

HUB NETWORK DESIGN MODELS FOR INTERMODAL LOGISTICS

by

RAFAY ISHFAQ

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Information Systems,
Statistics and Management Science
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2010

ABSTRACT

This research is motivated by the extraordinary increase in the use of intermodal shipments in recent years for both domestic and global movement of freight. Three mathematical models, which explore the dynamics of intermodal hub-and-spoke networks, are presented. These models use transportation costs, fixed location costs, modal connectivity costs, modal transit times and service time requirements, for three modes of transportation: road, rail and air. The objective is to compare the conventional over-the-road (OTR) networks with intermodal (IM) logistics networks over the financial, operational and service issues. This research extends the p-hub median approach of interacting hub network design to intermodal logistics domain. Intermodal shipments are classified as multiple class products with deterministic routings in a serial queuing network. Such an approach is used to model flow congestion at hubs and to study its impact on the hub network design. This research also develops and tests metaheuristic (Tabu Search) solution approaches. These solution approaches are benchmarked with tight lower bounds based on Lagrangian and linear relaxations. These benchmarking studies show that the metaheuristic solution approaches are suitable for solving large intermodal network design problems. In order to bring practical relevance to this work, real world data is used with the actual footprint of interstate highways, intermodal rail and air freight networks. Multiple research studies are conducted to gain useful insights into the issues related to the design and management of IM logistics networks. These studies show that the use of IM shipments can provide significant benefits in reducing total logistics costs. It is also shown that use of IM shipments requires re-alignment of a logistics network in terms of its network structure and service design. Overall, significant savings can be realized from the use of IM shipments, but operating an intermodal logistics network requires a high degree of planning and managerial control.

DEDICATION

This dissertation is dedicated to my best friend and life companion, Uzma, and to my kids, Scheherbano and Jahanzeb, who have persevered through dissertations of both their parents.

LIST OF ABBREVIATIONS AND SYMBOLS

T	=	$\{0,1,2\}$, set of modes
S	=	Set of cost lines in piecewise linear cost function
N	=	Set of all nodes/cities
R_O	=	$\{\text{unload, batch, load}\}$, Routing of Hub Outbound shipments
R_I	=	$\{\text{unload, break-bulk, load}\}$, Routing of Hub Inbound shipments
R_L	=	$\{\text{unload, load}\}$, Routing of Local shipments
i, j	=	Origin and destination cities
k, m	=	Origin and destination hub cities
C	=	Set of all cities
p	=	Number of hubs
α	=	Cost discount factor; $0 \leq \alpha \leq 1$
β	=	Delay factor for consolidation/breakbulk at hubs, $\beta \geq 1$
f_{ij}	=	Flow volume from the origin city i to the destination city j
c_{ij}^t	=	Unit transportation cost from city i to city j using mode t
c_{ij}	=	Unit transportation cost from node i to node j using road transportation
\hat{c}_{ij}	=	Unit transportation cost for a direct shipment from city i to city j
F_k	=	Fixed cost of opening and operating a hub in city k

MC_{kt}	=	Modal connectivity cost of serving mode t at hub k
τ	=	Time delay factor at hubs, $\tau \geq 1$
$C(k, m, t)$	=	Transportation costs of shipments between hubs k and m using mode
β_0^s, β_1^s	=	Intercept and slope of the cost line s , respectively
\acute{M}	=	$\sum_{i,j \in N} f_{ij}$, a large number
L^t	=	Size of a mode t container, $t \in \{1, 2\}$
d_{km}^t	=	Drayage cost of a container shipped by mode t from hub k to hub m
T_{ij}^t	=	Travel time between cities i and j using mode t
TW_{ij}	=	Service time requirement for a shipment between origin i and destination j
t_{ij}	=	Shipment transit time between node i and node j
y_k	=	Selection of city k as a hub
y_k	=	Direct shipment from city i to city j
X_{ijkm}^t	=	Assignment of hub pair (k, m) to origin-destination cities (i, j)
δ_{km}^s	=	Volume of flow between k and m corresponding to line segment s
S_{kt}	=	Use of mode t to serve hub ay city k
R_{km}^s	=	Total flows shipped between hubs k and m using cost segment s
CL_{km}^t	=	Total flows shipped between hubs k and m using mode t
λ_{ijk}	=	Lagrangian Multipliers 1
μ_{ijm}	=	Lagrangian Multipliers 2
D	=	Number of days in a business year
L^t	=	Batch size of inter-hub transfers using mode t
pHM	=	p hub median

CL	=	Size of an intermodal rail container
$MCNF$	=	Minimum cost network flow
TX	=	Transshipment
MH	=	Meta-heuristic
SP	=	Shortest Path
HUR	=	Heuristic
EX	=	Experimental study
CS	=	Case study
$LR(\bar{\lambda}, \bar{\mu})$	=	Lagrangian relaxation
Z	=	Optimal solution value of primal problem
$Z_{LR}(\bar{\lambda}, \bar{\mu})$	=	Optimal solution value of Lagrangian relaxation
$HL(\bar{\lambda}, \bar{\mu})$	=	Hub location sub-problem
$HA(\bar{\lambda}, \bar{\mu})$	=	Hub allocation sub-problem
LD	=	Lagrangian dual
UB	=	Upper bound
LB	=	Lower bound
OTR	=	Over-the-road logistics networks
IM	=	Inter-modal logistics networks
ϵ	=	Small constant value
\hat{Z}	=	Feasible solution value
ω_n^{ijk}	=	Sub-gradient
S	=	Step size

t_n	=	Multiplicative factor
H	=	Set of hub nodes
$N(H)$	=	Neighborhood of H
TS	=	Tabu search procedure
z_{TS}	=	Solution value of tabu search
z^*	=	Optimal solution value
CR	=	Cost ratio
MC	=	Modal connectivity cost
TXC	=	Total transportation costs
$IM - Hubs$	=	Number of intermodal hubs in the network
$IM - Flows$	=	Total flows that use intermodal shipments between hubs
$S - Flows$	=	Total flows that use single hub shipments
$\%IM - Hubs$	=	Percentage of IM hubs in a network
$W_k^O(\mathbf{X})$	=	Average waiting time for operation O at hub k
a_{ij}	=	Squared coefficient of variability of inter-arrival times at j from i
ca_{ij}	=	Expected arrival rate at station j from station i
s_j	=	Service time of station j
μ_j	=	Expected service rate of station j
cs_j	=	Squared coefficient of variability of service time of station j
n	=	Number of stations
R	=	Number of classes
n_r	=	Number of operations for class r

$\hat{\lambda}_r$	=	Expected external arrival rate of class r
$\hat{c}a_r$	=	Squared coefficient of variability of external inter-arrival times
n_{rl}	=	Station visited for operation l in class r routing
$E(s_{rl})$	=	Expected service time for operation l in class r routing
cs_{rl}	=	Squared coefficient of variability of service time for operation l in class r routing
d_j	=	Time between departures from station j
d_{rl}	=	Time between departures of class r from station j after operation l
z_{rl}	=	Number of aggregated class arrivals at station j between two arrivals of class r
μ_k^j	=	Service rate of station j at hub k
cs_k^j	=	Squared coefficient of variability of service time of station j at hub k
λ_k^j	=	Arrival rate at station j of hub k
ca_k^j	=	Squared coefficient of variability of arrival times at station j at hub k
cd_k^j	=	Squared coefficient of variability of departure times from station j at hub k
λ_{kmt}^O	=	Arrival rate of outbound hub shipments at hub k
λ_{kt}^I	=	Arrival rate of inbound hub shipments at hub k
λ_k^L	=	Arrival rate of local shipments at hub k
\tilde{c}_t	=	Transit time variability of mode t shipments
\hat{c}	=	Transit time variability of pickup/drop off shipments

ACKNOWLEDGEMENTS

I gratefully acknowledge the support and encouragement of my mentor and advisor, Dr. Charles Sox, who helped me successfully navigate the highs and lows of this dissertation. During the course of this dissertation, his continuous encouragement and help was invaluable. He facilitated many professional opportunities which helped me develop my skills both as a researcher and as an instructor.

I also acknowledge Dr. Charles Schmidt and Dr. John Mittenthal who along with Dr. Sox, taught me so much, both inside and outside of the classroom. Their high quality academic teaching and research advising is an asset for the Operations Management program at University of Alabama. It was an honor for me to get the opportunity to learn from these outstanding professors.

I wish to thank Dana Merchant for her efficient and caring ways. She took care of all the administrative needs and handled the matters with an ease that amazes me to this day.

I owe a great deal of gratitude to my peers at University of Alabama. Over the past few years, I met and studied with some of the best people I have known. Many thanks to Stephen Hill, Sudarsan Rangan, Nate Palmer, Ben Neve and James Malm. Without your knowing, I learnt many valuable lessons as an observer of your own trek through the doctoral program. Thanks to all of you for your guidance and support.

Most of all, I acknowledge with utmost love and respect, the role of my parents, M. Ishfaq Ali Khan and Shahida Ishfaq, who raised me and my siblings with the moral and ethical values that define me, both as a person and as a professional. I am also grateful to Raja Shafaatullah, my father-in-law, who kept a keen interest in my progress and encouraged me all through the graduate school.

CONTENTS

ABSTRACT	ii
DEDICATION	iii
LIST OF ABBREVIATIONS AND SYMBOLS	iv
ACKNOWLEDGEMENTS	ix
LIST OF TABLES	xi
LIST OF FIGURES	xiii
INTRODUCTION	1
ARTICLE 1: HUB LOCATION-ALLOCATION IN INTERMODAL LOGISTICS NETWORKS	7
ARTICLE 2: INTERPLAY OF FINANCIAL, OPERATIONAL AND SERVICE ISSUES IN INTERMODAL LOGISTICS	49
ARTICLE 3: DESIGN OF INTERMODAL NETWORKS UNDER HUB CONGESTION	99
OVERALL CONCLUSIONS	146

LIST OF TABLES

1	Intermodal Hub Network Literature	12
2	Network Scenarios	16
3	Subgradient Optimization	24
4	Pseudocode for Tabu Search	26
5	Data Sets for Computational Study	29
6	Parameters for Cost Matrix	29
7	Average Percent Optimality Gaps (Tabu Search)	31
8	Average Percent Duality Gaps (Lagrangian Relaxation)	31
9	Running Times (sec) for CPLEX and Tabu Search	32
10	Summary of Average Percent Solution Gap	33
11	Average Percent Solution Gap ($10 \leq n \leq 100$)	34
12	Experimental Design	36
13	Percentage Solution Gaps	36
14	Prior Literature - Intermodal Hub Networks	55
15	Pseudocode for Tabu Search Procedure	68
16	UFL Problem Parameters	71
17	Matrix Table	73
18	Cost & Time Matrices	73
19	Flow Matrices	73
20	Parameter Values for Computational Study	74
21	Average Percent Optimality Gaps for 10-city Network Problems	74
22	Average Percent Optimality Gaps for 15-city Network Problems	75
23	Average Percent Optimality Gaps for 20-city Network Problems	75

24	Regions and Flows in the Case Study	78
25	Factor Levels	80
26	Number of Hubs	83
27	Types of Shipments and their Use (Average Percentage of Total Flows)	85
28	Use of Direct and Single Hub Shipments	88
29	Modal Flows (% of Total Flows)	89
30	Prior Literature - Intermodal Hub Networks Under Congestion	103
31	Data Sets for Benchmarking Study	130
32	Average Percent Solution Gaps (Benchmarking Study)	131
33	Factor Levels	134
34	Selection Frequency of Hub Cities	138
35	Effects of TW and μ on Number of Hubs	138

LIST OF FIGURES

1	Performance of Tabu Search	33
2	Fixed Cost Plots	37
3	Modal Connectivity Plots	39
4	Discount Factor Plots	40
5	Service Time Plots	42
6	Types of Shipments	60
7	Piecewise Linear Cost Function	62
8	Empirical Study of 35-City Network	79
9	Hub Locations	82
10	Total Network Costs	84
11	Hub Shipments	87
12	Effects of Economies-of-Scale	93
13	Types of Shipments	105
14	Hub Operations	106
15	Hub Operations Queuing System	118
16	Tabu Search Solution Approach	128
17	Solution Procedure for Allocation Sub-problem	129
18	25 Cities used in Case Study	134
19	Total Network Costs	136
20	Best Locations for Logistics Hub	137
21	Modal Flows	140
22	Shipment Waiting Times	141

INTRODUCTION

This dissertation is focused on the design and management of intermodal logistics networks. For a third-party logistics service provider, an efficient logistics network provides a source of competitive advantage in the marketplace. In a quest to continually improve operational efficiencies and service levels, logistics service providers seek to develop a logistics strategy which can help in achieving this goal. One such opportunity is provided by the use of intermodal shipments. An intermodal shipment refers to the integrated use of two or more modes of transportation for delivering goods from origin to destination in a seamless flow. The inherent benefit of intermodal shipments lies in the cost advantage obtained by the joint use of multiple modes of transportation: road, intermodal rail, air and ocean freight. The operational efficiencies in an intermodal network are obtained by combining the load carrying capacity of intermodal rail, accessibility of interstate highways, speed of air freight and economies of density of ocean freight. With increasing demand for intermodal shipments, both private and public sectors have invested heavily in developing the infrastructure necessary for the seamless flow of freight between different modes of transportation. The use of standardized intermodal containers has acted as a catalyst in improving the operational and service performance of intermodal logistics networks.

In the early days of intermodal logistics, the use of intermodal shipments was impeded by the performance gap between intermodal rail and road carriers. The shorter average transit times and lesser transit time variability in road shipments favored the use of trucks or road-based logistics networks. However, the globalization of the marketplace brought about a paradigm shift in the logistics service industry. The need to ship freight across oceans and between continents in global supply chains required the use of intermodal logistics. The rise in demand for intermodal shipments for global movement of goods also generated domestic demand for such shipments. The growth in demand for intermodal shipments encouraged rail carriers to

improve their service performance and compete with road-based carriers. Rail carriers prioritized intermodal trains, added rail tracks and improved their overall transit time performance. The changing market dynamics and customer demand for better service performance, brought carriers in different segments of the logistics industry together. This resulted in business collaborations between road, rail and air freight carriers and the development of intermodal service offerings for domestic customers.

The expanding scale of activities and size of intermodal logistics networks in the United States has presented a new set of challenges. These challenges stem from the increase in complexity due to inherent differences among modal networks, in terms of their capability, cost structure, accessibility and service performance. To utilize the maximum cost advantage of intermodal shipments, logistics service providers need a realignment of their logistics strategy by updating the design of their logistics networks and logistics management practices.

DISSERTATION OBJECTIVES AND CONTRIBUTIONS

The primary objective of this research is to develop an understanding of the dynamics of intermodal logistics in terms of network design and logistics management, and to study the differences between the intermodal and the traditional road-based logistics networks. This understanding will help supply chain professionals in making informed decisions regarding appropriate assignment of resources and implementation of management controls for the critical parameters in domestic as well as global logistics networks.

This research seeks to develop multiple modeling frameworks and mathematical models to represent different aspects of the intermodal logistics network design. This research will also develop efficient and accurate solution procedures which can be used to solve real-life network design problems. The other objective of this dissertation is to conduct research studies based on real-world logistics data and modal networks. Through these studies, this research aims to develop relevant insights regarding the logistics strategy and management of logistics operations.

Based on these objectives, the major contributions of this research to the literature are:

- The hub network design problem is extended to the intermodal logistics network domain,

yielding a new mathematical model for the design of intermodal hub-and-spoke network under service time requirements.

- A metaheuristic solution approach and a Lagrangian relaxation procedure to find tight lower bounds of the problem are presented. In a computational study, the metaheuristic solution approach and lower bounds are tested over large network problems.
- The problem structure of the intermodal network design is explored in a research study which identifies the effects of different network parameters on the intermodal hub-and-spoke network structure.
- A new mathematical model is proposed which incorporates many features of the intermodal networks in the logistics industry, such as three transportation modes (road, rail, air), economies of scale, different modal rate structures and transit times, and modal service cost.
- A modified metaheuristic solution approach to the new model is developed and tested in a computational study.
- The logistics data from US Department of Transportation, logistics service providers and transportation carriers are compiled. The data are used in a research study along with the actual footprint of interstate highways, intermodal network of Class 1 rail carriers and air freight network of United Parcel Service (UPS).
- A comparative analysis of the differences between traditional road-based and intermodal logistics networks is conducted. The results demonstrate that these networks are different from each other in terms of network structure and service performance.
- A new model is proposed which integrates the hub location-allocation model with a hub operations queuing model. This integrated model incorporates the effect of hub congestion in a hub-based logistics network design.

- A new metaheuristic solution approach is developed which can handle the complexities of the integrated model. A linear relaxation procedure is presented which is used to test the satisfactory performance of the metaheuristic solution approach.
- A research study is conducted which investigates the impact of service time requirements and hub resource levels on the structure of an intermodal logistics network under hub congestion.

DISSERTATION FORMAT

The research work in this dissertation is spread over three articles. These articles are inter-related and build upon the knowledge and insights developed in the preceding work. The first article develops the background and introduces the dynamics of intermodal logistics networks. This article, entitled “Hub location-Allocation in Intermodal Logistics Networks” discusses the complexities of intermodal logistics networks. It presents a modeling framework which is used to develop a mathematical model for designing a hub-based intermodal logistics network. This mathematical model extends the *p-hub median* approach for the interacting hub location-allocation problem to the intermodal logistics domain. The model encompasses the dynamics of individual modes of transportation through transportation costs, modal connectivity costs, under service time requirements. A tabu search metaheuristic is used to solve large (100-node) problems. The solutions obtained using this metaheuristic, are compared with tight lower bounds developed using a Lagrangian relaxation approach. Using the model and the data from the Civil Aeronautics Board (CAB), this article provides insights into the structure of the intermodal network design problem. Using a sensitivity approach, a research study is designed to explore the effects and interactions of several factors on the design of intermodal hub networks subject to service time requirements.

The second article, entitled “Interplay of Financial, Operational and Service Issues in Intermodal Logistics”, discusses the increase in the use of intermodal shipments in recent years for both the domestic and the global movement of freight. This article is focused on a real-world logistics network based on the actual footprint of interstate highways, intermodal rail tracks of

class I rail carriers and the air freight network of United Parcel Service, one of the leading small shipment/package carriers in the US. The objective is to compare conventional over-the-road logistics networks with intermodal logistics networks in terms of their financial, operational and service performance. The mathematical model developed for this article adds to the work of the first article by incorporating an enhanced representation of an intermodal network. This model uses transportation costs of different modes of transportation, economies of scale, shipment transit times, fixed operational costs and cost of providing intermodal service for three modes of transportation: road, rail and air, under service time requirements.

Due to the added complexity of this model, a sophisticated solution approach is developed. This approach uses components of the work in the previous article and adds a module for handling the additional features of the model in the second article. This article conducts a large scale research study based on data from logistics service providers and intermodal rail/air freight carriers. This leads to an empirical investigation of the differences between the traditional over-the-road and intermodal logistics networks. This study demonstrates that the use of intermodal shipments can have a significant financial impact on the logistics costs of a company. The results of the empirical study show that the intermodal logistics networks are different from the over-the-road networks in terms of their hub locations and the network structure. The study shows that the over-the-road and intermodal networks differ in the use of direct and inter-hub shipments. In addition, the results demonstrate the significant role of service requirements and cost parameters in the location of hubs, the allocation of customers and the use of transportation modes in intermodal networks. The findings of this investigation will help supply chain professionals in decisions regarding appropriate assignment of resources and the implementation of management controls for the critical parameters in domestic as well as global logistics networks.

The third article is focused on the issue of congestion which is of significance in an intermodal logistics network, where shipments need to be delivered within a specific service time. The total shipment time in a network is affected by the time spent in a logistics hub and the variability in modal transit times. This article, entitled “Design of Intermodal Network under

Hub Congestion”, investigates the impact of hub congestion on the design of intermodal logistics networks. This article presents the case that in a road-rail intermodal hub network, cost benefits can be achieved through the use of intermodal shipments and the economies of scale achieved by the consolidation of flows at the hubs. However due to limited resources at the logistics hubs, congestion delays may affect service performance. In this article hub operations are modeled as a GI/G/1 queuing network and the shipments are modeled as multiple job classes with deterministic routings. By integrating the hub operation queuing model with a hub location-allocation model, this research investigates the effects of congestion on the design of intermodal networks. A research study of a 25-city intermodal logistics network finds that the availability of sufficient hub resources plays a big role in determining the network structure. The use of inter-hub shipments, which accounts for significant cost benefits, may be discouraged due to increase in dwell time at logistics hubs, if there are insufficient resources at these hubs. The results also show that a logistics network compensates for insufficient hub resources by using more hubs and redistributing flows serviced by each hub to a level which matches the available hub resources. However, the increase in the number of logistics hubs results in higher fixed operating and intermodal service costs, which increases the total network costs.

This dissertation concludes with a summary of results and managerial insights developed through this research. The conclusion section also provides potential extensions to the work done in this dissertation.

ARTICLE 1: HUB LOCATION-ALLOCATION IN INTERMODAL LOGISTICS NETWORKS

Intermodal transportation refers to the integrated use of two or more modes of transportation for delivering goods from origin to destination in a seamless flow (Crainic et al., 2007; Slack, 2001). Intermodal transportation differs from multimodal transportation, where the latter refers to the choice of a single mode of transportation among the available modes. The increased use of intermodal transportation started out as a direct result of globalization of the marketplace. This globalization was facilitated by regional and global trade agreements such as GATT and NAFTA (McCalla, 1999).

The emergence of global supply chains has led to the sustained demand for intermodal transportation. This revenue stream has encouraged all segments of the logistics industry in the US to collaborate in the provisioning of resources (such as infrastructure development, partnerships between carriers and prioritizing intermodal trains on rail tracks). These synergistic market forces have brought about the development of a logistics network which consists of multiple sub-networks of respective modes. Using standardized containers, shipments can seamlessly transfer between modes at transfer points (Slack, 1990).

Although global movement of freight has been the major force behind intermodal infrastructure development, the use of intermodal services has been on the rise in the U.S. domestic market. Even under lagging global economic conditions, domestic intermodal usage has been steadily increasing in recent years (IANA, 2008). The intermodal services in the domestic market include moving containerized loads over road-rail (and barges, although its use depends on geography), while packages and other smaller size, high value items are moved through road-air. Looking at the benefits (low cost, high capability and reach, competitive transit times) of intermodal transportation, it is not surprising that many shippers have started to move in this direction. Many LTL/TL carriers such as J.B. Hunt, Schneider National and Swift have reaped

high dividends from these services (Schwartz, 1992). Rail carriers such as Burlington Northern Santa Fe Railway (BNSF), Norfolk Southern (NS) and Union Pacific (UP) have seen a rise in intermodal freight traffic in the past ten years. Groothedde et al. (2005) discuss cases where 3PL companies have successfully organized the intermodal transportation part of their client's supply chain.

The ongoing research in this area relates to the complexity of an intermodal network which transcends the benefits and shortfalls of its respective transportation modes (Macharis and Bontekoning, 2004). Where some modes such as rail and water offer significant cost benefits, they also carry within their structure, inherent limitations in terms of transit times and accessibility. While road offers the most accessible transportation links, it cannot compete with rail and water in terms of large load movements (economies of density). However the integration of modes, provides the means to move shipments between origins and destinations which is both economical and operationally viable (Slack, 1990). The use of containerized shipments enables efficient transfer from one mode to the other.

Over the years a hub-based network structure has evolved for moving intermodal shipments. The emergence of hub-based intermodal networks indicates that economies of scale is the principal force behind their use (Slack, 1990). Because intermodal networks are combinations of their respective modal networks, it is natural that the hub network has emerged as the most suitable network structure for intermodal logistics (Bookbinder and Fox, 1998). The hub network consists of a small number of hubs, which serve their assigned (supply and demand) regions. Smaller sized shipments (packages, units and pallets) are consolidated at intermodal hubs and shipped between hubs using intermodal containers. These container-loads are akin to truck-loads in the conventional road transportation operations. The economical choice between intermodal and OTR (over-the-road) depends on the total logistics costs which includes fixed location costs, modal connectivity costs and underlying modal networks. The differences between modal networks in terms of transportation cost structure, network connectivity, location of transfer points and transit time performance indicate the level of complexity involved in intermodal networks. For a classification of research problems and applications of operations

research methods in intermodal transportation, see Macharis and Bontekoning (2004).

This work extends the existing research in the field of intermodal logistic networks by developing a novel modeling framework and a solution methodology for this class of problems. Given the structure of a hub network, this work extends the multiple-allocation p-hub median approach to the road-rail intermodal logistics domain. A modeling framework is presented which accommodates the operational structure of individual modes of transportation, the effect of shipment consolidation at hubs on transportation costs, the interactions between modes, the transit time delays and the service time requirements. It also uses a fixed cost of locating intermodal hubs and modal connectivity costs as a tradeoff between opening new facilities and reducing total transportation costs. The model compares the intermodal option with an OTR (over-the-road) option when choosing modal connectivity at hubs. The modal connectivity costs are specified by the type of modes used by a specific hub. In earlier research, solution techniques used to solve intermodal network problems had been restricted to the iterative application of the shortest path method and/or local search heuristics. Those techniques can be improved through the use of metaheuristics which have shown great promise in the realm of large scale networks. In this research, a tabu search metaheuristic is developed to solve large scale problem instances. The solutions are evaluated using tight lower bounds obtained through a Lagrangian relaxation approach.

The rest of this paper is organized as follows: First prior research in the intermodal logistics network domain is presented which also identifies the contributions of this work. Next, a modeling framework and the resulting mathematical model for the interacting intermodal hub location-allocation (IHLA) problem is discussed. This section also develops a lower bound procedure based on Lagrangian relaxation. Next, a tabu search based metaheuristic solution approach is presented. The tabu search procedure is tested over a range of data sets. The results of this experimental study are also reported. Next, managerial insights are discussed and followed by the conclusions of this article.

LITERATURE

The constructs of this research lie in the following major components: identifying optimal hub locations, assignment of hub(s) for each origin-destination pair, modal choice for flows between hubs, and service time requirements. The development of these constructs is based on the literature in the following broad research categories: (a) intermodal logistics, (b) interacting hub location-allocation problem, and (c) solution approaches. The relevant literature from these areas is discussed below.

Intermodal logistics has developed into a research stream in the past couple of decades within the transportation literature (Bontekoning et al., 2004). In one of the initial works in the design of intermodal hub networks, Arnold et al. (2001) developed formulations in which a fixed number of intermodal hubs are selected among candidate locations. These candidate locations are the nodes which are common to the respective modes. Arnold et al. (2004) presented alternate formulations by representing each constituent modal network as a sub-graph with nodes and arcs. The connectivity between the sub-graphs is represented by transfer arcs. The problem is solved using a heuristic approach involving the solution of a shortest path problem for each commodity (origin-destination pair). The shortest path approach is similar to Lozano and Storchi (2001) who used a minimum cost network flow formulation to find the shortest paths for the origin-destination pairs.

Racunica and Wynter (2005) evaluated the concept of hubs in a rail network. The consolidation in rail network results in rapid and reliable freight trains between a few hubs with high load factors (economies of density). The cost savings are realized due to higher equipment rotation, reduced staffing costs and higher train frequencies. The problem is solved by relaxing the integrality on the decision variables and solving the relaxed problem iteratively. In each iteration, some of the decision variables are set to zero or one, thus increasing the number of variables that have a fixed value.

Groothedde et al. (2005) studied the implementation of hub-based distribution networks in the consumer goods market. A road-barge intermodal option is compared with the road-only network. A heuristic solution is developed which starts with two hubs and iteratively selects a

hub based on improvement in the objective function value. The findings indicated that barge transportation is suitable for the stable part of the demand while the use of road transportation is required to handle variations over a short time horizon.

More recently, Limbourg and Jourquin (2009) discussed the location of terminals in a road-rail network. Their use of terminal is synonymous with hub networks. Their solution approach is based on a heuristic which solves the hub location problem completely over-the-road network first. The solution identifies optimal hub locations over the road. The flows between the hubs are assigned to a rail link, if such a link reduces the transportation costs. The problem is solved iteratively until the solution cannot be improved.

In above mentioned papers, the performance of a network is evaluated based on minimizing logistics costs. However aggregation of flows at hubs may result in shipments with large transit times. Competitive market conditions do not allow for ignoring this downside of the network scheme. Companies are looking for logistics network designs that reduce transit times while also reducing costs (Aykin, 1995). Ziliaskopoulos and Wardell (2000) developed an intermodal optimum path algorithm which accounts for transit time delays in intermodal networks. The algorithm computes the optimal assignments for origin-destination pairs using the available modes. Groothedde et al. (2005) included the time aspect in the model through variable costs which depend on transit time. In this manner, heuristic solution seeks to find hub locations which minimizes costs incurred due to transit time.

In order to highlight the contributions of this research, Table 1 summarizes the contents of the above mentioned articles and compares them with this current research. It can be seen that this current research develops a unified approach to the intermodal hub network problem. One contribution of this work is the use of service time requirements in the location and assignment decisions. These requirements may either be imposed by customers or be offered as delivery options. A premium service can be offered which delivers shipments within a preset service time while minimizing logistics costs. Although Groothedde et al. (2005) included transit times in the scope of their work, no specific service time requirements were used in the location/allocation decisions. Another aspect of this work is the use of modal

Table 1: Intermodal Hub Network Literature

Research Component	Current Paper	(Arnold et al., 2001)	(Arnold et al., 2004)	(Racunica and Wynter, 2005)	(Groothedde et al., 2005)	(Limbourg and Jourquin, 2009)
Modeling Approach	pHM	MCNF	MCNF	HLP	-	pHM
Fixed location costs	✓	✓	-	✓	-	-
Consolidation effect	✓	-	-	✓	✓	✓
Service times	✓	-	-	-	✓	-
Types of mode	Ro-R	Ro-R	Ro-R	Ro-R	Ro-B	Ro-R
Type of hubs	Ro,R	TX	TX	Rail	Barge	Ro,R
Solution approach	MH	-	SP	HUR	HUR	HUR
Solution validation	EX	-	CS	CS	CS	CS
Benchmarking (small problems)	Optimal	-	-	Optimal	-	-
Benchmarking (large problems)	LB	-	-	-	-	-

LEGEND:

pHM - p hub median; MCNF - Minimum cost network flow; HLP - Hub location problem
TX - Transshipment; Ro - Road; R - Rail; B - Barge; MH - Metaheuristic; SP - Shortest Path;
HUR - Heuristic; LB - Lower bound; EX - Experimental study; CS - Case study

connectivity costs. In Arnold et al. (2001) and Racunica and Wynter (2005) a fixed cost is used only in terms of the expense in opening a hub facility. This research expands the role of fixed costs to include the impact of resources, management and control issues which are involved in intermodal shipments. In Racunica and Wynter (2005) the scope of the hub network is limited to one (rail) network, whereas in this work both road and rail options are evaluated when establishing modal connectivity at the hubs. Though Groothedde et al. (2005) used two networks (road and barge) in its scope, the focus was more towards developing a heuristic rather than modeling. The modeling approach of Limbourg and Jourquin (2009) is similar to this work but lacks the use of fixed costs (location and modal connectivity) and service time restrictions. Furthermore, the p-hub median model is only applied to one mode (road), whereas this research models the complete intermodal network. From the solution methodology perspective, it can be seen that heuristics have been the dominant approach so far. These heuristics are used in

specific case studies without any measure of the quality of solutions obtained. This research presents a metaheuristic solution approach which is tested over a wide range of problems. This allows for detailed evaluation of the solution methodology through benchmarking with optimal solutions for small problems and lower bounds for larger problems.

The modeling approach used in this research is to formulate the intermodal hub network design as an interacting hub location-allocation problem. In general, a hub network is represented by an undirected graph $G(\mathbb{N}, \mathbb{A})$, where \mathbb{N} is a collection of nodes connected through a set of arcs \mathbb{A} . The arcs have weights which may represent distance between nodes, unit cost to travel between nodes, or travel time between nodes. The network is composed of a few central locations (relative to the total number of locations) which act as hubs. Each non-hub location is assigned to one or more hub(s). Such network designs are graphically represented as a wheel with the hub at the center connected to the non-hub locations through the spokes of the wheel. For this reason hub networks are also called hub-and-spoke networks. The flows between origin-destination pairs travel in the network through at most two hubs.

The interacting hub location-allocation research started with the pioneering work of O’Kelly (1986) and O’Kelly (1987) in which he presented a quadratic integer formulation. He showed that this is an NP-hard problem, and presented two enumeration-based heuristics. Klincewicz (1991) used a local neighborhood search to develop heuristics based on the clustering of nodes. Campbell (1994) provided a linear formulation by redefining the flow variables. He is among the first to use the term *p-hub median* to describe this problem (Campbell, 1992). This research is related to the multiple-allocation version of the p-hub median problem. As defined by Campbell (1994), in a multiple-allocation solution, each origin may use multiple hubs to send its shipments. Skorin-Kapov et al. (1996) obtained exact solutions to the p-hub median problem by developing tight linear relaxations of the formulation given by Campbell (1994). New MILP formulations of the problem that involve fewer variables and constraints were given in Ernst and Krishnamoorthy (1998) based on the idea proposed in Ernst and Krishnamoorthy (1996). For a more detailed review of hub location problems, see Alumur and Kara (2008).

The solution approaches for p-hub median problem were initially based on heuristic methods.

Campbell (1996) used a greedy-interchange heuristic. Later Ernst and Krishnamoorthy (1998) used an LP based branch-and-bound method to obtain exact solutions. O’Kelly et al. (1996) presented a linearization scheme which reduced the size of the problem formulation. The lower bounds used in branch-and-bound methods were improved in Boland et al. (2004) by developing preprocessing techniques and tightening constraints.

The metaheuristic solution approaches for solving large scale p-hub median problems are based on tabu search (Klincewicz, 1992; Skorin-Kapov and Skorin-Kapov, 1994; Skorin-Kapov et al., 1996; Carello et al., 2004), simulated annealing (Ernst and Krishnamoorthy, 1996; Abdinnour-Helm, 2001; Rodriguez et al., 2007) and genetic algorithms (Abdinnour-Helm and Venkataramanan, 1998; Topcuoglu et al., 2005; Kratica et al., 2007; Cunha and Silva, 2007). In Skorin-Kapov et al. (1996), the optimality (or $< 1\%$ optimality gap) of tabu search solutions was established for problem instances based on the CAB data set. Additionally, Abdinnour-Helm (2001) in a separate study showed that for p-hub median problems tabu search found better solutions than simulated annealing in 30% of the problem instances tested (the solutions were identical in the remaining problems).

Based on the prior literature, the tabu search methodology was selected for this research. Tabu search directs the solution procedure to break out of a local optimum and to move to previously unexplored areas of the solution space (Klincewicz, 1992). A short-term memory (tabu list) is used which records node interchanges (moves) previously undertaken. Such moves are deemed tabu for some duration (number of iterations). This helps avoid re-evaluating solutions that were recently visited. An aspiration criterion is used to override a tabu move so as not to miss a better solution during the search. A long-term memory list is used to restart the search process from a previously unexplored part of the solution space (Glover, 1989, 1990).

In order to evaluate the quality of solutions obtained by heuristics/metaheuristics, lower bounds are used. In this research lower bounds are computed using a Lagrangian relaxation approach. Using this approach some complicating set of constraints in the model are identified. These constraints are added to the objective function and weighted by multipliers, known as Lagrangian multipliers. This Lagrangian relaxation problem is solved iteratively while the

multipliers are adjusted along the way. Solving the Lagrangian dual problem yields a solution which provides a tight lower bound to the original problem. There have been several good surveys of Lagrangian relaxation technique, such as Geoffrion (1974) and Fisher (1981).

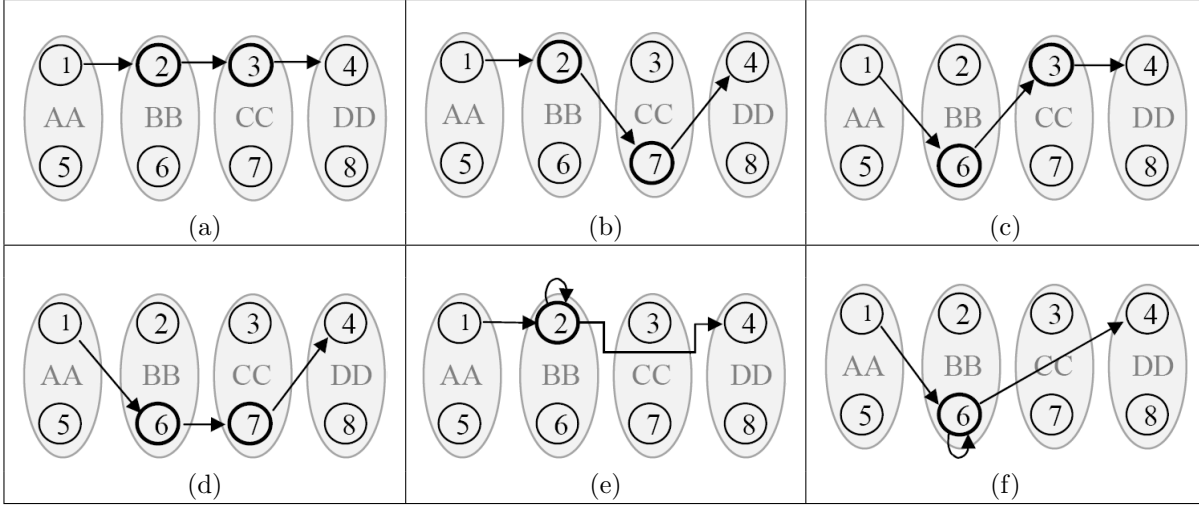
In the next section, a modeling framework is developed which incorporates the constructs discussed above. This modeling framework is used to develop a mathematical model for the intermodal hub location-allocation problem. A Lagrangian relaxation is proposed to find a tight lower bound on the optimal solution of this problem.

MODELING FRAMEWORK

Within the context of this research, an *intermodal hub* is a shipment handling facility which has the access and the capability of handling atleast two modes of transportation. The modeling framework discussed in this section falls within the context of a road-rail intermodal network. However, this framework is also applicable to other intermodal networks, such as road-air and road-barge. The intermodal hub handles shipments received from originating locations (cities) delivered through trucks. The packages are sorted and grouped with other packages destined for a particular geographical region. The hubs consolidate LTL shipments into truck/car loads and thus realize economies of scale in transportation costs. The shipments sent by rail are loaded in an intermodal rail container which is moved to an intermodal ramp at a railway yard for dispatch. The shipments sent by road are loaded onto a long-haul truck. The shipments received at the destination hub are separated by customers and shipped via truck to their final destinations. Thus a regional intermodal hub is an alternate shipment handling facility which manages shipments similar to a traditional regional road hub, except that it offers the choice between multiple modes of transportation.

In its general form, an intermodal network is a collection of distinct modal networks. In this research, road and rail networks are considered. Road and rail networks are represented by undirected graphs $\mathbb{G}_1(\mathbb{N}_1, \mathbb{A}_1)$ and $\mathbb{G}_2(\mathbb{N}_2, \mathbb{A}_2)$, respectively. The node sets \mathbb{N}_1 and \mathbb{N}_2 represent cities, and the arc sets \mathbb{A}_1 and \mathbb{A}_2 represent transportation links between cities for each of the two modes. The intermodal network is thus represented by $\mathbb{G}(\mathbb{N}, \mathbb{A})$, where $\mathbb{N} = \mathbb{N}_1 \cup \mathbb{N}_2$

Table 2: Network Scenarios



and $\mathbb{A} = \mathbb{A}_1 \cup \mathbb{A}_2$. Due to the nature of intermodal networks, the size of the network graph is generally large (Arnold et al., 2001; Racunica and Wynter, 2005; Groothedde et al., 2005; Limbourg and Jourquin, 2009). However the computational results of this research show that the proposed solution approach is capable of handling this representation of an intermodal network.

