened at node *i*. Alternatively, we can state that if a truck travels from node *i* to pick up node $p(q)$ and the link type is 0, indicating a depot-pickup pair, node *i* should function as a depot point. Constraint (6.20) enforces that each truck is located in only one depot and serves only one pickup node. When examining the *x* variables in these constraints, it can be observed that the type is set to 0, indicating that the link type is a depot-to-pickup arc. Constraints (6.21) ensure that a truck goes from a pickup node to node *j* on an arc with type 2, if a relay is opened at node *j*. Similarly, if a truck travels from pick up node $p(q)$ to node *j* and the link type is 2, representing a pickup-relay pair, node *j* should serve as a relay point.

Constraint (6.22) and (6.23) impose upper bounds on the number of depots and relays in the network, respectively. Constraints (6.24) state that after each delivery, each truck must return to the depot it originated from. Constraints (6.25), (6.26), and (6.27) represent the conservation flow constraints at pickup, relay, and delivery nodes, respectively. These constraints ensure that the total net flow out of these nodes is zero, as they are neither sources nor sinks. Constraints (6.28) ensure that a truck visits a relay point solely for the purpose of exchanging its load. Constraints (6.29)- (6-31) require the variables to be binary.

Rais et al. introduced three sets of binary decision variables, two with three indices and another with four indices (Rais, Alvelos, & Carvalho, 2014). In our model, since the binary decision variables D_i and R_j correspond to depot-pickup and pickup-relay links, respectively, we can replace them with x_{ii}^{k0} and x_{jj}^{k2}, respectively. Index *k* can take a fixed value. By adopting this approach, we can effectively decrease the number of variables and streamline the model's complexity.

To address the potential occurrence of subtours in the solution, Rais et al. incorporated a set of subtour elimination constraints into the model. Moreover, the researchers assumed that the depot locations are pre-determined, and the selection of relay points is restricted to a subset of nodes (Rais, Alvelos, & Carvalho, 2014). In our research, we introduce a novel concept called "link type," which enables us to treat the depot and relay locations as decision variables. Moreover, by employing link types, the concern about the presence of subtours is effectively eliminated. Consequently, there is no need to incorporate subtour elimination constraints into the model.

6.3 Experimental results

In this study, the Australia Post (AP) data set was utilized, which captures the mail flows within an Australian city. Figure 6-3 displays the first 50 nodes of the data set. Each node in the data set is represented by *X*-*Y* coordinates, indicating the respective locations within the network.

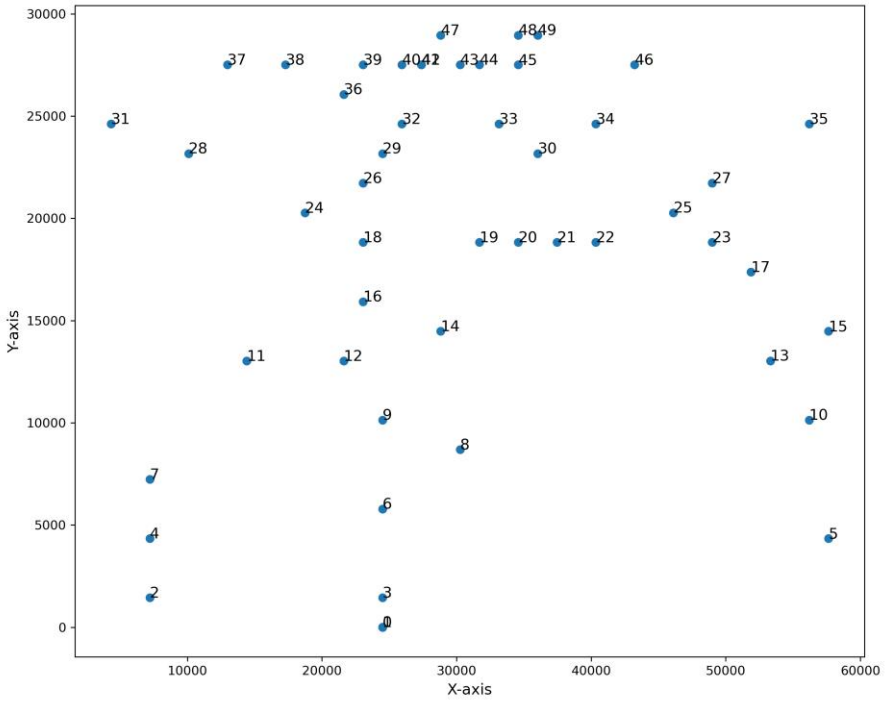

Figure 6-3 AP dataset nodes

The arcs in the network are characterized by Euclidean distances, representing the lengths between the nodes. The total number of nodes, including potential depots, potential relays, pickups, and deliveries, was fixed at 50, while the number of requests varied across experiments, specifically set to 2, 5, 8, and 10. Each request comprises a randomly selected pair of pickup and delivery nodes. Two cases were investigated: one without considering relays and the other considering relays. However, it is important to note that in both problems, the locations of depots were treated as decision variables. The proposed model was implemented for both cases using Python 3.9 programming language, with GUROBI 9.1.2 serving as the solver. Notably, across all scenarios, the inclusion of relays led to a substantial reduction in the total distance covered by the trucks. For instance, in the absence of relays, drivers were required to travel approximately 8250 kilometers to handle 10 requests, whereas the utilization of relays resulted in a decreased distance