In the following analysis, the following assumptions are made:

- Any city can have a road hub,
- Any city that has rail access can have either a road hub or an intermodal hub, and
- Pickup from origins and drop-off to destinations are done over the road.

For further discussions, a simple example of a four-city problem is used. In this example, City AA is represented by nodes 1 and 5, city BB by nodes 2 and 6, city CC by nodes 3 and 7 and city DD is represented by nodes 4 and 8. Different network scenarios are shown in Table 2 below. The terms “originating hub” and “dispatching hub” are used to identify hub pairs assigned to a particular shipment. Note that nodes 1,...,4 represent road nodes and nodes 5,...,8 represent rail nodes. Also note that in these scenarios, thick circles represent hub locations.

The scenarios (a)-(d) in Table 2 present cases for two-hub shipments, while scenarios (e) and (f) represent cases for single-hub shipments. In scenario (a), arcs are shown for origin-

destination pair (city AA, city DD) when hubs are located at road nodes in city BB and CC. The weight of arc (1,2) represents the cost of shipment pickup from city AA and travel to the originating hub at city BB by road. The weight of arc (2,3) represents the cost of shipment handling at hubs in cities BB and CC and road transportation cost from city BB to city CC. The weight of arc (3,4) represents the cost of road transportation to destination city DD.

Scenario (b) shows the case where a road hub is located in city BB and a rail hub in city CC. For an origin-destination pair (city AA, city DD), arc (1,2) represents pickup from origin city AA to originating hub city BB by road. The shipments travel between hubs in cities BB and CC by road, represented by arc (2,7). Note that the weight of arc (2,7) is the same as that of arc (2,3), as discussed in scenario (a). The weight of arc (7,4) represents the road dropoff cost to city DD.

A variation of the case discussed above is shown in scenario (c), in which originating hub is located at the rail node of city BB, while the dispatching hub is located at the road hub in city CC. The weight of arc (1,6) represents the cost of shipment pickup from city AA to originating hub at city BB by road. The weight of arc (6,3) represents the costs of shipment handling at cities BB and CC and road transportation costs from city BB to city CC. Arc (3,4) represents the cost of road transportation to city DD and local delivery to the destination.

Scenario (d) shows the case when both hubs are located at rail nodes in cities BB and CC, respectively. The weight of arc (1,6) is pickup cost for shipments originating in city AA. The weight of arc (6,7) represents the processing cost at hub facilities and interhub travel over rail. The weight of arc (7,4) is road dropoff cost from city CC to destination city DD.

The next two scenarios show cases where only one hub is used for shipments between origin-destination pair (city AA, city DD). In scenario (e), hub located at road node in city BB is used for shipments originating in city AA for destination city DD. Arc (1,2) represents the pickup from origin city AA to hub city BB, by road. The weight of arc (2,2) is the processing cost at the hub. The road dropoff to destination city DD is represented by the arc (2,4). In the last Scenario (f), hub used for shipments is located at a rail node. In such a network structure, arcs (1,6) and (6,6) have same meanings and weights as that of arcs (1,2) and (2,2) in scenario

(e). The weight of arc (6,4) represents the dropoff cost to the destination city DD, which is the same as arc (2,4) in scenario (e).

MODEL (IHLA) FORMULATION

This section presents a mathematical formulation for a road-rail intermodal hub location-allocation (IHLA) problem. The model uses a node identification scheme in which the road nodes are numbered consecutively from 1 to $|C|$, where C is the set of all cities in the problem. The rail nodes are numbered $|C| + 1$ to $2|C|$. In a typical intermodal network, shipments are picked up by truck and brought to the originating hub location. At the hub, shipments are consolidated according to their destinations and sent to their respective dispatching hubs near the destination. Owing to the economies of scale, transportation costs are discounted over the hub-to-hub link. In this research, a linear approximation (α) is used to represent the discounted inter-hub cost structure (Campbell, 2009; Limbourg and Jourquin, 2009; Alumur and Kara, 2008). In this model, a shipment may flow from origin to destination either entirely on the road network (road only case) or a combination of rail and road (intermodal case). The choice of modes lies in the economics and transit time restrictions for the respective shipments. Any change in mode for a shipment can only happen at an intermodal hub.

In any logistics network, the fixed costs of a hub relates to the capital cost incurred in opening a facility. In an intermodal logistics network, hub facilities are more complex because of the additive effect of managing multiple modes of transportation. These *modal connectivity costs* are directly linked to the type of modes serviced by the hub. For example the docking/loading/unloading arrangement for an intermodal rail container may be different from say a road trailer. Similarly in the case of airfreight, containers have different size and loading/unloading needs. This model associates a different fixed cost to opening different types of hub facilities. In the case of road hubs, the fixed cost include only the location costs. The fixed cost of an intermodal hub reflects both location costs as well as modal connectivity costs. Thus the fixed cost of an intermodal hub are higher than a road-only hub.

The service time restrictions are defined for each origin-destination pair. The location and

allocation decisions impact the shipment transit times which are composed of the travel time during pickup, inter-hub travel and drop off. Furthermore, shipments spend additional time at the hubs, waiting for consolidation (at originating hub) and break bulk (at dispatching hub), represented by the factor β . The model seeks to make location-allocation decisions which are both economical and feasible with respect to the service requirements.

The problem is modeled using the following notation and mathematical formulation.

Indices and Sets:

- C = Set of all cities, $C = \{1, 2, \dots, |C|\}$, indexed by c ,
- N = Set of all nodes, indexed by i or j , $N = \{1, 2, \dots, 2|C|\}$, and
- k, m = Origin and destination hub nodes, respectively.

Parameters:

- p = Number of hubs,
- f_{ij} =
$$\begin{cases} \text{Flows from origin } i \text{ to destination } j, \{i \neq j \text{ and } i, j \leq |C|\} \\ 0, \{i = j \text{ or } i > |C| \text{ or } j > |C|\}, \end{cases}$$
- c_{ij} = Unit transportation cost from node i to node j ,
- t_{ij} = Shipment transit time between node i and node j ,
- F_k = Fixed cost of opening a hub at node k ,
- α = Cost discount factor; $0 \leq \alpha \leq 1$,
- β = Delay factor for consolidation/breakbulk at hubs, $\beta \geq 1$, and
- TW_{ij} = Service time requirements for shipments between i and j .

Decision Variables:

$$X_{ijkm} = \begin{cases} 1 & ; \text{if shipments from } i \text{ to } j \text{ are assigned to hub pair } (k, m) \\ 0 & ; \text{otherwise} \end{cases}$$

$$y_k = \begin{cases} 1 & ; \text{if node } k \text{ is a hub} \\ 0 & ; \text{otherwise} \end{cases}$$

(IHLA) Model Formulation:

$$\text{minimize } \sum_{i,j,k,m \in N} f_{ij} X_{ijkm} (c_{ik} + \alpha c_{km} + c_{mj}) + \sum_{k \in N} F_k y_k \quad (1)$$

subject to:

$$\sum_{k \in N} y_k = p \quad (2)$$

$$\sum_{k,m \in N} X_{ijkm} = 1 \quad , \forall i, j \in N \quad (3)$$

$$\sum_{m \in N} X_{ijkm} \leq y_k \quad , \forall i, j, k \in N \quad (4)$$

$$\sum_{k \in N} X_{ijkm} \leq y_m \quad , \forall i, j, m \in N \quad (5)$$

$$y_c + y_{c+|C|} \leq 1 \quad , \forall c \in C \quad (6)$$

$$X_{ijkm} (t_{ik} + \beta t_{km} + t_{mj}) \leq TW_{ij} \quad , \forall i, j, k, m \in N \quad (7)$$

$$X_{ijkm} \in \{0, 1\} \quad , \forall i, j, k, m \in N$$

$$y_k \in \{0, 1\} \quad , \forall k \in N$$

In the model formulation there are two sets of decision variables. The first set of decision variables X_{ijkm} , are binary and take a value of 1 if the flow from i to j passes through hubs k and m , and 0 otherwise. The second set of variables is for the hub location decision at any node k , given by y_k . This binary variable is set to a value of 1 if node k is selected as a hub; else it is assigned a value of 0. The objective function (1) is to minimize the sum of total transportation

costs for all origin-destination flows and the fixed cost of hub facilities. In the first term of the objective function, c_{ik} represents the unit transportation cost from origin i to originating hub k ; c_{km} represents the unit transportation cost for inter-hub (k, m) travel; and c_{mj} represents the unit transportation cost from dispatching hub m to destination j . The transportation cost for the second leg of the travel (inter-hub) is discounted by α , where $0 \leq \alpha \leq 1$. In the second term of the objective function, F_k is the fixed cost of locating a hub at node k . In this problem, there are two types of hubs: road hubs (represented by nodes $1 \dots |C|$) and intermodal hubs (represented by nodes $|C| + 1 \dots 2|C|$). The fixed costs of an intermodal hub are higher than a road hub because the assets used in an intermodal hub are more extensive in order to handle multiple modes of transportation and require complex planning and control functions.

The constraint (2) ensures that exactly p hubs are used. Constraints (3) ensure that every origin-destination city pair is assigned to a hub pair. The choice of hub pair assignments is based on the best economic option. An origin-destination flow f_{ij} can be assigned to node pair (k, m) only if k and m are hubs. This restriction is enforced with X_{ijkm} allowed to be 1, only if corresponding hub location variables y_k and y_m are assigned a value of 1. Constraints (4) and (5) implement this requirement. Constraints (6) limit one node per city to be selected as a hub. Constraints (7) ensure that all flows are assigned such that total transit time is within the service time window. The terms on the left hand side of constraints (7) correspond to the transit times for the three legs of travel.

LAGRANGIAN LOWER BOUNDS

The Lagrangian relaxation procedure is based on identifying a complicating set of constraints in a model. These constraints are added to the objective function, weighted by associated multipliers, known as Lagrange multipliers. This Lagrangian relaxation problem is solved iteratively while the multipliers are adjusted along the way. The subgradient optimization method (Fisher, 1985) is used to update Lagrange multipliers in each iteration. The procedure stops when it does not improve the solution in a pre-determined number of iterations. There have been several surveys of this technique, such as Fisher (1981) and Geoffrion (1974).

A Lagrangian relaxation is used to find a lower bound for the IHLA problem. An important aspect of the application of Lagrangian relaxation is the judicious choice of constraints to relax. The resulting relaxation should be easier to solve than the original problem and should provide tight lower bounds; otherwise, there is no advantage in considering the relaxation. Note that constraints (4) and (5) are the only constraints in the IHLA formulation that link the allocation variables, X_{ijkm} , with the location variables, y_k and y_m . Therefore, relaxing these constraints leads to a decomposition of the IHLA problem into a hub location and an allocation subproblem. Two sets of Lagrange multipliers, $(\lambda_{ijk}, \mu_{ijm} \geq 0)$ are used for constraints (4) and (5), respectively. These constraints are added to the objective function as the following term:

$$\sum_{i,j,k \in N} \left(\sum_m X_{ijkm} - y_k \right) \lambda_{ijk} \text{ and } \sum_{i,j,m \in N} \left(\sum_k X_{ijkm} - y_m \right) \mu_{ijm}.$$

The resulting problem is called $\text{LR}(\bar{\lambda}, \bar{\mu})$ which can be written as follows;

$$\mathbf{LR}(\bar{\lambda}, \bar{\mu})$$

minimize

$$\sum_{ijkm} X_{ijkm} (f_{ij} C_{ijkm} + \lambda_{ijk} + \mu_{ijm}) + \sum_k \left[F_k - \sum_{ij} (\lambda_{ijk} + \mu_{ijm}) \right] y_k \quad (8)$$

subject to:

$$\sum_y y_k = p \quad (9)$$

$$y_c + y_{c+|C|} \leq 1 \quad ; \forall c \quad (10)$$

$$\sum_{km} X_{ijkm} (t_{ik} + \beta t_{km} + t_{mj}) \leq TW_{ij} \quad ; \forall i, j \quad (11)$$

$$\sum_{k,m} X_{ijkm} = 1 \quad ; \forall i, j \quad (12)$$

$$X_{ijkm} \in \{0, 1\} \quad ; \forall i, j, k, m$$

$$y_k \in \{0, 1\} \quad ; \forall k$$

Note that $C_{ijkm} = c_{ik} + \alpha c_{km} + c_{mj}$ is the transportation cost of sending a shipment from origin i to destination j , through hub pair (k, m) .

Let Z and $Z_{LR}(\bar{\lambda}, \bar{\mu})$ be the optimal values of P and $LR(\bar{\lambda}, \bar{\mu})$, respectively, where P is the IHLA model formulation. From the theory of Lagrangian duality, it is known that $Z \geq Z_{LR}(\bar{\lambda}, \bar{\mu})$. The formulation of $LR(\bar{\lambda}, \bar{\mu})$ is separable into a hub location problem and an allocation problem. Note that the hub location variables (y_k) and the assignment variables (X_{ijkm}) are no longer related through the constraint set of $LR(\bar{\lambda}, \bar{\mu})$.

The hub location subproblem is stated as follows:

HL($\bar{\lambda}, \bar{\mu}$)

minimize

$$\sum_{k \in N} \left[F_k - \sum_{i, j \in N} (\lambda_{ijk} + \mu_{ijk}) \right] y_k \quad (13)$$

subject to:

$$\sum_{y \in N} y_k = p \quad (14)$$

$$y_c + y_{c+|C|} \leq 1, \forall c. \quad (15)$$

The allocation subproblem is stated as follows:

HA($\bar{\lambda}, \bar{\mu}$)

minimize

$$\sum_{i, j, k, m \in N} X_{ijkm} (f_{ij} C_{ijkm} + \lambda_{ijk} + \mu_{ijm}) \quad (16)$$

subject to:

$$\sum_{k, m \in N} X_{ijkm} = 1, \forall i, j \in N \quad (17)$$

$$\sum_{k, m \in N} X_{ijkm} (t_{ik} + \beta t_{km} + t_{mj}) \leq TW_{ij}, \forall i, j \in N. \quad (18)$$

The problem $HL(\bar{\lambda}, \bar{\mu})$ can be solved by selecting the p hubs having the lowest net cost, $F_k - \sum_{i, j} (\lambda_{ijk} + \mu_{ijk})$, whereas $HA(\bar{\lambda}, \bar{\mu})$ is an *All-Pair Shortest Path* problem. Known algorithms, e.g., Floyd-Warshal algorithm (Floyd, 1962), can be used to determine the optimal allocations (paths). The solution procedure involves solving the corresponding Lagrangian dual (LD) problem. That is, to find the Lagrange multipliers that maximize the value of the lower bound given by the $LR(P)$. The dual problem can be stated as follows:

(LD)

$$\max_{\bar{\lambda}, \bar{\mu} \geq 0} LR(\bar{\lambda}, \bar{\mu}). \quad (19)$$

The Lagrangian dual problem is solved using *Subgradient Optimization* method. The master problem is the Lagrangian dual (LD) with location HL($\bar{\lambda}, \bar{\mu}$) and allocation HA($\bar{\lambda}, \bar{\mu}$) subproblems. At each iteration, the algorithm computes the optimal solution of LR($\bar{\lambda}, \bar{\mu}$) and its value based on current multipliers $\bar{\lambda}, \bar{\mu}$ by solving subproblems HL($\bar{\lambda}, \bar{\mu}$) and HA($\bar{\lambda}, \bar{\mu}$) independently.

Table 3: Subgradient Optimization

Step 0:	$n = 0$; Initialize multipliers $\bar{\lambda}_n = \bar{0}$ and $\bar{\mu}_n = \bar{0}$ Set lower bound $LB = 0$ and upper bound $UB = \widehat{Z}$, where \widehat{Z} is any feasible solution, e.g., final solution obtained from tabu search.
Step 1:	Solve HA($\bar{\lambda}_n, \bar{\mu}_n$) and HL($\bar{\lambda}_n, \bar{\mu}_n$). Calculate $Z_{LR}^n = Z_{LR}(\bar{\lambda}_n, \bar{\mu}_n)$. Set $LB = \max\{LB, Z_{LR}^n\}$. If $UB - LB \leq \epsilon$ or $n = MAX_n$ then STOP else go to Step 2
Step 2:	Compute subgradients and step size: Subgradients: $\omega_n^{ijk} = \sum_m X_{ijkm} - y_k$; $\delta_n^{ijm} = \sum_k X_{ijkm} - y_m$ Step size $S_n = t_n \left(\frac{(\widehat{Z} - Z_{LR}^n)}{\ \omega_n\ ^2 + \ \delta_n\ ^2} \right)$ where t_n is a multiplicative factor which is reduced every N iterations, and $\ \omega_n\ ^2 = \sum_{ijk} (\omega_n^{ijk})^2$; $\ \delta_n\ ^2 = \sum_{ijm} (\delta_n^{ijm})^2$
Step 3:	Update each multiplier as $\lambda_{n+1}^{ijk} = \max\{0, \lambda_n^{ijk} + S_n \omega_n^{ijk}\}$; $\mu_{n+1}^{ijm} = \max\{0, \mu_n^{ijm} + S_n \delta_n^{ijm}\}$
Step 4:	Set $n = n + 1$ and go to Step 1

While solving the HA($\bar{\lambda}, \bar{\mu}$) problem, ties may be encountered. Ties are broken based on the total number of shipments passing through each hub in the HL($\bar{\lambda}, \bar{\mu}$) solution, by favoring hubs which already have large flows. The solution of LR($\bar{\lambda}, \bar{\mu}$) is a lower bound for the solution of primal problem (P). Next, subgradients, step size and new multipliers are computed. The process repeats until the lower bound is not improved in some preset number of iterations. The steps included in the subgradient search procedure are summarized in Table 3.

TABU SEARCH METAHEURISTIC

In this section a metaheuristic solution approach for solving the IHLA model is presented. This approach is particularly suitable for solving larger problem instances. Tabu search (Glover, 1989, 1990) is a general iterative metaheuristic for solving combinatorial problems. It has been shown that tabu search can find very good solutions ($< 1\%$ optimality gaps) for the p -hub median problem (Skorin-Kapov et al., 1996) compared to other metaheuristics such as simulated annealing (Abdinnour-Helm, 2001).

Tabu search is composed of two phases, known as intensification phase and diversification phase. In the *intensification phase*, tabu search starts at a randomly generated initial solution. This initial solution is comprised of a set of hub nodes, H . The neighborhood, $N(H)$, of the current solution H is comprised of solutions that can be reached with a swap of a non-hub node with a hub node. This exchange is called a *move*. As a move is made, the move attributes, i.e., entering node and exiting node are recorded in a *tabu list*. The search moves from one solution to the next and stops when a local optimum is reached. After the intensification phase is completed, tabu search enters into the *diversification phase*. In this phase the tabu search restarts the search process from a new starting solution.

The pseudo-code for the tabu search developed for this research is shown in Table 4. The procedure starts with setting the tabu search parameters. Each restart in the diversification phase is called a *run*. The while loop (line 5) executes the diversification phase *max run* number of times. The search starts with a solution in which hub locations are randomly selected. In successive iterations, the set of hub locations change according to a particular move. The optimal solution can be generated from a given set of hub locations by finding the least-cost path for each origin-destination pair (Sohn and Park, 1998). The path must be time feasible according to the service time requirements and pass either through one or two hub nodes.

Using the starting solution as the best solution, tabu search enters the intensification phase implemented in the while loop (line 8). In each iteration, a move is generated by evaluating neighboring solutions. The entering node is selected by evaluating solutions generated by

Table 4: Pseudocode for Tabu Search

1	Tabu Search() {
2	$maxrun \leftarrow p; maxcount \leftarrow 3n; freqlimit \leftarrow n/3; listsize \leftarrow p - 1, nodefrequency \leftarrow 0$
3	$currentsolution = generatestartingsolution()$;
4	$bestsolution = currentsolution$;
5	while($runcount < maxrun$){
6	$count = 0$;
7	$nodefrequency \leftarrow 0$;
8	while($count < maxcount$ AND $iterationnum < maxiter$){
9	$acceptmove = 0$;
10	while($acceptmove = 0$){
11	generate move;
12	if($move \notin tabulist$)
13	$acceptmove = 1$;
14	update tabulist;
15	update currentsolution;
16	elseif($move$ satisfies aspiration criterion)
17	$acceptmove = 1$;
18	update currentsolution;
19	else
20	$acceptmove = 0$;
21	}
22	if($currentsolution < bestsolution$){
23	update bestsolution;
24	$count = 0$;
25	$runcount = 0$;
26	}
27	update nodefreq;
28	$count ++$;
29	}
30	$runcount ++$;
31	if($nodefrequency \geq freqlimit$)
32	exclude nodes from starting solution;
33	$currentsolution = generate starting solution()$;
34	clear tabu list;
35	}
36	report bestsolution;
37	}

pairwise interchanges between hub nodes and non-hub nodes. The entering node is the one that corresponds to a move with the smallest cost solution among all the available moves.

Before a move can be accepted, its tabu status is evaluated (line 12). The tabu criteria used in this research are: (a) a node cannot be added to the solution if it was removed from the solution during the tabu tenure, and (b) a node that is added cannot be removed from the solution for the duration of the tabu tenure. The tenure of the tabu status, called *list size*, is the duration (in terms of the number of iterations) for which a move once added to the tabu list, cannot be reversed. The tabu status of the move can be overruled if the move will generate a solution which passes the *aspiration criterion*. The use of an aspiration criterion is to accept a tabu move if it results in a solution that is the best found so far in the entire search. A move is rejected if it fails the tabu check as well as the aspiration criterion check (line 16). An accepted move is made and the tabu list is updated. As a move is made, the move attributes, i.e., entering node and exiting node are recorded in a tabu list. The nodes enter and leave the tabu list in a first-in-first-out (FIFO) order.

The solution that results from the move is recorded as the current solution. If the current solution is the best found so far, the best solution is updated (line 22). If in any iteration the current solution does not improve the best solution, the count of consecutive non-improving moves is incremented (line 28). The tabu search also keeps a record for each node that counts the number of times that node was included in the solutions generated by the search. This record is referred to as *node frequency*. The intensification phase of the search terminates after “max count” number of non-improving moves.

After the intensification phase is completed, tabu search enters into the *diversification phase*. In this phase the tabu search restarts the search process from a new starting solution using *node frequency* (lines 31-33). The threshold for a node to be disqualified from being a part of the new initial starting solution is called the *freq limit*. This aspect of diversification phase enables the search process to proceed to unexplored areas of the solution space. Tabu search repeats the search for *max run* number of consecutive non-improving runs before terminating the entire search and reporting results.

A major concern in using heuristics to solve large-scale problems is the quality of the solutions obtained. In the absence of optimal solutions to use as benchmarks, lower bounds are needed. These lower bounds are required to be tight for the solution evaluation to be accurate. In this research the maximum size problems (measured in terms of number of nodes, n) solved to optimality is limited to $n \leq 30$. The solutions obtained by tabu search for larger size problems are evaluated using the lower bound from the Lagrangian relaxation procedure described in the previous section.

COMPUTATIONAL STUDY

The purpose of the computational study is to evaluate the performance of metaheuristic solution approach by comparing it with optimal solutions (where available) and lower bounds. The performance is evaluated based on the optimality gap, the distance from lower bound and the computation time. The study of the effectiveness of a heuristic approach requires that realistic test instances are used and that these instances span a range of problem characteristics. Such wide spanning test instances are not available from a single source and may take significant time to gather from multiple real world cases. Furthermore, the proprietary nature of corporate data does not allow most business organizations to share their data publicly. Under these circumstances, randomly generated test instances offer researchers a mechanism to explore the effect of various problem characteristics in the form of controlled test instances. However such test instances should conform as close to reality as possible (Rardin and Uzsoy, 2001). The use of a well designed set of randomly generated problem instances enables the researcher to generate varied instances which encompasses the dynamic environment of the problem domain.

While generating random data for this research, it was important that the generated data conform to the realities of real-world logistic networks. For example, the transportation cost between two arbitrary cities through a road link is typically higher (per unit weight) compared to the rail link. If the transportation costs for both the modes are generated through, say uniform distribution with different parameters, a verification was made which ensured that any randomness does not violate that relationship. This verification was also applied to transit times

Table 5: Data Sets for Computational Study

<i>Nodes</i> (<i>n</i>)	<i>Data Sets</i> (<i>A – E</i>)	<i>Hubs</i> (<i>p</i>)	<i>Discount</i> <i>Factor</i> (α)	<i>Total</i>
10	1-5	{2,4}	{0.5,0.9}	20
20	6-10	{2,4,6}	{0.5,0.9}	30
30	11-15	{2,4,6}	{0.5,0.9}	30
40	16-20	{2,4,6}	{0.5,0.9}	30
50	21-25	{2,4,6}	{0.5,0.9}	30
60	26-30	{2,4,6}	{0.5,0.9}	30
70	31-35	{2,4,6}	{0.5,0.9}	30
80	36-40	{2,4,6}	{0.5,0.9}	30
90	41-45	{2,4,6}	{0.5,0.9}	30
100	46-50	{2,4,6}	{0.5,0.9}	30

Table 6: Parameters for Cost Matrix

<i>Data Sets</i>	ρ	π
A	5	10
B	15	20
C	25	30
D	35	40
E	45	50

of the rail mode which typically exceeds those of the road mode. The fixed costs were computed as $D\nu$ (where $D = \sum_{i,j \in N} f_{ij}$, and ν is a multiplier) for the purpose of experimentation. The value of ν is set to 1 for road hubs and to 1.5 for intermodal hubs.

To test the performance of the metaheuristic, data sets are used that differ in problem size (number of nodes) and problem scale (range of input data values) to avoid misrepresented cases (Rardin and Uzsoy, 2001). The data sets used in this research are listed in Table 5. Each row corresponds to a set of problem instances associated with size n . Recall, that the number of cities in a problem is equal to half of n . For each n there are five data sets, $\{A, B, C, D, E\}$, marked as 1-5, 6-10, etc. For each n and each data set (A through E), four different random samples (replications) of problem parameters (flows, costs and transit times) are generated from the corresponding probability distributions. The road transportation costs are sampled from a $\text{Uniform}(1, \rho)$ and rail transportation costs are sampled from $\text{Uniform}(1, \pi)$. The values of ρ and π for each data set are tabulated in Table 6. The origin-destination flows are generated by sampling from $\text{Uniform}(1, 10)$ distribution. $\text{Uniform}(1, 4)$ distribution is used to generate origin-destination road transit times and a $\text{Uniform}(1, 8)$ distribution to sample origin-destination rail transit times.

Hence for each of the fifty data sets in Table 5, there are six problem instances based on

different combinations of $p \in \{2, 4, 6\}$ and $\alpha \in \{0.5, 0.9\}$. These account for 300 ($= 50 \times 6$) problem instances. For $n = 10$ (5-city case) $p = 6$ cannot be tested since it exceeds the number of cities in the problem. Thus Table 5 represents a total of 290 ($= 300 - 5 \times 2$) problem instances. With four random samples each, a total of 1160 ($= 290 \times 4$) problem instances are tested in the computational study.

RESULTS: BENCH MARKING WITH OPTIMAL SOLUTIONS

The tests were run on a Pentium 4 CPU with 3.20 GHz clock and 4GB RAM. The size of a problem is represented by the number of nodes, n . The problems of size $n = 10, 20, 30$ were solved to optimality using CPLEX 10.0 and OPL Studio version 5.2. The script language of OPL Studio was used to iteratively run the problem instances in an efficient manner. Four independent samples of each data set were generated and solved using CPLEX. The same problems were solved using tabu search. The solutions obtained from tabu search were used as upper bounds in the Lagrangian relaxation procedure. Sub-gradient optimization was used to obtain lower bounds on these problems. The tabu search and lower bound methods are benchmarked with the optimal solutions.

The percent optimality gap is calculated as $(\frac{z_{TS} - z^*}{z^*})$, where z_{TS} is the value of the tabu search solution and z^* is the value of the optimal solution. The percent optimal gaps for the four replications of each problem instance are averaged and recorded as the average percent optimality gap. The average percent optimality gaps obtained by solving the problem instances using tabu search are reported in Table 7. The average of optimality gaps over all test instances is 0.71%. The minimum average gap for tabu search is 0.00% with 73% of the problem instances solved to within 1% of the optimal solution and 99% of the problem instances are solved to within 3% of the optimal solution. The maximum average optimality gap is 3.19%.

Using the solutions obtained above through tabu search, subgradient optimization was then used to compute lower bounds. The percent duality gap (for Lagrangian relaxation) is calculated as $\frac{z^* - LB}{z^*}$, where LB is the value of the lower bound. The percent duality gaps for the four replications of each problem instance are averaged and recorded as the average percent duality

Table 7: Average Percent Optimality Gaps (Tabu Search)

	n	$\alpha = 0.5$			$\alpha = 0.9$		
		$p = 2$	$p = 4$	$p = 6$	$p = 2$	$p = 4$	$p = 6$
Data Set 1	10	0.34	0.59	-	0.17	0.00	-
Data Set 2	10	0.00	0.99	-	0.00	0.89	-
Data Set 3	10	0.00	0.77	-	0.00	0.30	-
Data Set 4	10	0.00	0.00	-	0.00	0.15	-
Data Set 5	10	0.00	0.00	-	0.00	0.00	-
Data Set 6	20	0.00	0.00	0.00	0.01	0.00	0.00
Data Set 7	20	0.00	0.18	0.23	0.00	1.14	1.59
Data Set 8	20	0.10	0.99	2.17	0.77	1.16	1.94
Data Set 9	20	0.17	0.60	0.78	0.42	0.71	0.83
Data Set 10	20	0.19	0.98	1.51	0.00	1.86	2.74
Data Set 11	30	0.00	0.00	0.04	0.00	0.00	0.00
Data Set 12	30	0.50	0.00	0.12	0.11	1.64	1.77
Data Set 13	30	0.32	2.40	1.20	0.14	3.19	2.09
Data Set 14	30	0.39	1.09	2.70	0.12	1.75	2.51
Data Set 15	30	0.51	1.78	2.71	0.12	1.51	2.77

Table 8: Average Percent Duality Gaps (Lagrangian Relaxation)

	n	$\alpha = 0.5$			$\alpha = 0.9$		
		$p = 2$	$p = 4$	$p = 6$	$p = 2$	$p = 4$	$p = 6$
Data Set 1	10	0.41	1.11	-	0.78	1.65	-
Data Set 2	10	1.07	1.38	-	1.14	1.11	-
Data Set 3	10	0.95	0.94	-	1.43	1.68	-
Data Set 4	10	0.93	2.16	-	0.94	1.26	-
Data Set 5	10	2.36	2.11	-	1.43	0.47	-
Data Set 6	20	0.33	1.16	0.50	0.19	1.52	0.73
Data Set 7	20	1.62	1.30	1.32	1.40	2.26	0.64
Data Set 8	20	1.79	1.01	0.91	0.79	0.76	0.37
Data Set 9	20	0.65	1.74	1.47	0.78	2.07	1.54
Data Set 10	20	1.65	1.48	1.69	1.58	0.87	0.61
Data Set 11	30	1.65	0.65	1.04	1.73	1.18	1.80
Data Set 12	30	2.10	1.37	0.96	1.50	0.88	0.78
Data Set 13	30	1.96	0.95	1.42	2.25	1.59	1.08
Data Set 14	30	2.33	1.29	1.08	2.02	1.04	0.72
Data Set 15	30	0.83	0.97	1.47	1.63	1.28	1.49

gap in Table 8. The average of percent duality gaps over all test instances is 1.26%. The minimum average duality gap is 0.19% with 35% of the problem instances solved within 1% of optimal solution and 100% of the problem instances have lower bounds within 3% of optimal solution. The maximum average duality gap is 2.36%.

The running times for CPLEX and tabu search are displayed in Table 9. It can be seen that in each case the running time for CPLEX is increasing exponentially, whereas the running time for tabu search is increasing linearly. This observation is verified by fitting an exponential curve to the CPLEX times (Ordinary least squares fit equation: $y = 0.6785e^{0.0683x}$ with $R^2 = 0.8632$).

RESULTS: BENCHMARKING WITH LOWER BOUND

The smaller problems, $n = \{10, 20, 30\}$, can be solved to optimality, and the results reported

Table 9: Running Times (sec) for CPLEX and Tabu Search

		$\alpha = 0.5$			$\alpha = 0.9$		
		$p = 2$	$p = 4$	$p = 6$	$p = 2$	$p = 4$	$p = 6$
Data Set 1	CPLEX	1.02	1.26	-	1.39	1.52	-
	TS	0.05	0.03	-	0.03	0.02	-
Data Set 2	CPLEX	1.26	1.13	-	1.45	1.27	-
	TS	0.04	0.02	-	0.03	0.03	-
Data Set 3	CPLEX	1.41	1.20	-	1.36	1.37	-
	TS	0.05	0.03	-	0.02	0.02	-
Data Set 4	CPLEX	1.49	1.18	-	1.65	1.42	-
	TS	0.03	0.02	-	0.04	0.02	-
Data Set 5	CPLEX	1.40	1.05	-	1.24	0.97	-
	TS	0.06	0.03	-	0.03	0.03	-
Data Set 6	CPLEX	5.59	6.37	3.25	3.92	8.52	6.88
	TS	0.07	0.17	0.53	0.07	0.20	0.53
Data Set 7	CPLEX	5.84	6.80	8.34	5.40	8.30	10.28
	TS	0.07	0.18	0.59	0.06	0.24	0.63
Data Set 8	CPLEX	11.84	7.27	6.25	8.68	4.56	4.80
	TS	0.08	0.20	0.56	0.06	0.23	0.52
Data Set 9	CPLEX	7.03	7.03	4.81	4.20	5.83	3.67
	TS	0.08	0.31	0.45	0.07	0.26	0.52
Data Set 10	CPLEX	8.27	6.33	4.98	6.83	4.17	3.01
	TS	0.09	0.21	0.47	0.07	0.23	0.42
Data Set 11	CPLEX	50.46	32.74	35.94	46.68	38.86	38.84
	TS	0.23	0.69	2.57	0.17	0.73	2.36
Data Set 12	CPLEX	131.71	79.18	51.86	79.64	84.32	80.98
	TS	0.22	0.77	2.38	0.21	0.95	2.13
Data Set 13	CPLEX	146.97	99.49	88.95	130.54	84.62	66.64
	TS	0.22	0.85	2.42	0.15	0.76	2.23
Data Set 14	CPLEX	151.72	110.06	103.61	117.29	86.66	65.32
	TS	0.23	0.75	2.13	0.20	0.76	2.33
Data Set 15	CPLEX	107.84	108.41	82.98	123.24	104.70	90.37
	TS	0.22	0.62	2.37	0.20	0.89	2.84

Table 10: Summary of Average Percent Solution Gap

<i>Nodes(n)</i>	(p, α)					
	(2, 0.5)	(2, 0.9)	(4, 0.5)	(4, 0.9)	(6, 0.5)	(6, 0.9)
10	1.25	1.20	2.12	1.53	-	-
20	1.33	1.21	1.92	2.51	2.15	2.21
30	2.16	1.97	2.13	2.85	2.58	3.03
40	2.09	1.70	2.55	2.42	2.09	2.60
50	2.60	1.95	2.68	2.95	3.31	4.11
60	2.89	2.30	3.94	3.15	3.29	3.01
70	4.37	3.44	4.70	3.61	3.61	3.39
80	3.77	3.43	5.08	4.34	4.27	3.93
90	4.55	3.88	6.61	5.25	5.54	4.70
100	5.24	4.69	7.18	6.25	6.13	5.31
min	1.25	1.20	1.92	1.53	2.09	2.21
max	5.24	4.69	7.18	6.25	6.13	5.31
average	3.03	2.58	3.89	3.49	3.66	3.59

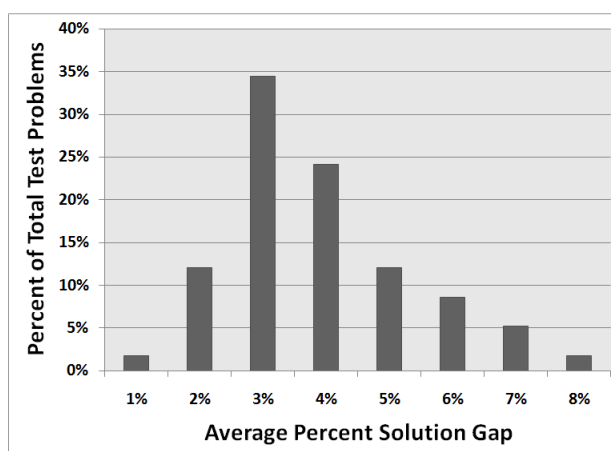


Figure 1: Performance of Tabu Search

above indicate good performance for both tabu search and the lower bound method. However, real world problems can be much larger in size. To evaluate the performance of the tabu search procedure, larger problem instances ($10 \leq n \leq 100$) are tested. These problems are tabulated in Table 5. For each of these problem instances, the percent *solution gap* is defined as the difference in the objective function value of the tabu search solution and the lower bound. The percent solution gap is computed as $\frac{z_{TS} - LB}{LB}$. The percent solution gap for the four replications of each problem instance are averaged and recorded as the average percent solution gap in columns 3 through 8 of Table 11. The last column of Table 11 averages these results over each data set. The values in the last column are also shown graphically in Figure 1.

Table 11 shows that the average percentage solution gaps were typically small. Figure 1 shows that in 48% of the problem instances, tabu search found solutions which were within 3%

Table 11: Average Percent Solution Gap ($10 \leq n \leq 100$)

	n	$\alpha = 0.5$			$\alpha = 0.9$			Avg
		$p = 2$	$p = 4$	$p = 6$	$p = 2$	$p = 4$	$p = 6$	
Data Set 1	10	0.76	1.72	-	0.95	1.68	-	1.28
Data Set 2	10	1.08	2.43	-	1.15	2.05	-	1.68
Data Set 3	10	0.96	2.07	-	1.47	2.02	-	1.63
Data Set 4	10	0.94	2.24	-	0.95	1.43	-	1.39
Data Set 5	10	2.5	2.17	-	1.46	0.48	-	1.65
Data Set 6	20	0.33	1.17	0.51	0.2	1.54	0.74	0.75
Data Set 7	20	1.66	1.51	1.58	1.44	3.51	2.24	1.99
Data Set 8	20	1.96	2.03	3.11	1.58	1.94	2.32	2.16
Data Set 9	20	0.83	2.39	2.28	1.21	2.84	2.4	1.99
Data Set 10	20	1.88	2.51	3.26	1.62	2.75	3.37	2.57
Data Set 11	30	1.69	0.66	1.09	1.78	1.2	1.83	1.38
Data Set 12	30	2.66	1.4	1.09	1.66	2.55	2.56	1.99
Data Set 13	30	2.33	3.4	2.66	2.45	4.85	3.21	3.15
Data Set 14	30	2.8	2.43	3.83	2.18	2.83	3.25	2.89
Data Set 15	30	1.35	2.78	4.25	1.78	2.84	4.32	2.89
Data Set 16	40	1.94	0.58	0.57	1.26	0.75	0.54	0.94
Data Set 17	40	1.57	2.7	2.14	1.35	2.21	2.23	2.03
Data Set 18	40	1.48	2.3	2.35	1.14	2.53	3.16	2.16
Data Set 19	40	1.99	2.49	2	1.92	2.38	3.05	2.31
Data Set 20	40	3.48	4.66	3.41	2.85	4.2	4	3.77
Data Set 21	50	1.19	1.45	0.83	0.85	1.03	1.02	1.06
Data Set 22	50	2.34	2.95	1.91	1.27	2.1	1.66	2.04
Data Set 23	50	3.93	4.11	3.11	3.27	3.83	3.2	3.58
Data Set 24	50	4.11	5.67	4.66	3.27	4.86	4.32	4.48
Data Set 25	50	2.6	2.68	3.31	1.95	2.95	4.11	2.93
Data Set 26	60	2.82	2.19	1.14	2.18	1.4	1.09	1.80
Data Set 27	60	2.61	3.52	3.11	2.07	2.43	3.03	2.80
Data Set 28	60	3.05	3.92	2.12	1.56	2.98	2.33	2.66
Data Set 29	60	2.79	4.77	5.11	2.96	4.43	4.39	4.08
Data Set 30	60	3.17	5.31	4.98	2.73	4.52	4.18	4.15
Data Set 31	70	3.91	2.87	2.02	2.9	2.07	1.69	2.58
Data Set 32	70	3.54	3.25	2.76	2.31	2.37	2.29	2.75
Data Set 33	70	5.62	5.91	4.16	4.86	5.1	4	4.94
Data Set 34	70	5.48	5.78	4.16	4.37	4.56	4.34	4.78
Data Set 35	70	3.29	5.71	4.97	2.78	3.95	4.65	4.23
Data Set 36	80	3.12	2.5	1.78	2.81	2.2	1.52	2.32
Data Set 37	80	2.51	4.74	3.61	1.87	3.43	2.97	3.19
Data Set 38	80	3.67	5.1	4.19	2.89	5.02	4.21	4.18
Data Set 39	80	4.86	5.84	5.54	5.06	4.8	5.53	5.27
Data Set 40	80	4.71	7.22	6.23	4.53	6.29	5.42	5.73
Data Set 41	90	1.76	2.59	2.06	1.61	1.78	1.57	1.90
Data Set 42	90	5.61	5.17	3.9	4.3	4.37	2.88	4.37
Data Set 43	90	3.76	7.5	6.36	3.96	6.54	5.88	5.67
Data Set 44	90	6.74	9.7	7.74	5.56	7.41	6.76	7.32
Data Set 45	90	4.88	8.11	7.62	3.98	6.15	6.42	6.19
Data Set 46	100	4.7	3.27	2.26	3.61	2.63	1.76	3.04
Data Set 47	100	4.75	6.27	5.49	3.28	4.91	4.14	4.81
Data Set 48	100	6.21	7.88	7.14	5.34	6.39	5.89	6.48
Data Set 49	100	4.01	8.91	7.23	3.66	7.64	6.9	6.39
Data Set 50	100	6.51	9.59	8.52	7.58	9.66	7.88	8.29

of the lower bound, while 84% of the problems were solved within 5% of the lower bound. In order to get a more detailed view of the results, Table 10 organizes the results by problem size (n) and scope (p, α). It can be seen that for a given value of n , the overall average percent solution gap is stable over the range of (p, α) problems, e.g., average gap ranges over 1.21% to 2.51% for $n = 20$. Across the problem size (n), the percent solution gap shows a small increasing trend. This trend indicates that there may be a limit in the problem size for good solutions obtained by the metaheuristic. However, the percent solution gaps observed are reasonable, given the large problem size of 100 nodes.

MANAGERIAL INSIGHTS

This section presents the results of a study in which the structure of the intermodal hub location-allocation problem was explored. This study used the Civil Aeronautics Board (CAB) data (O’Kelly et al., 1996) which contains origin-destination flows and air transportation costs for 25 cities in the US air network and can be downloaded from the OR-Library (Beasley, 1990). This data set was modified to generate data for an intermodal network using the following changes. Each city may be serviced by two modes, thus the size of the potential intermodal network is 50 nodes. The cost in the data set was scaled down by 1000 to reflect appropriate unit road transportation costs.

In order to generate the rail transportation costs an experimental parameter, *Cost Ratio* (CR) was used. A value of $CR = 1.0$ implies equal road and rail unit transportation costs. A higher cost ratio e.g., $CR = 1.1$, implies that the unit road transportation cost is 10% higher than the unit rail transportation cost. The unit rail transportation costs were generated by multiplying the unit road transportation costs by 0.9 for $CR = 1.1$, by 0.8 for $CR = 1.2$ and by 0.7 for $CR = 1.3$. The origin-destination transit time data (not given in the CAB data set) was sampled from a Uniform (1,4) distribution for the road transportation and from a Uniform(1,8) distribution for the rail transportation. This range of transit times were used for computational purposes to reflect the average transit times for the road networks which are shorter than the rail networks.

Table 12: Experimental Design

<i>Factors</i>	<i>Levels</i>
α	{0.5, 0.7, 0.9}
CR	{1.0, 0.9, 0.8, 0.7}
FC	{\\$200K, \\$400K, \\$600K}
MC	{5%, 10%, ..., 40%} of FC
TW	{8, 10, 12, 14, 16}

Table 13: Percentage Solution Gaps

<i>Range</i>	<i>Frequency</i>
1.00% to 1.99%	66%
2.00% to 2.99%	31%
3.00% to 3.99%	3%

This study investigates the effect of the following factors on the structure of the intermodal hub network: inter-hub discount (α), cost ratio (CR), fixed cost (FC), modal connectivity cost (MC) and service time requirement (TW). The structure of the network is characterized by the type of hubs (OTR/IM), mode usage, transportation costs and total network costs. In this computational study, a full factorial design was used in which the factors were set at different levels as shown in Table 12. A full factorial design of the above mentioned factors resulted in 1440 problem instances. Each problem instance was solved multiple times using tabu search by varying the parameter $p \in \{2, 3, \dots, 8\}$, where p is the number of hubs in the network. Among the solutions generated by different values of p , the solution with the lowest network cost and the corresponding p (*Num-of-Hubs*) was reported.

In addition to recording the best number of hubs (*Num-of-Hubs*) for each problem instance, the following values were also recorded: network costs, which records the objective function value; total transportation costs (TXC), which includes pickup, inter-hub transfer and dropoff costs; total number of intermodal hubs in the network (*IM-Hubs*); total flows that use intermodal shipments between hubs (*IM-Flows*); total flows that use single-hub shipments (*S-Flows*); and total flows that ship using two hubs (*Hub-to-Hub Flows*). The solutions obtained by tabu search are compared with the lower bound. Table 13 shows that the solution gaps are very small. In 66% of the experiments, tabu search found solutions between 1.00% and 1.99% of the corresponding lower bound for each problem instance. In the remaining 34% of the experiments, tabu search solutions are less than 4.00% from the lower bound. The average percent

solution gap is 1.92% with a minimum gap of 1.16% and a maximum gap of 3.88%. These results suggest that the solutions used in this study are close to optimal.

RESULTS AND DISCUSSION

The first issue investigated in this study is the effect of fixed cost on the intermodal network. The fixed cost plays a significant role in the structure of a hub network as there may exist a tradeoff between the number of hubs in a network and the fixed cost of operating the hubs. This tradeoff impacts the structure and the transportation costs of a hub network.

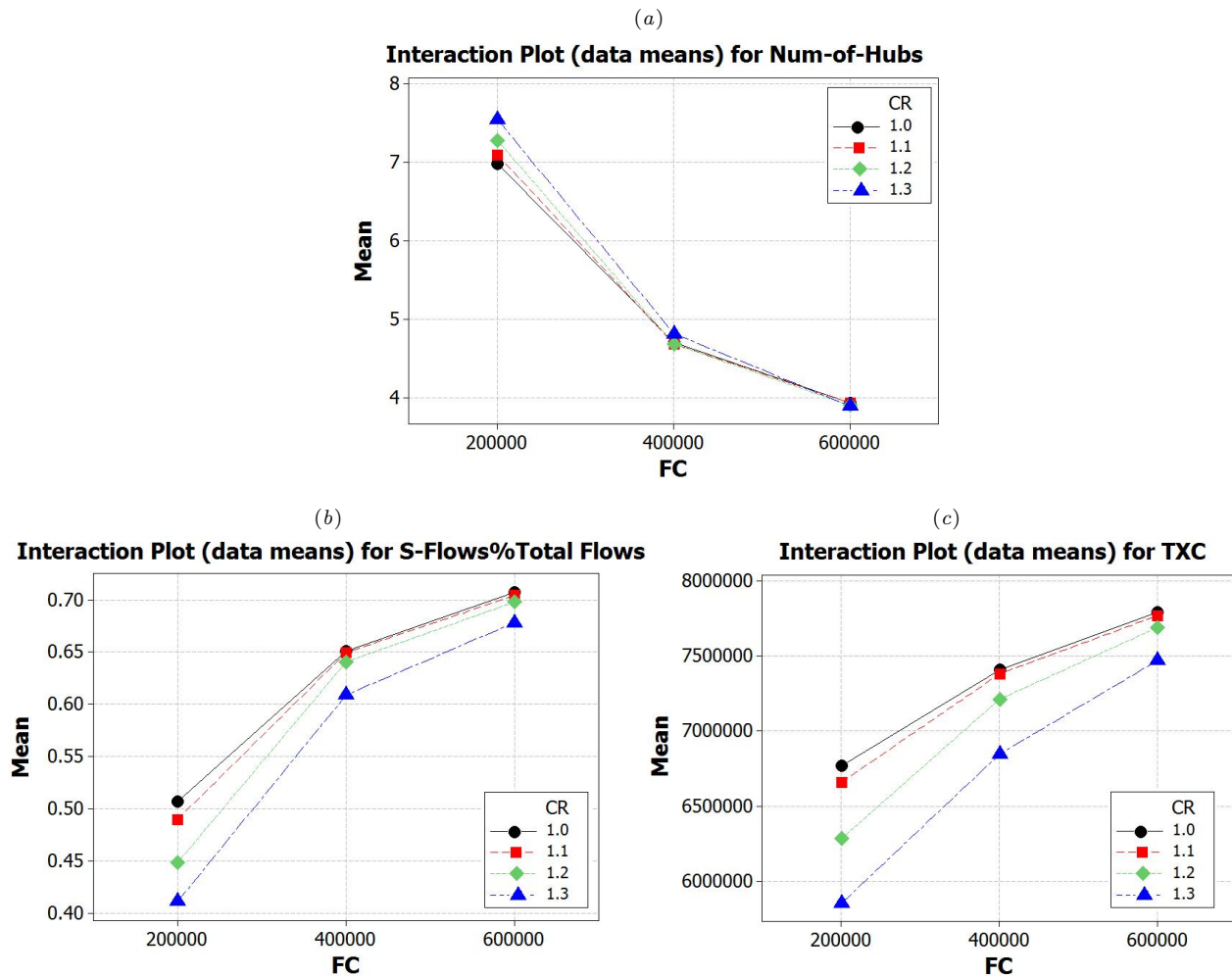


Figure 2: Fixed Cost Plots

The direct effect of higher fixed costs (FC) is a reduction in the number of hubs in a network. This effect is seen in Figure 2a where fewer hubs are used in the network when FC increases from \$200,000 to \$600,000. It is interesting to note that at a high FC , the *Num-of-Hubs* in both OTR ($CR = 1.0$) and IM ($CR > 1.0$) networks converge. Because of the reduced number of hubs in the network when FC increases from \$200,000 to \$600,000, more flows use single-hub shipments (see Figure 2b), e.g., the percentage of the total flows using single hub shipments increases from 42% to 67% (for $CR=1.3$). Since the number of hub-to-hub shipments are reduced, the benefit of lower IM rates is not fully utilized, which results in an increase of transportation costs (see Figure 2c). Figure 2c also shows that the total transportation costs of the intermodal networks ($CR > 1.0$) are smaller compared to an OTR network ($CR = 1.0$), e.g., the total transportation costs of an OTR network (for $FC=\$200,000$) is 15.7% higher compared to the total transportation costs of an IM ($CR=1.3$) network.

An interesting insight gained by these results is that a higher fixed cost of operating a hub increases the transportation costs of a network. Such a condition exists since a network has fewer hubs when the fixed cost of operating a hub is high. In a hub network with fewer hubs, the service regions tend to be larger. This means that more origin-destination pairs are assigned to the same hub, thereby increasing the number of single hub shipments. Such single-hub flows do not allow for the use of intermodal shipments that have a smaller unit transportation cost. This causes the total transportation costs of the network to increase.

The next issue investigated in this study is the effect of modal connectivity cost on the intermodal network. The modal connectivity costs provide a tradeoff between the cost advantage of intermodal transportation and the cost of providing intermodal service. Figure 3a shows the cost advantage of an intermodal (IM) network. Compared to an over-the-road (OTR) network (represented by $CR=1.0$), the total network costs of the intermodal networks (represented by $CR > 1.0$) are smaller. As the cost ratio increases the total network costs decrease. This cost advantage decreases as the cost of intermodal service, i.e., *modal connectivity costs* (MC), increases. This cost advantage is larger for intermodal networks in which the CR is larger.

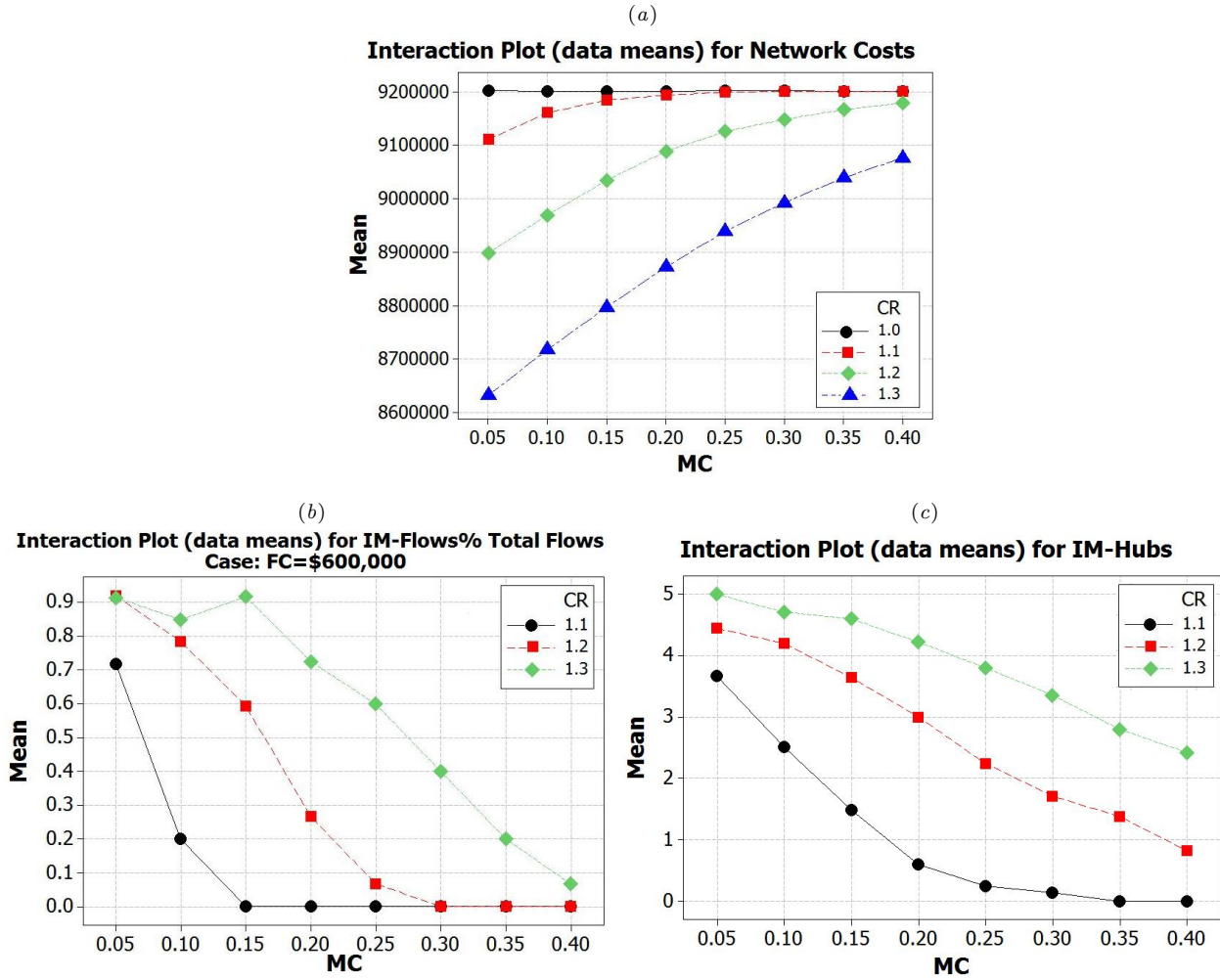


Figure 3: Modal Connectivity Plots

The modal connectivity costs affect the use of intermodal shipments. Figure 3b shows the changes in IM flows when MC changes for the case of $FC = \$600,000$. When the fixed cost of the intermodal service increases, the volume of intermodal shipments decreases. In Figure 3b, it is seen that for the case of $CR=1.1$, the IM flows reduce from 72% of the total flows to 0% as the MC increase from 5% to 15%. The value of MC at which there are no intermodal flows in the network is defined as the “threshold value”. This value refers to the modal connectivity costs at which the costs of providing intermodal services are higher than the cost advantage derived from the use of intermodal transportation. This means that for ($MC \geq Threshold\ value$), an OTR network has a lower total network cost than an IM network.

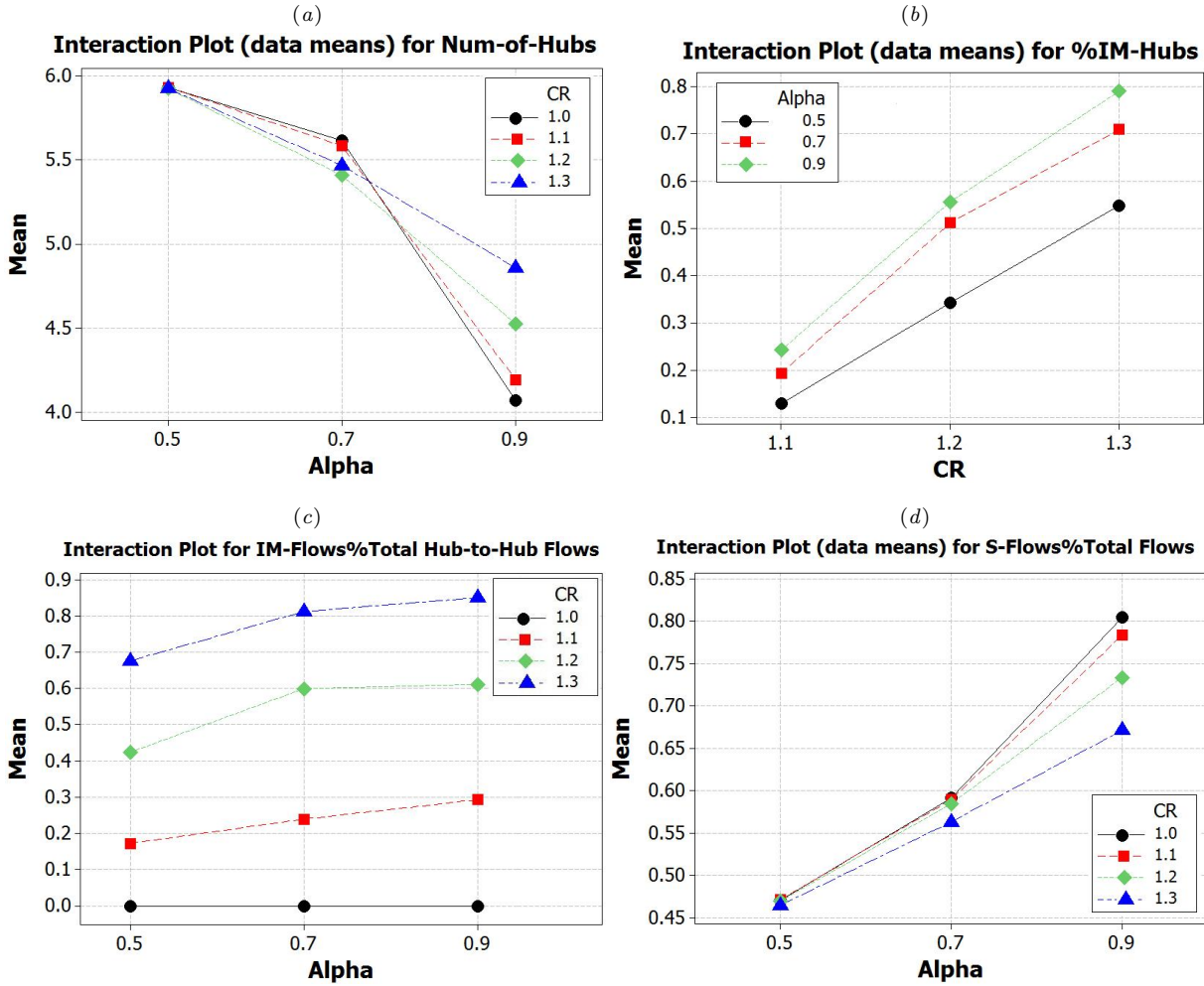


Figure 4: Discount Factor Plots

The *threshold level* is higher for the IM networks in which the difference between the road and rail transportation costs is bigger, i.e., higher CR value. As seen in Figure 3b, the threshold value is reached at $MC=15\%$ for $CR = 1.1$, at $MC=30\%$ for $CR = 1.2$ and at $MC \geq 40\%$ for $CR = 1.3$. Similar results are also observed for other fixed costs used in the study. The modal connectivity costs also affect the number of intermodal hubs in a network. Figure 3c shows that as MC increases, the number of intermodal hubs in the network decreases. This effect diminishes for larger CR .

The above mentioned results show that the cost advantage derived from the use of intermodal shipments depends on the modal connectivity cost (MC). The intermodal networks provide the most cost advantage when FC and MC are low. These results also suggest that MC affects

the number of intermodal hubs in the network as well as the use of intermodal shipments. In a network in which the road and rail transportation rates differ more, i.e., a higher CR , the effect of MC on the intermodal shipments and the number of intermodal hubs is less than in a network with a smaller cost ratio.

The next issue investigated in this study is the effect of inter-hub discount factor on the intermodal network. The hub networks leverage economies of scale which result in reduced transportation costs for shipments traveling between the hubs. This reduction in transportation costs is modeled through a multiplier called the *discount factor* (α). A discount factor (α) of 0.9 means that the transportation costs between hub cities are discounted by $(1 - \alpha) = 0.1$ or 10%. In this study three levels of the discount factor were used, i.e., 0.5, 0.7 and 0.9.

Figure 4a shows that when transportation costs between the hubs are discounted more (smaller α), the network should use more hubs. The number of hubs in the network decreases as the discount is reduced (larger α), e.g., in Figure 4a, the average number of hubs in the network drops from 5.92 to 4.86 for the IM ($CR = 1.3$) network and to 4.06 for the OTR ($CD = 1.0$) network. This result suggests that the discount factor has a greater effect on the number of hubs in an OTR network than it does in an IM network.

The type of hubs in a network is also affected by the scale of the cost benefit of intermodal shipments. Figure 4b shows the effect of CR on the number of IM hubs for different intermodal networks ($CR=1.1, 1.2$ and 1.3). This plot uses $\%IM-Hubs$ to report the average proportion of hubs in each case. It is seen that the proportion of IM hubs increases significantly with increasing CR , e.g., in the case of $\alpha = 0.9$, $\%IM-Hubs$ increases from 25% to 80% as CR changes from 1.1 to 1.3. This effect is smaller when the hub-to-hub shipments are discounted more, e.g., for the same case as discussed above, $\%IM-Hubs$ increases from 13% to 55% for a higher discount factor, $\alpha = 0.5$. Figure 4b further shows that the discount factor affects the number of IM hubs less when CR is lower, e.g., the $\%IM-Hubs$ reduces from 0.80 to 0.55 for $CR = 1.3$ when α changes from 0.9 to 0.5, whereas $\%IM-Hubs$ reduces from 0.24 to 0.13 for a change in α from 0.9 to 0.5.

The higher proportion of IM hubs leads to more use of intermodal shipments between the

hubs (see Figure 4c). It is seen that the proportion of flows using IM shipments increases in an intermodal network with higher CR , e.g., at $\alpha = 0.9$ the proportion of flows using IM shipments increases from 30% to 60% between the IM network with $CR = 1.1$ and $CR = 1.2$, respectively. Figure 4c also shows that when the discount for hub-to-hub shipments is high, more flows tend to use OTR shipments. In this figure, the proportion of hub-to-hub flows using IM shipments decreases by as much as 20% (for $CR = 1.1$) when the discount increases from 30% ($\alpha = 0.7$) to 50% ($\alpha = 0.5$). This result suggests that even in networks which have high intermodal usage, a significant volume of flows may still use OTR shipments.

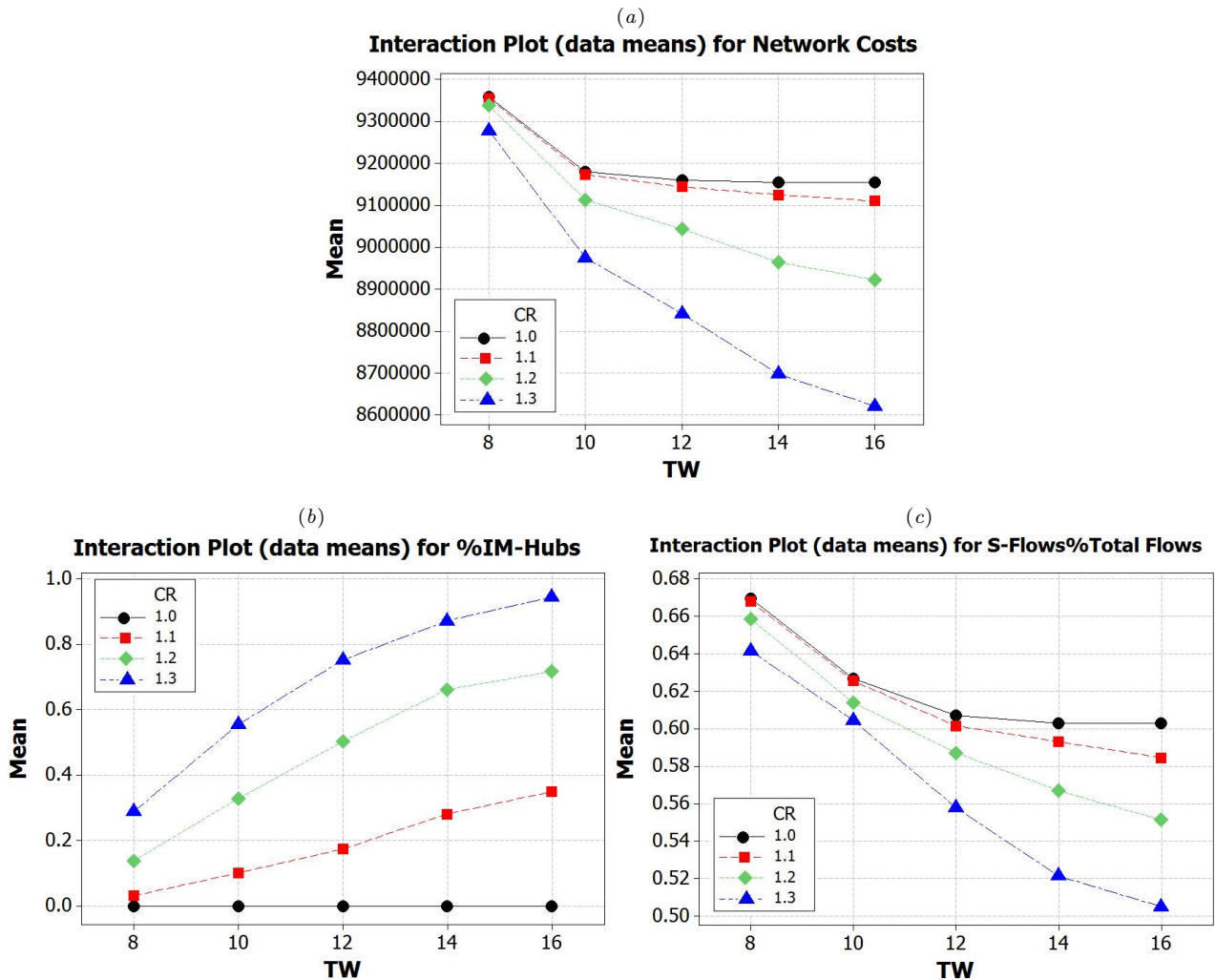


Figure 5: Service Time Plots

Figure 4*d* shows the effect of the discount factor (α) on the use of single hub shipments. When the inter-hub discount is small, a higher proportion of the flows use single hub shipments. Note that this proportion is smaller in the IM networks, e.g., in Figure 4*d*, the proportion of total flows using single hub shipments, is 66% for the IM ($CR = 1.3$) network compared to 80% for the OTR ($CR = 1.0$) network. However this difference is negligible at $\alpha = 0.5$.

These results suggest that higher economies of scale in a hub network encourage the use of OTR shipments over intermodal shipments. Although for lower discounts ($1 - \alpha$), more flows use intermodal shipments, but as the discount increases, more and more flows shift to road shipments. This insight suggests that the increasing economies of scale offset the cost benefit of intermodal shipments.

Another issue investigated in this study is the effect of service time requirements on the intermodal network. The total shipment time from the origin to its destination depends on the transit time as well as β , which accounts for the time from consolidation and break-bulk delays at the hubs. A feasible network design ensures that the total shipment times are less than the service time requirements, see constraint (7). In this study only TW was used to study the impact of service time requirements because it has a similar effect as β . More specifically, increasing the value of β has a similar effect as reducing the service time requirements.

The study used different service time values, $TW = \{8, 10, 12, 14, 16\}$. Note that the problem is infeasible for $TW < 8$. The results show that the impact of lower intermodal transportation rates on total network costs diminishes as shorter service time requirements are imposed (see Figure 5*a*). As the service time requirements gets shorter, the longer transit times of rail shipments do not satisfy the service times allowed. Under such conditions, the shipments switch to the faster road transportation. This is reflected in the decreasing number of IM hubs in the network as TW is reduced (see Figure 5*b*). This figure shows that the proportion of IM hubs falls from 70% to 17% for an IM ($CR = 1.2$) network as TW decreases from 16 to 8 time units. The shorter service time requirement also affects the number of hub-to-hub shipments. As shown in Figure 5*c*, shorter service time requirements cause an increase in the single hub shipments, e.g., 51% of the total flows in an IM ($CR = 1.3$) network use single-hub shipments

whereas this percentage increases to 64% when the service time requirements are reduced to 8 time units. Similarly, an increase in the value of β increases the total shipment time for the hub-to-hub routes thereby shifting some shipments to single-hub routes. Such a shift also increases the total network costs.

CONCLUSIONS

This research explored the impact of using intermodal shipments within the context of a hub logistics network. The contributions of this research lies in the development of a modeling framework, a fast and accurate solution approach and a procedure to compute tight lower bounds. This research developed a modeling framework which incorporates the fixed cost of operating a hub, the cost of providing intermodal services and service time requirements in a road-rail intermodal network. This research also presented a metaheuristic (tabu search) solution approach and developed Lagrangian lower bound. The computational study demonstrated that the solution approach finds high quality solutions (average optimality gap of 0.71%) in a reasonable length of time. This research also showed that the Lagrangian lower bounds for this problem are very tight (average duality gap of 1.26%). Large test problems (up to 100 nodes) were used to demonstrate the performance of the tabu search metaheuristic. The solutions obtained by tabu search were compared with the corresponding lower bounds, and the results showed that for a large number (84%) of these problems, the solution approach yielded solutions which were within 5% of the lower bound.

The model developed for this problem was used in a study to gain relevant managerial insights into the intermodal logistics networks. This research highlighted the impact of the cost of providing the intermodal services (modal connectivity cost) at a hub. The results showed that the modal connectivity cost not only affects the number of intermodal hubs in the network but also affects the use of intermodal shipments. This research identified a threshold level for the modal connectivity costs beyond which the cost benefits of intermodal shipments are outweighed by the costs of providing intermodal services at the hubs. This research also showed that the difference in road and rail transportation rates, called cost ratio, affects the structure of

the intermodal network. The cost ratio impacts the interplay between fixed costs, hub discount factor and service time requirements and their effect on the resulting network structure. The results also showed that the transportation costs are affected by the service time requirement. Under shorter service time requirements, some of the lowest cost network paths are infeasible. If the service time requirement is too short, the cost benefit of intermodal transportation cannot be realized and the network structure shifts to an over-the-road network.

This research can be extended by investigating additional issues such as the effect of different network coverage of each mode of transportation, the effect of both the number and location of modal transfer points, the effect of limited hub capacity and the use of actual transportation rate structures for different modes of transportation. Furthermore, a single transportation mode can be differentiated in terms of capacity such as vans, small trucks, 20 feet trailers and 48 feet trailers for road freight and in terms of speed such as regular service and premium service for intermodal rail. Such options can be handled as separate modes for comparison and analysis. Another direction for this research is the use of additional modes, such as water or air, which would extend the scope of this research to the domain of global logistics.

REFERENCES

- Abdinnour-Helm, S., 2001. Using simulated annealing to solve the p-hub median problem. *International Journal of Physical Distribution and Logistics Management* 31 (3), 203–220.
- Abdinnour-Helm, S., Venkataramanan, M., 1998. Solution approaches to hub location problems. *Annals of Operations Research* 78 (0), 31–50.
- Alumur, S., Kara, B. Y., 2008. Network hub location problems: The state of the art. *European Journal of Operational Research* 190 (1), 1–21.
- Arnold, P., Peeters, D., Thomas, I., 2004. Modelling a rail/road intermodal transportation system. *Transportation Research Part E: Logistics and Transportation Review* 40 (3), 255–270.
- Arnold, P., Peeters, D., Thomas, I., Marchand, H., 2001. For an optimum location of the intermodal centers of transshipment : Formulation and extensions. *The Canadian Geographer* 45 (3), 427–436.
- Aykin, T., 1995. Networking policies for Hub-and-Spoke systems with applications to the air transportation systems. *Transportation Science* 29 (3), 201–221.

- Beasley, J. E., 1990. Or-library: Distributing test problems by electronic mail. *The Journal of the Operational Research Society* 41 (11), 1069–1072.
- Boland, N., Krishnamoorthy, M., Ernst, A. T., Ebery, J., 2004. Preprocessing and cutting for multiple allocation hub location problems. *European Journal of Operational Research* 155 (3), 638–653.
- Bontekoning, Y. M., Macharis, C., Trip, J. J., 2004. Is a new applied transportation research field emerging? A review of intermodal railtruck freight. *Transportation Research Part A: Policy and Practice* 38 (1), 1–34.
- Bookbinder, J. H., Fox, N. S., 1998. Intermodal routing of Canada-Mexico shipments under NAFTA. *Transportation Research Part E: Logistics and Transportation Review* 34 (4), 289–303.
- Campbell, J. F., 1992. Location and allocation for distribution systems with transshipments and transportation economies of scale. *Annals of Operations Research* 40 (1), 1572–9338.
- Campbell, J. F., 1994. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research* 72 (2), 387–405.
- Campbell, J. F., 1996. Hub locations and p-hub median problem. *Operations Research* 44 (6), 923–935.
- Campbell, J. F., 2009. Hub location for time definite transportation. *Computers & Operations Research* 36 (12), 3107–3116.
- Carello, G., Croce, F. D., Ghirardi, M., Tadei, R., 2004. Solving the hub location problem in telecommunication network design: A local search approach. *Networks* 44 (2), 94–105.
- Crainic, T. G., Kim, K. H., Barnhart, C., Laporte, G., 2007. Chapter 8 Intermodal Transportation. Vol. 14. Elsevier, pp. 467–537.
- Cunha, C. B., Silva, M. R., 2007. A genetic algorithm for the problem of configuring a hub-and-spoke network for a LTL trucking company in Brazil. *European Journal of Operational Research* 179 (3), 747–758.
- Ernst, A. T., Krishnamoorthy, M., 1996. Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location Science* 4 (3), 139–154.
- Ernst, A. T., Krishnamoorthy, M., 1998. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *European Journal of Operational Research* 104 (1), 100–112.
- Fisher, M. L., 1981. The lagrangian relaxation method for solving integer programming problems. *Management Science* 27 (1), 1–18.
- Fisher, M. L., 1985. An applications oriented guide to lagrangian relaxation. *Interfaces* 15 (2), 10–21.
- Floyd, R., 1962. Algorithm 97: Shortest path. *Commun. ACM* 5 (6), 345.

- Geoffrion, A. M., 1974. Lagrangian relaxation for integer programming. *Mathematical Programming Study* 2, 82–114.
- Glover, F., 1989. Tabu search - part I. *ORSA Journal on Computing* 1 (3), 190–206.
- Glover, F., 1990. Tabu search - part II. *ORSA Journal on Computing* 2 (1), 4–32.
- Groothedde, B., Ruijgrok, C., Tavasszy, L., 2005. Towards collaborative, intermodal hub networks: A case study in the fast moving consumer goods market. *Transportation Research Part E: Logistics and Transportation Review* 41 (6), 567–583.
- IANA, Aug. 2008. Domestic Intermodal Gains Accelerate. Intermodal Association of North America.
- Klincewicz, J. G., 1991. Heuristics for the p-hub location problem. *European Journal of Operational Research* 53 (1), 25–37.
- Klincewicz, J. G., 1992. Avoiding local optima in the p-hub location problem using tabu search and GRASP. *Annals of Operations Research* 40 (1), 283–302.
- Kratka, J., Stanimirovic, Z., Tomic, D., Filipovic, V., 2007. Two genetic algorithms for solving the uncapacitated single allocation p-hub median problem. *European Journal of Operational Research* 182 (1), 15–28.
- Limbourg, S., Jourquin, B., 2009. Optimal rail-road container terminal locations on the european network. *Transportation Research (E)* 45 (4), 551–563.
- Lozano, A., Storchi, G., Mar. 2001. Shortest viable path algorithm in multimodal networks. *Transportation Research Part A: Policy and Practice* 35 (3), 225–241.
- Macharis, C., Bontekoning, Y. M., 2004. Opportunities for OR in intermodal freight transport research: A review. *European Journal of Operational Research* 153 (2), 400–416.
- McCalla, R. J., 1999. Global change, local pain: intermodal seaport terminals and their service areas. *Journal of Transport Geography* 7 (4), 247–254.
- O’Kelly, M. E., 1986. The location of interacting hub facilities. *Transportation Science* 20 (2), 92–106.
- 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., Bryan, D., Skorin-Kapov, D., Skorin-Kapov, J., 1996. Hub network design with single and multiple allocation: A computational study. *Location Science* 4 (3), 125–138.
- Racunica, I., Wynter, L., 2005. Optimal location of intermodal freight hubs. *Transportation Research Part B: Methodological* 39 (5), 453–477.
- Rardin, R. L., Uzsoy, R., 2001. Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics* 7, 261–304.