of 6368 kilometers. We also notice that the solution time increases exponentially as the number of requests increases. The details are displayed in Table 6-1.

Table 6-1 Comparison of total distance traveled in network with and without relays

# of Nodes	# of Requests	# of Depots	# of Relays	DRPLTR without Relays			DRPLTR with Relays				Improvement (%)
				Depots	Obj	Time	Depots	Relays	Obj	Time	
50	2	2	1	7,13	1867	<1	7,13	8	1044	<1	44.09
	5	2	2	13,14	4186	<1	11,13	8,14	2980	<1	28.81
	8	2	2	8,18	6047	2	11,13	8,19	4601	300	23.91
	10	2	2	13,19	8250	10	11,17	14,30	6368	745	22.81

We also compared the DRPLTR with the PDPT for 10 requests. In PDPT, the unit cost of transportation from node $i \in V$ to node $j \in V$ using vehicle $k \in K$ was represented as c_{ij}^{k}. However, to facilitate a comparison between both models, we have opted to utilize the distance between nodes *i* and *j*, denoted by d_{ij}, without distinguishing between vehicles. As the PDPT assumes that the depot locations are pre-determined and the selection of relay points is restricted to a subset of nodes, we provide the PDPT with two sets of depots and relays. These sets include the depots and relays selected by the DRPLTR. Interestingly, we observed that both models yield the same objective function. However, in terms of solution time, the DRPLTR outperforms the PDPT. One possible reason for this is the smaller number of decision variables in the DRPLTR. Moreover, in order to tackle the possibility of subtours routes within the solution, Rais et al. integrated a set of constraints for eliminating subtours into their model. However, by employing link types, we effectively eliminate the concern about the presence of subtours. Therefore, there is no need for us to incorporate subtour elimination constraints into our model. The results are shown in Table 6-2.

Table 6-2 Comparison of DRPLTR and PDPT (PDPT uses a fixed set of relays and depots selected by DRPLTR)

# of Nodes	# of Requests	# of Depots	# of Relays	PDPT				DRPLTR			
				Depots	Relays	Obj	Time	Depots	Relays	Obj	Time
50	10	2	2	11,17	14,30	6368	3520	11,17	14,30	6368	745

In addition to feeding the PDPT with the depots and relays selected by the DRPLTR, we also provided alternative depots and relays that were distinct from those chosen by DRPLTR. The main objective of this comparison is to investigate the influence of the assumption in the PDPT, where depot locations are pre-determined, and the selection of relay points is restricted to a subset of nodes, on the performance of the PDPT model. By examining the impact of these assumptions, we can gain insights into how the PDPT operates under various scenarios and assess its overall performance. We explored three distinct scenarios in this regard:

a) Deploying different depots and relays in close proximity to those selected by DRPLTR.
    - Candidate locations for depots: 7, 9, 12, 13, 15, 23
    - Candidate locations for relays: 8, 16, 18, 19

b) Utilizing different depots and relays located far away from those chosen by DRPLTR.
    - Candidate locations for depots: 0, 3, 5, 28, 34, 46
    - Candidate locations for relays: 3, 5, 37, 46

c) Employing different depots and relays selected at random.
    - Candidate locations for depots: 5, 12 17, 24, 38, 41
    - Candidate locations for relays: 7, 19, 33, 42

The primary research question addressed in these scenarios is as follows: How does the performance of the PDPT model compare to the DRPLTR model when the model is provided with a subset of relays and depots candidate locations different from those selected by the DRPLTR? In all these three scenarios, we selected more than two candidate depots and two candidate relays, allowing the model to determine the optimal two depots and two relays.

As illustrated in Table 6-3, in the first scenario, when the PDPT is constrained to depots and relays that are close to those selected by the DRPLTR, the objective value is slightly higher than that of the DRPLTR (6492 versus 6368). However, in the second scenario, when the PDPT is served by depots and relays that are located far from the ones selected by the DRPLTR, the objective value experiences a significant increase in comparison to that of the DRPLTR (7385 versus 6368). Additionally, when depots and relays are randomly selected, the objective value falls between the first two scenarios (6601 versus 6368). When examining the solution times, we also notice a significant difference in the speed of obtaining the optimal solution between PDPT and DRPLTR. The proposed model, DRPLTR, demonstrates a clear advantage in terms of efficiency.

Table 6-3 Comparison of DRPLTR and PDPT (PDPT uses a fixed set of relays and depots that differ from those selected by DRPLTR)

# of Nodes	# of Requests	# of Depots	# of Relays	Scenario	PDPT (Fixed)				DRPLTR			
					Depots	Relays	Obj	Time	Depots	Relays	Obj	Time
50	10	2	2	a	7,23	8,19	6492	3520	11,17	14,30	6368	745
				b	28,34	3,46	7385	3518				
				c	17,24	19,33	6601	3535				

It is worth mentioning that allowing any point to be a depot or relay point in the trucking industry has both advantages and disadvantages from a management perspective. Here are some key points to consider:

Advantages:

Flexibility and Adaptability: Allowing any point to be a depot or relay point provides greater flexibility in responding to changes in demand patterns, supply sources, or operational constraints. It enables the network to adapt quickly to unforeseen changes and optimize routing options.

Cost Optimization: With the freedom to choose any point as a depot or relay point, the management can potentially identify cost-efficient locations that minimize transportation costs and reduce overall supply chain expenses.

Disadvantages:

Complexity: With the freedom of choice for depot or relay points, the logistics network can become more complex to manage. Optimizing such a network becomes computationally intensive and may require advanced algorithms and modeling techniques.

Sensitivity to Demand Fluctuations: The decisions regarding depot or relay points can be highly sensitive to fluctuations in demand. If the demand patterns change significantly, the previously optimized network may no longer be efficient. Frequent fluctuations in demand may require constant adjustments to the depot or relay point locations, which could lead to higher operational costs and complexities.

The management should consider conducting sensitivity analyses to understand how different demand scenarios can impact the proposed network design. Additionally, they should

monitor market trends and technological advancements to ensure the chosen model remains relevant and effective over time.

Ultimately, the decision to allow any point to be a depot or relay point should align with the company's strategic objectives, customer requirements, and operational capabilities. Regular review and continuous improvement are crucial to maintaining an efficient and adaptive logistics network.

6.4 Conclusion

In this chapter, we discussed a critical issue in the trucking industry: the alarming rate of driver turnover and chronic shortage of truckers. The demanding working conditions, which require drivers to be away from home for extended periods, have contributed to this problem. Factors such as transit delays and lifestyle-related health pressures have further exacerbated the high turnover rate among drivers. To alleviate this challenge, we introduced a novel integer programming model for the pickup and delivery problem with relay points. Unlike existing models, our approach considered the location of depots and relays as decision variables. Our introduction of the concept of "link types" improved the model's flexibility by providing a framework to address the different movements and actions of trucks in the problem.

By adopting our approach, we effectively reduced the number of variables and eliminated subtours in the solution, thereby reducing complexity. We conducted investigations involving two scenarios: one without considering relays and another considering relays. Notably, in all scenarios, the inclusion of relays resulted in a significant reduction in the total distance covered by the trucks. This reduction demonstrated the practical benefits of our model.

Furthermore, we compared our proposed model, the DRPLTR, with the PDPT under various scenarios. Our model consistently outperformed in terms of objective value and solution time, highlighting its effectiveness and superiority in addressing the problem at hand.

Chapter 7 Conclusion and future research

The efficiency of transportation systems is of paramount importance as it significantly contributes to improved access to markets, employment opportunities, and increased investments. However, the presence of inefficiencies within the transportation sector, such as inadequate resource management, presents substantial challenges that must be addressed to achieve enhanced efficiency throughout the sector and reap economic benefits.

This dissertation endeavored to develop effective optimization techniques and algorithms to tackle diverse transportation challenges, aiming to optimize resource allocation and enhance resource utilization within the transportation sector. This chapter highlights the significant research contributions and key findings of this dissertation.

7.1 Conclusion and future directions on hub covering problems considering service availability

We delved into a critical research gap present in the existing literature on the hub covering location problems. Specifically, we identified that there has been limited attention given to incorporating uncertainties and congestion effects into the study of hub covering networks. This gap is particularly significant when considering the scenario where hub servers, such as runways in an airport hub, become unavailable due to a high volume of take-offs and landings. This can lead to extended waiting times for planes and negatively impact the overall efficiency of the system. Recognizing the importance of addressing this research gap, Chapter 3 of our study was dedicated to the development of hub maximal covering location problems that take into account the unavailability of hub servers during flow service.