- Rodriguez, V., Alvarez, M. J., Barcos, L., 2007. Hub location under capacity constraints. *Transportation Research Part E: Logistics and Transportation Review* 43, 495–505.
- Schwartz, M., 1992. *J.B. Hunt: The Long Haul to Success*. The University of Arkansas Press.
- Skorin-Kapov, D., Skorin-Kapov, J., 1994. On tabu search for the location of interacting hub facilities. *European Journal of Operational Research* 73 (3), 502–509.
- Skorin-Kapov, D., Skorin-Kapov, J., O’Kelly, M. E., 1996. Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research* 94 (3), 582–593.
- Slack, B., 1990. Intermodal transportation in north america and the development of inland load centers. *Professional Geographer* 42 (1), 72–83.
- Slack, B., 2001. Intermodal transportation. In: Brewer, A., Button, K. J., Hensher, D. A. (Eds.), *Handbook of Logistics and Supply Chain Management*, 1st Edition. Pergamon, p. 700.
- Sohn, J., Park, S., 1998. Efficient solution procedure and reduced size formulations for p-hub location problems. *European Journal of Operational Research* 108 (1), 118–126.
- Topcuoglu, H., Corut, F., Ermis, M., Yilmaz, G., Apr. 2005. Solving the uncapacitated hub location problem using genetic algorithms. *Computers & Operations Research* 32 (4), 967–984.
- Ziliaskopoulos, A., Wardell, W., 2000. An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. *European Journal of Operational Research* 125 (3), 486–502.

ARTICLE 2: INTERPLAY OF FINANCIAL, OPERATIONAL AND SERVICE ISSUES IN INTERMODAL LOGISTICS

The competitive marketplace requires an efficient and effective logistics strategy. Such a strategy aims at managing shipments across geographically dispersed supply and demand areas within a reasonable time and at a competitive cost. To fulfill those needs, intermodal logistics networks offer a viable option (Gooley, 1997; Arnold et al., 2004). In an intermodal network, a shipment uses multiple modes of transportation in its journey from the origin to the destination in a seamless manner through the use of intermodal containers (Crainic et al., 2007; Slack, 2001, 1990). However, the design and the management of such a logistics network is restricted by the existing transportation infrastructure, location of modal transfer points and logistics cost structure (Warsing et al., 2001).

The growth of intermodal logistics initially resulted from the globalization of the marketplace. This globalization was facilitated by the regional and global trade agreements such as GATT and NAFTA (McCalla, 1999). The use of intermodal shipments has also been on the rise in the US domestic freight market. Even under current lagging economic conditions, domestic intermodal usage has been steadily increasing (IANA, 2008). The sustained demand for intermodal shipments has paid high dividends for logistics companies, such as J.B. Hunt, Schneider National and Swift, which offer domestic intermodal services to their customers (Schwartz, 1992).

Intermodal logistics uses the benefits of its constituent transportation modes to deliver a competitive service compared to the traditional over-the-road (OTR) networks (Macharis and Bontekoning, 2004). The integration of modes within an intermodal logistics network provides the means to move shipments between the origins and the destinations that is both economical and operationally viable (Slack, 1990). The competitiveness of the intermodal networks is not only based on the lower cost, but also on the shipment capabilities. These capabilities

are characterized by weight, volume, access and transit time performance. The limitations of the OTR network in terms of shipment weight and volume is overcome by the weight/volume capability of rail; limitations in terms of access to rail is overcome by using road for pickup and delivery of an intermodal shipment (i.e., drayage).

Intermodal logistics offer opportunities that go beyond an alternative for truckload (TL) shipments. Intermodal services are designed for less-than-truckload (LTL) shipments as well. Services such as EMP Domestic Container Program offered by the rail carriers Union Pacific and Norfolk Southern provide LTL carriers an opportunity to integrate the intermodal component to their LTL/package handling operations. The prioritization of the intermodal trains on the rail tracks under programs such as the Blue Streak Service of Union Pacific and the Premium Container Service of BNSF Railway offer competitive transit times to many major markets in direct competition to the long haul (more than 500 miles per day) OTR services. The intermodal services are already being used by the package carrier United Parcel Service (UPS). UPS, which started as a package carrier using air transportation, has added road and rail shipments to its operations. Their use of road/rail/air has expanded to an extent that UPS is now the largest U.S. customer of intermodal services.

Over the years, a hub based network structure has evolved for moving intermodal shipments. The emergence of hub-based intermodal networks indicates that economy of scale is the principal force behind their use (Slack, 1990). Because an intermodal network is an extension of its respective single mode networks, it is natural that the hub network has emerged as the most suitable network structure for intermodal logistics (Bookbinder and Fox, 1998).

In a hub-and-spoke logistics network, cost savings are realized due to the concentration of flows between the hubs. This concentration of flows creates economies of scale and density. Economies of scale are realized through the consolidation of less-than-truckload shipments into containerized shipments which reduce the unit transportation cost. The economies of density are realized by high load factors for the road/rail/air shipments over fixed distances. The economies of scale create a non-linear rate structure where the unit transportation cost is a non-increasing function of the volume shipped. Intermodal rail and air freight have cost structures which are

based on a container-load and are inherently different from the road transportation costs.

This research focuses on the design and management of a hub-based logistics network for a logistics service provider that serves a multiregional customer base. Such a service provider manages shipments between origins and destinations through the use of different modes of transportation such as road, intermodal rail, and air. At a strategic planning level, such a logistics service provider must develop general, long-term policies about the location of logistics hubs for shipment handling (consolidation, staging, and break-bulk), the routing policies for shipments, the choice of transportation modes and associated freight flow levels, the planning of resource requirements, and the service design.

The hub network design for such an organization involves identifying the number and location of logistics hubs and the assignment of shipments that are served by each hub. In the context of this research, a logistics hub is a shipment handling facility that receives, consolidates or breaks down, and dispatches shipments. Such a logistics hub may have local access to road, intermodal-rail, or air freight terminals that may be operated by other carriers or service providers. The choice of mode (road, rail, or air) for moving a shipment between hubs is determined by the tradeoff between the service requirements quoted to the customer and the transportation cost associated with each mode. The intermodal transfer of shipments is managed through the services offered by the carriers that serve a specific mode. The movement of shipments over the road may be done through private assets owned by the logistics service provider as in the case of a firm such as J.B. Hunt or through the use of a third-party carrier as in the case of a firm such as Landstar Logistics.

The analysis conducted in this research is focused on the financial, operational and service aspects of intermodal logistics networks. For a large number of companies, the design of the logistics network is based on the road freight. This research seeks to identify the economic benefits gained by using intermodal shipments in a logistics network. Any economic benefit gained by using intermodal shipments may also alter the structure of the optimal logistics network. This research analyzes the changes in the hub locations and service area allocations due to the use of intermodal shipments. The cost of providing intermodal services, i.e., modal

connectivity cost, at a logistics hub acts as a tradeoff for any potential economic benefits. The interplay between the modal connectivity cost and the modal choice is analyzed.

The performance of a logistics network can be gauged on different metrics such as cost, service frequency, service time, delivery reliability, flexibility and safety (Beuthe and Bouffieux, 2008). In this research, total network costs and service time are used as the performance criteria as previously done by Crainic and Laporte (1997). More specifically, the model used in this research minimizes total network costs while satisfying maximum service time requirements. Using this model framework, the research investigates the effect of service time requirements on the network flows and the hub network design.

To conduct the above mentioned analysis, this research develops a novel modeling framework and a solution methodology. Extending the multiple-allocation p-hub median approach to the intermodal logistics domain, the modeling framework accommodates the operational structure of individual modes of transportation, the effect of shipment consolidation at hubs on transportation costs, the interaction between modes, transit time delays and service time requirements. It also uses the fixed cost of locating the intermodal hubs and the modal connectivity cost as a tradeoff between opening new hub facilities and reducing the total transportation costs. The model compares the intermodal option with an OTR (over-the-road) option when selecting modal connectivity at the hubs. The modal connectivity cost is specified by the type of modes serviced at a specific hub.

The rest of this article is organized as follows: First a review of prior research in intermodal logistics network domain is presented, followed by a discussion of the contributions of this work. Next, a modeling framework and a mathematical model for an intermodal logistics hub network is described. A solution methodology is presented which re-characterizes the original problem as an uncapacitated facility location problem and solves it using a tabu search metaheuristic. This methodology is tested over a range of data sets. The results of this experimental study are reported. Next, managerial insights developed through a case study based on real world logistics data, followed by the conclusions drawn in this article.

LITERATURE

This section presents a review of the intermodal logistics network literature that provides the background and establishes a framework for this research. The research presented in this paper draws from the concepts in the literature related to the interacting hub location-allocation problem and concave cost network design. This section also identifies the contributions of the research presented in this paper in the context of intermodal logistics.

An intermodal hub network can be represented by a graph in which the nodes represent the demand and supply points and the arcs represent the transportation links between the nodes (Guelat et al., 1990). The best location of the intermodal hubs in this network depend on many factors such as, the flows between origins and destinations serviced by the network, the transportation costs, economies of scale, service time performance, modal connectivity and fixed costs (Merrina et al., 2007). The traditional single mode network design models ignore the interplay between multiple modes, the difference in cost structure, the modal connectivity and the service time tradeoffs (Macharis and Bontekoning, 2004). With the inherent disjunction between modes in terms of their strengths and weaknesses, an integrated modeling approach is warranted (Merrina et al., 2007).

In the previous literature, researchers have studied regional and national freight network design with an emphasis on the transportation network infrastructure and its constituent components, such as terminals and links (Crainic and Laporte, 1997). This stream of research is focused on the design of multi-modal networks to route interregional commodity flows. One of the earlier works in this area was Harker (1987) which extended the general spatial price equilibrium model to include additional links which correspond to different levels of service. Crainic et al. (1990) developed STAN, an interactive graphic system for planning freight transportation with provisions for using multiple modes. Guelat et al. (1990) developed a minimum cost multi-commodity, multi-mode network flow problem with assignment of multiple products to multiple modes. This model and its algorithmic solution based on “shortest path with transfers” were embedded in the STAN software mentioned above. The hub-location allocation approach for intermodal network design used in this research builds upon this multi-mode flow assignment

approach. However, the hub location-allocation approach used in this research incorporates the hub location aspect of the problem along with the routing of multi-modal flows.

Slack (1990) developed the description of a regional intermodal hub as a shipment handling facility which consolidates the shipments in its service region and ships to other regions using different modes of transportation. This consolidation permits economies of scale in the inter-hub transfer of shipments. Each hub serves a regional market using trucks for pickup and delivery to respective customers. One of the relevant works in intermodal hub network design is that of Arnold et al. (2001). Within road-rail transportation networks, they define transfer points as nodes where the respective transportation networks intersect. Using a binary linear programming approach within a linear transportation cost framework, their model selects transfer points for each origin-destination pair. In an update of their previous work, Arnold et al. (2004) developed a new formulation based on a fixed charge network design. This arc based modeling approach reduced the size of the previous formulation. Their problem is solved using a shortest path approach within a local search heuristic. This shortest path based approach is similar to Lozano and Storchi (2001) who used the minimum cost network flow formulation to develop shortest paths for the origin-destination pairs.

For road-barge transportation networks, Groothedde et al. (2005) integrated the large freight capacity of the barge transportation with the faster speed of road transportation. Under a time varying demand situation, the larger stable proportion of the demand is sent via a road-barge hub network, while the highly varying part of the demand is sent direct through trucks. The network design problem is solved by using a local search procedure based on the shortest path method.

The concept of economies of scale in rail networks is addressed in Racunica and Wynter (2005). Though this work is based on a rail network only, it highlights the significance of economies of scale in a rail hub network. The cost savings are realized due to higher equipment rotation, reduced staffing costs and higher train frequencies. Combined with the fact that economies of scale exist in the road and air hub networks, road-rail-air intermodal hub networks, discussed in this paper, can also make use of these economies of scale.

Table 14: Prior Literature - Intermodal Hub Networks

Research Component	Current Paper	(Arnold et al., 2001)	(Arnold et al., 2004)	(Racunica and Wynter, 2005)	(Groothedde et al., 2005)	(Limbourg and Jourquin, 2009)
Fixed location costs	✓	✓	-	✓	-	-
Non-linear transportation costs	✓	-	-	✓	-	-
Modal connectivity costs	✓	-	-	-	-	-
Service time requirements	✓	-	-	-	✓	-
Direct shipment option	✓	✓	-	✓	✓	✓
Types of mode	Ro-R-A	Ro-R	Ro-R	R	Ro-B	Ro-R
Solution approach	MH	-	SP	HUR	SP	HUR
Solution validation	EX/CS	-	CS	EX/CS	CS	CS

LEGEND:

Ro - Road; R - Rail; B - Barge; A - Air; MH - Metaheuristic

SP - Shortest Path; HUR - Heuristic; EX - Experimental study; CS - Case study

In their work on the design of a road-rail intermodal network, Limbourg and Jourquin (2009) used variable costs of handling flows at the hubs, called transshipment costs, in addition to the transportation costs. The unit transshipment cost for a location depends on the volume of flows that pass through that location. The solution approach starts by solving the hub network design problem over the road transportation network using an initial estimate of the unit transshipment costs. This step identifies optimal hub locations which are used to route all origin-destination flows based on the least cost mode (road or rail). The unit transshipment costs are updated based on the resulting flow assignments. The network design is re-optimized with the updated transshipment costs. These steps are repeated iteratively until the total costs cannot be reduced any more.

In modern supply chains, companies gain competitive edge through efficient and effective logistics operations. The performance of logistics networks is evaluated based not only on logistics costs but also on service times. Companies seek to employ logistics networks which reduce costs while providing satisfactory service (Aykin, 1995). Groothedde et al. (2005) used

a penalty cost for the transit times in the objective function of their model formulation. The objective is to find the hub locations which minimize costs including the time related costs. A similar approach was taken by Chen et al. (2008) for a road-only network. Another work in the area of time sensitive hub networks is that of Campbell (2009), in which service time requirements are used in the constraint set to ensure conformance to a service time requirement. This work is related to road-only networks.

In order to highlight the contributions of this research, Table 14 summarizes the contents of relevant articles and compares them with this research. It can be seen that this research extends the intermodal hub network literature while considering a larger road-rail-air intermodal network. One contribution of this work is to integrate the service time requirements into the decisions about the location of the intermodal hubs and the customer assignments to hubs. Though Groothedde et al. (2005) used transit time within the scope of their work, they did not impose any service time requirements on the network design. Another contribution of this research is in the use of non-linear transportation costs. Racunica and Wynter (2005) used non-linear transportation costs, however no service time or modal connectivity costs were considered. Compared to Racunica and Wynter's work, this work uses three modes of transportation (road, rail and air) between the hubs instead of only rail. Groothedde et al. (2005) have studied the road-rail intermodal network, but their focus was on developing a heuristic, rather than using a mathematical modeling approach as is done in this work. Limbourg and Jourquin (2009) used a similar modeling approach (p-hub median) as this current work. However, this work differs from theirs in the use of non-linear economies-of-scale, fixed hub location costs, modal connectivity costs and service time requirements. Furthermore their use of the p-hub median approach is restricted to one mode (road), while this work applies it to an intermodal network. This work also differs from Arnold et al. (2001) and Arnold et al. (2004), in the use of nonlinear transportation costs, modal connectivity costs and service time requirements.

The modeling approach used in this research is to formulate the design of intermodal hub network as an interacting hub location-allocation problem. This problem was first studied by O'Kelly (1986) for single-hub and two-hub express delivery networks. Later on he presented

a quadratic integer formulation for this class of problems (O’Kelly, 1987). He showed that this is an NP-hard problem, presenting two enumeration-based heuristics. Klincewicz (1991) developed a local neighborhood search heuristic which was based on the the clustering of nodes. Campbell (1992, 1994) coined the term *p-hub median* to describe this problem. He developed a linear formulation to the quadratic formulation of O’Kelly (1987) by redefining the flow variables. Campbell also identified two versions of the p-hub median problem: single-allocation and multiple-allocation. In the single-allocation version, each non-hub node is assigned to a unique hub node, irrespective of the destination. The other version, multiple-allocation allows non-hub nodes to use different hubs depending on the destination of the shipment. This research uses the multiple-allocation version of the hub network design approach. Skorin-Kapov et al. (1996) showed that the exact solution to the multiple-allocation p-hub median problem can be found using a tight linear relaxation of the formulation given by Campbell (1994). In order to reduce the size of the Campbell (1994) formulation, Ernst and Krishnamoorthy (1998) developed new MILP formulations of the problem based on the idea proposed by Ernst and Krishnamoorthy (1996). For a more detailed review of hub location problems, see Alumur and Kara (2008).

In any hub based logistics network, the economies of scale allow for the cost savings stemming from the bundling of the inter-hub flows (Bryan and O’Kelly, 1999). The quantification of these savings is done through a concave function which models the transportation cost rate as non-increasing with the increasing flow volume. In this case the transportation cost increase at a decreasing rate as flows increase (O’Kelly and Bryan, 1998). This realistic approach improves upon the simplistic linear discount rate structure which is unaffected by the flow volume. For freight transportation, Horner and O’Kelly (2001) explored the effects of various concave cost structures and their impact on the hub network design.

The non-linear cost function can be effectively approximated by a piecewise linear function. O’Kelly and Bryan (1998) presented a linearized formulation of a concave cost network. Their work showed that such an approximation is a good substitute for the concave cost curve, allowing the use of linear/interger programming techniques to solve the model to optimality.

In many applications, the above mentioned piecewise linear approximation may require a large number of binary and linear variables. A way around this limitation is the use of a technique which is based on re-characterizing the hub network design problem as an uncapacitated facility location problem (Klincewicz, 2002). Algorithmically, the facility location problem (UFLP) is NP-complete. However, in practice large problems have been solved efficiently (Galvo and Raggi, 1989; Erlenkotter, 1978; Bilde and Krarup, 1977).

Another aspect of logistics network design deals with external costs, such as the costs of congestion, accidents, highway maintenance, air and noise pollution, and land opportunity costs which are not borne directly by the logistics service provider. The use of intermodal shipments and efficient logistics network designs can potentially alleviate such externalities (Yamada et al., 2009). The model framework used here does not consider the external costs associated with the logistics and transportation network because of the difficulty of estimating these costs. For more discussion on the impact of external costs on logistics and transportation networks, see Calthrop and Proost (1998), Forkenbrock (1991), Forkenbrock (2001) and Janic (2007).

In order to solve the hub network design problem, various heuristic-based approaches have been used. These include the greedy-interchange heuristic (Campbell, 1996), simulated annealing (Ernst and Krishnamoorthy, 1999) and tabu search (Skorin-Kapov and Skorin-Kapov, 1994; Klincewicz, 1992, 2002). Ernst and Krishnamoorthy (1998) used an LP based branch-and-bound method to obtain exact solutions, though their approach was only able to solve smaller size problems within a reasonable amount of time. Among these solution approaches, tabu search has been very successful in finding close to optimal solutions for large size problems. Tabu search (Glover, 1989, 1990) is a general iterative metaheuristic for solving combinatorial problems. Tabu search directs the solution search to break out of the local optima and move to the previously unexplored areas of the solution space (Klincewicz, 1992). Using a short-term and a long-term memory of the solutions visited, the search explores the solution space in an efficient manner, yielding good solutions in a reasonable time.

In the next section, a modeling framework is discussed which provides the foundation for the development of a mathematical model for the intermodal hub network. The mathematical

model extends the p-hub median and concave cost network design approaches to the domain of intermodal logistics network.

MODELING FRAMEWORK

In this research an intermodal network is represented by a graph $\mathbb{G}(\mathbb{N}, \mathbb{A})$ in which the set of nodes \mathbb{N} represents the cities and the set of arcs \mathbb{A} represents the different modes of transportation between the cities. The set of arcs \mathbb{A} can be partitioned into three disjoint subsets: \mathbb{A}_0 , \mathbb{A}_1 and \mathbb{A}_2 , which represent the three modes of transportation: road, rail, and air, respectively. These subset of arcs describe the modal network connectivity between the cities. A shipment between two cities may travel over any of the available arcs representing the choice of the mode of transportation.

In this modeling framework a shipment between an origin city i may travel directly to the destination city j using direct road shipments. The shipment may also travel through a pair of hubs (k, m) . These hub-based shipments use three legs of travel. The first leg is the pickup from the origin city i and travel to the origin hub k . The second leg is the inter-hub travel between the origin and destination hubs, k and m . This is the line haul where the shipments move in larger quantities and with higher frequencies. The third leg is the drop off from hub m to the destination city, j . In an intermodal network the pickup and the drop off is performed by trucks. The inter-hub shipments can travel from the origin hub k to the destination hub m using different modes of transportation, t . In this research t has three possible values: $t=0$ for road, $t=1$ for rail and $t=2$ for air. It is also possible that an origin-destination pair may be served by one hub, i.e., $k = m$. In such a case there is no inter-hub travel.

In the context of this research, an intermodal logistics hub has local access to road, intermodal-rail and air freight terminals. A shipment may be sent between hubs through the use of any one of the available modes. The choice of mode (road, rail, or air) for moving a shipment between hubs is determined by the tradeoff between the service requirement quoted to the customer and the transportation cost associated with each mode. Figure 6 shows the different ways that a shipment may travel between its origin and its destination. A direct shipment from the origin

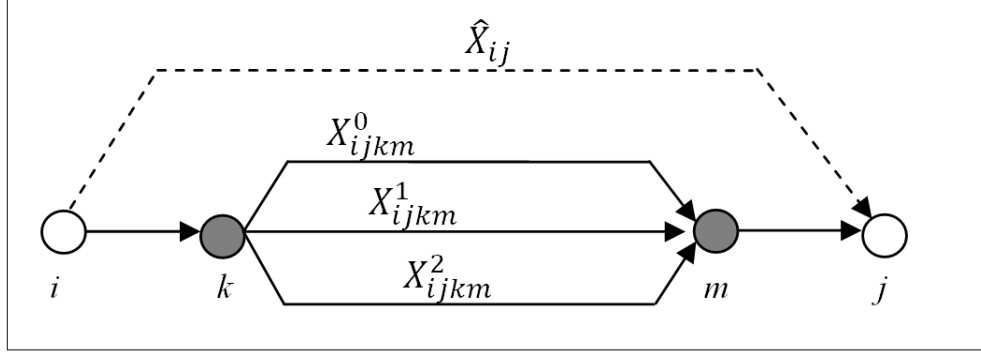


Figure 6: Types of Shipments

city i to the destination city j is represented by the binary decision variable \hat{X}_{ij} . The shipment may also travel through the hubs k and m using transportation mode t , which is represented by the binary variable X_{ijkm}^t . Every hub shipment travels from its origin city i to its origin hub k using road transportation. The travel between the origin hub k and the destination hub m may occur over any one of the arcs which represent the different modes of transportation between the hubs. The choice of a specific arc represents the use of the corresponding transportation mode. Although a specific mode is selected for an inter-hub shipment for a specific origin-destination pair, other origin-destination pairs may flow through the same hubs using a different mode of transportation. Thus, for a specific origin-destination pair (i, j) and hub pair (k, m) , $X_{ijkm}^0 = 1$ implies road travel between the hub cities, $X_{ijkm}^1 = 1$ implies rail travel, and $X_{ijkm}^2 = 1$ implies air travel. For each origin-destination pair, only one mode is selected. The final leg of travel from the destination hub k to the destination city j always uses road transportation.

MODEL FORMULATION

An intermodal logistics network in this model is described by the parameters related to flow, cost and transit time. There are three types of costs in an intermodal network: fixed hub operating costs, fixed modal connectivity costs, and transportation costs. In a traditional road logistics network, the fixed cost of operating a hub (F_k) includes the costs of utility, leasing, manpower, equipment and cost of capital. In an intermodal network, a hub facility is more complex because it services multiple modes of transportation. As a result, there are additional fixed costs that need to be considered. These fixed costs, called *modal connectivity*

costs (MC_{kt}), directly depend on the types of modes t serviced by the hub k . For example the docking/loading/unloading arrangement for an intermodal rail container is different from a road trailer. Similarly airfreight containers have different sizes and loading/unloading needs. Thus the fixed modal connectivity costs of an intermodal hub depend on the number and type of modes it services. The modal connectivity at a hub is represented by the binary decision variables S_{kt} , which identify the modes t which are serviced at hub k .

The variable costs of a logistics network are comprised of its modal transportation costs. Each mode of transportation considered in this research has a different rate structure. The differences are based on the unit transportation cost of moving a shipment between cities. The intermodal rail and air freight rates are based on moving container loads. The intermodal rail transportation rates for shipping one unit of flow are the lowest among the transportation modes considered in this research. On the other hand air transportation rates are the highest which reflect the premium for fast movement of time sensitive shipments. In the model c_{ij}^1 and c_{ij}^2 represent the unit transportation cost of shipping an intermodal container between cities i and j using rail and air, respectively. The model uses CL_{km}^1 and CL_{km}^2 , which represent the intermodal rail and air container flows between hubs k and m , to calculate the inter-hub rail and air transportation costs.

The road transportation rate structure is inherently different from the rail and air rate structure. For road transportation, the freight rates depend on the total number of shipments. For a small flow volume, transportation rates known as less-than-truckload (LTL) rates apply. In this model the pickup, drop-off and small volume inter-hub road transportation costs are based on the LTL rates. In this research the unit cost of shipping a pallet by LTL road shipment is represented by the parameter c_{ij}^0 . The shipments may also travel directly from the origin to the destination. The rate structure for direct shipments are higher than the LTL rates. This difference is due to the special arrangements for specific pickup/delivery schedule, shipment frequency and origins/destinations outside the regular lanes serviced by the carrier. The model represents the unit cost of direct shipments between cities i and j with the parameter \hat{c}_{ij} .

If the flow volume of road transportation for inter-hub shipments is high then the unit

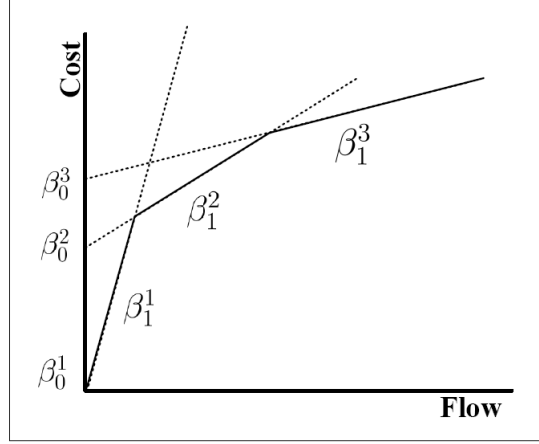


Figure 7: Piecewise Linear Cost Function

transportation costs are reduced due to the economies of scale. These rates, called truckload (TL) rates, have lower unit transportation costs than the LTL rates. This economy of scale in road transportation is represented by a concave cost function. O’Kelly and Bryan (1998) used a piecewise linear cost function to approximate this concave function (see Figure 7). The function comprises of a series of cost lines with non-increasing slopes. The approximation is based on the lower envelope of the linear cost line segments. As more and more flows are routed through the hub-to-hub link, the transportation costs increase at a lower rate, as represented by the cost line segment which falls on the lower envelope. Each cost line is represented by two parameters: an intercept (β_0) and a slope (β_1). The choice of the best cost segment s for a hub pair (k, m) is indicated by the decision variables δ_{km}^s . The model uses R_{km}^s variables, which compute the total flows passing through hubs k and m for the cost segment s to calculate the inter-hub road transportation costs.

The freight flows between origin city i and destination city j is represented by the parameter f_{ij} . The model uses binary decision variables, y_k to represent the selection of a city for locating a hub. The transit time of moving a shipment between city i and j using mode t is represented by the parameter T_{ij}^t . The mathematical formulation is as follows:

Sets:

T = $\{0,1,2\}$, set of modes,

S = Set of cost lines in piecewise linear cost function, and

N = Set of cities.

Decision Variables:

$$y_k = \begin{cases} 1 & , \text{ if city } k \text{ is a hub} \\ 0 & , \text{ otherwise,} \end{cases}$$

$$\hat{X}_{ij} = \begin{cases} 1 & , \text{ if direct shipments are made from } i \text{ to } j \\ 0 & , \text{ otherwise,} \end{cases}$$

$$X_{ijkm}^t = \begin{cases} 1 & , \text{ if shipments from } i \text{ to } j \text{ flow through hub pair } (k,m) \text{ using mode } t \\ 0 & , \text{ otherwise,} \end{cases}$$

$$\delta_{km}^s = \begin{cases} 1 & , \text{ if the volume of flow between } k \text{ and } m \text{ uses line segment } s \\ 0 & , \text{ otherwise,} \end{cases}$$

$$S_{kt} = \begin{cases} 1 & , \text{ if hub } k \text{ is served by mode } t \in \{1,2\} \\ 0 & , \text{ otherwise,} \end{cases}$$

R_{km}^s = Total flows shipped between hubs k and m using cost segment s , and

CL_{km}^t = Total flows shipped between hubs k and m using mode $t \in \{1,2\}$.

Parameters:

p = Number of hubs,

f_{ij} = Flow volume from the origin city i to the destination city j ,

c_{ij}^t = Unit transportation cost from city i to city j using mode t ,

\hat{c}_{ij} = Unit transportation cost for a direct shipment from city i to city j ,

F_k = Fixed cost of opening and operating a hub in city k ,

- MC_{kt} = Modal connectivity cost of serving mode t at hub k for $t \in \{1, 2\}$,
 τ = Time delay factor at hubs, $\tau \geq 1$,
 $C(k, m, t)$ = Transportation costs of shipments between hubs k and m using mode $t \in T$,
 β_0^s, β_1^s = Intercept and slope of the cost line s , respectively,
 \hat{M} = $\sum_{i,j \in N} f_{ij}$, a large number,
 L^t = Size of a mode t container, $t \in \{1, 2\}$,
 d_{km}^t = Drayage cost of a container shipped by mode t from hub k to hub m , $t \in \{1, 2\}$,
 T_{ij}^t = Travel time between cities i and j using mode t , and
 TW_{ij} = Service time requirement for a shipment between origin i and destination j .

Model Formulation:

$$\begin{aligned}
\text{minimize} \quad & \sum_{k \in N} F_k y_k + \sum_{k \in N, t \in \{1, 2\}} MC_{kt} S_{kt} + \sum_{i, j \in N} f_{ij} \hat{c}_{ij} \hat{X}_{ij} \\
& + \sum_{k, m \in N, t \in T} C(k, m, t) + \sum_{i, j, k, m \in N, t \in T} f_{ij} (c_{ik}^0 + c_{mj}^0) X_{ijkm}^t \quad (20)
\end{aligned}$$

subject to:

$$\sum_{k \in N} y_k = p \quad (21)$$

$$X_{ijkm}^t \leq y_k, \quad \forall i, j, k, m \in N, t \in T \quad (22)$$

$$X_{ijkm}^t \leq y_m, \quad \forall i, j, k, m \in N, t \in T \quad (23)$$

$$\sum_{k, m \in N, t \in T} X_{ijkm}^t = 1 - \hat{X}_{ij}, \quad \forall i, j \in N \quad (24)$$

$$\sum_{i, j, m \in N} X_{ijkm}^t \leq |N|^2 S_{kt}, \quad \forall k \in N, t \in \{1, 2\} \quad (25)$$

$$\sum_{i, j, k \in N} X_{ijkm}^t \leq |N|^2 S_{mt}, \quad \forall m \in N, t \in \{1, 2\} \quad (26)$$

$$C(k, m, 0) = \sum_{s \in S} (\beta_0^s \delta_{km}^s + \beta_1^s R_{km}^s) c_{km}^0, \quad \forall k \neq m \in N \quad (27)$$

$$\sum_{s \in S} \delta_{km}^s = 1, \quad \forall k \neq m \in N \quad (28)$$

$$\sum_{s \in S} R_{km}^s = \sum_{i, j \in N} f_{ij} X_{ijkm}^0, \quad \forall k \neq m \in N \quad (29)$$

$$R_{km}^s \leq \hat{M}\delta_{km}^s, \forall k \neq m \in N, s \in S \quad (30)$$

$$C(k, m, t) = \frac{CL_{km}^t}{L^t} (c_{km}^t + d_{km}^t), \forall k \neq m \in N, t \in \{1, 2\} \quad (31)$$

$$CL_{km}^t = \sum_{i,j \in N} f_{ij} X_{ijkm}^t, \forall k \neq m \in N, t \in \{1, 2\} \quad (32)$$

$$\sum_{k,m \in N, t \in T} X_{ijkm}^t (T_{ik}^0 + \tau T_{km}^t + T_{mj}^0) + \hat{X}_{ij} T_{ij}^0 \leq TW_{ij}, \forall i, j \in N \quad (33)$$

$$y_k, X_{ijkm}^t, \hat{X}_{ij}, S_{kt}, \delta_{km}^s \in \{0, 1\}, \forall i, j, k, m \in N, s \in S, t \in T$$

The first term in the objective function (20) is the fixed costs of locating and operating the hubs. The hub locations are identified by the decision variables y_k . The second term in the objective function is the modal connectivity costs of servicing modes at the hubs. The intermodal rail and intermodal air connectivity at each hub are identified by the decision variables S_{k1} and S_{k2} , respectively. The variables S_{k1} and S_{k2} are set equal to 1 when a hub k uses intermodal rail and intermodal air services, respectively. Each hub in the network is assumed to have a default road service for pickup and drop-off of shipments as well as for long haul inter-hub shipments. The fixed costs (F_k) includes the cost of road service. The third term is the transportation costs of the direct shipments from origin i to destination j . The fourth term $C(k, m, t)$ in the objective function deals with the transportation costs of the hub-to-hub shipments between the hub pair (k, m) using a transportation mode t . The last term in the objective function includes the transportation costs of pickup from the origin and the transfer to the origin hub and the transportation costs of drop off from the destination hub to the final destination.

Constraint (21) requires that a total of p hubs are opened. Constraints (22) and (23) ensure that any hub pair assignment for an origin-destination pair is limited to opened hubs. Constraint (24) allows for a choice between using a direct shipment and a hub-based shipment for any origin-destination pair. If it is economical to send flows from i to j directly ($\hat{X}_{ij} = 1$), the corresponding hub shipment decision variables X_{ijkm}^t are set equal to zero. In the case of hub-based shipments, each origin-destination pair is assigned a specific (k, m) pair and a specific mode t .

If a hub sends or receives shipments using rail or air then the modal connectivity costs reflect the resources required to support that service. Constraint (25) represents the case of outbound shipments from a hub. If any of the X_{ijkm}^t variables related to a (k,t) pair is equal to 1, then the corresponding S_{kt} variable must be set equal to 1 in the solution. Constraint (26) represents the case of a destination hub where shipments may arrive through a mode t . Note that the model allows for a hub to use different modes for inbound and outbound shipments.

The inter-hub road transportation costs are computed using the $C(k,m,0)$ cost function, as given by (27). The binary variables δ_{km}^s represent the choice of selecting the appropriate cost line in the piecewise linear cost function. The choice of a specific cost line s is indicated by a value of 1 assigned to its corresponding δ_{km}^s variable. Corresponding to some specific volume of flow, the cost line which lies on the lower envelope of the cost function is selected. In other words, $\delta_{km}^s = 1$ represents the selection of an intercept (β_0^s) and a slope (β_1^s) for use in computing the inter-hub road transportation costs.

The model selects the correct cost line s by computing the total flows passing through the hub using road transportation. Recall that each shipment (as specified by its origin-destination pair) has the choice of using any of the available modes. The model computes the total hub-to-hub flows traveling by road ($t = 0$) using the set of equations given by constraint (29). The right hand side computes the total flows that pass through a hub over the road network. This volume of flows corresponds to one of the cost lines s lying on the lower envelope of the cost function. Thus all but one of the terms in the left hand side of these constraints will be zero. The model selects the non-zero terms using the inequalities given by constraint (30). According to constraint (28) only one of the cost lines can be selected.

The intermodal rail and air freight transportation costs are computed using the $C(k,m,1)$ and $C(k,m,2)$ cost functions, as given by (31). CL_{km}^1 and CL_{km}^2 represent the total volume of flows that are shipped between the hubs k and m using rail and air, respectively. These flows are divided by the size of the container, L^1 or L^2 , to convert into the equivalent number of container units. The model computes the mode specific transportation costs in constraint (31) using the flows computed in (32).

This model restricts the transit time of a shipment to be within a certain service time window. The left side of constraint (33) computes the transit time for each origin-destination pair based on its shipment method, i.e., direct shipment or hub-based shipment. The right side represents a service time restriction for each origin-destination pair. This approach models different service options offered to the customers who may select a service time specific to their needs. This also allows for the design of a logistics network with a geographical footprint derived from the service time requirements.

SOLUTION PROCEDURE

The solution procedure used in this research is based on tabu search (Glover, 1989, 1990) which is a solution search procedure which explores the solution space in an efficient manner. The search moves from one solution to the next using a short and a long term memory of previously visited solutions. Within the context of this research, a solution is comprised of a set of hub cities and the service assignment of a single-hub, two-hub or direct shipment for each origin-destination city pair. The search procedure moves from one solution to the next by a swap of a hub city with a non-hub city. The candidate city to enter the set of hubs H is selected from the collection of nodes that constitute the neighborhood $N(H)$ of the current solution. The neighborhood is composed of those solutions which can be reached by a pairwise interchange between the hub nodes and the non-hub nodes. This exchange is called a *Move*. The attributes of the move are stored in a short term memory, called the *Tabu List*, for the duration of some pre-specified number of iterations, called the *Tabu Tenure*. This memory is used to avoid the moves that will return to recently visited solutions. The search moves from one solution to the next, recording the best solution found. Most often this phase of the search identifies a local optimum. In the second phase of the tabu search, the search re-starts from a new starting solution. The starting solution is generated using a long term search memory, called *Node Frequency*, which records the number of times a node was selected as a hub in the visited solutions. In each restart, the starting solution ignores the nodes with high node frequency, allowing the search to proceed to a previously unexplored area of the solution space.

Table 15: Pseudocode for Tabu Search Procedure

1	Tabu Search () {
2	Initialize search parameters;
3	Generate <i>Starting Solution</i> ;
4	<i>Best Solution</i> = <i>Starting Solution</i> ;
5	while (<i>Run Count</i> < <i>Max Run</i>)
6	<i>Count</i> = 0, <i>Node Frequency</i> \leftarrow 0;
7	while (<i>Count</i> < <i>Max Count</i> AND <i>Iteration Num</i> < <i>Max Iter</i>)
8	Generate move \rightarrow Select a new set of hub locations;
9	Convert original problem to UFL problem;
10	Solve UFL problem;
11	<i>Move Solution</i> \leftarrow Convert UFL solution to solution of original problem;
12	if (<i>Move</i> \notin <i>Tabu List</i> OR <i>Move</i> satisfies aspiration criterion)
13	Update <i>Current Solution</i> = <i>Move Solution</i> ;
14	else
15	Go to line 8;
16	if (<i>Current Solution</i> < <i>Best Solution</i>)
17	Update <i>Best Solution</i> = <i>Current Solution</i> ;
18	<i>Count</i> = 0;
19	<i>Count</i> ++, <i>Iteration Num</i> ++;
20	Update <i>Node Frequency</i> ;
21	end while
22	<i>Run Count</i> ++;
23	Generate new <i>Starting Solution</i> ;
24	if (<i>Node Frequency</i> > <i>Freq Limit</i>)
25	Exclude node from starting solution;
26	Update <i>Current Solution</i> = <i>Starting Solution</i> ;
27	Clear <i>Tabu List</i> ;
28	end while
29	Report <i>Best Solution</i> ;
30	}

The search terminates after a pre-specified number of re-starts that do not improve the best solution.