To tackle this uncertainty, we formulated two models that aimed to maximize the expected coverage of flow, building upon the classical covering location problems. In the first model, we assumed each hub has a fixed number of servers, while in the second model, the number of servers was treated as a variable to be determined. By formulating these models, our objective was to locate *p* hub facilities in a manner that ensures coverage for origin-destination pairs of non-hub nodes through a pair of hub nodes. Both of our proposed models were solved using exact solutions as well as metaheuristics, specifically genetic algorithms and Tabu Search. This approach allowed us to provide effective solutions to the hub maximal covering location problems.

To validate our models and algorithms, we conducted experiments using real-world data from American Airlines domestic flights in 2019. The results obtained from these experiments shed light on the importance of considering the busy fractions of servers when determining the coverage and efficiency of hub networks. Moreover, our algorithms showcased their capability to handle large-scale problems efficiently and yielded satisfactory outcomes.

In order to enhance the efficiency of the HCLP-(q_h ,x_h), incorporating valid cuts can be a promising avenue of future study. By exploring the potential benefits of valid cuts, we aim to optimize the model's performance and potentially overcome the existing limitations. Valid cuts have shown to be effective in improving the solution quality and reducing computational effort in similar problem domains. By integrating appropriate valid cuts into our approach, we anticipate a substantial improvement in solving the HCLP-(q_h, x_h) problem for larger instances such as $n = 50$, paving the way for more efficient solutions.

7.2 Conclusion and future directions on the proposed MAHMCP

Chapter 4 of this dissertation addressed the challenges in solving the MAHMCP and has proposed a novel approach to tackle these issues. The existing works by Campbell (1994) and Qu and Weng (2009) provided valuable insights into the problem domain but fell short in delivering precise numerical results and exact solutions for larger instances.

By employing integer programming techniques, Chapter 4 introduced a new model for the MAHMCP, with a particular focus on improving the root relaxation value through tighter constraints. Theoretical analysis demonstrated that the proposed model outperforms previous approaches, utilizing a significantly smaller number of binary variables while achieving stronger results. Remarkably, the new model exhibited exceptional performance in solving larger instances, surpassing the capabilities of the model proposed by Qu and Weng (2009).

Notably, the experiments conducted in Chapter 4 showcased remarkable improvements in computational efficiency. Compared to the model by Qu and Weng (2009), our approach achieved a staggering 99.15 percent reduction in branch and bound running time for a problem instance with 70 nodes and 5 hubs. Furthermore, while the previous model failed to produce a solution within 3 hours for instances with $n \geq 90$, the proposed model successfully solved a problem with 100 nodes and 5 hubs in just 174.51 seconds.

These findings highlight the significant contributions made by Chapter 4 in advancing the MAHMCP model. The proposed model, with its enhanced computational efficiency and improved solution quality, opens new avenues for addressing real-world transportation planning challenges. Moreover, the success achieved in solving larger instances demonstrates the scalability and applicability of the new approach.

However, it is important to acknowledge that conducting further investigations is warranted to thoroughly explore the model's performance under different scenarios and validate its effectiveness in diverse real-world settings. As part of our future research, we intend to assess the impact of both stochastic demand and stochastic distances on hub locations. By incorporating these stochastic elements into the analysis, we aim to gain a deeper understanding of how uncertainties in both demand and distances can influence the optimal hub placement decisions. This comprehensive investigation will provide valuable insights into the robustness and adaptability of the proposed model, ultimately enhancing its practical applicability in real-world transportation planning scenarios.

7.3 Conclusion and future directions on chassis inventory management

An in-depth review of the existing literature on intermodal transportation reveals a notable bias towards studying the maritime industry, leaving a significant gap in understanding the unique challenges faced by intermodal transportation at truck-rail terminals. The literature also falls short in considering crucial factors that are integral to the efficient functioning of intermodal transportation, such as the presence of multiple terminals, the demand for loaded containers, the influence of multiple time periods, and the impact of chassis shortages. These gaps in knowledge highlight the pressing need for a comprehensive and innovative approach that addresses the complex issues surrounding chassis inventory management in intermodal transportation.

Motivated by these identified research gaps, Chapter 5 of our study introduced a novel mathematical model designed to optimize chassis inventory management within the realm of intermodal transportation. This innovative approach took into account a multi-time period and multiple terminal perspective, enabling effective management and efficient relocation of chassis

and empty containers based on demand. The primary objective of our research was to minimize the total cost incurred throughout the entire process. By developing a mathematical model, our aim was to tackle the challenges associated with chassis shortages and provide practical solutions that fill the existing research gaps within the field of intermodal transportation.

To enhance the realism of our model, we also made a significant modification to the initial approach. We incorporated a tracking binary variable to monitor the movement of trains on a daily basis, enabling us to determine the specific terminals where the trains originate and conclude their journeys. This modification not only ensured that the capacity for transporting containers by trains is not exceeded but also provided a more accurate representation of the real-world dynamics of intermodal transportation.