The pseudo-code for the tabu search procedure is shown in Table 15. After initializing the search parameters, the tabu search procedure randomly selects a set of hub nodes and generates the starting solution (line 3). The tabu search procedure restarts *Max Run* number of times. In each restart (line 5), the search continues until *Max Count* number of non-improving moves are made (line 7). Each move involves identifying a non-hub node which replaces a hub node (line 8). This node pair is selected by evaluating all possible moves within the current neighborhood. The entering and exiting nodes are selected based on the move which has the smallest objective function value among all the available moves. Each move generates a new set of hub nodes which are used to compute the new solution. The computation of the solution requires that

the original problem (lines 9, 10 and 11) is re-characterized, which will be discussed later. A selected move may be rejected (line 12) if it violates the following tabu criteria: (a) a node cannot enter the set of hubs if it was removed recently (i.e., during the tabu tenure) and (b) a node cannot be removed from the existing set of hubs if it was added recently (i.e., during the tabu tenure). A move that is rejected because of its tabu status can still be accepted if it satisfies the aspiration criterion, i.e., it yields a solution that is the best one found so far in the search. An accepted move is made and the tabu list is updated (line 13), otherwise another move is generated (line 15). The solution which results from an accepted move is recorded as the current solution. If the current solution is the best found so far, the best solution is updated (line 16). If in any iteration, the current solution does not improve the best solution, the count of the consecutive non-improving moves is incremented (line 19). After *Max Count* number of non-improving moves occur, the tabu search procedure restarts the search (line 23) from a new starting solution. In each restart the starting solution is generated from a set of randomly selected hub nodes. This set is selected from the nodes which have the *Node Frequency* value less than a threshold value called, *Freq Limit*. With a starting solution, the procedure repeats the steps in line 6 through 20. The tabu search procedure terminates after *Max Run* number of re-starts.

During each iteration of the tabu search procedure, a set of hub nodes is selected (line 8). Based on these hub nodes, the complete solution is developed by determining the least cost assignment of all origin-destination pairs to the hubs by solving the assignment subproblem (lines 9, 10 and 11). If the discount for the inter-hub travel is linear, the assignment problem is separable by the origin-destination (i, j) pairs and can be solved as a set of shortest path problems. However in the case of concave costs, the discount for inter-hub shipments depend on the total flow volume between the two hubs. The links with more flows may achieve a greater discount due to the economies of scale. In this case freight flows may be directed away from the lowest LTL cost path to a path which has sufficient flows to qualify for the higher truckload discount. This situation corresponds to increasing the flow volume on a link to a point where a cost segment (s) with smaller slope (β_1^s) in the piecewise linear cost function is reached

thereby reducing the transportation costs. Hence in generating a solution to the assignment problem, it is necessary that the solution finds X_{ijkm}^t values for each origin-destination pair (i, j) , as well as the δ_{km}^s values for each hub pair (k, m) . The latter is needed to know which segment (s) of the piecewise linear cost function will minimize the cost on each hub-to-hub link. These complexities can be addressed by converting the problem into an *Uncapacitated Facility Location* (UFL) problem (Klincewicz, 2002).

In general, the UFL problem is stated as follows. There is a set of customers, indexed by l , and a set of possible facility locations, indexed by n . Each opened facility incurs a fixed cost f_n and it costs c_{nl} to assign customer l to facility n . The objective is to minimize the total cost of selecting facilities and assigning customers to those facilities. The UFL problem can be used to represent the original assignment problem, where customers and facilities in the UFL problem represent origin-destination pairs and different types of shipments in the original problem, respectively. The UFL problem is mathematically stated as follows.

$$(UFL) \quad \text{minimize} \quad \sum_n f_n y_n + \sum_{n,l} c_{nl} x_{nl} \quad (34)$$

subject to:

$$\sum_n x_{nl} = 1 \quad , \forall l \quad (35)$$

$$x_{nl} \leq y_n \quad , \forall n, l \quad (36)$$

$$y_n, x_{nl} \in \{0, 1\} \quad , \forall n, l$$

Define t_{nl} as the transit time associated with a customer-to-facility assignment. Based on this parameter, a service time constraint is added to the UFL problem, as follows:

$$\sum_n t_{nl} x_{nl} \leq TW_l \quad , \forall l. \quad (37)$$

In the above formulation, the binary variable x_{nl} is set to 1 if customer l is assigned to facility n , and it is set to 0, otherwise. The binary variable y_n is set to 1 if a facility n is

Table 16: UFL Problem Parameters

Mode	Facility Count	Fixed Cost (f_n)	Assignment Cost (c_{nl})	Assignment Time (t_{nl})
Road	sP_2^p	$\beta_0^s c_{km}^0$	$f_{ij}(c_{ik}^0 + c_{mj}^0 + \beta_1^s c_{km}^0)$	$t_{ik}^0 + \tau t_{km}^0 + t_{mj}^0$
Rail	P_2^p	0	$f_{ij}(c_{ik}^0 + c_{mj}^0) + \left(\frac{f_{ij}}{L^1}\right)(c_{km}^1 + d_{km}^1)$	$t_{ik}^0 + \tau t_{km}^1 + t_{mj}^0$
Air	P_2^p	0	$f_{ij}(c_{ik}^0 + c_{mj}^0) + \left(\frac{f_{ij}}{L^2}\right)(c_{km}^2 + d_{km}^2)$	$t_{ik}^0 + \tau t_{km}^2 + t_{mj}^0$
Single Hub	p	0	$f_{ij}(c_{ik}^0 + c_{kj}^0)$	$t_{ik}^0 + t_{kj}^0$
Direct	1	0	$f_{ij}\hat{c}_{ij}$	t_{ij}^0

opened, and to 0, otherwise. Constraint (35) ensures that each customer is assigned to some facility. Constraint (36) assigns the customers only to open facilities. Although it is simple to state, this problem is NP-Hard. However, in practice large size problems can be solved efficiently (Bilde and Krarup, 1977; Erlenkotter, 1978).

This model can be applied to the intermodal network by transforming the original (assignment) problem into a corresponding UFL problem. Each customer in the UFL problem represents an origin-destination pair in the original problem. Each facility in the UFL problem represents one of the ways a shipment may travel from the origin to the destination in the original problem, such as passing through two hubs (using road, rail or air), passing through a single hub, or as a direct shipment. The parameters of the UFL problem, i.e., the fixed cost of opening a facility, cost and time of a customer assignment to a facility, are derived from the original problem, as shown in Table 16.

In the context of the original problem, a shipment may use two hubs (k, m). The number of ways in which an ordered subset of two elements can be obtained from a set of p hubs is given by $P_2^p = \frac{p!}{(p-2)!}$. Furthermore, for each such hub pair, there are s different cost segments which represent the inter-hub road transportation cost. Thus, there are (sP_2^p) number of (road) facilities in the UFL problem which represent the two-hub road shipments. These road facilities are identified by the triple (s, k, m) . Additionally for each hub pair, a shipment may choose rail transportation or air transportation. These choices are represented by P_2^p facilities for each mode. Another way a shipment may travel from the origin to the destination in the original problem is by using a single hub. Thus, there are p additional facilities in the UFL problem. Finally, one facility is used to represent the direct shipments. For illustration, a three-hub

($p = 3$) network is considered. If there are two cost lines used ($s = 2$), there are $sP_2^p = 12$ facilities for road, $P_2^p = 6$ each, for rail and air, 3 for single hub and 1 for direct shipments, i.e., a total of 28 facilities.

Table 16 shows the parameters of the UFL problem as they are computed from the parameters of the original problem. The fixed cost of a road facility (s, k, m) in the UFL problem is given by $\beta_0^s c_{km}^0$. Note that only the road hub shipments include a fixed cost which is needed to model its concave cost structure. The cost of assigning a customer (i, j) to a road facility (s, k, m) is given by $f_{ij}(c_{ik}^0 + c_{mj}^0 + \beta_1^s c_{km}^0)$. Furthermore, $f_{ij}(c_{ik}^0 + c_{mj}^0) + \left(\frac{f_{ij}}{L^1}\right)(c_{km}^1 + d_{km}^1)$ and $f_{ij}(c_{ik}^0 + c_{mj}^0) + \left(\frac{f_{ij}}{L^2}\right)(c_{km}^2 + d_{km}^2)$ represent the assignment cost of a customer to the rail and air facilities, respectively. Similarly, the assignment cost for a customer to a direct ship or a single-hub facility is given by $f_{ij}\hat{c}_{ij}$ and $f_{ij}(c_{ik}^0 + c_{kj}^0)$, respectively. Table 16 also provides the values for the assignment times.

For a given set of hub locations, solving the UFL problem yields a solution which assigns each customer to a facility. The assignment of a customer to a facility in the UFL solution corresponds to a minimum cost path in the original problem. For example the assignment of customer (i, j) to road facility (s, k, m) identifies the hub pair and the parameters of the cost segment, i.e., β_0^s and β_1^s . This information is used to generate the solution values for the corresponding X_{ijkm}^t , y_k , S_{kt} , and \hat{X}_{ij} variables. These values are used to compute the remaining decision variables, R_{km}^s and CL_{km}^t . Using the fixed location cost and the modal connectivity cost, the objective function value of the original problem is then computed.

COMPUTATIONAL STUDY

The purpose of this computational study is to evaluate the performance of the tabu search procedure over a variety of test problems. To effectively test the performance of the procedure, the data sets differ in size (number of nodes) and scale (range of input data values) to avoid biased cases (Rardin and Uzsoy, 2001). The test problems used in the computational experiments span networks with 10, 15 and 20 cities. For each of these networks different combinations of flow, cost and time matrices are used, as shown in Table 17. The elements of these

Table 17: Matrix Table

Combination	Cost	Flow	Time
1	C_1	F_1	T_1
2	C_1	F_1	T_2
3	C_1	F_2	T_1
4	C_1	F_2	T_2
5	C_2	F_1	T_1
6	C_2	F_1	T_2
7	C_2	F_2	T_1
8	C_2	F_2	T_2

Table 18: Cost & Time Matrices

Cost Data	Road Costs	Rail Costs	Air Costs
C_1	(1, 10)	(40, 200)	(8, 120)
C_2	(1, 30)	(40, 1000)	(8, 360)
Time Matrix	Road Time	Rail Time	Air Time
T_1	(1, 4)	(1, 8)	(1, 1)
T_2	(1, 6)	(1, 12)	(1, 1)

matrices are randomly generated from a Uniform probability distribution with the parameters (a, b) . These parameters are set at different values for the computational experiments. Table 18 shows the parameter values used for generating the cost and time data, and Table 19 shows the parameter values used for the flow data. The values of L^1 and L^2 are arbitrarily set at 40 and 8, respectively, for the computational experiments. With three network sizes and eight combinations each, a total of 24 data sets are generated.

These data sets are used to solve different test problems based on the model parameters, as shown in Table 20. The parameter values are number of hubs, $p = \{3, 5\}$; service time requirement, $TW = \{12, 16\}$; and modal connectivity cost, $MC = \{10\%, 30\%\}$ as a percentage of the fixed location cost. The fixed location cost is computed as $\sum_{(i,j) \in N} f_{ij}$. The economy-of-scale is modeled with a two-segment piecewise linear cost function, with the parameters $\{(\beta_0^1, \beta_1^1), (\beta_0^2, \beta_1^2)\}$. The two different sets of values used in the test problems are $\{(0, 0.9), (100, 0.8)\}$ and $\{(0, 0.8), (200, 0.6)\}$. Those parameter values and the 24 data sets yield 384 distinct test problems. Using four random samples for each of these test problems, a total of 1536 problem instances were solved.

Table 19: Flow Matrices

Flow Matrix	Flows
F_1	(1, 50)
F_2	(1, 100)

Table 20: Parameter Values for Computational Study

Nodes (n)	Hubs (p)	Modal Cost (% of F_k)	Slope (β_1^s)	Service Time (TW)
{10,15,20}	{3,5}	{0.1,0.3}	{(0.9,0.8),(0.8,0.6)}	{12,16}

Table 21: Average Percent Optimality Gaps for 10-city Network Problems

Hubs (p)	MC (% of F_k)	Slope (β_1^1, β_1^2)	Service (TW)	Data Set								Avg
				1	2	3	4	5	6	7	8	
3	10%	(0.9,0.8)	12	0.00%	0.27%	0.00%	0.24%	0.00%	0.00%	0.58%	0.30%	0.17%
3	10%	(0.9,0.8)	16	0.00%	0.00%	0.00%	0.00%	0.00%	0.29%	0.57%	0.00%	0.11%
3	10%	(0.8,0.6)	12	0.00%	0.34%	0.02%	0.31%	0.12%	0.00%	0.53%	0.41%	0.22%
3	10%	(0.8,0.6)	16	0.00%	0.00%	0.02%	0.02%	0.00%	0.04%	0.53%	0.22%	0.11%
3	30%	(0.9,0.8)	12	1.22%	3.16%	0.79%	2.68%	0.29%	0.35%	0.57%	0.92%	1.25%
3	30%	(0.9,0.8)	16	0.68%	1.33%	0.52%	1.69%	0.00%	0.35%	0.55%	0.53%	0.71%
3	30%	(0.8,0.6)	12	1.79%	3.27%	1.24%	2.72%	0.64%	0.48%	0.69%	0.74%	1.45%
3	30%	(0.8,0.6)	16	0.96%	1.94%	0.72%	2.11%	0.00%	0.18%	0.66%	0.77%	0.92%
5	10%	(0.9,0.8)	12	0.30%	0.43%	0.28%	0.20%	0.42%	0.18%	0.00%	0.00%	0.23%
5	10%	(0.9,0.8)	16	0.16%	0.00%	0.33%	0.00%	0.10%	0.00%	0.00%	0.00%	0.07%
5	10%	(0.8,0.6)	12	0.41%	0.00%	0.36%	0.21%	0.47%	0.42%	0.07%	0.00%	0.24%
5	10%	(0.8,0.6)	16	0.16%	0.00%	0.33%	0.00%	0.10%	0.00%	0.17%	0.00%	0.10%
5	30%	(0.9,0.8)	12	1.08%	2.59%	1.20%	2.61%	0.40%	0.49%	0.16%	0.14%	1.08%
5	30%	(0.9,0.8)	16	0.70%	0.46%	0.89%	0.79%	0.10%	0.09%	0.16%	0.00%	0.40%
5	30%	(0.8,0.6)	12	2.08%	3.68%	2.05%	3.89%	0.61%	0.90%	0.42%	0.70%	1.79%
5	30%	(0.8,0.6)	16	1.28%	1.25%	1.65%	1.80%	0.23%	0.14%	0.52%	0.00%	0.86%

COMPUTATIONAL RESULTS

The computational experiments were run on a Pentium 4 CPU with a 3.20 GHz clock and 4GB RAM. The problems were solved to optimality using CPLEX ver 10.0 with OPL Studio ver 5.2. The script language of OPL Studio was used to iteratively run the problem instances in an efficient manner. The test problems were also solved using the tabu search procedure. The tabu search solutions are compared with the optimal solutions obtained from CPLEX. The percent optimality gap is calculated as $(\frac{z_{TS}-z^*}{z^*})$. The percent optimal gaps of the four replications for each test problem are averaged and recorded as *Average Percent Optimality Gap*.

The results of computational study are tabulated in Table 21 for 10-node problems, Table 22 for 15-node problems and Table 23 for 20-node problems. The columns in each table show the number of hubs, modal connectivity cost as a percentage of fixed location cost, slope of the cost segments in the piecewise linear cost function and the service time requirement, respectively. The next eight columns report the average percent optimality gap over four replications for each data set. The last column averages the results over all data sets in each row.

Table 22: Average Percent Optimality Gaps for 15-city Network Problems

Hubs	MC	Slope	Service	Data Set								Avg
(p)	(% of F_k)	(β_1^+, β_1^-)	(TW)	9	10	11	12	13	14	15	16	
3	10%	(0.9,0.8)	12	0.11%	0.32%	1.00%	1.13%	0.03%	0.22%	0.51%	0.57%	0.49%
3	10%	(0.9,0.8)	16	0.00%	0.15%	0.72%	0.13%	0.01%	0.93%	1.86%	0.38%	0.52%
3	10%	(0.8,0.6)	12	0.21%	0.62%	1.10%	1.24%	0.69%	0.29%	0.44%	0.26%	0.61%
3	10%	(0.8,0.6)	16	0.00%	0.27%	0.73%	0.14%	0.04%	0.94%	1.60%	0.00%	0.47%
3	30%	(0.9,0.8)	12	2.63%	5.60%	3.84%	5.62%	0.88%	2.38%	0.86%	1.56%	2.92%
3	30%	(0.9,0.8)	16	1.45%	3.30%	2.84%	4.02%	0.21%	0.17%	1.73%	0.32%	1.75%
3	30%	(0.8,0.6)	12	3.01%	5.31%	4.06%	5.16%	1.03%	2.51%	0.93%	1.88%	2.99%
3	30%	(0.8,0.6)	16	1.86%	3.77%	3.25%	4.27%	1.01%	0.36%	1.87%	0.60%	2.13%
5	10%	(0.9,0.8)	12	0.00%	0.00%	0.21%	0.31%	0.26%	0.07%	0.25%	0.67%	0.22%
5	10%	(0.9,0.8)	16	0.03%	0.42%	0.00%	0.00%	0.01%	0.10%	0.35%	0.39%	0.16%
5	10%	(0.8,0.6)	12	0.00%	0.23%	0.23%	0.46%	0.22%	0.25%	0.18%	0.63%	0.28%
5	10%	(0.8,0.6)	16	0.03%	0.39%	0.00%	0.04%	0.00%	0.53%	0.31%	0.34%	0.20%
5	30%	(0.9,0.8)	12	3.55%	7.23%	4.41%	7.82%	0.24%	1.39%	0.29%	2.18%	3.39%
5	30%	(0.9,0.8)	16	2.29%	4.38%	3.04%	4.49%	0.01%	0.10%	0.34%	0.37%	1.88%
5	30%	(0.8,0.6)	12	4.64%	8.10%	5.42%	8.52%	0.36%	2.01%	0.38%	2.96%	4.05%
5	30%	(0.8,0.6)	16	3.41%	5.47%	4.10%	4.74%	0.02%	0.60%	0.32%	0.36%	2.38%

Table 23: Average Percent Optimality Gaps for 20-city Network Problems

Hubs	MC	Slope	Service	Data Set								Avg
(p)	(% of F_k)	(β_1^+, β_1^-)	(TW)	17	18	19	20	21	22	23	24	
3	10%	(0.9,0.8)	12	0.92%	1.37%	0.83%	1.23%	0.40%	0.14%	0.09%	0.70%	0.71%
3	10%	(0.9,0.8)	16	0.51%	0.36%	0.15%	0.46%	0.55%	0.27%	0.09%	0.88%	0.41%
3	10%	(0.8,0.6)	12	0.98%	1.54%	0.85%	1.43%	0.38%	0.15%	0.00%	0.75%	0.76%
3	10%	(0.8,0.6)	16	0.57%	0.52%	0.16%	0.52%	0.60%	0.36%	0.00%	0.89%	0.45%
3	30%	(0.9,0.8)	12	4.49%	5.22%	4.75%	4.33%	0.75%	0.85%	0.43%	1.17%	2.75%
3	30%	(0.9,0.8)	16	3.25%	4.67%	2.54%	4.44%	0.73%	1.31%	0.27%	1.91%	2.39%
3	30%	(0.8,0.6)	12	4.18%	4.40%	5.03%	4.58%	0.59%	0.93%	0.45%	0.70%	2.61%
3	30%	(0.8,0.6)	16	3.58%	4.98%	2.94%	4.77%	0.59%	1.27%	0.00%	1.66%	2.47%
5	10%	(0.9,0.8)	12	0.57%	0.91%	0.55%	0.40%	0.46%	0.00%	0.79%	0.26%	0.49%
5	10%	(0.9,0.8)	16	0.57%	0.34%	0.52%	0.25%	0.00%	0.17%	0.46%	0.00%	0.29%
5	10%	(0.8,0.6)	12	0.61%	1.00%	0.42%	0.03%	0.46%	0.00%	0.83%	0.10%	0.43%
5	10%	(0.8,0.6)	16	0.57%	0.41%	0.52%	0.33%	0.00%	0.15%	0.46%	0.00%	0.30%
5	30%	(0.9,0.8)	12	5.81%	8.28%	6.38%	8.08%	0.71%	1.22%	1.18%	1.82%	4.18%
5	30%	(0.9,0.8)	16	5.83%	5.39%	4.58%	5.26%	0.25%	0.51%	0.77%	0.51%	2.89%
5	30%	(0.8,0.6)	12	6.26%	8.99%	6.31%	7.21%	0.97%	1.27%	1.44%	1.50%	4.24%
5	30%	(0.8,0.6)	16	3.17%	5.94%	4.71%	6.04%	0.40%	0.81%	0.90%	0.69%	2.83%

The average of optimality gaps over all the test problems is 1.30%. The minimum average gap for the tabu search procedure is 0.00% with 70% of all the problem instances solved within 1% of the optimal solution and 93% of the test problems solved to within 5% of the optimal solution. It is seen that generally the solutions of the large size network problems have bigger average percent optimality gaps, e.g., for the case of 20-city network with $p=5$, $MC=30\%$ and slope values of (0.8.0.6) the average percent optimality gap for data sets 17 - 20 are relatively large. However it is not always the case, as shown by the small gaps of the other data sets, 21 - 24 for the same network parameters. The results of the computational study show low average percent optimality gaps achieved by the tabu search procedure which indicate that the procedure works well over test problems of varying size and scale.

EMPIRICAL STUDY

The purpose of this empirical study is to investigate issues related to the design and management of intermodal logistics networks. This study investigates the effect of using intermodal shipments on the network structure which is characterized by the number and location of hubs, the use of transportation modes, transportation and network costs. This study also explores the effect of economies-of-scale on inter-hub shipments and the mix of modes when a logistics network must satisfy a specific service time requirement.

The data for this study was collected from various sources. The cities and the freight flows used in this study were extracted from the 2002 Freight Analysis Framework Commodity Origin-Destination (FAF) database. This database is maintained by the United States Department of Transportation and contains data from a variety of sources that are integrated to estimate the annual freight transportation activity among the states, regions, and major international gateways. Use of the commodity flow data in an actual geographical setting provides an accurate representation of a U.S. logistics network and allows for realistic interpretation of the results (Southworth and Peterson, 2000). The freight flows in the FAF database are based on the sample data collected by the Commodity Flow Survey (CFS) of more than 100,000 freight shippers within the United States. This database reports origin, destination, commodity, weight

and mode for the shipments in the manufacturing, wholesale and retail industries. The origin and destination data correspond to the 114 Metropolitan Statistical Areas. These areas are defined by the U.S. Office of Management and Budget (OMB) for use by the federal agencies in collecting, tabulating, and publishing federal statistics. The 42 commodity codes used in the database are based on the Standard Classification of Transported Goods (SCTG). There are four transportation modes covered in the database: highway, railroad, water and air.

The FAF database was used to generate a representative sample of flows for the case study. Recall that the focus of this research is on logistics service providers that use intermodal containers and/or trailers to move shipments through their logistics networks. The range of goods handled by such organizations span commodities that are readily containerized, and the types of goods that cannot be readily containerized are not handled by such logistics service providers. For this reason, the freight flows corresponding to commodities such as agricultural products, minerals, coal, and petroleum products were excluded from consideration in the study. The freight flows were computed in terms of the number of pallet loads that must be moved between the origin and destination cities. The total commodity flows between the cities in the FAF database were converted to the pallet level by using a conversion factor of 1 pallet = 350 pounds. These pallet load quantities were then used as the demand flow between each pair of cities in the study. Thus, the unit of analysis for this research is a “pallet-load,” and the demand flows represent the total pallet-loads that must be shipped between respective cities. In compiling the data, inbound and outbound flows for each city/metropolitan area were aggregated to compute the total flow volume for each city. The cities were then rank ordered according to their total flow volume and the top 35 cities were selected for this study (see Figure 8). The cities in the case study data are geographically dispersed across five U.S. regions: Northeast, Southeast, Midwest, Southwest and West. Table 24 shows that each of these five regions in the U.S. are adequately represented in the case study both in the number of cities as well as in the distribution of flows.

The transportation costs for each lane (origin-destination pair) and mode were estimated using a number of different sources of actual freight rates and converted into a cost per pallet.

Table 24: Regions and Flows in the Case Study

Region	No. of Cities	% of Total Flows
Northeast	6	17.0%
Southeast	11	19.3%
Midwest	8	30.4%
Southwest	4	13.0%
West	6	20.3%

Less-than-truckload (LTL) and direct delivery freight rates were obtained from a private logistics service provider that is among the top ten U.S. freight carriers. The local transportation costs for the pick-up and delivery of a shipment (c_{ij}^0) are based on the actual LTL freight rates. These local transportation costs are computed based on the distance and the LTL rate for each lane. The cost of the direct shipment for a pallet-load (\hat{c}_{ij}) was calculated in a similar fashion using the direct delivery freight rates which are higher than the LTL rates (see section ?? for more details). The long-haul full truckload (TL) costs between hubs are less than the local LTL costs and are scaled from the LTL costs using the parameter, $\beta \leq 1$, which is an experimental factor in the case study. By varying β , the study generates different long haul costs and examines their impact on the network design.

The transportation costs for intermodal rail are based on the published freight rates offered by the various U.S. rail carriers: Union Pacific (UP), Norfolk Southern (NS), Burlington Northern Santa Fe (BNSF) and CSX Railways. The intermodal rail costs are based on the freight rates for shipping one 48-foot domestic intermodal-rail container. The air transportation costs are based on the air freight rates of United Parcel Service (UPS) for shipping one pallet using their next-day delivery service. Both intermodal rail and air shipments also incur drayage costs which were included in their respective intermodal transportation costs for each lane.

The average transit times for road shipments were computed using the commercial software PCMiller with consideration for the provisions of the *Hours-of-Service* rule from the Federal Motor Carrier Safety Administration (FMCSA, 2008). The average intermodal rail transit times used in this study were obtained from the published data of the same U.S. rail carriers used to obtain the freight rates. A transit time of one day (next-day delivery) is used for all air shipments.

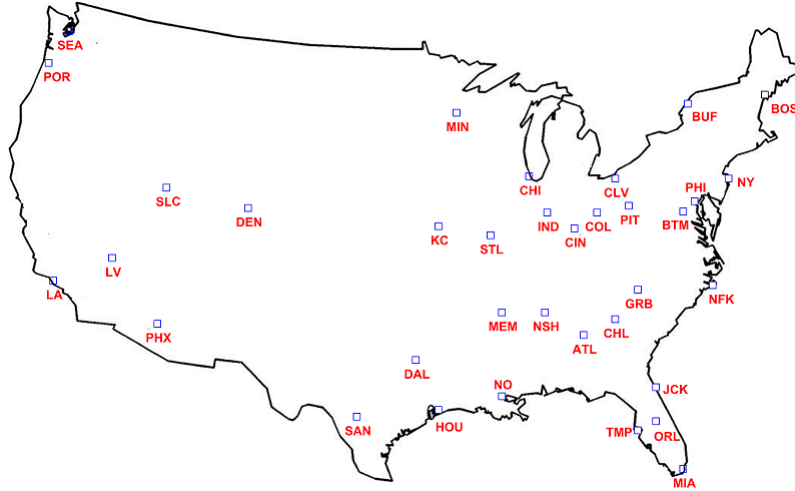


Figure 8: Empirical Study of 35-City Network

The other data required for the model include the parameters for the piecewise linear cost function, the fixed cost of operating a hub and the modal connectivity cost. The parameters describing the piecewise linear cost function for each cost line s are (β_0^s, β_1^s) . The value of these parameters are based on Giordano (2008) in which the values of $(\beta_1^1, \beta_1^2) = (1, 0.9)$ were reported. The flow volume at which the unit transportation cost changes is set at 5,000 units for computational purposes. The other parameter used in the model is the fixed cost of operating a logistics hub (F_k). Waller (2003) provides an estimate that the average fixed costs account for 40% of the average total facility cost. The components of the fixed costs are: building costs, which accounts for 50%, energy costs, which account for 12% and equipment costs, which account for 38% of the average total fixed costs per facility. In order to estimate the building costs (and subsequently the total fixed cost of a hub facility), the size of the facility must be specified. This study sets the size of the hub facility at 100,000 ft^2 which is comparable to the size of the United Parcel Services (UPS) freight facilities (UPS, 2008). In computing the building cost, the *Rent Survey of US Commercial Real Estate* (Colliers Research Services, 2009) was used. This survey reported that the average rental rate for warehouse space was $\$5.51/ft^2$ in 2008. Using the data discussed above, the fixed operating costs of a 100,000 ft^2 logistics facility was computed to be $\$1,000,000$. The last parameter of the model is the modal connectivity costs (MC). These costs were modeled as a fraction of the fixed operating costs.

EXPERIMENTAL DESIGN

The problem instances used in this case study were based on five different factors: type of network (*Type*), number of hubs (p), modal connectivity costs (MC), service time requirement (TW), and economies of scale (β_1^s). The study was based on a full factorial design in which these factors were set at different levels as shown in Table 25. The full factorial design of these factor levels resulted in 216 problem instances.

Table 25: Factor Levels

<i>Factors</i>	<i>Levels</i>
<i>Type</i>	{ <i>OTR, IM</i> }
<i>MC</i>	{10%, 20%, 30%} of FC
<i>TW</i>	{4, 6, 8}
β_1^s	{(1, 0.90), (1, 0.85), (1, 0.80)}
p	{3, 4, 5, 6}

The factor *Type* was set at two levels which represent the OTR and IM networks. The factor *MC* was set at three levels which reflect different modal connectivity costs at a hub. These costs were set as a percentage of the fixed cost of a hub. The factor *TW* was set at three levels; 4, 6 and 8 days, respectively. The problem instances generated by the different combinations of these factor levels were solved with four different values of $p = \{3, 4, 5, 6\}$. Among the solutions generated by different values of p , the solution with the lowest total network cost was reported. In addition, the following values were also recorded: network costs which records the objective function value; total transportation costs (*TXC*), which includes pickup, inter-hub transfer and drop off costs; total number of intermodal hubs in the network (*IM-Hubs*); total flows that use intermodal shipments between hubs (*IM-Flows*); total flows that use single hub shipments (*S-Flows*); total flows that use two hubs (*Hub-to-Hub Flows*); total flows that use direct shipments (*D-Flows*) and the location of the chosen hubs.

The computational study in section ?? showed that the largest problem size that can be solved to optimality is 20 cities. Because the size of the problem instances in this case study (35 cities) exceeds that limit, optimal solutions cannot be used to investigate this case study. Therefore, the tabu search solution approach was used to solve the problem instances considered in this case study.

RESULTS AND DISCUSSION: HUB LOCATIONS

This section describes the hub locations in the OTR and IM networks. Using the problem instances of the case study a list of hub locations in the OTR and the IM networks was compiled. This list identified the frequency with which a particular city was selected as a hub in different problem instances of the case study. This list is graphically presented in Figure 9. This graph shows all the hub cities and the frequency with which these cities were selected for locating a hub. For each hub city, the graph also indicates the percentage of total flows which originate or terminate in that city. The graph shows that Chicago (CHI) which has 17% of the total flow volume was selected in 85% of the intermodal problem instances, whereas it was never selected as a hub location in the OTR problem instances.

It can be seen that among all the cities of the northeastern region in this case study, Philadelphia (PHI) was the best choice in locating a hub. This choice reflect the strategic location and access to different modal networks which made the city Philadelphia a suitable location for a hub in both OTR and IM networks in this case study. While a city may be a suitable location for a hub in an OTR network its location or intermodal network connectivity may not be sufficient to justify a hub in an IM network. For example, Memphis (MEM) was selected as a hub city in 100% of the OTR network problem instances but was never selected as a hub location in the IM network problem instances. On the other hand Atlanta (ATL) appeared in 75% of the IM network problem instances and in none of the OTR network problem instances.

In the IM problem instances, the hub locations selected in a majority of the IM networks were Philadelphia (PHI), Chicago (CHI), Atlanta (ATL), Dallas (DAL) and Los Angeles (LA). In 66% of the IM network problem instances, hubs were located in Dallas (DAL) and Los Angeles (LA) to service customers in the southwestern and western regions, respectively. In the remaining 34% of the problem instances this pair of hubs was replaced by hubs in Las Vegas (LV) and Portland (POR). In those cases the flows to/from the southwestern cities of Dallas (DAL), Houston (HOU), San Antonio (SAN) and Phoenix (PHX) were assigned to hubs closer to their respective locations. In the case of the OTR network problem instances, the hubs

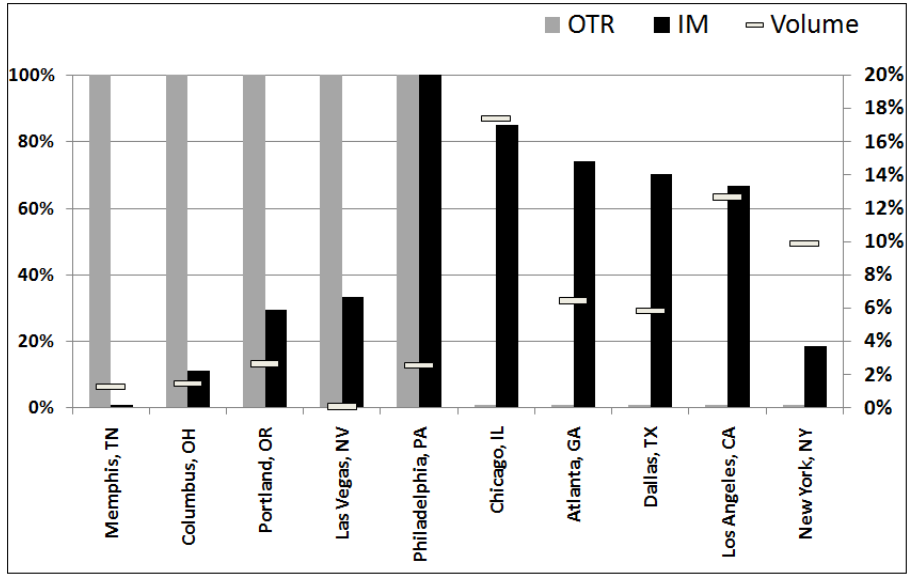


Figure 9: Hub Locations

locations were very consistent. The best locations for hubs in the OTR networks were always Memphis (MEM), Columbus (COL), Portland (POR), Las Vegas (LV) and Philadelphia (PHI).

The results of this case study suggest that by using intermodal shipments, the location of the hubs in a logistics network may change. These changes may require moving some of the OTR hubs to new locations where access to intermodal networks can be utilized. Such shifts in location will affect the footprint of the service area, consequently affecting the network costs as discussed in the next section. The results show that cities that are frequently chosen for IM hubs typically have larger total flow volumes than cities that are chosen for OTR hubs. These results also suggest that the location of intermodal hubs may be more sensitive to changes in the network parameters compared with OTR networks.

NETWORK COSTS

This section compares the OTR and IM networks based on their network costs. The network costs comprise of fixed location costs, inter-hub transportation costs, pickup and delivery costs, direct shipment costs and modal connectivity costs. Table 26 shows the different values of p and the percentage of test problems for which each p yielded solutions with the lowest total network costs. The table shows that the best number of hubs in all OTR problem instances

Table 26: Number of Hubs

Type	Frequency			
	$p = 3$	$p = 4$	$p = 5$	$p = 6$
<i>OTR</i>	0%	0%	100%	0%
<i>IM</i>	0%	9%	80%	11%

was $p = 5$. The best p for the IM networks varied between 4 and 6, with $p = 5$ being the best number of hubs in a majority (80%) of the problems.

Figure 10 plots the average network costs for different factor levels used in this case study. These plots show that the average network costs of the intermodal networks were less than those of the OTR networks in a majority of the problem instances in this study. An increase in the cost of providing intermodal services at the hubs (MC) caused an increase in the IM network costs (see Figure 10a) because the benefit of the lower costs of the intermodal service is offset by the higher modal connectivity costs. This causes a change in the intermodal flows and the total network costs. Another important factor in this study is the economies-of-scale for long haul road transportation. The economies-of-scale reduce the transportation costs of the road hub shipments and consequently reduce the total network costs. It can be seen in Figure 10b that the average total network costs of the problem instances decreased with higher economies-of-scale. The direct effect of higher economies of scale was to attract more flows to hub-based shipments, and this effect lowered the total transportation costs of the network. Because the OTR networks use more road shipments compared to IM networks, the economies of scale affected the OTR networks more than the IM networks.

Due to the long transit times of intermodal shipments, the use of IM shipments can be greatly affected by the service time requirements. As seen in Figure 10c the network costs for the IM networks increased with shorter service time requirements. The plot shows that service time requirements had a significant effect on the network costs for the IM networks. The direct effect of the smaller service time requirements was a reduction in the use of intermodal shipments. This reduction was caused by the long transit times of the intermodal rail shipments, which were infeasible for the shorter service time requirements. As a result the intermodal flows used higher cost road shipments, thereby increasing the network costs. Under shorter service time

requirements the network structure changes from IM to OTR. In this sense an OTR network is a limiting case of an IM network where all inter-hub shipments use road network. This means that the total network costs of an IM network can be no larger than a corresponding OTR network. Due to the heuristic nature of the solution procedure an anomaly is observed in the $TW=4$ result in Figure 10c. The effect of shorter service time requirements on the OTR networks was marginal. This suggests that the OTR network structure may not be significantly effected for $TW \geq 4$ days. However as the analysis of the next sections shows, this effect may be significant for shorter service time requirements.

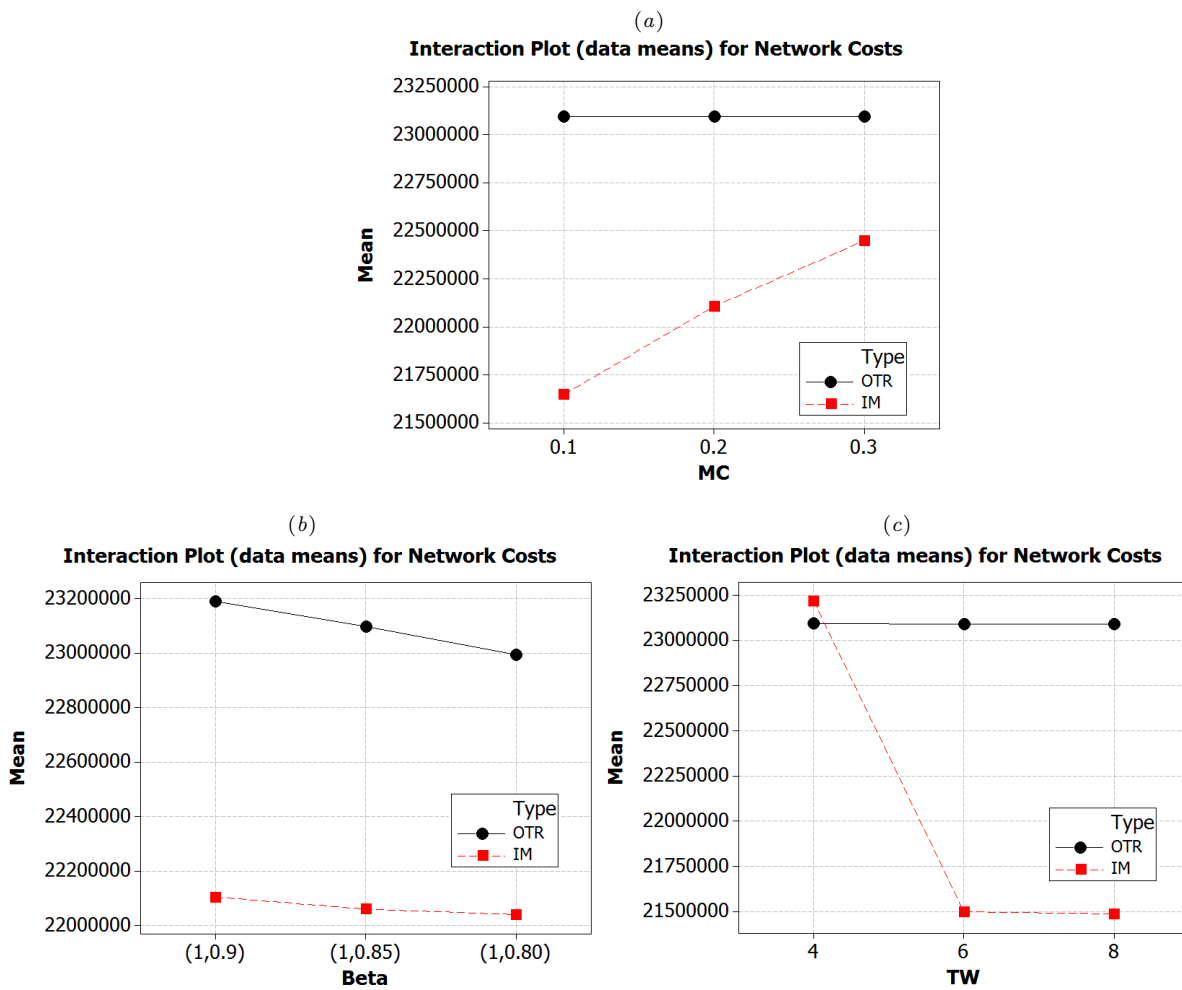


Figure 10: Total Network Costs

HUB SHIPMENT AND MODAL USAGE

This section analyzes the use of different types of shipments in the OTR and IM networks. The shipments in a hub network can travel in three distinct ways: direct travel from origin to destination, travel through a single hub and travel through two hubs. The last option allows for consolidation of flows and the use of different modes of transportation. The first option is more suitable for time sensitive shipments where time performance may be more valuable than transportation costs. The single hub shipments are used when both origin and destination cities are served by the same hub.

Table 27: Types of Shipments and their Use (Average Percentage of Total Flows)

Network Type	Single Hub Shipments	Two Hub (Road) Shipments	Two Hub (Rail) Shipments	Direct Shipments
<i>OTR</i>	34%	47%	0%	19%
<i>IM</i>	33%	12%	48%	7%

In the experimental study, the proportion of flows were categorized by single-hub shipments, two-hub road shipments, two-hub rail shipments and direct shipments. Table 27 shows the average proportion of flows in each of these categories. The single-hub shipments are used for the flows between cities which are located within the service region of a hub. It is interesting to note that the use of single-hub shipments in both the OTR and the IM networks is similar. This result suggests that the hub service regions have similar intra-regional flows in the OTR and IM networks. The data in Table 27 also show that the flows in OTR and IM networks use inter-hub shipments differently. The inter-hub shipments in the OTR networks accounted for 47% of the total flows, while this proportion increased to 60% in the IM networks. This increase reflects the benefit of lower transportation rates for intermodal shipments which attracts more volume to the inter-hub links.

The results showed that direct shipments are beneficial in two cases. The first case is when the service time requirements necessitate the use of direct shipments because they are faster than hub shipments. The second case is when the total transportation cost of a direct shipment is less than the transportation cost of the hub-based shipments for a specific origin-destination pair. The data shows that a higher proportion (19%) of the total flows used direct shipments

in the OTR networks compared to only 7% of the total flows which used direct shipments in the IM networks.

The reduced use of direct shipments in IM networks is due to the lower intermodal rates which make the transportation costs of hub-based shipments more competitive with direct shipments. Hence the availability of intermodal service at hubs results in a smaller volume of direct shipments. However, there are still flows that use direct shipments in the IM networks due primarily to the service time requirements.

IMPACT OF SERVICE TIME REQUIREMENTS ON MODAL FLOWS

In a hub network, shipments traveling between two hubs take advantage of the lower transportation costs which are possible due to the economies of scale. However that cost advantage comes at the expense of longer shipment transit time. When service time requirements are short, this service aspect of the hub network design cannot be ignored. Therefore this section investigates the impact of the service time requirement on the network structure.

The results of the case study showed that in problem instances with smaller service time requirement the use of intermodal hub shipments was significantly reduced (see Figure 11*b*), while the use of road hub shipments increased (see Figure 11*a*). For example, the proportion of flows which used intermodal (rail) shipments fell from 60% to 18% and road hub shipments increased to 35% when the service time requirement was reduced from 6 days to 4 days. Figure 11*c* shows that the service time requirements also affect the use of direct shipments in IM networks which increased from 4.5% to 10.5% when TW dropped from 8 days to 4 days. For the case of single hub shipments, Figure 11*d* shows that as the service time requirements decreased the proportion of flows using single-hub shipments in an IM network increased by approximately 10%.

The plots in Figure 11 show that the structure of an intermodal network is significantly affected by the service requirements. In order to develop more insights about the effect of service requirements, the case of an IM network with $p = 5$, $MC = 10\%$ and $\beta = (1, 0.9)$ is solved with different values of $TW = \{3, 4, 5, 6, 7, 8\}$. For each TW value, modal flows, hub

location, inbound and outbound hub flows and their modes of transportation are recorded and discussed.

Table 28 shows the effect of changing service time requirements (TW) on the proportion of total flows that use direct and single-hub shipments. It is clear that the proportion of flows which use direct shipments increases significantly when the service time requirements decrease. This data also suggests that direct shipments become necessary for short service time requirements (see cases of $TW = 4$ and $TW = 3$ days).

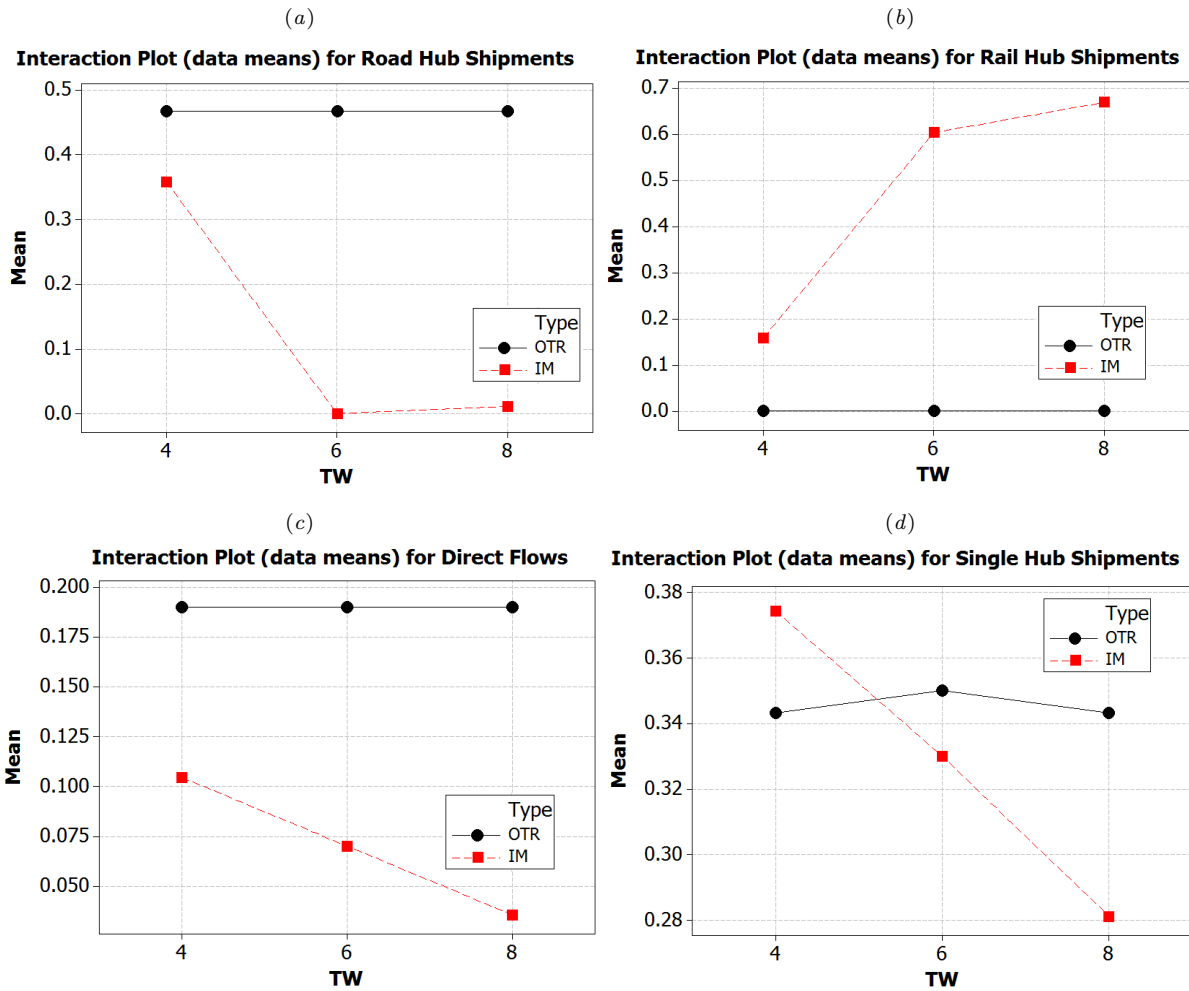


Figure 11: Hub Shipments

The service time requirements also affect the use of different transportation modes. Table 29 shows the hub locations, inbound hub flows, outbound hub flows, single hub flows and direct shipment flows for different service time (TW) values. For each service time requirement (TW),

Table 28: Use of Direct and Single Hub Shipments

Shipments	$TW(\text{days})$					
	8	7	6	5	4	3
Direct	4.8%	4.8%	4.8%	4.9%	6.2%	18.7%
Single-hub	31.9%	31.9%	32.4%	37.5 %	41.3%	37.7%

the first column shows different types of shipments (hub inbound, hub outbound, single hub and direct shipments) and the mode used. The next five columns list the best hub locations and the flows as a percentage of the total flows in the network. The last column adds up all the flows in each row. As a check on the conservation of flows, the total percentage of the inbound flows and the outbound flows sums up to the same value.

The results in Table 29 for $TW = \{8, 7, 6\}$ show a similar pattern, which suggests that the change in service time requirement from 8 days to 6 days does not change the intermodal network structure. The following observations can be made at this stage: The best hub locations were Los Angeles (LA), Dallas (DAL), Chicago (CHI), Atlanta (ATL) and Philadelphia (PHI). The transit times of intermodal rail shipments were found to be feasible for this range of service time requirements, as no road shipments were used between the hubs. The majority of the flows (63%) moving between the hubs used intermodal rail shipments. The hub in Chicago handled the most flows (inbound and outbound) compared to any other hub. The majority (68%) of the out bound flows from Chicago hub had destinations in only three cities, i.e., Los Angeles, Dallas and Atlanta. A major proportion (40%) of the inbound flows at the Chicago hub were destined for the city itself. These results suggest that cities where a high volume of flows originate or terminate are good candidates for locating hubs in an intermodal network.

It is also seen that there is an imbalance between the inbound and outbound flows at all hubs except Atlanta. The inbound intermodal rail volume at Los Angeles hub was twice the outbound intermodal rail volume. A similar flow imbalance can be seen for the hubs at Chicago and Philadelphia. Most of the shipments handled by the hub at Philadelphia were outbound with a very small inbound flow. A majority (70%) of these outbound flows originated in New York and Boston.

The effects of service time requirements on the network structure started to appear with the

Table 29: Modal Flows (% of Total Flows)

Modes	Hubs					
TW=8 days	Los Angeles	Dallas	Chicago	Atlanta	Philadelphia	Total %Flows
Road (Hub Outbound)	-	-	-	-	-	-
Rail (Hub Outbound)	8.07	12.72	13.93	8.81	19.77	63.29
Road (Hub Inbound)	-	-	-	-	-	-
Rail (Hub Inbound)	17.48	9.63	23.12	11.01	2.05	63.29
Single Hub	4.66	3.35	16.17	7.70	-	31.89
Direct						4.81
TW=7 days	Los Angeles	Dallas	Chicago	Atlanta	Philadelphia	Total %Flows
Road (Hub Outbound)	-	-	-	-	-	-
Rail (Hub Outbound)	8.07	12.72	13.93	8.81	19.77	63.29
Road (Hub Inbound)	-	-	-	-	-	-
Rail (Hub Inbound)	17.48	9.59	23.84	10.25	2.13	63.29
Single Hub	4.66	3.35	16.17	7.70	-	31.89
Direct						4.81
TW=6 days	Los Angeles	Dallas	Chicago	Atlanta	Philadelphia	Total %Flows
Road (Hub Outbound)	-	-	-	-	-	-
Rail (Hub Outbound)	7.50	12.74	14.13	8.96	19.50	62.83
Road (Hub Inbound)	-	-	-	-	-	-
Rail (Hub Inbound)	16.22	10.48	23.86	10.01	2.26	62.83
Single Hub	4.66	3.35	16.63	7.70	-	32.35
Direct						4.81
TW=5 days	Los Angeles	Dallas	Chicago	Atlanta	Philadelphia	Total %Flows
Road (Hub Outbound)	-	-	-	-	2.85	2.85
Rail (Hub Outbound)	4.78	13.50	13.57	8.93	14.04	54.82
Road (Hub Inbound)	-	2.85	-	-	-	2.85
Rail (Hub Inbound)	8.89	10.33	22.27	10.15	3.18	54.82
Single Hub	8.17	3.48	18.10	7.70	-	37.45
Direct						4.87
TW=4 days	Portland	Las Vegas	Chicago	Atlanta	Philadelphia	Total %Flows
Road (Hub Outbound)	9.04	0.29	0.59	-	13.57	23.50
Rail (Hub Outbound)	1.25	-	6.76	7.94	12.97	28.92
Road (Hub Inbound)	0.41	13.63	0.21	-	9.24	23.50
Rail (Hub Inbound)	0.62	-	13.88	12.27	2.15	28.92
Single Hub	-	8.00	18.81	14.53	-	41.34
Direct						6.24
TW=3 days	Portland	Las Vegas	Columbus	Memphis	Philadelphia	Total %Flows
Road (Hub Outbound)	8.13	0.14	10.41	3.07	21.53	43.27
Air (Hub Outbound)	-	0.28	-	-	-	0.28
Road (Hub Inbound)	2.93	14.08	6.53	11.61	8.13	43.27
Air (Hub Inbound)	0.11	-	0.07	0.10	-	0.28
Single Hub	-	8.06	10.96	18.72	-	37.74
Direct						18.70

case of $TW=5$ days. The total hub shipments fell from 62.83% for the case of $TW=6$ days to 54.82% for the case of $TW=5$ days. This indicates a shift in flows from hub-based shipments to single-hub shipments as the transit time of some of the intermodal rail routes did not satisfy the service requirement of 5 days. This effect was distinctly visible on the inbound intermodal rail flows at the Los Angeles hub which were reduced by 45%. Upon examination of the solution it was found that these flows originated from the cities of New York, Portland, Seattle and Tampa, all of which shifted to road shipments. The solution also showed that outbound rail service from Los Angeles to the Philadelphia and Atlanta hubs was no longer used. However the intermodal rail routes from Los Angeles to Chicago and Dallas were still feasible.

When the service time requirements were reduced to $TW=4$ days, a major shift to road transportation occurred. At this service level, the road shipments constituted 64.84% of hub shipments, i.e., 41.3% and 23.5% of the total flows used single-hub and two-hub road shipments, respectively. Due to this modal shift, a change in the hub locations was also observed. The Los Angeles and Dallas hubs were closed and replaced with hubs in Portland (POR) and Las Vegas (LV). A major change was also observed in the outbound road shipments from the Philadelphia hub which increased from 2.85% to 13.57%. The increased use of road shipments did not affect the rail shipments at the Philadelphia hub which decreased only marginally. These shifts in the flow volumes made the Philadelphia hub a central location for the entire network. The inbound and outbound flows at the Chicago hub decreased by 50% and 40%, respectively. The flows at the new hubs in Portland and Las Vegas reflect the nature of their regional flows. The Portland hub was used primarily for outbound shipments, while the Las Vegas hub handled a higher number (13.6%) of inbound shipments compared to the outbound shipments (<1%).

The case of $TW=3$ days is interesting as it showed that the use of direct shipments increased significantly. Note that the use of direct shipments went up from 6.24% for $TW=4$ days to 18.70% for $TW=3$ days. The use of direct shipment increased because it offers the smallest transit time. The hub locations were also affected by these changes. Compared to the previous case, the hubs in Chicago and Atlanta were replaced with hubs in Columbus and Memphis. None of the selected hubs used rail shipments at this level of service requirement. An interesting

characteristic about the direct shipments was that 60% of the direct shipments had either an origin or a destination in Chicago or Jacksonville. The direct shipments for these cities account for 15% of the total network flows. This situation showed that even though Chicago was no longer a hub, it still had significant flow volumes passing through the city. Some outbound flows from Los Angeles and Seattle used air shipments through the Las Vegas hub for their destinations on the east coast, i.e., Boston and Miami.

The above analysis showed that under shorter service time requirements the network undergoes major structural changes. These changes affect hub locations, use of transportation modes, use of single-hub shipments and direct shipments. For this case study the results showed that a hub did not necessarily have a balance in its inbound and outbound flows. In some cases hubs were set up to be used for only inbound or outbound flows. Similarly, hubs may differ in the use of transportation modes, e.g., the results showed that a hub may extensively use intermodal rail for hub shipments and have no inbound/outbound inter-hub flows that use road shipments. For this case study the results also showed that under shorter service time requirements, the flows transitioned from intermodal rail to road to single hub shipments to direct/air shipments. The results of the case study showed that intermodal rail service was the dominant mode of hub shipments down to $TW=5$ days. There was still some intermodal rail usage for $TW=4$ days because the transit times of some intermodal rail shipments were within the allowed service time. The study also showed that the hub location were greatly affected by the service time requirements. Some hubs were relocated with different modal connectivity as the service requirements changed.

EFFECT OF ECONOMIES-OF-SCALE ON HUB SHIPMENTS

In hub networks, flows are consolidated at the hubs. This consolidation allows for economies-of-scale which reduces the unit transportation costs for road travel between the hubs. This reduction in unit road transportation costs depends on the operational efficiency of the OTR network. An efficiency gain in an OTR network increases the benefit of the economies-of-scale. In this section the effect of higher economies-of-scale on the modal flows and the use of different

types of shipments are investigated. From discussion in the previous section it is clear that use of different modes and types of shipments is affected by the service requirements. In order to develop more insights about the effect of economies-of-scale, case of an IM network with $p = 5$, $MC = 10\%$ is solved with different values of $TW = \{4, 5, 6\}$. The values of TW are selected in light of discussions in the previous section. For each TW value the problem is solved for different values of $\beta = (\beta_1^1, \beta_1^2) = \{(1, 0.85), (1, 0.80), (1, 0.75), (1, 0.70), (1, 0.65), (1, 0.60), (1, 0.55), (1, 0.50), (1, 0.45)\}$. For each of these problem instances the modal flows and types of shipments are recorded.

For the case of $TW=6$ (see Figure 12a), the road hub shipments started to increase at $\beta = (1, 0.60)$, where some inter-hub shipments shifted from rail to road. However the higher economies-of-scale were not sufficient to attract a significant volume of flows, showing only marginal modal shifts. The case of $TW=5$ (see Figure 12b) is more interesting where the service time requirements cause a shift towards a higher use of road hub shipments. In this case the effect of economies-of-scale on road hub shipments is observed at $\beta = (1, 0.65)$ where the gains in road hub shipments come from a reduction in rail hub shipments. The single-hub shipments also showed a downward trend at this point. The road hub shipments accounted for a greater percentage of the total flows than the rail hub shipments at approximately $\beta = (1, 0.53)$.

These results identify an interplay between the effects of economies-of-scale and the service time requirements. When service times are large (see $TW=6$ case), there is a significant use of intermodal rail shipments, and much larger increases in efficiency gains are required to bring road transportation rates at par with intermodal rail rates. When service time requirements get smaller (see $TW=5$), the use of road shipments increases. The efficiency gains for such networks have a bigger impact on the the modal flows where significant shifts from intermodal rail shipments to road shipments is observed. This effect is also seen in the case of $TW=4$ where the initial proportion of road hub shipments is significantly larger (23%) than the previous two cases ($< 5\%$), and the use of road hub shipments started to increase at $\beta = (1, 0.60)$. The gains in road hub shipments is accompanied again by reductions in rail hub shipments and single-hub shipments.

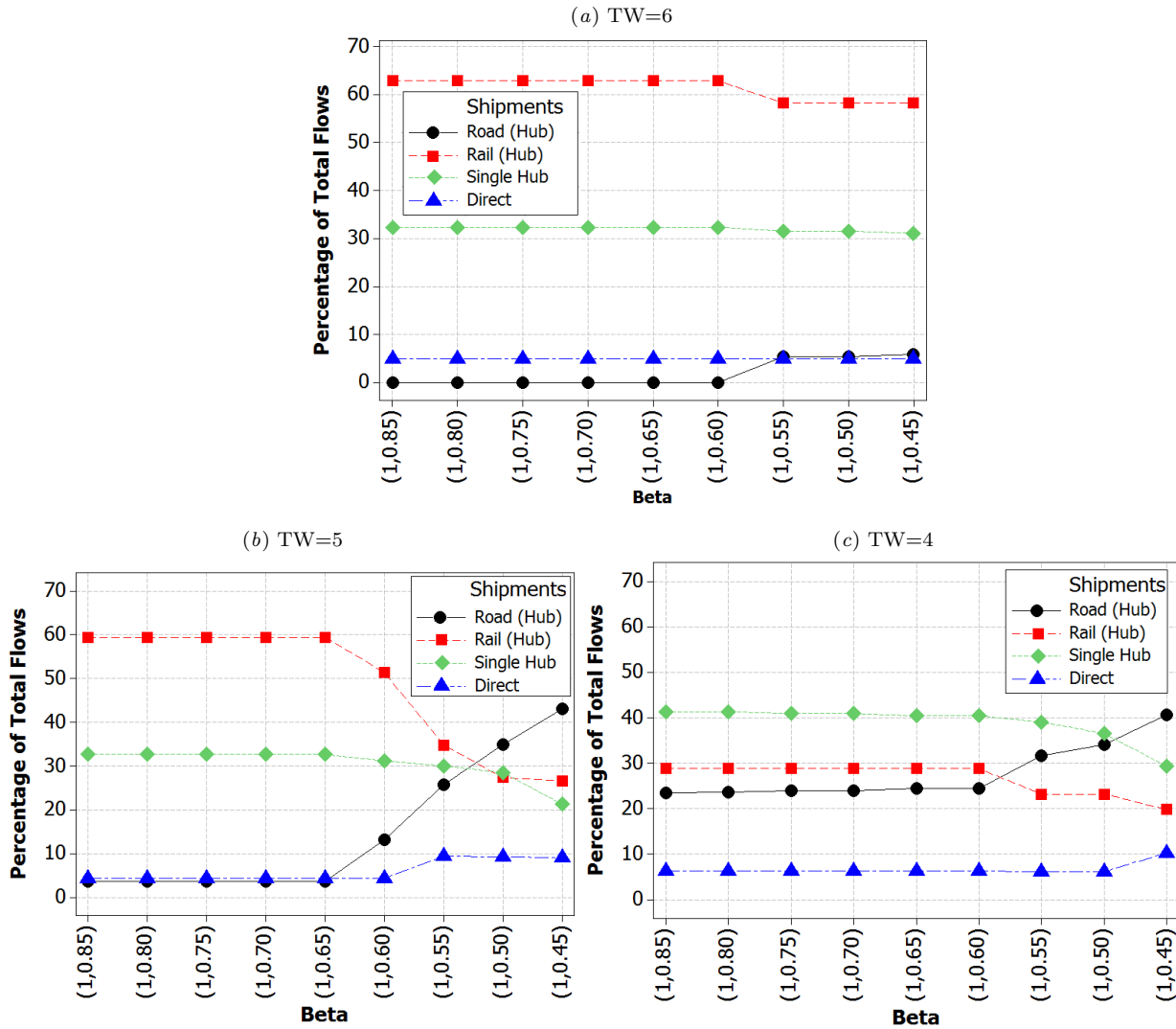


Figure 12: Effects of Economies-of-Scale

These results showed that operational efficiency gains in an OTR network can help increase the use of inter-hub road shipments. But the actual impact of operational efficiency gains depend on the service requirements in the network. With longer service time requirements, hub shipments accounted for only a small proportion of the total flows and the efficiency gains did not affect the modal choice. The OTR efficiency gains allowed road hub shipments to be competitive and at a high enough level attracted more flows than rail hub shipments. However the prohibitive level of effort required to achieve a 40% efficiency gain highlights the inherent benefits of intermodal shipments over road shipments.

CONCLUSIONS

This research has explored the interplay between the financial, operational and service aspects of an intermodal logistics hub network. The contributions of this paper to the literature are in the development of a modeling framework for an intermodal logistics hub network, a metaheuristic solution approach, and managerial insights into the design and management of an inter-regional, intermodal logistics hub network. The modeling framework incorporates different types of shipments (direct, inter-hub and single-hub), different modes of transportation (road, rail and air), fixed hub operating costs, modal connectivity costs, economies-of-scale and service time requirements. This research developed a metaheuristic (tabu search) solution approach for the intermodal hub network design problem. A computational study used the measure of percent optimality gap to demonstrate the performance of the metaheuristic by comparing its solutions with the corresponding optimal solutions. The average percent optimality gap over all the test problems was 1.30%.

An empirical study was used to gain relevant managerial insights into the intermodal logistics networks. This study highlighted the differences in the structure of over-the-road (OTR) and intermodal (IM) logistics networks. The study found that the location of hubs in the OTR and IM networks differed greatly. The choice of the best hub locations in the IM networks was found to be affected by the network parameters, whereas the hub locations in OTR networks were very robust. The results identified cities which have a strategic geographical location and adequate modal network connectivity that make them very suitable candidates for hub locations. The results also showed that while the use of intermodal shipments reduced the transportation costs, the modal connectivity costs and service time requirements limited the benefits of those intermodal shipments. It was shown that the location of hubs and the design of service regions in the IM networks allowed for more hub-based shipments than the OTR networks. The study also found that changes in service time requirements required reorganizing the intermodal hub network by moving hubs and using different types of shipments and transportation modes. Under short service time requirements, the modal usage transitioned from intermodal rail to

road to single hub shipments to direct shipments and air shipments. The study also found large variations in the inbound and outbound flows at the hubs with changing service requirements. The efficiency gains in OTR networks can reduce the cost benefit of intermodal shipments, however it would require a prohibitive level of effort to achieve the necessary efficiency gain.

Additional research can provide further understanding about the design of intermodal hub networks. Future research that considers different service offerings, existing hub locations, the effect of limited hub capacity, and the impact of congestion on the use of different modes would provide more insights into this problem. Some other possible extensions are the modeling of multi-commodity flows, the use of differentiated service, the consideration of link capacities, and the use of multi-modal freight flows along each lane instead of the all-or-nothing approach used in this research. Additional empirical studies on the global movement of freight would also extend the scope of this research into the global logistics arena.

REFERENCES

- Alumur, S., Kara, B. Y., 2008. Network hub location problems: The state of the art. *European Journal of Operational Research* 190 (1), 1–21.
- Arnold, P., Peeters, D., Thomas, I., 2004. Modelling a rail/road intermodal transportation system. *Transportation Research Part E: Logistics and Transportation Review* 40 (3), 255–270.
- Arnold, P., Peeters, D., Thomas, I., Marchand, H., 2001. For an optimum location of the intermodal centers of transshipment : Formulation and extensions. *The Canadian Geographer* 45 (3), 427–436.
- Aykin, T., 1995. Networking policies for Hub-and-Spoke systems with applications to the air transportation systems. *Transportation Science* 29 (3), 201–221.
- Beuthe, M., Bouffieux, C. H., 2008. Analysing qualitative attributes of freight transport from stated orders of preference experiment. *Journal of Transport Economics and Policy* 42, 105–128.
- Bilde, O., Krarup, J., 1977. Sharp lower bounds and efficient algorithms for the simple plant location problem. *Annals of Discrete Mathematics* 1, 79–97.
- Bookbinder, J. H., Fox, N. S., 1998. Intermodal routing of Canada-Mexico shipments under NAFTA. *Transportation Research Part E: Logistics and Transportation Review* 34 (4), 289–303.

- Bryan, D. L., O'Kelly, M. E., 1999. Hub-and-Spoke networks in air transportation: An analytical review. *Journal of Regional Science* 39 (2), 275–295.
- Calthrop, E., Proost, S., 1998. Road transport externalities. *Environmental and Resource Economics* 11 (3/4), 335–348.
- Campbell, J. F., 1992. Location and allocation for distribution systems with transshipments and transportation economies of scale. *Annals of Operations Research* 40 (1), 1572–9338.
- Campbell, J. F., 1994. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research* 72 (2), 387–405.
- Campbell, J. F., 1996. Hub locations and p-hub median problem. *Operations Research* 44 (6), 923–935.
- Campbell, J. F., 2009. Hub location for time definite transportation. *Computers & Operations Research* 36 (12), 3107–3116.
- Chen, H., Cambell, A. M., Thomas, B. W., 2008. Network design for time-constrained delivery. *Naval Research Logistics* 55 (6), 493–515.
- Colliers Research Services, 2009. A survey of commercial real estate market. Tech. rep.
- Crainic, T. G., Florian, M., Guelat, J., Spiess, H., 1990. Strategic planning of freight transportation: STAN, an interactive graphic system. *Transportation Research Record* (1283), 97–124.
- Crainic, T. G., Kim, K. H., Barnhart, C., Laporte, G., 2007. Chapter 8 Intermodal Transportation. Vol. 14. Elsevier, pp. 467–537.
- Crainic, T. G., Laporte, G., 1997. Planning models for freight transportation. *European Journal of Operational Research* 97 (3), 409–438.
- Erlenkotter, D., 1978. A dual-based procedure for uncapacitated facility location. *Operations Research* 26 (6), 992–1010.
- Ernst, A. T., Krishnamoorthy, M., 1996. Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location Science* 4 (3), 139–154.
- Ernst, A. T., Krishnamoorthy, M., 1998. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *European Journal of Operational Research* 104 (1), 100–112.
- Ernst, A. T., Krishnamoorthy, M., 1999. Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research* 86, 141–159.
- FMCSA, 2008. Hours of service of drivers. *Federal Register* 73 (224), 69567–69586.
- Forkenbrock, D. J., 1991. External costs of intercity truck freight transportation. *Transportation Research Part A: Policy and Practice* 33 (7/8), 505–526.

- Forkenbrock, D. J., May 2001. Comparison of external costs of rail and truck freight transportation. *Transportation Research Part A: Policy and Practice* 35 (4), 321–337.
- Galvo, R. D., Raggi, L. A., 1989. A method for solving to optimality uncapacitated location problems. *Annals of Operations Research* 18 (1), 225–244.
- Giordano, J. N., 2008. Economies of scale after deregulation in LTL trucking. *Managerial and Decision Economics* 29, 357–370.
- Glover, F., 1989. Tabu search - part I. *ORSA Journal on Computing* 1 (3), 190–206.
- Glover, F., 1990. Tabu search - part II. *ORSA Journal on Computing* 2 (1), 4–32.
- Gooley, T. B., 1997. One location, many options. *Logistics Management and Distribution Report* 37 (11), 81–89.
- Groothedde, B., Ruijgrok, C., Tavasszy, L., 2005. Towards collaborative, intermodal hub networks: A case study in the fast moving consumer goods market. *Transportation Research Part E: Logistics and Transportation Review* 41 (6), 567–583.
- Guelat, J., Florian, M., Crainic, T. G., 1990. A multimode multiproduct network assignment model for strategic planning of freight flows. *Transportation Science* 24 (1), 25–39.
- Harker, P. T., 1987. Predicting intercity freight flows. VNU Science Press BV.
- Horner, M. W., O’Kelly, M. E., 2001. Embedding economies of scale concepts for hub network design. *Journal of Transport Geography* 9 (4), 255–265.
- IANA, Aug. 2008. Domestic Intermodal Gains Accelerate. Intermodal Association of North America.
- Janic, M., 2007. Modelling the full costs of an intermodal and road freight transport network. *Transportation Research Part D: Transport and Environment* 12 (1), 33–44.
- Klincewicz, J. G., 1991. Heuristics for the p-hub location problem. *European Journal of Operational Research* 53 (1), 25–37.
- Klincewicz, J. G., 1992. Avoiding local optima in the p-hub location problem using tabu search and GRASP. *Annals of Operations Research* 40 (1), 283–302.
- Klincewicz, J. G., 2002. Enumeration and search procedures for a hub location problem with economies of scale. *Annals of Operations Research* 110 (1), 107–122.
- Limbourg, S., Jourquin, B., 2009. Optimal rail-road container terminal locations on the european network. *Transportation Research (E)* 45 (4), 551–563.
- Lozano, A., Storchi, G., Mar. 2001. Shortest viable path algorithm in multimodal networks. *Transportation Research Part A: Policy and Practice* 35 (3), 225–241.
- Macharis, C., Bontekoning, Y. M., 2004. Opportunities for OR in intermodal freight transport research: A review. *European Journal of Operational Research* 153 (2), 400–416.

- McCalla, R. J., 1999. Global change, local pain: intermodal seaport terminals and their service areas. *Journal of Transport Geography* 7 (4), 247–254.
- Merrina, A., Sparavigna, A., Wolf, R., 2007. The intermodal networks: A survey on intermodalism. *World Review of Intermodal Transportation Research* 1 (3), 286–299.
- O’Kelly, M. E., 1986. The location of interacting hub facilities. *Transportation Science* 20 (2), 92–106.
- 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., Bryan, D. L., 1998. Hub location with flow economies of scale. *Transportation Research Part B: Methodological* 32 (8), 605 – 616.
- Racunica, I., Wynter, L., 2005. Optimal location of intermodal freight hubs. *Transportation Research Part B: Methodological* 39 (5), 453–477.
- Rardin, R. L., Uzsoy, R., 2001. Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics* 7, 261–304.
- Schwartz, M., 1992. J.B. Hunt: The Long Haul to Success. The University of Arkansas Press.
- Skorin-Kapov, D., Skorin-Kapov, J., 1994. On tabu search for the location of interacting hub facilities. *European Journal of Operational Research* 73 (3), 502–509.
- Skorin-Kapov, D., Skorin-Kapov, J., O’Kelly, M. E., 1996. Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research* 94 (3), 582–593.
- Slack, B., 1990. Intermodal transportation in north america and the development of inland load centers. *Professional Geographer* 42 (1), 72–83.
- Slack, B., 2001. Intermodal transportation. In: Brewer, A., Button, K. J., Hensher, D. A. (Eds.), *Handbook of Logistics and Supply Chain Management*, 1st Edition. Pergamon, p. 700.
- Southworth, F., Peterson, B. E., 2000. Intermodal and international freight network modeling. *Transportation Research Part C: Emerging Technologies* 8 (1-6), 147–166.
- UPS, 2008. Facts sheet: UPS operations. <http://www.pressroom.ups.com/Fact+Sheets>.
- Waller, D. L., 2003. *Operations Management: A Supply Chain Approach*, 2nd Edition. Cengage Learning Business Press, p. 1088.
- Warsing, D. P., Souza, G. C., Greis, N. P., 2001. Determining the value of dedicated multimodal cargo facilities in a multi-region distribution network. *European Journal of Operational Research* 133 (1), 81–93.
- Yamada, T., Russ, B. F., Castro, J., Taniguchi, E., 2009. Designing Multimodal Freight Transport Networks: A Heuristic Approach and Applications. *Transportation Science* 43 (2), 129–143.

ARTICLE 3: DESIGN OF INTERMODAL NETWORKS UNDER HUB CONGESTION

A hub network is composed of a few central locations (relative to the total number of candidate locations) which act as hubs. Such network designs are graphically represented by a wheel with the hub at the center, connected to the non-hub cities through the spokes of the wheel. For this reason hub networks are also known as *hub-and-spoke* networks. The use of hub-and-spoke designs for many-to-many distribution networks has increased significantly in the last thirty years. This network structure has been the operational structure of small package carriers and the modern airline industry (Pirkul and Schilling, 1998).

A hub-and-spoke network derives its benefits from the economies of scale achieved through the aggregation of flows between origins and destinations. The flows are attracted to hubs due to the reduced unit transportation cost achieved by consolidating large flows which are shipped between the hubs. This scheme reduces total transportation costs of the logistics system. Though shipments pass through no more than two hubs, the hub-and-spoke scheme does increase the total distance traveled and the total shipment time compared to a direct move of a shipment to its destination. Current logistics service providers cannot ignore this limiting effect on customer service. Companies are looking for logistics network designs that reduce total shipment times while also reducing costs (Aykin, 1995).

The overall performance of hub-and-spoke networks depends on the operational performance of the hubs. Due to limited resources (material handling equipment, workforce, number of loading/un-loading docks and the size of staging areas) and waiting times to consolidate smaller shipments into truck-load (TL) or container-load (CL) quantities, shipments may stay at a hub for a significant time. Such delays caused by congestion at the hubs may have a dire effect on the service performance of a logistics service provider.

One of the sources of congestion at logistics hubs is the limited availability of material

handling and storage resources. This limitation causes congestion at the unloading docks, staging areas and the loading docks of a logistics hub. A mismatch between the arrival rate of shipments and the processing rate of these resources may cause queues to build. Thus an arriving shipment may have to wait for unloading. This situation, if not managed properly, can contribute to significant delays, adversely affecting the service due to an increase in the total shipment time. The second potential source of congestion at a logistics hub is the transit time variability of different transportation modes. The two most common transportation modes used for domestic freight movement in the US are road and intermodal rail. Both these modes of transportation have different average transit times as well as different transit time variability between the same origins and destinations. Large variations in the shipment transit times increase the uncertainty in the number of shipments that will need handling at a logistics hub over a given period of time.

Due to these issues, congestion at logistics hubs is a significant concern for logistics service providers. This situation is further exacerbated in a hub network structure which seeks to allocate large flow volumes to the hubs to reduce total network costs (Grove and O’Kelly, 1986). Thus the decisions regarding the location of hubs and the allocation of customer flows to these hubs directly affect the rate of shipment arrivals at a hub. If the arrival rate of shipments is close to the processing rate of the hub resources, queues start to build up and cause delays due to congestion.

This research explores the effect of resource limitations and modal transit time variability on a hub network design under service time requirements. This research integrates the multiple-allocation p-hub median modeling approach for the interacting hub location-allocation problem with a queuing model for hub operations. The former approach selects the location of intermodal logistics hubs and allocates origin-destination pairs to design a hub-and-spoke logistics network. The latter approach is based on a queuing network to model the operations of an intermodal logistics hub and develops estimates of expected waiting times for individual shipments (treated as product classes with deterministic routings). Through the use of multiple product classes, the waiting times can be associated with specific shipments and transportation modes. In combining

these above mentioned aspects, this research investigates the tradeoffs between logistics costs, hub flows, use of intermodal service, total shipment times and service time requirements.

The rest of this article is organized as follows: First, prior research in the area of logistics network design under congestion is presented. Next, the modeling framework and a mathematical model for an intermodal logistics hub network is presented. A framework for modeling operations in a logistics hub is presented which is used to develop a multi-class serial queuing system with deterministic routings. Next, a solution approach which uses a tabu search metaheuristic is developed. Managerial insights are developed into the design of intermodal networks under hub congestion through a case study based on real world logistics data. The article concludes by summarizing the findings of this research.

LITERATURE

This section presents a review of the literature on hub network design under congestion and establishes the contributions of this research. This research draws from existing work in the following areas; (a) use of queuing models in hub location-allocation decisions, (b) steady state analysis of queuing systems, and (c) multiple job classes with routings in a queuing network. The research from these areas is extended to the intermodal hub network design arena.