Furthermore, our study delved into a potential scenario that could disrupt the operations of one of the busiest terminals in the country—an occurrence of a tornado. By considering this extreme event, we sought to emphasize the critical need for robust contingency plans. We highlighted that adequate preparation and proactive measures are essential to minimize the impact of such disruptions, ensuring the continuity of operations.

Through our analysis, we identified several key insights and findings. First and foremost, we highlighted the significance of proper chassis and empty containers allocation and relocation. Implementing appropriate allocation and relocation decisions can significantly enhance the utilization of resources, thereby reducing unnecessary costs and ensuring optimal transportation. Building upon these findings, we recommended designating specific terminals as intermodal hubs, both in normal operations and during disruptions.

Ultimately, the implementation of the insights and recommendations gleaned from our study will undoubtedly contribute to substantial cost reduction, enhanced operational efficiency,

and improved overall performance within the global supply chain. By addressing the gaps in existing research and providing practical solutions, our findings have the potential to revolutionize the field of intermodal transportation and shape its future trajectory.

For future avenues of research, there are several aspects that can be explored to further advance the study of intermodal transportation and chassis inventory management.

Firstly, this thesis assumed that only one terminal is affected by a disruption. However, in reality, it is possible for multiple terminals to be simultaneously affected. Therefore, it would be interesting to consider a set of terminals when addressing disruptions. Exploring the impact of disruptions on multiple terminals would provide a more comprehensive understanding of the system's resilience and enable the development of effective contingency plans.

Additionally, this study assumed that a container is transported from one terminal to another following a direct path, without considering the possibility of containers visiting multiple stops before reaching their destination terminal and being unloaded from the train. Investigating the dynamics of container movement involving multiple stops would be an intriguing avenue to explore. By considering the complexities of multi-stop container routes, researchers can develop more sophisticated models and optimization strategies for intermodal transportation.

Furthermore, this study primarily focused on uncertainty related to travel times. To enhance the model's applicability and realism, it would be valuable to consider stochastic demand as well. Incorporating uncertainty in demand can capture the variations and fluctuations in market conditions, providing insights into the robustness of the proposed inventory management strategies.

By pursuing these future avenues of research, scholars can contribute to the advancement of intermodal transportation and chassis inventory management. Exploring the impact of

disruptions on multiple terminals, investigating multi-stop container movements, and considering stochastic demand will lead to more comprehensive and practical models and solutions.

7.4 Conclusion and future directions on relay networks

In Chapter 6 of the thesis, a critical issue plaguing the trucking industry was discussed: the alarming rate of driver turnover and chronic shortage of truckers. The demanding working conditions, which required drivers to be away from home for extended periods, contributed to this problem. These conditions have significantly contributed to the persistent problem of high driver turnover. Moreover, factors such as transit delays and the health-related pressures associated with the trucker lifestyle have further exacerbated this issue.

To tackle the challenge of driver turnover and shortage, the thesis proposed a novel approach in the form of an integer programming model for the pickup and delivery problem with relay points. Unlike existing models, this innovative approach took into account the location of depots and relays as decision variables. By introducing the concept of "link types," the model was able to enhance its flexibility and provide a comprehensive framework for addressing the different movements and actions of trucks in the problem.

The adoption of this new approach yielded significant benefits. Firstly, it effectively reduced the number of variables involved in the model, streamlining the solution process. Additionally, it successfully eliminated subtours. This reduction in complexity simplified the problem-solving process.

The thesis conducted thorough investigations into two scenarios: one scenario without considering relays and another scenario that incorporated relays. Notably, in all scenarios, the inclusion of relays proved highly advantageous. It resulted in a substantial reduction in the total

distance covered by the trucks, indicating practical benefits and enhancing the overall efficiency of the proposed model.

Furthermore, the proposed model, referred to as the DRPLTR, underwent a comprehensive comparison with a well-established existing model in the literature known as PDPT. In the scenarios examined, the DRPLTR consistently outperformed the PDPT in terms of objective value and solution time, except when the subset of locations included those that were optimal in DRPLTR. In that specific case, the PDPT achieved the same objective function value as the DRPLTR, though it still took longer for the PDPT to find the solution.

In this study, we focused on a deterministic problem, assuming that travel times from node to node in the trucking network are fixed. However, real-world conditions are dynamic and subject to various factors such as traffic congestion or the absence of traffic. These uncertainties can significantly impact the optimal selection of relay and depot locations, as well as the routing decisions made by the model. Thus, to enhance the model's applicability and realism, it becomes valuable to incorporate stochastic travel times into the analysis. By considering stochastic travel times, we can capture the inherent variability in the transportation system. This can be achieved by incorporating historical data on travel times, real-time traffic information, or even predictive models based on machine learning techniques. By accounting for the uncertainty in travel times, the model can make more informed decisions on relay and depot locations and optimize the routing strategies, taking into consideration the probabilistic nature of the travel times. This stochastic approach adds a layer of complexity and realism to the model, reflecting the challenges faced by trucking companies and their operations in the real world.