The integration of queuing models in the hub location-allocation decisions is a relatively new area. The steady state analysis of the queuing systems can be used to study the impact of congestion on the design of hub networks. The relationship between hub-and-spoke networks and congestion was initially investigated by Grove and O’Kelly (1986). Using a simulation of a hub-and-spoke network they showed that when the location of the hubs are fixed, the assignment of flows to the hubs determine the delays at the hubs and consequently the total shipment time. To the best of our knowledge, among the earliest attempts at including congestion endogenously within a hub network design is by Marianov and Serra (2003) in which the airport hubs are modeled as a *single* stage M/D/c queuing system, i.e., poisson arrivals, deterministic service times and multiple servers. A probabilistic constraint is imposed that restricts the probability of the number of waiting airplanes to be greater than the queue capacity, to be within a low

(pre-specified) value. This approach selects hub locations and allocates flows such that the average number of airplanes in the queue is within an acceptable range. Elhedhli and Hu (2005) also modeled the effect of congestion in a network design by adding a convex (power) cost function to the objective function. If more flows are assigned to a hub the value of the cost function increases non-linearly. This approach balances the assignment of flows to the hubs such that the total congestion costs in the network are reduced. A similar approach is used by de Camargo et al. (2008) in which the network design seeks to balance the transportation cost savings and the costs associated with congestion effects.

It should be noted that the approach used in modeling congestion in this research differs from others, which have used capacity constraints (Aykin, 1994; Ernst and Krishnamoorthy, 1999; Ebery et al., 2000). The use of capacity constraints does not account for the variable nature of congestion effects related to variations in volume of flows passing through the hubs (de Camargo et al., 2008). Instead this research uses the approach of queuing models and steady state approximations of waiting times to study the network design problem under conditions of congestion.

Queuing models have been used before in the context of general network design. Although the papers discussed below deal only with the assignment sub-problem within a hub network design and do not include the location aspect, they do demonstrate the use of a queuing system to model congestion delays at the hubs. One of these works is Warsing et al. (2001) in which the average waiting time of a shipment in a hub is computed from the steady state approximations of a GI/G/c queuing system. The steady state requirement in the use of these approximations is implemented in the constraints through the use of the utilization factor ($\rho < 1$). Another work is that of Rodriguez et al. (2007) in which a hub is modeled as an M/M/1 queuing system, where trucks arrive and wait in the queue if the unloading resources are busy. The arrival rate depends on the number of trucks assigned to the hub. A hub with a higher number of assigned trucks has higher waiting times. A tradeoff exists between the assigned number of trucks and the waiting times, under a maximum shipment time requirement. The model uses penalty costs for violating the maximum travel time limit for flows which are not delivered in time.

Table 30: Prior Literature - Intermodal Hub Networks Under Congestion

Research Component	Current Paper	Marianov and Serra (2003)	Elhedhli and Hu (2005)	de Camargo et al. (2008)
Multiple Transportation Modes	✓	-	-	-
Modal Transit Variability	✓	-	-	-
Hub-and-Spoke Structure	✓	✓	✓	✓
Modeling Approach (Network)	p-Hub Median	p-Hub Median	p-Hub Median	p-Hub Median
Modeling Approach (Congestion)	GI/G/1 Queue	M/D/c Queue	Non-linear Cost	Non-linear Cost
Multiple Product Classes	✓	-	-	-
Type of Routing	Deterministic	-	-	-
Service time requirements	✓	-	-	-

This research also draws from the area of queuing systems with multiple job classes and routings. Bitran and Tirupati (1988) are among the first who worked with deterministic routing of job classes in a GI/G/1 system. Their work was based on the aggregate class approach of Whitt (1983), who developed mathematical expressions for approximating the mean arrival rate, squared coefficient of variability in arrival times, mean service time and squared coefficient of variability in departure times of an M/M/1 system with probabilistic routings. Whitt (1988) later updated his 1983 work to include the deterministic routing case. For a detailed survey of GI/G/1 and GI/G/m queuing systems with probabilistic and deterministic routings, see Bitran and Morabito (1996).

The other area of research which is relevant to this work is that of interacting hub location-allocation research domain. O’Kelly (1986) presented the first known quadratic integer formulation of an interacting hub location-allocation problem. Campbell (1994) later developed a linear version of this problem. This research is related to the multiple-allocation version of Campbell’s work. In a multiple-allocation solution, each origin may use different hub-pairs to send its shipments. New MILP formulations of the hub location-allocation problem with fewer variables and constraints were developed by Ernst and Krishnamoorthy (1998). For a more detailed review of hub location-allocation problems, see Alumur and Kara (2008).

In order to highlight the contributions of this research, Table 30 summarizes the contents of relevant articles and compares them with this research. It can be seen that this research extends the literature on hub network design under congestion by including multiple modes of transportation, variability in transit times associated with different transportation modes, modeling shipments as distinct classes with deterministic routings through a logistics hub and service time requirements.

In the next section, a modeling framework is discussed that provides the foundation to develop a mathematical model for the intermodal hub network design and a queuing system that represents the flow of shipments in a logistics hub.

MODELING FRAMEWORK

This section presents a modeling framework for the design of an intermodal hub-and-spoke logistics network. This framework also models the operations within a hub using a queuing system. The hub operations include processes such as unloading, consolidation, break-bulk and loading. This section also lays the ground work for developing traffic rate and variability equations to estimate the expected waiting times for the shipments passing through each hub. Using this framework, a mathematical model is presented which integrates the hub location-allocation formulation with waiting time estimates from the queuing system.

An intermodal hub-and-spoke logistics network is represented by a graph $\mathbb{G}(\mathbb{N}, \mathbb{A})$ in which \mathbb{N} is the set of nodes (i.e., cities) and \mathbb{A} is the set of arcs which represent different modes of transportation between different cities. The set of arcs \mathbb{A} can be partitioned into two disjoint subsets: \mathbb{A}_1 and \mathbb{A}_2 , representing the two modes of transportation: road and rail, respectively. These subsets of arcs describe the modal network connectivity between the cities. A shipment between two cities may travel over any of the available arcs representing the choice of the mode of transportation.

In this modeling framework a shipment between an origin city i and a destination city j can travel either through a pair of hubs (k, m) or through a single hub. Figure 13 shows the different ways a shipment may travel between the origin and the destination. In the case of

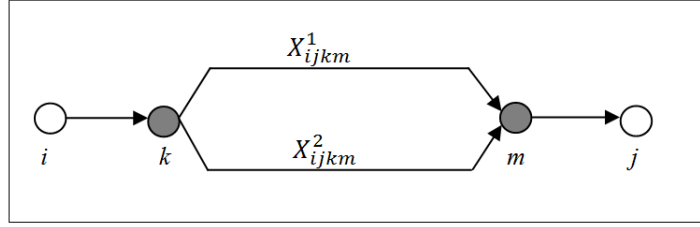


Figure 13: Types of Shipments

a two-hub shipment, the journey between the origin and the destination involves three legs of travel. The first leg is the pickup from the origin city i and travel to the origin hub k using road transportation. The second leg is the inter-hub travel between the origin and destination hubs, k and m . This is the line haul where the shipments move in larger quantities, with higher frequencies and may use any one of the arcs. The choice of the specific arc represents the use of the corresponding transportation mode, represented by the binary variable X_{ijkm}^t , where t represents the two modes of transportation: $t=1$ for road and $t=2$ for rail. The third leg is the drop off from hub m to the destination city, j using road transportation. In the case of a single hub shipment ($k = m$), the journey consist of only two legs: travel from the origin city to the hub and the drop off from the hub to the destination.

Although a specific mode may be selected for the inter-hub transfer of shipments for a specific origin-destination pair, other origin-destination pairs may ship through the same hubs using a different mode of transportation. Thus for a specific origin-destination pair (i, j) with hub pair (k, m) , $X_{ijkm}^1 = 1$ implies road travel and $X_{ijkm}^2 = 1$ implies rail travel between the hub cities. For each origin-destination pair, only one mode is selected.

In a logistics network the total shipment time is a combination of transit times between origin and destination cities and the time taken to handle and process shipments at the hub(s). A shipment arriving at a hub is classified as one of the following: (a) *Hub Outbound*, which is a shipment which arrives from its origin city and will depart for another hub, (b) *Hub Inbound*, which is a shipment which arrives from another hub in the network, and (c) *Local*, which is a shipment which has its origin and destination serviced by the same hub. Each shipment classification is associated with a specific hub operations routing scheme.

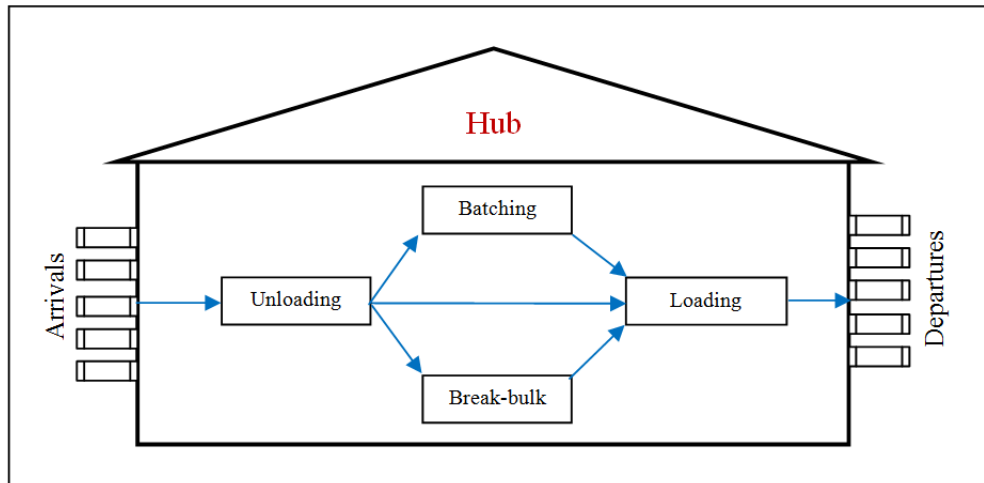


Figure 14: Hub Operations

In a logistics hub, four different types of operations are undertaken. These operations are: (a) unloading, (b) batching, (c) break-bulk, and (d) loading. The *unloading* operation deals with activities such as the assignment of shipment classifications, the assignment of unloading docks, and the physical unloading of shipments. The *batching* operation deals with the activities of moving the unloaded shipments to the staging areas according to their respective classifications, where the shipments wait until truck-load or container-load quantities are accumulated. The *break-bulk* operation is composed of separating the truck-loads/container-loads into individual shipments according to their destinations. The *loading* operation deals with the movement of shipments to the loading docks and the physical loading activities. In the case of road transportation, shipments leave directly from the loading docks to destination hubs (as in the case of hub outbound shipments) or to their final destinations (as in the case of local shipments). In the case of intermodal rail transportation, a drayage company picks up the load (container) and takes it to an intermodal rail terminal. The limited resources available for each of these operations may cause congestion at the hubs if the average rate of shipment arrivals approaches or exceeds the processing rate of these resources. The congestion increases as more and more flows are attracted to the hub to take advantage of the economies of scale.

The movement of shipments through a hub is illustrated in Figure 14. When a shipment arrives at a hub, it is classified as one of the following types: *Hub Outbound*, *Hub Inbound* or

Local. At the receiving (unloading) dock, a shipment may have to wait for the material handling equipment for unloading. After a shipment is unloaded, its flow through the hub depends on its routing. A *Hub Outbound* shipment moves to a staging area (batching) where it waits for other shipments with the same destination hub and mode of transportation. Once a *moving load* (i.e., truck-load/container-load) is completed, it is dispatched to the loading docks for loading onto an outbound truck or an intermodal-rail container. The routing of a *Hub Inbound* shipment is different. Such a shipment is moved to the break-bulk operation. These shipments then move to the loading docks for dispatch to their final destinations. In the case of a *local shipment*, it is moved directly to the loading docks for dispatch. Note that in this research no consolidation is modeled for the final leg of travel.

The total time spent by a shipment in a logistics hub (called *dwell time*) is a combination of the processing times of each operation it visits and the time a shipment may have to wait for its turn to be processed. The average waiting (queuing) times are represented by $W_k^O(\mathbf{X})$, where k represents a hub, O represents unloading, batching, break-bulk or loading operations and \mathbf{X} is a vector of X_{ijkm}^t values. The average waiting times depend on the service rate of hub operations and the flows allocated to hubs, represented by the X_{ijkm}^t values.

The unit transportation costs (c_{ij}^t) of sending a shipment between an origin-destination pair depends on the mode of transportation. The unit road transportation costs (c_{ij}^1) are given on a per shipment basis and the unit intermodal rail costs c_{ij}^2 are given on a per container basis. The demand flows between an origin-destination pair of cities (i, j) and the service time requirements are given by f_{ij} and TW_{ij} , respectively. Owing to the economies of scale, transportation costs are discounted over the hub-to-hub link using a linear approximation (α) that represents the discounted inter hub cost structure (Campbell, 2009; Limbourg and Jourquin, 2009; Alumur and Kara, 2008). The fixed costs of operating a logistics hub k is represented by F_k , and the cost of providing intermodal services is given by MC_{kt} .

MODEL FORMULATION

The mathematical formulation is as follows:

Sets:

T = {1,2}, set of modes,

N = Set of cities,

R_O = {unload, batch, load}, Routing of Hub Outbound shipments,

R_I = {unload, break-bulk, load}, Routing of Hub Inbound shipments,

R_L = {unload, load}, Routing of Local shipments.

Decision Variables:

$$y_k = \begin{cases} 1 & , \text{ if city } k \text{ is a hub} \\ 0 & , \text{ otherwise,} \end{cases}$$

$$X_{ijkm}^t = \begin{cases} 1 & , \text{ if shipments from } i \text{ to } j \text{ flow through hub pair } (k,m) \text{ using mode } t \\ 0 & , \text{ otherwise,} \end{cases}$$

$$S_{kt} = \begin{cases} 1 & , \text{ if hub } k \text{ is served by mode } t \in T \\ 0 & , \text{ otherwise,} \end{cases}$$

Parameters:

- p = Number of hubs,
 f_{ij} = Flow volume from the origin city i to the destination city j ,
 c_{ij}^t = Unit transportation cost from city i to city j using mode t ,
 α = Inter-hub transportation discount factor for road transportation, $0 < \alpha < 1$,
 F_k = Fixed cost of opening and operating a hub in city k ,
 MC_{kt} = Modal connectivity cost of serving mode t at hub k for $t \in \{1, 2\}$,
 CL = Size of an intermodal rail container,
 T_{ij}^t = Travel time between cities i and j using mode t ,
 $W_k^O(\mathbf{X})$ = Average waiting time for operation O at hub k , as a function of X_{ijkm}^t values
 TW_{ij} = Service time requirement for a shipment between origin i and destination j ,
 s_k^O = Average processing time of operation O at hub k .

Model Formulation:

$$\begin{aligned}
 \text{minimize} \quad & \sum_{k \in N} F_k y_k + \sum_{k \in N, t \in T} MC_{kt} S_{kt} + \sum_{i, j, k, m \in N} (\alpha c_{km}^1 f_{ij} X_{ijkm}^1 + c_{km}^2 \frac{f_{ij}}{CL} X_{ijkm}^2) \\
 & + \sum_{i, j, k, m \in N, t \in T} f_{ij} (c_{ik}^1 + c_{mj}^1) X_{ijkm}^t \quad (38)
 \end{aligned}$$

subject to:

$$\sum_{k \in N} y_k = p \quad (39)$$

$$X_{ijkm}^t \leq y_k, \quad \forall i, j, k, m \in N, t \in T \quad (40)$$

$$X_{ijkm}^t \leq y_m, \quad \forall i, j, k, m \in N, t \in T \quad (41)$$

$$\sum_{k, m \in N, t \in T} X_{ijkm}^t = 1, \quad \forall i, j \in N \quad (42)$$

$$X_{ijkm}^t \leq S_{kt} \quad , \forall i, j, k, m \in N, t \in T \quad (43)$$

$$X_{ijkm}^t \leq S_{mt} \quad , \forall i, j, k, m \in N, t \in T \quad (44)$$

$$\begin{aligned} \sum_{k \neq m \in N, t \in T} X_{ijkm}^t [T_{ik}^1 + \sum_{O \in R_O} W_k^O(\mathbf{X}) + \sum_{O \in R_O} s_k^O + T_{km}^t + \sum_{O \in R_I} W_m^O(\mathbf{X}) \\ + \sum_{O \in R_I} s_m^O + T_{mj}^1] \leq TW_{ij} \quad , \forall i, j \in N \end{aligned} \quad (45)$$

$$\sum_{k \in N} X_{ijkk}^1 [T_{ik}^1 + \sum_{O \in R_L} W_k^O(\mathbf{X}) + \sum_{O \in R_L} s_k^O + T_{kj}^1] \leq TW_{ij} \quad , \forall i, j \in N \quad (46)$$

$$y_k, X_{ijkm}^t, S_{kt} \in \{0, 1\} \quad , \forall i, j, k, m \in N, t \in T$$

The first term in the objective function (38) is the total fixed costs of operating the hubs. The hub locations are identified by the decision variables y_k . The second term in the objective function is the total cost of intermodal service at the hubs. The road and intermodal rail connectivity at each hub are identified by the decision variables S_{k1} and S_{k2} , respectively. The variables S_{k1} and S_{k2} are set equal to 1 when a hub k uses road or intermodal rail services, respectively. The third term in the objective function deals with the transportation costs of the hub-to-hub shipments. The last term in the objective function includes the transportation cost of pickup from the origin and the transportation cost of drop off to the final destination.

Constraint (39) requires that a total of p hubs are opened. Note that this constraint may be omitted to solve the model for finding the best number of hubs (p). Constraints (40) and (41) ensure that any hub pair assignment for an origin-destination pair is limited to open hubs. Constraint (42) selects one hub pair (k, m) for each origin-destination pair (i, j) and a specific mode t . Constraints (43) and (44) are used to set the values of decision variables S_{kt} which are used in computing the total cost of intermodal service. If any of the X_{ijkm}^t variables in constraints (43) and (44) is equal to 1, then the corresponding S_{kt} and S_{mt} variables must be set equal to 1 in the solution. Note these constraints in the model allow for a hub to use different modes for its inbound and outbound shipments.

This model restricts the total time of a shipment to be within a certain service time window

(TW_{ij}). Constraint (45) models this restriction for a shipment that uses two hubs. The total shipment time consists of transit times, processing and waiting times at hubs. Each shipment takes T_{ik}^1 time to travel from origin city to the origin hub by road, T_{km}^t time units for inter-hub travel using mode t , and T_{mj}^1 time to reach the final destination by road. The other component of total shipment time is the time spent by a shipment in logistics hub(s). These times depend on the routing of individual shipments. The routings of hub outbound, hub inbound and local shipments are given by $R_O = \{unload, batch, load\}$, $R_I = \{unload, breakbulk, load\}$ and $R_L = \{unload, load\}$, respectively. When a shipment arrives at origin hub k , it may have to wait for the unloading resources to be available, given by W_k^{unload} . The unloading operation on the average takes s_k^{unload} time units. After unloading, a shipment moves to the batch operation, where it waits for W_k^{batch} time units. The batch wait times depend on the arrival rate of other shipments with the same destination hub and mode of transportation. As TL/CL quantities are collected, shipments move to the loading operation and are loaded onto a trailer or an intermodal rail container. A shipment may have to wait for W_k^{load} time units if the loading resources are not available. The loading process on the average takes s_k^{load} time units. At the destination hub m , shipments are unloaded taking $(W_m^{unload} + s_m^{unload})$ time units and undergo a break-bulk process which takes $(W_m^{breakbulk} + s_m^{breakbulk})$ time units. Finally, the shipments are loaded onto the trucks and dispatched for final destination with $(W_m^{load} + s_m^{load})$ time units. Constraint (46) models the service time restriction for single-hub shipments. In this case, a shorter travel time is needed since there is no need for hub-to-hub travel. The total shipment time is a combination of transit times to and from the hub, processing and waiting times at unload and load hub operations.

QUEUEING SYSTEM FOR HUB OPERATIONS

A queueing system for logistics hub operations is composed of shipments (jobs) and processes (stations). Each station contains the following elements: (a) arrival process, (b) service process and (c) waiting queues. The arrival process at a station consists of jobs (shipments) arriving with a probabilistic inter-arrival time. It is assumed that shipments arrive with inter-arrival

times which are independent of other arrivals but are all identically distributed. Such types of arrivals are referred to as a GI-arrival process or a renewal process. A special case of a renewal process is a Markovian process where the inter-arrival times are exponentially distributed. The stations in this queueing system are loading, unloading, batching and break bulk. In a queueing system there may be single or multiple servers at each station with service times that are independent of jobs (shipment) types. The waiting queues are assumed to have infinite capacity, which implies that once a shipment enters the queueing system, it does not leave. The queues are processed on a first-come-first-serve basis without any priority scheme assigned to the waiting jobs or a preemption rule. A new shipment arriving at a station may have to wait because the server is busy processing an earlier shipment. Each shipment visits a station according to its routing scheme.

The existing literature provides different approaches for analyzing queueing networks, such as (a) exact methods, (b) approximation methods and (c) simulation techniques. Exact methods exist for Markovian processes. However in many real world applications the processes are less variable than a poisson process thus violating its assumptions. In the absence of exact methods for open queueing networks, approximation and simulation methods are used. The benefit of these approaches lie in their assumptions that are closer to reality.

The GI/G/1 queueing system used in this research is analyzed using the decomposition approach. The decomposition approach for open queueing networks allows for each station to be analyzed individually as a GI/G/1 system. The decomposition approach makes use of superposition of arrivals and splitting of departures. The former relates to the combining of individual arrivals from other stations into a single merged arrival, and the latter decomposes the merged departures from a station into individual departures to other stations. The flows in the queueing system are approximated using the first two moments of the probability distributions (mean and variance).

In the analysis, the following assumptions are made:

1. Arrivals at hubs have independent and identically distributed inter-arrival times.

2. The system has single server stations with first-come-first-serve (FCFS) queue discipline.
3. The service times at each station are assumed to be independent and identically distributed.
4. General distributions model arrival and service times.
5. Average daily arrival rate does not change over time and the annual arrival rate is the sum of the daily arrival rates.

ANALYZING OPEN QUEUING SYSTEMS

As mentioned above, the queuing system considered in this research consists of three processes: (a) arrival process, (b) service process and (c) departure process. Each station is characterized by the following parameters:

$$\begin{aligned}
 a_{ij} &= \text{inter-arrival time at station } j \text{ from station } i, \\
 \lambda_{ij} &= \text{expected arrival rate at station } j \text{ from station } i, \lambda_{ij} = \frac{1}{E(a_{ij})}, \\
 ca_{ij} &= \text{squared coefficient of variation of inter-arrival times at station } j \text{ for} \\
 &\quad \text{arrivals from station } i, ca_{ij} = \frac{\text{var}(a_{ij})}{E(a_{ij})^2}, \\
 s_j &= \text{service time of station } j, \\
 \mu_j &= \text{expected service rate of station } j, \mu_j = \frac{1}{E(s_j)}, \text{ and} \\
 cs_j &= \text{squared coefficient of variation of service time of station } j, cs_j = \frac{\text{var}(s_j)}{E(s_j)^2}.
 \end{aligned}$$

In the case of merged (superposition) arrivals at station j from multiple stations, a system of linear equations known as *Traffic Rate Equations* can be written as follows:

$$\lambda_j = \sum_{i=0}^n \lambda_{ij} \quad , \forall j. \tag{47}$$

In general, the expected arrival rate from station i to station j , $\lambda_{ij} = q_{ij}\lambda_i$, where q_{ij} is the probability of going from station i to station j . Note that λ_{0j} refers to external arrivals. The

variability in inter-arrival times for merged arrivals at station j (ca_j) is approximated by the convex combination of inter-arrival time variability for arrivals from all other stations (Sevcik et al., 1977),

$$ca_j = \sum_{i=0}^n \left(\frac{\lambda_{ij}}{\lambda_j} \right) ca_{ij} \quad , \forall j. \quad (48)$$

Whitt (1983) refined this approximation using the following equations,

$$\begin{aligned} ca_j &= w_j \left[\sum_{i=0}^n \frac{\lambda_{ij}}{\lambda_j} ca_{ij} \right] + 1 - w_j \quad , \text{ where} \\ w_j &= [1 + 4(1 - \rho_j)^2(\nu_j - 1)]^{-1} \quad , \text{ and} \\ \nu_j &= \left[\sum_{i=0}^n \left(\frac{\lambda_{ij}}{\lambda_j} \right)^2 \right]^{-1}. \end{aligned} \quad (49)$$

Note that $\rho_j = \frac{\lambda_j}{\mu_j}$ is the traffic intensity at station j ($0 \leq \rho < 1$). The departure process is a function of the arrival process (λ_j, ca_j) and the service process (μ_j, cs_j) . With a FCFS queue discipline and single server stations, the departure rate from a station is equal to the arrival rate (steady state condition). The variability in times between departures (cd_j) is approximated by Whitt (1984) as follows:

$$cd_j = \rho_j^2 cs_j + (1 - \rho_j^2) ca_j \quad ; \forall j. \quad (50)$$

In a departure splitting process, the merged expected departure rate λ_j and merged inter-departure time variability (cd_j) are decomposed into departure rate ($\lambda_{ji} = q_{ji} \lambda_j$) and inter-departure time variability (cd_{ji}). Sevcik et al. (1977) gives the following approximation,

$$cd_{ji} = q_{ji} cd_j + 1 - q_{ji}. \quad (51)$$

The system of equations based on (49),(50) and (51) is known as *Traffic Variability Equations*. The traffic rate equations and traffic variability equations allow for the characterization of each station j in terms of its parameters $(\lambda_j, ca_j, \mu_j, cs_j)$.

OPEN QUEUING SYSTEM WITH MULTIPLE JOB CLASSES

The use of job classes is suitable in queuing networks where jobs entering the system have distinctive characteristics. These characteristics may be specified by the routing, processing time, process and move batch sizes. The open queuing system with multiple job classes is analyzed by aggregating the job classes into a single class (Whitt, 1983). This single aggregated class system is analyzed using the GI/G/1 approach under the approximate assumption that nodes (stations) are stochastically independent (Bitran and Morabito, 1996).

In this research, a shipment (job) arriving at a hub is characterized by a specific class which depends on its destination, mode of inter-hub transportation and destination hub. Each of these job classes have a corresponding deterministic routing which describes its flow through the hub queuing system. The first-come-first-serve queuing rule and a single server system imply that the sequence of job departures is identical to that of the arrivals. The routing of each job class refers to the sequence of stations a job will visit.

The traffic rate equations for the case of multiple job classes can be expressed in the same way as discussed in the previous section. However the traffic variability equations require some more work. In queuing systems, variability is caused by the stochastic nature of the arrival and service processes. When the arrivals have multiple job classes, there is an added source of variability. This variability occurs when the departure process of a job stream at any station is distorted by the presence of other job classes. An approach to handle such a situation is discussed by Bitran and Morabito (1996) using the following notations,

n = number of stations,

R = number of classes,

n_r = number of operations for class r ,

λ_r = expected arrival rate of class r ,

ca_{rl} = squared coefficient of variability of inter-arrival times of class r at operation l ,

n_{rl} = station visited for operation l in class r routing,

In the case of deterministic routings, the expected arrival rate at station j can be computed by adding the arrival rates of all the classes which use station j :

$$\lambda_j = \sum_{r=1}^R \sum_{l=1}^{n_r} \lambda_r \mathbf{1}\{(r, l) : n_{rl} = j\}, \quad (52)$$

where $\mathbf{1}\{(r, l) : n_{rl} = j\}$ is an indicator function which takes the value of 1, if operation l of class r is done at station j , and 0 otherwise.

Also the proportion of class r arrivals out of all arrivals at station j is given by

$$q_{rj} = \frac{\lambda_r}{\lambda_j}. \quad (53)$$

Bitran and Tirupati (1988) showed that variability in inter-arrival times (ca_j) at station j and variability in inter-departure times (cd_{rj}) of job class r from station j , can be expressed as:

$$ca_j = w_j \left[\sum_{r=1}^R \sum_{l=1}^{n_r} \frac{\lambda_r \mathbf{1}\{(r, l), n_{rl} = j\} ca_{rl}}{\lambda_j} \right] + 1 - w_j, \quad (54)$$

where

$$w_j = [1 + 4(1 - \rho_j)^2(\nu_j - 1)]^{-1},$$

$$\nu_j = \left[\sum_{r=1}^R \sum_{l=1}^{n_r} \left(\frac{\lambda_r \mathbf{1}\{(r, l), n_{rl} = j\}}{\lambda_j} \right)^2 \right]^{-1},$$

where the variability in inter-arrival times of class r at operation l (ca_{rl}) is equal to the variability in the inter-departure times from the previous operation, i.e., $ca_{rl} = cd_{r,l-1}$, and

$$cd_{rj} = q_{rj} cd_j + (1 - q_{rj}) q_{rj} + (1 - q_{rj})^2 ca_{rj}, \quad (55)$$

where $j = n_{rl}$, and cd_j is given by (50).

INTERMODAL HUB QUEUING SYSTEM

The following notation is used in developing the traffic rate and variability equations for this intermodal hub queuing system.

- μ_k^s = Service rate of station s at hub k ,
- cs_k^s = Squared coefficient of variance of service time of station s at hub k ,
- λ_k^s = Arrival rate at station s of hub k ,
- ca_k^s = Squared coefficient of variance of arrival times at station s at hub k ,
- cd_k^s = Squared coefficient of variance of departure times from station s at hub k ,
- λ_{kmt}^O = Arrival rate of outbound hub shipments at hub k ,
- λ_{kt}^I = Arrival rate of inbound hub shipments at hub k ,
- λ_k^L = Arrival rate of local shipments at hub k ,
- \tilde{c}_t = Transit time variability of mode t shipments,
- \hat{c} = Transit time variability of pickup/drop off shipments,
- D = Number of days in a business year,
- L^t = Batch size of inter-hub transfers using mode t .

The queuing system for hub operations is shown in Figure 15. It is assumed that the service time of station s is independent of the job class r , i.e., $E(s_k^{sr}) = \mu_k^s$. It is also assumed that the service time variability (cs_k^s) of station s is independent of the job class. Note that two different measures of transit time variability are used, i.e., \tilde{c}_t and \hat{c} . These different measures reflect the distinction that variability in transit times of local (LTL) pickup and dropoff may be different from the variability of inter-hub shipments which use different type and size of vehicles and modes of transportation. Furthermore, note that the variability of the arrival process for the hub-inbound shipments is affected by the inter-hub transit time variability, the variability of the departure process at the origin hub and the use of truck-load/container-load (batch) deliveries.

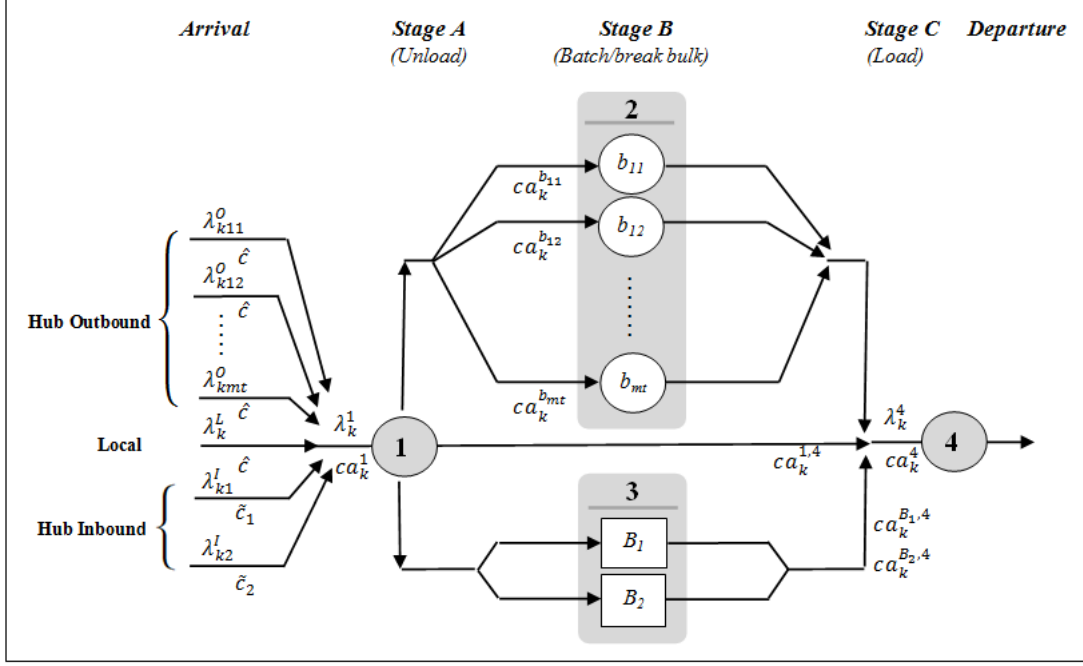


Figure 15: Hub Operations Queuing System

A simplifying assumption is made that the squared coefficient of variation of times between the arrivals of a job class (shipment) delivered through hub shipments is dominated by the transit time variability of the mode of transportation (road or intermodal rail) used. This simplification allows for the analysis to be focused on one aspect of variability that is of interest in this research. The consideration of other types of variation, as described above, are possible extensions of this research.

The shipments in the queuing system are assigned to their respective job classes as follows. The hub outbound shipments are classified by the tuple $\{k, m, t\}$, which corresponds to an origin hub k , a destination hub $m \neq k$ and a mode of transportation t . The hub inbound shipments are classified by the pair $\{k, t\}$, which corresponds to the destination hub k and the mode of transportation t used for inter-hub transfer.

The arrival rates of different job classes can be computed from X_{ijkm}^t values. Recall that for any origin-destination pair (i, j) there is a unique X_{ijkm}^t variable which is set to a value of 1. This X_{ijkm}^t describes the type of shipment used by an origin-destination pair. If $k = m$, then the shipments are sent using one hub (i.e., local shipments), and if $k \neq m$ then the shipments

are sent through hub-pair (k, m) using mode t (i.e., hub outbound/hub inbound shipments). The daily arrival rate (number of shipments/day) of hub outbound, hub inbound and local shipments (denoted by λ_{kmt}^O , λ_{kt}^I and λ_k^L , respectively) can be computed as follows:

$$\lambda_{kmt}^O = \frac{1}{D} \sum_{i,j \in N} f_{ij} X_{ijkm}^t, \quad \forall k, m \in N (k \neq m), t \in T \quad (56)$$

$$\lambda_{kt}^I = \frac{1}{D} \sum_{i,j, m \neq k \in N} f_{ij} X_{ijmk}^t, \quad \forall k \in N, t \in T \quad (57)$$

$$\lambda_k^L = \frac{1}{D} \sum_{i,j \in N} f_{ij} X_{ijkk}^1, \quad \forall k \in N \quad (58)$$

The traffic rate and variability equations for each operation in a hub queuing system are developed below.

UNLOADING OPERATION

The arrivals at the unloading station are comprised of three different types of shipments: hub outbound, local and hub inbound which are identified by their type and corresponding routing, as discussed earlier. The waiting time of a shipment at the unloading station is affected by the total arrival rate and the processing rate of unloading operations. The total arrival rate at this station is equal to the sum of all external arrivals as given by (56)-(58). The total arrival variability at the unloading station is affected by the variability in arrival times of different types of shipments. The traffic rate and variability equations for this station are:

$$\lambda_k^1 = \sum_{m \in N, t \in T} \lambda_{kmt}^O + \lambda_k^L + \sum_t \lambda_{kt}^I; \quad \forall k, \quad (59)$$

$$ca_k^1 = w_k^1 \left[\sum_{m \in N, t \in T} \left(\frac{\lambda_{kmt}^O}{\lambda_k^1} \hat{c} \right) + \frac{\lambda_k^L}{\lambda_k^1} \hat{c} + \sum_t \left(\frac{\lambda_{kt}^I}{\lambda_k^1} \tilde{c}_t \right) \right] + 1 - w_k^1, \quad \text{with} \quad (60)$$

$$w_k^1 = [1 + 4(1 - \rho_k^1)^2 (\nu_k^1 - 1)]^{-1},$$

$$\nu_k^1 = \left[\sum_{m,t} \left(\frac{\lambda_{kmt}^O}{\lambda_k^1} \right)^2 + \left(\frac{\lambda_k^L}{\lambda_k^1} \right)^2 + \sum_t \left(\frac{\lambda_{kt}^I}{\lambda_k^1} \right)^2 \right]^{-1}, \quad \text{and} \quad \rho_k^1 = \frac{\lambda_k^1}{\mu_k^1}$$

$$ca_k^1 = (\rho_k^1)^2 cs_k^1 + (1 - (\rho_k^1)^2) ca_k^1. \quad (61)$$

Note that the first term in (60) captures the effect of inter-arrival time variability of different types of shipments handled at the unloading station.

BATCH OPERATION

After unloading, the hub outbound shipments move to the batching stage. In the queuing system, there is a batch station, designated by b_{mt} , for each class $\{k, m, t\}$. A shipment of a particular class waits in this queue for other shipments of the same class to arrive. The total waiting time to form a batch is given by the sum $[(L^t - 1) + (L^t - 2) + \dots + 1] \frac{1}{\lambda_{kmt}^O} = \frac{(L^t - 1)L^t}{2\lambda_{kmt}^O}$ (using the identity $\sum_{i=1}^n i = \frac{n(n+1)}{2}$), where each term is the waiting time of $1^{st}, 2^{nd}, \dots, L^t - 1$ arrival. Thus the average waiting time for a shipment at a batch station b_{mt} is $\frac{(L^t - 1)}{2\lambda_{kmt}^O}$. With zero processing time at each batch station, the accumulated shipments move to the next stage immediately without delay. The variability of arrivals at each batch station is determined by the departure process at the unloading station and the splitting of departures of different job classes. The arrival variability at a batch station b_{mt} is given as follows:

$$ca_k^{b_{mt}} = q_k^{1,b_{mt}} cd_k^1 + (1 - q_k^{1,b_{mt}}) q_k^{1,b_{mt}} + (1 - q_k^{1,b_{mt}})^2 \hat{c}, \quad (62)$$

where $q_k^{1,b_{mt}} = \frac{\lambda_{kmt}^O}{\lambda_k^1}$.

BREAK-BULK OPERATION

After unloading, the hub inbound shipments move through the break-bulk stage. The break-bulk stage is composed of two staging areas. The hub inbound shipments which use road transportation pass through staging area B_1 and shipments using intermodal rail pass through staging area B_2 . Each shipment which passes through this stage is delayed by a fixed break-bulk time. This fixed time is used to represent the additional delay encountered when a shipment is unloaded at a destination hub from a truck or an intermodal rail container. This fixed time delay is related to the activity of moving the unloaded shipments to a temporary staging area before moving to the loading docks for dispatch to the final destination. The variability in time between arrivals at the next (loading) station is computed as follows:

$$ca_k^{B_t,4} = q_k^{1,B_t} cd_k^1 + (1 - q_k^{1,B_t}) q_k^{1,B_t} + (1 - q_k^{1,B_t})^2 \tilde{c}_t, \quad (63)$$

$$\text{with } q_k^{1,B_t} = \frac{\lambda_{kt}^I}{\lambda_k^1}.$$

LOADING OPERATION

The loading process receives arrivals from three streams, i.e., from the batch process (in the case of outbound hub shipments), the unloading process (in the case of local shipments) and the break-bulk process (in the case of hub inbound shipments). The arrival rate at the loading station is given by the sum of departure rates from each of the three preceding stations. Note that in the case of a single server queue with FCFS priority rule, the departure rate is the same as the arrival rate.