Furthermore, in the current study, we did not explicitly consider the scheduling of drivers. However, in practice, drivers have regulatory constraints, such as hours of service regulations,

which limit their driving time and require them to take breaks and rest periods. By incorporating driver scheduling and considering time windows for pickups and deliveries, the model becomes more representative of the operational constraints faced by trucking companies. By optimizing driver schedules while respecting these constraints, the model can provide practical and feasible solutions that consider both operational efficiency and compliance with regulations.

Additionally, addressing the uncertainty in demand patterns adds yet another layer of complexity to the problem. As orders and shipment volumes fluctuate over time, it becomes crucial to strategically position trucks and make informed decisions about accepting additional orders. Balancing the trade-off between meeting future demand from high-volume regions and potential disruptions to the existing network requires careful analysis. By incorporating statistical learning methods, the model can predict future demand probabilities and optimize the driver schedule and network design accordingly.

Solving the stochastic problem with probabilistic travel times, driver scheduling considerations, and uncertainty in demand can be accomplished using techniques such as the L-shaped method. The L-shaped method iteratively solves the stochastic problem by decomposing it into a master problem and a sub-problem, allowing us to capture and optimize for uncertainty effectively.

Overall, these future directions involving stochastic travel times, driver scheduling, and uncertain demand open up new research opportunities and allow for a more comprehensive analysis of the trucking industry's challenges. By incorporating these elements into the model and utilizing advanced optimization techniques, we can develop more robust and effective solutions that align with the complexities and uncertainties faced by the trucking industry.

References

Alvarez, J. F. (2009). Joint routing and deployment of a fleet of container vessels. *Maritime Economics & Logistics, 1*(11), 286-208.

American Airlines. (2020). *American Airlines 2020 Workforce Demographics.*

Andreas T. Ernst, & Krishnamoorthy, M. (1999). Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research, 86*(0), 141–159.

Bashiri, M., Mirzaei, M., & Randall., M. (2013). Modeling fuzzy capacitated p-hub center problem and a genetic algorithm solution. *Applied mathematical modelling, 37*(5), 3513-3525.

Bin, W., & Zhongchen, W. (2007). Research on the optimization of intermodal empty container reposition of land-carriage. *Journal of Transportation Systems Engineering and Information Technology, 7*(3), 29-33.

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

Brimberg, J., Mišković, S., Todosijević, R., & Uroševic, D. (2022). The uncapacitated r-allocation p-hub center problem. *International Transactions in Operational Research, 29*(2), 854-878.

Bruce C.Hartman and Christopher Clott. (2015). Intermodal chassis supply in the US—A Bayesian game model. *Research in Transportation Business & Management, 14*(1), 66-71.

Bureau of Transportation Statistics. (2021, March 11). Retrieved from https://www.bts.gov/newsroom/full-year-2020-and-december-2020-us-airline-traffic-data

Campbell, A. M., Lowe, T. J., & Zhang, L. (2007). The p-hub center allocation problem. *European journal of operational research, 176*(2), 819-835.

Campbell, J. F. (1992). Location and allocation for distribution systems with transshipments and transportation economies of scale. *Annals of operations research, 40*(1), 77-99.

Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research, 72*(2), 387-405.

Chassiakos, A., Jula, H., VanderBeek, T., Shellhammer, M., & Dona, S. (2017). Analysis and Optimization Methods for Centralized Processing of Chassis.

Collins, K. O. (2021). Optimizing empty container repositioning in a truck-rail intermodal network. *Master's dissertation, Kansas State University*.

Corry, P., & Kozan, E. (2006). An assignment model for dynamic load planning of intermodal trains. *Computers & Operations Research, 33*(1), 1-17.

Couchman, C. A. (2020). A two-stage stochastic inventory management model for an intermodal trucking company. *Master's dissertation, Kansas State University*.

Daskin, M. S. (1997). Network and discrete location: models, algorithms, and applications. *Journal of the Operational Research Society, 48*(7), 763-764.

Donegan, B. (2022, March 18). (Fox News) Retrieved from https://www.foxweather.com/extreme-weather/april-may-june-most-active-months-tornadoes-united-states

Drezner, Z., & Hamacher, H. W. (2004). *Facility location: applications and theory*. Springer Berlin, Heidelberg.

E.O'kelly, M. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research, 32*(3), 393-404.

Ebery, J. (2001). Solving large single allocation p-hub problems with two or three hubs. *European journal of operational research, 128*(2), 447-458.

Ebery, J., Krishnamoorthy, M., Ernst, A., & Boland, N. (2000). The capacitated multiple allocation hub location problem: Formulations and algorithms. *European journal of operational research, 120*(3), 614-631.

Ernst, A. T., Jiang, H., Krishanmoorthy, M., & Baatar, D. (2018). Reformulations and Computational Results for the Uncapacitated Single Allocation Hub Covering Problem. *Data and Decision Sciences in Action*, 133–148.

Federal Aviation Administration. (2020). *Federal Aviation Administration*. Retrieved from https://www.faa.gov/air_traffic/by_the_numbers/media/Air_Traffic_by_the_Numbers_2020.pdf

Gao, Y., & Qin, Z. (2016). A chance constrained programming approach for uncertain p-hub center location problem. *Computers & Industrial Engineering, 102*, 10-20.

Glover, F. (1989). Tabu Search—Part I. *INFORMS Journal on Computing, 1*(3), 190-206.

Glover, F. (1990). Tabu Search-Part II. *INFORMS Journal on Computing, 2*(1), 4-32.

Hanh D.Le-Griffin, Lam Mai, Mark Griffin. (2011). Impact of container chassis management practices in the United States on terminal operational efficiency: An operations and mitigation policy analysis. *Research in Transportation Economics, 39*(1), 90-99.

Hatice Calık, S. A. (2009). A tabu-search based heuristic for the hub covering problem over incomplete hub networks. *Computers & Operations Research, 36*(12), 3088-3096.

Holland, J. H. (1992). Genetic Algorithm. *Scientific American, 267*(1), 66-73.

Hwang, Y. H., & Lee, Y. H. (2012). Uncapacitated single allocation p-hub maximal covering problem. *Computers & Industrial Engineering, 63*(2), 382-389.

Idris, H. R., Ioannis Anagnostakis, B. D., Hansman, R. J., Clarke, J.-P., Feron, E., & Odoni, A. R. (2016). Observations of Departure Processes at Logan Airport to Support the Development of Departure Planning Tools. *Air Traffic Control Quarterly, 7*(4), 229-257.

J. F. Campbell, A. T. Ernst, M. Krishnamoorthy. (2005). Hub Arc Location Problems: Part I—Introduction and Results. *Management Science, 51*(10), 1540-1555.

Janković, O., & Stanimirović, Z. (2017). A general variable neighborhood search for solving the uncapacitated r-allocation p-hub maximal covering problem. *Electronic Notes in Discrete Mathematics, 58*(1), 23-30.

Justice, E. D. (1995). Optimization of chassis reallocation in doublestack container transportation systems. *University of Arkansas, PhD dissertation*.

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

Kara, B. Y., & Tansel., B. C. (2000). On the single-assignment p-hub center problem. *European Journal of Operational Research, 125*(3), 648-655.

Kewcharoenwong, P., & Üster, H. (2017). Relay Network Design with Capacity and Link-Imbalance Considerations: A Lagrangean Decomposition Algorithm and Analysis. *Transportation Science, 51*(4), 1177-1195.

Köksalan, M., & Soylu., B. (2010). Bicriteria p-hub location problems and evolutionary algorithms. *INFORMS Journal on Computing, 22*(4), 528-542.

Lin, A. (2015). *Rail vs Truck: Which Mode of Transport is Better?* Retrieved from https://www.freightcourse.com/rail-vs-truck/

Lin, C.-C., Lin, J.-Y., & Chen, Y.-C. (2012). The capacitated p-hub median problem with integral constraints: An application to a Chinese air cargo network. *Applied Mathematical Modelling, 36*(6), 2777-2787.

Lopez, E. (2003). How do ocean carriers organize the empty containers reposition activity in the USA? *Maritime Policy & Management , 3*(4), 339-355.

Macambira, E. M., & Souza, C. C. (2000). The edge-weighted clique problem: valid inequalities, facets and polyhedral computations. *European Journal of Operational Research, 123*(2), 346-371.

Maleki, M., Majlesinasab, N., & Sepehri, M. M. (2014). Two new models for redeployment of ambulances. *Computers & Industrial Engineering, 58*, 271-284.

Maleki, M., Majlesinasab, N., & Sinha, A. K. (2023). An efficient model for the multiple allocation hub maximal covering problem. *Optimization Methods and Software*, 1-22.

Marianov, V., & Serra, D. (2003). Location models for airline hubs behaving as M/D/c queues. *Computers & Operations Research, 30*(7), 983-1003.

Marianov, V., Serra, D., & ReVelle, C. (1999). Location of hubs in a competitive environment. *European journal of operational research, 114*(2), 363-371.

Marín, A., Cánovas, L., & Landete, M. (2006). New formulations for the uncapacitated multiple allocation hub location problem. *European Journal of Operational Research, 172*(1), 274-292.

Mark S. Daskin. (1983). A Maximum Expected Covering Location Model: Formulation, Properties and Heuristic Solution. *Transportation Science, 17*(1), 48-70.

Megiddo, N., Zemel, E., & Hakimi, S. L. (1983). The maximum coverage location problem. *SIAM Journal on Algebraic Discrete Methods, 4*(2), 253-261.

Melton, K., & Ingalls, R. (2012). Utilizing relay points to improve the truckload driving job. *International Journal of Supply Chain Management, 1*(3), 1-10.

Murray, J. (2020, November 20). *www.denverpost.com*. Retrieved from https://www.denverpost.com/2020/11/20/denver-airport-new-gates-united/

Ng, M. (2021). Strategies for chassis dislocation management at container ports: repositioning and yard consolidation. *Transportation Research Part C: Emerging Technologies, 124*, 1-9.

Ng, M., & Talley, W. K. (2017). Chassis inventory management at US container ports: Modelling and case study. *International Journal of Production Research, 55*(18), 5394–5404.

Nga, M., & Talley, W. K. (2020). Rail intermodal management at marine container terminals: Loading double stack trains. *Transportation Research Part C, 112*, 252-259.

O'kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European journal of operational research, 32*(3), 393-404.

O'Kelly, M. E. (1992). Hub facility location with fixed costs. *Papers in Regional Science, 71*(3), 293-306.

Padberg, M. (1989). The boolean quadric polytope: some characteristics, facets and relatives. *Mathematical programming, 45*, 139-172.

Peker, M., & Kara, a. B. (2015). The P-Hub maximal covering problem and extensions for gradual decay functions. *Omega, 54*, 158-172.

Philadelphia International Airport. (2017, May 23). Retrieved from https://www.phl.org/drupalbin/media/PHLAmerican517.pdf

Qu, B., & Weng, K. (2009). Path relinking approach for multiple allocation hub maximal covering problem. *Computers & Mathematics with Applications, 57*(11), 1890-1894.

Rais, A., Alvelos, F., & Carvalho, M. S. (2014). New mixed integer-programming model for the pickup-and-delivery problem with transshipment. *European Journal of Operational Research, 235*(3), 530-539.

Sepehri, M. M., Maleki, M., & Majlesinasab, N. (2013). Designing a Redeployment Model for Located Ambulances. *International Journal of Industrial Engineering & Production Management, 24*(2), 171-182.

Shintani, K., Imai, A., Nishimura, E., & Papadimitriou, S. (2007). The container shipping network design problem with empty container repositioning. *Transportation Research Part E: Logistics and Transportation Review, 43*(1), 39-59.

Silva, M. R., & Cunha, C. B. (2017). A tabu search heuristic for the uncapacitated single allocation p-hub maximal covering problem. *European Journal of Operational Research, 262*(3), 954-965.

Skorin-Kapov, D., Skorin-Kapov, J., & O'Kelly, M. (1996). Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research, 94*(3), 582-593.

Song, D.-P., & Dong, J.-X. (2012). Cargo routing and empty container repositioning in multiple shipping service routes. *Transportation Research Part B: Methodological, 46*(10), 1556-1575.

T.Ernst, A., & Mohan, K. (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location Science, 4*(3), 139-154.

T.Ernst, A., Hamacher, H., Jiang, H., Krishnamoorthy, M., & Woeginger, G. (2009). Uncapacitated single and multiple allocation p-hub center problems. *Computers & Operations Research, 36*(7), 2230-2241.

Taha, T. T., & Taylor, G. D. (1994). An integrated modeling framework for evaluating hub-and-spoke networks in truckload trucking. *Logistics and Transportation Review, 30*(2), 141-166.

Üster, H., & Kewcharoenwong, P. (2011). Strategic design and analysis of a relay network in truckload transportation. *Transportation Science, 45*(4), 505-523.

Üster, H., & Maheshwari, N. (2007). Strategic network design for multi-zone truckload shipments. *IIE Transactions, 39*(2), 177-189.

Wagner, B. (2008). Model formulations for hub covering problems. *Journal of the Operational Research Society, 59*(7), 932-938.

Wanger, B. (2007). An exact solution procedure for a cluster hub location problem. *European Journal of Operational Research, 178*(2), 391-401.

Yaman, H., Y.Kara, B., & Ç.Tansel, B. (2007). The latest arrival hub location problem for cargo delivery systems with stopovers. *Transportation Research Part B: Methodological, 41*(8), 906-919.

Yang, K., Liu, Y., & Yang, G. (2013). An improved hybrid particle swarm optimization algorithm for fuzzy p-hub center problem. *Computers & Industrial Engineering, 64*(1), 133-142.