The traffic rate and variability equations for the loading station are given as follows:

$$\lambda_k^4 = \sum_{m \in N, t \in T} \lambda_{kmt}^O + \lambda_k^L + \sum_t \lambda_{kt}^I, \quad \forall k, \quad (64)$$

$$ca_k^4 = w_k^4 \left[\sum_{m,t} \left(\frac{\lambda_{kmt}^O}{\lambda_k^4 L^t} ca_k^{bmt} \right) + \left(\frac{\lambda_k^L}{\lambda_k^4} ca_k^{1,4} \right) + \sum_t \left(\frac{\lambda_{kt}^I}{\lambda_k^4} ca_k^{B_t,4} \right) \right] + 1 - w_k^4, \quad (65)$$

$$w_k^4 = [1 + 4(1 - \rho_k^4)^2 (\nu_k^4 - 1)]^{-1},$$

$$\nu_k^4 = \left[\sum_{m,t} \left(\frac{\lambda_{kmt}^O}{\lambda_k^4} \right)^2 + \left(\frac{\lambda_k^L}{\lambda_k^4} \right)^2 + \sum_t \left(\frac{\lambda_{kt}^I}{\lambda_k^4} \right)^2 \right]^{-1}, \quad \text{and } \rho_k^4 = \frac{\lambda_k^4}{\mu_k^4}$$

$$ca_k^{1,4} = q_k^{1,4} cd_k^1 + (1 - q_k^{1,4}) q_k^{1,4} + (1 - q_k^{1,4})^2 \hat{c}, \quad (66)$$

$$\text{with } q_k^{1,4} = \frac{\lambda_k^L}{\lambda_k^1}.$$

WAITING TIMES

In the hub queuing system, each shipment may have to wait at an operation/station. The total waiting time of a shipment depends on its classification, i.e., hub-outbound, local and hub-inbound, and its corresponding routing R_O , R_L , and R_I . The total average waiting time of a shipment is a sum of average waiting times at the stations it visits. The average waiting times for unloading (W_k^{unload}) and loading (W_k^{load}) stations is due to queuing wait time

for shared resources. The average waiting time at a batch station (W_k^{batch}) is due to the time needed to form a move batch, while there is a fixed time delay, ξ , at the break-bulk station, i.e., ($W_k^{break} = \xi$).

The expected waiting time due to queuing at the unloading and loading stations is given by the following approximation (Bitran and Morabito, 1996):

$$W_k^{unload} = \left(\frac{ca_k^1 + cs_k^1}{2} \right) \left(\frac{\lambda_k^1}{\mu_k^1 (\mu_k^1 - \lambda_k^1)} \right) g(\rho_k^1, ca_k^1, cs_k^1) \quad \forall k, \quad (67)$$

where $\rho_k^1 = \frac{\lambda_k^1}{\mu_k^1} < 1 \Rightarrow \lambda_k^1 < \mu_k^1$, and

$$g(\rho_k^1, ca_k^1, cs_k^1) \begin{cases} e^{-\frac{2(1-\rho_k^1)(1-ca_k^1)^2}{3\rho_k^1(ca_k^1+cs_k^1)}} & , \text{ if } ca_k^1 < 1 \\ 1 & , \text{ if } ca_k^1 \geq 1. \end{cases}$$

$$W_k^{load} = \left(\frac{ca_k^4 + cs_k^4}{2} \right) \left(\frac{\lambda_k^4}{\mu_k^4 (\mu_k^4 - \lambda_k^4)} \right) g(\rho_k^4, ca_k^4, cs_k^4) \quad \forall k, \quad (68)$$

where $\rho_k^4 = \frac{\lambda_k^4}{\mu_k^4} < 1 \Rightarrow \lambda_k^4 < \mu_k^4$, and

$$g(\rho_k^4, ca_k^4, cs_k^4) \begin{cases} e^{-\frac{2(1-\rho_k^4)(1-ca_k^4)^2}{3\rho_k^4(ca_k^4+cs_k^4)}} & , \text{ if } ca_k^4 < 1 \\ 1 & , \text{ if } ca_k^4 \geq 1. \end{cases}$$

The expected waiting time to form a move batch at batching operation (b_{mt}) is given by:

$$W_k^{b_{mt}} = \frac{L^t - 1}{2\lambda_{kmt}^O} \quad \forall k, m, t \quad (69)$$

SOLUTION PROCEDURES

The model formulation in this article can be solved to optimality using a general purpose mixed integer nonlinear optimization software for small size problems. For larger size problems a more robust solution procedure is devised using a metaheuristic approach. The performance of the metaheuristic approach is benchmarked with a lower bound based on linear relaxation. The lower bound and metaheuristics solution procedures are discussed below.

LOWER BOUND PROCEDURE

The model presented earlier in this article is a mixed integer-nonlinear mathematical formulation. The nonlinearity is due to the waiting time approximation (67) – (69) used in constraints (45) and (46). The solution of a mixed integer-nonlinear mathematical model is a challenging task. The use of optimization techniques such as branch-and-bound, branch-and-cut, cutting planes, branch-and-price and interior point algorithms can only yield solutions in a reasonable time if there is an underlying problem structure which can be leveraged (Floudas, 1995). Otherwise such approaches are limited in their performance. The above mentioned techniques are generally based on solving continuous nonlinear programming problems by relaxing the integrality of some of the discrete variables (Floudas, 1995).

In the realm of commercial solvers for this class of problems, algorithmic and technological advances have yielded solver packages such as BONMIN, KNITRO, FILMINT and MINLP. In a pilot study, these solvers were used to solve some small size (5 cities) problems. These experiments resulted in long solution times to find the optimal solutions. This situation highlighted the need to devise a lower bound procedure which exploits the structure of the problem and the computational power of commercial solvers, albeit for small problem sizes. The details of the lower bound procedure are presented below.

Definition Let \mathbb{P} be the mixed integer non-linear model formulation of Section (??). Relaxing the integrality of X_{ijkm}^t and S_k^t yields the linear relaxation $\mathbf{LR}(\mathbb{P})$. The optimal value of $\mathbf{LR}(\mathbb{P})$, $z_{\mathbf{LR}(\mathbb{P})}^*$, is a lower bound for the optimal value of \mathbb{P} , $z_{\mathbb{P}}^*$, *i.e.*, $z_{\mathbb{P}}^* \geq z_{\mathbf{LR}(\mathbb{P})}^*$.

Lemma 4.1 *Let the variability in inter-arrival and service times at all operations at hub k be fixed, *i.e.*, independent of the decision variables, then the waiting time approximations used in (67) and (68) are convex in λ_k^1 and $\lambda_k^4 \in \mathbb{R}^+$, respectively.*

Proof The waiting time expression in (67) can be written as follows:

$$W_k^{unload} = C_k^1 \left[\frac{\lambda_k^1}{\mu_k^1(\mu_k^1 - \lambda_k^1)} \right] e^{-\frac{K_k^1(\mu_k^1 - \lambda_k^1)}{\lambda_k^1}}, \text{ where} \quad (70)$$

$$C_k^1 = \left(\frac{ca_k^1 + cs_k^1}{2} \right) \text{ and } K_k^1 = \frac{2(1 - ca_k^1)^2}{3(ca_k^1 + cs_k^1)}.$$

Note that C_k^1 and K_k^1 are constants.

Dropping the constant term C and the indices without loss of generality, it can be shown that,

$$\begin{aligned} \frac{dW}{d\lambda} &= \frac{\lambda}{\mu(\mu - \lambda)} e^{-\frac{K(\mu - \lambda)}{\lambda}} \frac{d}{d\lambda} \left[-\frac{K(\mu - \lambda)}{\lambda} \right] \\ &+ e^{-\frac{K(\mu - \lambda)}{\lambda}} \frac{d}{d\lambda} \left[\frac{\lambda}{\mu(\mu - \lambda)} \right], \text{ where} \end{aligned} \quad (71)$$

$$\begin{aligned} \frac{d}{d\lambda} \left[-\frac{K(\mu - \lambda)}{\lambda} \right] &= -\frac{\lambda(-K) - K(\mu - \lambda)(1)}{\lambda^2} \\ &= -\frac{-K\lambda - K\mu + K\lambda}{\lambda^2} \\ &= \frac{K\mu}{\lambda^2}, \text{ and} \end{aligned}$$

$$\begin{aligned} \frac{d}{d\lambda} \left[\frac{\lambda}{\mu(\mu - \lambda)} \right] &= \frac{\mu(\mu - \lambda)(1) - \lambda(-\mu)}{\mu^2(\mu - \lambda)^2} \\ &= \frac{\mu^2 - \mu\lambda + \mu\lambda}{\mu^2(\mu - \lambda)^2} \\ &= \frac{1}{(\mu - \lambda)^2} > 0. \end{aligned}$$

Replacing the terms in (71), the first order derivative becomes

$$\begin{aligned} \frac{dW}{d\lambda} &= \frac{\lambda}{\mu(\mu - \lambda)} e^{-\frac{K(\mu - \lambda)}{\lambda}} \frac{K\mu}{\lambda^2} + \frac{1}{(\mu - \lambda)^2} e^{-\frac{K(\mu - \lambda)}{\lambda}} \\ &= \frac{K}{\lambda(\mu - \lambda)} e^{-\frac{K(\mu - \lambda)}{\lambda}} + \frac{1}{(\mu - \lambda)^2} e^{-\frac{K(\mu - \lambda)}{\lambda}} > 0, \text{ for } \lambda < \mu. \end{aligned}$$

Furthermore, the second order derivative can be written as

$$\begin{aligned} \frac{d^2W}{d\lambda^2} &= \frac{d}{d\lambda} \left[\frac{K}{\lambda(\mu - \lambda)} e^{-\frac{K(\mu - \lambda)}{\lambda}} + \frac{1}{(\mu - \lambda)^2} e^{-\frac{K(\mu - \lambda)}{\lambda}} \right] \\ &= e^{-\frac{K(\mu - \lambda)}{\lambda}} \frac{d}{d\lambda} \left[\frac{K}{\lambda(\mu - \lambda)} \right] + \frac{K}{\lambda(\mu - \lambda)} \frac{d}{d\lambda} \left[e^{-\frac{K(\mu - \lambda)}{\lambda}} \right] \\ &+ e^{-\frac{K(\mu - \lambda)}{\lambda}} \frac{d}{d\lambda} \left[\frac{1}{(\mu - \lambda)^2} \right] + \frac{1}{(\mu - \lambda)^2} \frac{d}{d\lambda} \left[e^{-\frac{K(\mu - \lambda)}{\lambda}} \right] \end{aligned} \quad (72)$$

Solving the derivative terms separately,

$$\begin{aligned}\frac{d}{d\lambda} \left[\frac{K}{\lambda(\mu - \lambda)} \right] &= \frac{-K(\mu - 2\lambda)}{\lambda^2(\mu - \lambda)^2}, \\ \frac{d}{d\lambda} \left[\frac{1}{(\mu - \lambda)^2} \right] &= \frac{-2(\mu - \lambda)(-1)}{(\mu - \lambda)^4} \\ &= \frac{2}{(\mu - \lambda)^3}, \text{ and} \\ \frac{d}{d\lambda} \left[e^{-\frac{K(\mu - \lambda)}{\lambda}} \right] &= e^{-\frac{K(\mu - \lambda)}{\lambda}} \frac{d}{d\lambda} \left[-\frac{K(\mu - \lambda)}{\lambda} \right] \\ &= \frac{K\mu}{\lambda^2} e^{-\frac{K(\mu - \lambda)}{\lambda}}.\end{aligned}$$

Replacing the derivative terms in (72),

$$\begin{aligned}\frac{d^2W}{d\lambda^2} &= \frac{-K(\mu - 2\lambda)}{\lambda^2(\mu - \lambda)^2} e^{-\frac{K(\mu - \lambda)}{\lambda}} + \frac{K}{\lambda(\mu - \lambda)} \frac{K\mu}{\lambda^2} e^{-\frac{K(\mu - \lambda)}{\lambda}} \\ &+ \frac{2}{(\mu - \lambda)^3} e^{-\frac{K(\mu - \lambda)}{\lambda}} + \frac{1}{(\mu - \lambda)^2} \frac{K\mu}{\lambda^2} e^{-\frac{K(\mu - \lambda)}{\lambda}} \\ &= e^{-\frac{K(\mu - \lambda)}{\lambda}} \left[\frac{-K(\mu - 2\lambda)}{\lambda^2(\mu - \lambda)^2} + \frac{K^2\mu}{\lambda^3(\mu - \lambda)} + \frac{2}{(\mu - \lambda)^3} + \frac{K\mu}{\lambda^2(\mu - \lambda)^2} \right] \\ &= e^{-\frac{K(\mu - \lambda)}{\lambda}} \left[\frac{-K\mu + 2K\lambda + K\mu}{\lambda^2(\mu - \lambda)^2} + \frac{K^2\mu}{\lambda^3(\mu - \lambda)} + \frac{2}{(\mu - \lambda)^3} \right] \\ &= e^{-\frac{K(\mu - \lambda)}{\lambda}} \left[\frac{2K}{\lambda(\mu - \lambda)^2} + \frac{K^2\mu}{\lambda^3(\mu - \lambda)} + \frac{2}{(\mu - \lambda)^3} \right] > 0, \text{ for } 0 < \lambda < \mu.\end{aligned}$$

Using the same approach for waiting time expression in (68), the result follows. ■

Lemma 4.2 *The batch waiting time in (69) is convex in $\lambda_{kmt}^O \in \mathbb{R}^+$.*

Proof The waiting time in (69) is of the form $f(u) = \frac{1}{u}$. For any $x, y \in \mathbb{R}^+$ and $\mu \in [0, 1]$, $f(u)$ is convex when:

$$\begin{aligned}f[\mu x + (1 - \mu)y] &\leq \mu f(x) + (1 - \mu)f(y) \\ \frac{1}{\mu x + (1 - \mu)y} &\leq \frac{\mu}{x} + \frac{1 - \mu}{y} \\ \frac{1}{\mu x + (1 - \mu)y} &\leq \frac{\mu y + (1 - \mu)x}{xy}\end{aligned}$$

Rearranging the terms,

$$\begin{aligned}
(\mu x + (1 - \mu)y)(\mu y + (1 - \mu)x) &\geq xy \\
(\mu x + y - \mu y)(\mu y + x - \mu x) &\geq xy \\
xy + (\mu - \mu^2)(x - y)^2 &\geq xy \\
xy + \mu(1 - \mu)(x - y)^2 &\geq xy \quad \blacksquare
\end{aligned}$$

Lemma 4.3 *If the inter-arrival and service time variability at operation O of hub k is fixed and independent of the decision variables, constraints (45) and (46) in $\mathbf{LR}(\mathbb{P})$ are convex in $\mathbf{X} \in [0, 1]$.*

Proof If a function $g(\mathbf{Y})$ is convex in \mathbf{Y} , and $\mathbf{h}(\mathbf{X})$ is an affine function of \mathbf{X} , then $g[\mathbf{h}(\mathbf{X})]$ is convex in \mathbf{X} , see Pg. 80, Bazaraa et al. (1979). Note that the waiting time, $W_k^{unload}(\lambda_k^1)$ is convex in λ_k^1 , the waiting time, $W_k^{load}(\lambda_k^4)$ is convex in λ_k^4 (by Lemma 4.1) and the waiting time, $W_k^{batch}(\lambda_{kmt}^O)$ is convex in λ_{kmt}^O (by Lemma 4.2). Since λ_k^1 , λ_k^4 and λ_{kmt}^O are linear functions of \mathbf{X} in $\mathbf{LR}(\mathbb{P})$ (by definition), it follows that the waiting times in constraints (45) and (46) in $\mathbf{LR}(\mathbb{P})$ are convex in \mathbf{X} . The remaining terms in these constraints are linear and thus constraints (45) and (46) in $\mathbf{LR}(\mathbb{P})$ are convex in $\mathbf{X} \in [0, 1]$. \blacksquare

Theorem 4.4 *If the variability in inter-arrival and service times in the logistics hub operations are assumed fixed, then any local optimal solution to $\mathbf{LR}(\mathbb{P})$ is also a global optimum.*

Proof $\mathbf{LR}(\mathbb{P})$ defines a non-linear formulation with a convex feasible region (see Lemma 4.1, 4.2 and 4.3). A local optimal solution to $\mathbf{LR}(\mathbb{P})$ is also a global optimum, see Pg. 181, Luenberger (1989). \blacksquare

Corollary 4.5 *If the inter-arrival and service times are exponentially distributed, i.e., $ca_k^O = cs_k^O = 1$, for each operation O at hub k , then constraints (45) and (46) in $\mathbf{LR}(\mathbb{P})$ are convex and a local optimal solution to $\mathbf{LR}(\mathbb{P})$ is also a global optimum.*

The results discussed above show that $\mathbf{LR}(\mathbb{P})$ can be solved to optimality for the case of $ca_k^O = cs_k^O = 1$, which yields a lower bound solution to the \mathbb{P} problem. A procedure is developed

to solve the $\mathbf{LR}(\mathbb{P})$ problem which uses a special characteristic of the model formulation. Any selection of hub locations fixes the corresponding decision variables y_k to a value of 1. With a fixed set of hub locations, the original problem is reduced to solving the nonlinear assignment sub-problem over the decision variables X_{ijkm}^t and S_k^t . In this procedure, the assignment sub-problem is solved for all C_p^n combinations of hubs using a commercial solver (BONMIN). The smallest cost solution yields the optimal value to $\mathbf{LR}(\mathbb{P})$ which is a lower bound to the original hub location-allocation problem (\mathbb{P}).

METAHEURISTIC SOLUTION PROCEDURE

As mentioned earlier, small size problems can be solved to optimality. However the analysis of real world logistics networks requires a solution approach that can solve large size problems. In this research a solution approach is developed which is based on a metaheuristic known as tabu search (Glover, 1989). Tabu search is a solution search procedure that efficiently explores the solution space by moving from one solution to the next. In the context of this research, a solution is comprised of two decisions: location of hub cities, called the location sub-problem, and the allocation of freight flows for each origin-destination pair to at most two hubs, called the allocation sub-problem.

The location sub-problem is solved iteratively by selecting a set of hub cities. Tabu search uses this hub set and solves the corresponding allocation sub-problem using a solution procedure based on a time-constrained shortest path problem. Tabu search moves from one solution to the next by a pairwise interchange of a hub city with a non hub city, called a *Move*. The attributes of a move are stored in a short term memory, called *Tabu List*, for the duration of some pre-specified number of iterations, called *Tabu Tenure*. Tabu search uses this short term memory to avoid returning to recently visited solutions. The search ends when a local optimum is found. In the second phase of tabu search, the search re-starts from a new starting solution. The starting solution is generated using a long term search memory, called *Node Frequency*, which records the number of times a city was selected as a hub in the visited solutions. This procedure allows the search to ignore visited solutions and to proceed to previously unexplored

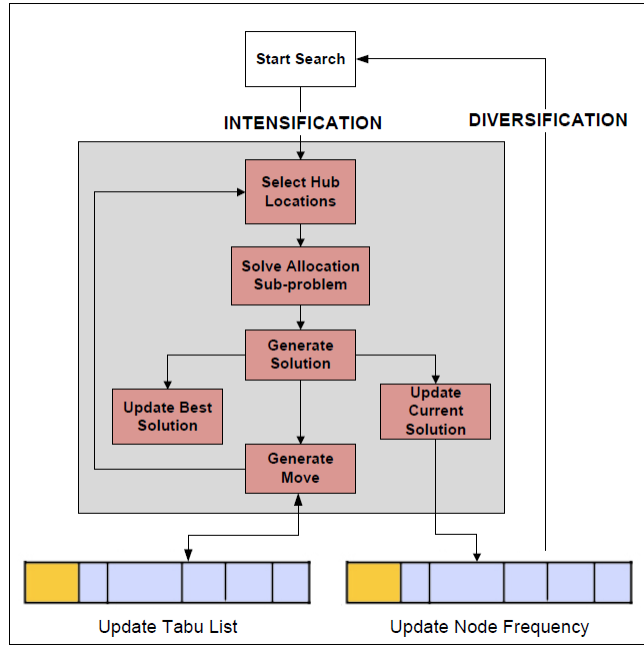


Figure 16: Tabu Search Solution Approach

areas of the solution space. The search terminates after a pre-specified number of re-starts that do not improve the best solution.

As mentioned above, each iteration of tabu search procedure selects a set of hub locations (cities) which is used to solve the allocation sub-problem. The solution of allocation sub-problem requires that flows from each origin-destination pair are assigned to one or two hubs. These allocations are based on the least total transportation cost criterion using the shortest-path approach with considerations for service time requirements. The latter requirements dictate that the total shipment time of any allocation falls within the service time requirements. These requirements are stated in constraints (45) and (46) of the model presented in Section ??.

The outline of the solution procedure (see Figure 17) for the allocation sub-problem is discussed below. The solution of the allocation sub-problem requires computing waiting times for each origin-destination pair using expressions given in (67) – (69). A closer inspection of these expressions shows that the waiting times can be calculated when the values of the X_{ijkm}^t decision variables and the corresponding values of λ_j , ca_j , μ_j and ρ_j are known. Based on the selection of hub cities in each iteration of the tabu search, the solution of the allocation sub-problem starts by setting all the waiting time estimates, W_k^O to zero. The problem is solved

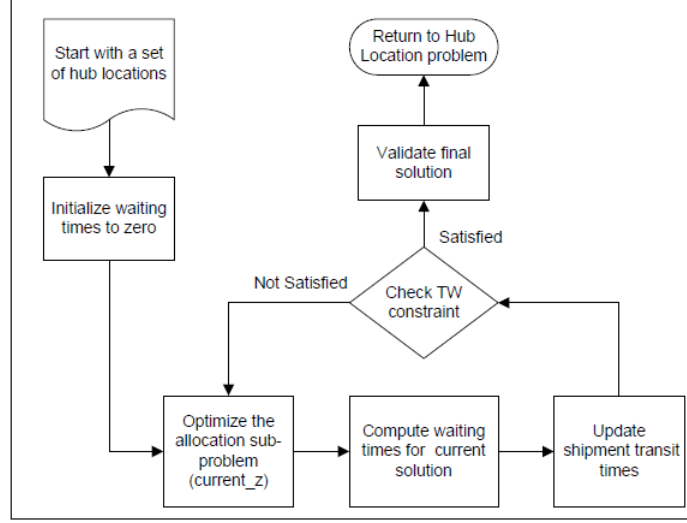


Figure 17: Solution Procedure for Allocation Sub-problem

using the shortest path approach, yielding a solution to the allocation sub-problem in which the X_{ijkm}^t decision variables are assigned a value of 0 or 1. Based on these values, the waiting time estimates are updated using the traffic rate and variability equations given in Section ???. The waiting times, processing times of hub operations and the transit times of different legs of travel are added to compute the total shipment time of each shipment. The allocation sub-problem is then re-solved with the updated values of W_k^O in constraints (45) and (46). This process is repeated until the objective function value of a feasible solution cannot be improved any further.

The iterative procedure described above is formalized as follows:

Step 0: $n = 1$

Set $W_k^O = 0, \forall k$ hubs

Solve revised model and obtain X_{ijkm}^t values from the solution z^1 . Set $n = 2$

Step n Compute and update W_k^O for each (i, j) pair based on z^{n-1}

Re-solve updated model to obtain X_{ijkm}^t values from the solution z^n

If $\frac{|z^n - z^{n-1}|}{|z^{n-1}|} < \epsilon$ and the solution is feasible, STOP

Else $n = n + 1$ and go to Step n

Table 31: Data Sets for Benchmarking Study

<i>Data Set</i>	<i>Number of Cities (n)</i>	<i>Flow (γ)</i>	<i>Cost (π, ϕ)</i>	<i>Transit Times (τ, ψ)</i>
1	5	50	(10,05)	(04,08)
2	5	100	(20,15)	(08,16)
3	5	150	(30,25)	(12,24)
4	5	200	(40,35)	(16,32)
5	5	250	(50,45)	(20,40)
6	8	50	(10,05)	(04,08)
7	8	100	(20,15)	(08,16)
8	8	150	(30,25)	(12,24)
9	8	200	(40,35)	(16,32)
10	8	250	(50,45)	(20,40)
11	10	50	(10,05)	(04,08)
12	10	100	(20,15)	(08,16)
13	10	150	(30,25)	(12,24)
14	10	200	(40,35)	(16,32)
15	10	250	(50,45)	(20,40)

BENCHMARKING STUDY

The effectiveness of the metaheuristic solution approach is tested in a benchmarking study over randomly generated data sets. These data sets differ in problem size (number of cities) and problem scale (range of input data values) to avoid biased cases (Rardin and Uzsoy, 2001). The data sets used in this research are listed in Table 31. Each row corresponds to a data set of n cities and tabulates the parameters of the uniform distributions used to sample the data. The origin-destination flows are generated by sampling from a $\text{Uniform}(1, \gamma)$ distribution. The road transportation costs are sampled from a $\text{Uniform}(1, \pi)$, and rail transportation costs are sampled from $\text{Uniform}(1, \phi)$. $\text{Uniform}(1, \tau)$ distribution is used to generate origin-destination road transit times and a $\text{Uniform}(1, \psi)$ distribution to sample origin-destination rail transit times. For each data set, five different random samples (replications) of problem parameters (flows, costs and transit times) are generated from the corresponding probability distributions. For each of the fifteen data sets in Table 31, there are four problem instances based on different combinations of $p \in \{3, 5\}$ and $\alpha \in \{0.5, 0.9\}$. These account for 60 ($= 4 \times 15$) test problems. With five random samples each, a total of 300 ($= 60 \times 5$) problem instances are tested in the benchmarking study.

These tests were run on a Pentium 4 CPU with 3.20 GHz clock and 4GB RAM. The script language of AMPL was used to implement the optimization solution approach and each allocation subproblem was solved using BONMIN solver (Bonami et al., 2008) which was downloaded

Table 32: Average Percent Solution Gaps (Benchmarking Study)

<i>Data Set</i>	<i>n</i>	α	<i>p</i>	Replications					Avg % Gap	
				1	2	3	4	5		
1	5	0.5	3	0.06	0.07	1.16	1.76	0.09	0.63	
			5	0.01	0.01	1.16	1.69	0.02	0.58	
			3	0.00	0.00	0.00	0.00	0.00	0.00	
		0.9	3	0.00	0.00	0.00	0.00	0.00	0.00	
			5	0.00	0.00	0.00	0.00	0.00	0.00	
			3	0.00	0.00	0.00	0.00	0.00	0.00	
		2	0.5	3	0.79	0.12	0.13	0.07	1.43	0.51
				5	0.75	0.02	0.03	0.01	1.37	0.44
				3	0.00	0.00	0.00	0.00	0.00	0.00
0.9	3	0.00	0.00	0.00	0.00	0.00	0.00			
	5	0.00	0.00	0.00	0.00	0.00	0.00			
	3	0.00	0.00	0.00	0.00	0.00	0.00			
3	0.5	3	0.79	0.54	0.07	0.05	0.03	0.30		
		5	0.74	0.51	0.01	0.01	0.01	0.25		
		3	0.00	0.00	0.00	0.00	0.00	0.00		
0.9	3	0.00	0.00	0.00	0.00	0.00	0.00			
	5	0.00	0.00	0.00	0.00	0.00	0.00			
	3	0.00	0.00	0.00	0.00	0.00	0.00			
4	0.5	3	0.06	0.04	0.02	0.03	0.07	0.04		
		5	0.01	0.01	0.00	0.01	0.01	0.01		
		3	0.00	0.00	0.00	0.00	0.00	0.00		
0.9	3	0.00	0.00	0.00	0.00	0.00	0.00			
	5	0.00	0.00	0.00	0.00	0.00	0.00			
	3	0.00	0.00	0.00	0.00	0.00	0.00			
5	0.5	3	0.01	0.03	0.03	0.03	0.17	0.05		
		5	0.00	0.01	0.01	0.01	0.15	0.03		
		3	0.00	0.00	0.00	0.00	0.00	0.00		
0.9	3	0.00	0.00	0.00	0.00	0.00	0.00			
	5	0.00	0.00	0.00	0.00	0.00	0.00			
	3	0.00	0.00	0.00	0.00	0.00	0.00			
6	8	0.5	3	0.07	0.12	0.13	0.23	1.07	0.32	
			5	0.01	0.03	0.06	0.12	1.00	0.25	
			3	1.65	0.11	0.97	0.55	0.05	0.66	
		0.9	3	1.62	0.03	0.94	0.52	0.01	0.62	
			5	0.06	0.08	0.43	0.05	0.07	0.14	
			3	0.01	0.01	0.37	0.01	0.01	0.08	
		7	0.5	3	0.10	0.03	0.08	0.04	0.07	0.06
				5	0.09	0.01	0.02	0.01	0.01	0.03
				3	0.04	0.04	0.09	0.02	0.20	0.08
0.9	3	0.01	0.01	0.02	0.00	0.16	0.04			
	5	0.04	0.05	0.07	0.08	0.06	0.06			
	3	0.01	0.01	0.02	0.07	0.01	0.02			
8	0.5	3	0.06	0.03	0.03	0.10	0.06	0.06		
		5	0.01	0.01	0.01	0.07	0.01	0.02		
		3	0.04	0.09	0.05	0.05	0.08	0.06		
0.9	3	0.01	0.07	0.01	0.01	0.02	0.02			
	5	0.01	0.07	0.01	0.01	0.02	0.02			
	3	0.02	0.03	0.02	0.06	0.03	0.03			
9	0.5	3	0.00	0.01	0.00	0.01	0.01	0.01		
		5	0.03	0.02	0.02	0.04	0.03	0.03		
		3	0.01	0.00	0.00	0.01	0.01	0.01		
0.9	3	0.01	0.00	0.00	0.01	0.01	0.01			
	5	0.01	0.00	0.00	0.01	0.01	0.01			
	3	0.01	0.00	0.00	0.01	0.01	0.01			
10	10	0.5	3	0.21	0.15	0.12	0.15	0.43	0.21	
			5	0.12	0.03	0.03	0.03	0.33	0.11	
			3	0.31	0.12	0.96	0.59	0.68	0.53	
		0.9	3	0.27	0.02	0.84	0.53	0.67	0.47	
			5	0.05	0.08	0.07	0.06	0.07	0.07	
			3	0.01	0.02	0.01	0.01	0.02	0.01	
		11	0.5	3	0.05	0.07	0.06	0.28	0.05	0.10
				5	0.01	0.01	0.01	0.24	0.01	0.06
				3	0.05	0.03	0.03	0.04	0.05	0.04
0.9	3	0.01	0.01	0.01	0.01	0.01	0.01			
	5	0.01	0.01	0.01	0.01	0.01	0.01			
	3	0.11	0.03	0.03	0.03	0.03	0.05			
12	0.5	3	0.08	0.01	0.01	0.01	0.01	0.02		
		5	0.03	0.03	0.03	0.04	0.03	0.04		
		3	0.01	0.01	0.01	0.01	0.01	0.01		
0.9	3	0.04	0.03	0.02	0.03	0.03	0.03			
	5	0.01	0.01	0.00	0.01	0.01	0.01			
	3	0.02	0.03	0.03	0.02	0.02	0.03			
13	0.5	3	0.00	0.01	0.01	0.01	0.00	0.01		
		5	0.04	0.03	0.05	0.03	0.02	0.03		
		3	0.01	0.01	0.01	0.01	0.01	0.01		
0.9	3	0.04	0.03	0.05	0.03	0.02	0.03			
	5	0.01	0.01	0.01	0.01	0.01	0.01			
	3	0.01	0.01	0.01	0.01	0.01	0.01			
14	0.5	3	0.02	0.03	0.03	0.02	0.02	0.03		
		5	0.00	0.01	0.01	0.01	0.00	0.01		
		3	0.04	0.03	0.05	0.03	0.02	0.03		
0.9	3	0.01	0.01	0.01	0.01	0.00	0.01			
	5	0.01	0.01	0.01	0.01	0.01	0.01			
	3	0.01	0.01	0.01	0.01	0.01	0.01			
15	0.5	3	0.02	0.03	0.03	0.02	0.02	0.03		
		5	0.00	0.01	0.01	0.01	0.00	0.01		
		3	0.04	0.03	0.05	0.03	0.02	0.03		
0.9	3	0.01	0.01	0.01	0.01	0.00	0.01			
	5	0.01	0.01	0.01	0.01	0.01	0.01			
	3	0.01	0.01	0.01	0.01	0.01	0.01			

from the COIN-OR website (<http://www.coin-or.org/Bonmin/>). The same test problems were solved using tabu search procedure which was implemented using C++. The tabu search solutions are benchmarked with the lower bound solutions through the *percent solution gap* measure. The percent solution gap is calculated as $(\frac{z_{TS}-z_{LB}}{z_{LB}})$, where z_{TS} is the value of the tabu search solution and z_{LB} is the objective function value of the lower bound solution. The percent solution gaps for the five replications of each test problem were averaged and recorded as the *average percent solution gap*. The percent solution gaps obtained by solving the problem instances using tabu search are reported in Table 32. The average of percent solution gap over all test instances is 0.12%. The minimum solution gap for tabu search is 0.00% with 97% of the problem instances solved to within 1% of the lower bound solution. The maximum percent solution gap is 1.76%. The small solution gaps indicate that tabu search approach can find near optimal solutions.

EMPIRICAL STUDY

The purpose of this computational study is to investigate the effects of congestion at logistics hubs due to limited material handling resources on the hub network design of an intermodal logistics network. This study investigates these effects in terms of network costs, number and location of logistics hubs, hub flows and shipment waiting times under specific service time requirements. This study is based on a network of 25 US cities. These cities and inter-city freight flows were extracted from the 2002 Freight Analysis Framework (FAF) database maintained by the United States Department of Transportation. This database was used to generate a representative sample of flows for the case study. The freight flows were computed in terms of the number of pallet loads that must be moved between the origin and destination cities. The total commodity flows between the cities in the FAF database were converted to the pallet level by using a conversion factor of 1 pallet = 350 pounds. These pallet load quantities were then used as demand flows between each pair of cities in the study. In compiling the data, inbound and outbound flows for each city/metropolitan area were aggregated to compute the total flow volume for each city. The cities were then rank ordered according to their total flow

volume. The top 25 cities with geographical locations in the Northeast, Midwest and South were selected for this study (see Figure 18).

The less-than-truckload (LTL) freight rates were obtained from a private logistics service provider that is among the top ten U.S. freight carriers. The local transportation costs for the pick-up and delivery of a shipment (c_{ij}^1) are based on the actual LTL freight rates. These local transportation costs are computed based on the distance and the LTL rate for each lane (origin-destination pair). The long-haul full truckload (TL) costs between hubs are less than the local LTL costs and are scaled from the LTL costs using the parameter α , which is an experimental factor in the case study. The transportation costs for intermodal rail (c_{ij}^2) are based on the published freight rates offered by the various U.S. rail carriers: Union Pacific (UP), Norfolk Southern (NS), Burlington Northern Santa Fe (BNSF) and CSX Railways. The intermodal rail costs are based on the freight rates for shipping one 48 foot domestic intermodal-rail container.

The average transit times for road shipments were computed using the commercial software PCMiller with consideration for the provisions of the *Hours-of-Service* rule from the Federal Motor Carrier Safety Administration (FMCSA, 2008). The average intermodal rail transit times used in this study were obtained from the published data of the same U.S. rail carriers used to obtain the freight rates.

The other parameter used in this study is the fixed cost of operating a logistics hub (F_k). This study sets the size of the hub facility at 100,000 ft^2 which is comparable to the size of the United Parcel Services (UPS) freight facilities (UPS, 2008). Using the estimates given in Waller (2003) and Colliers Research Services (2009), the fixed operating cost per hub was set at \$1,000,000. For computational purpose, cost of intermodal service (MC) was set \$100,000.

EXPERIMENTAL DESIGN

The problem instances used in this case study were based on five different factors: number of hubs (p), service time requirement (TW), long-haul discount factor (α), processing rate of loading and unloading resources (referred as hub resource levels) used at a hub (μ_L, μ_U), and dwell time limit for shipments (Max_Wait). The dwell time limit represents the maximum time

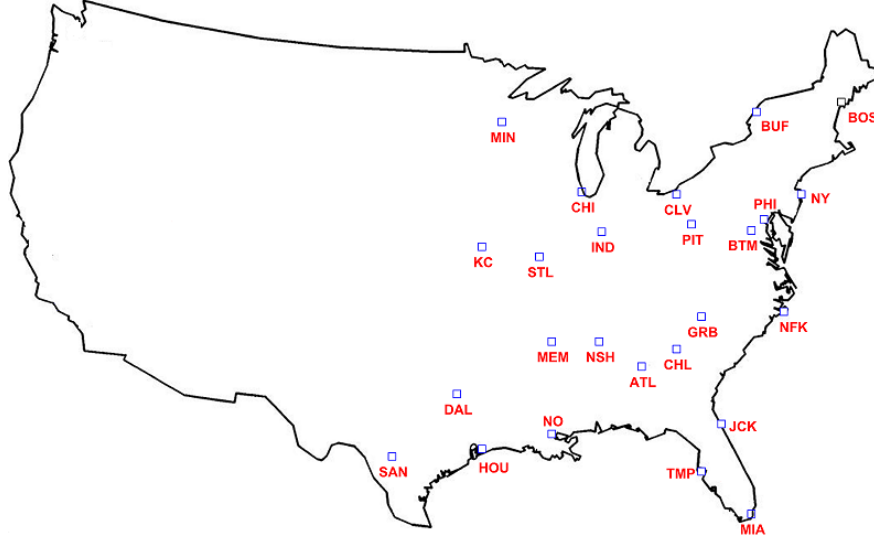


Figure 18: 25 Cities used in Case Study

Table 33: Factor Levels

<i>Factors</i>	<i>Levels</i>
α	{0.5, 0.7, 0.9}
μ_L, μ_U	{900, 1000, 1100}
TW	{3, 4, 5}
Max_Wait	{0.5, 1.0}
p	{3, 4, 5, 6}

a shipment will wait for collecting move batch quantities. For shipment classes which do not have sufficient total daily demand rates, batch waiting time may be too long to satisfy service requirements. Thus Max_Wait imposes a limit on how long a shipment is allowed to wait in consolidation/batch process before it is shipped.

The study was based on a full factorial design in which these factors were set at different levels as shown in Table 33. The factor α was set at three levels which represent the reduction in transportation costs when shipments travel between hubs. The factors (μ_L) and (μ_U) which represent the processing rate of the loading and unloading resources, respectively, at a logistics hub were set at levels of {900, 1000, 1100} pallets per day. The range of levels used for μ_L and μ_U were selected through a pilot study which tested different values of μ_L and μ_U . The range of levels selected provides resource utilizations in the range $0.50 < \rho < 0.90$ which allows for the use of traffic rate and variability equations developed in the previous section. The factor TW was set at three levels: 3, 4 and 5 days. Finally, the factor Max_Wait was set at two levels:

0.5 and 1 day. The inter-arrival time and service time variability i.e., ca and cs were set at the value of 1.

The problem instances generated by the different combinations of the factor levels described above were solved with four different values of $p = \{3, 4, 5, 6\}$. All the test problems were solved using the tabu search approach. Among the solutions generated by different values of p , the solution with the lowest total network cost was reported. For each test problem, the following values were recorded: Network costs which records the objective function value; total transportation costs (TXC), which includes pickup, inter-hub transfer and drop off costs; total number of intermodal hubs in the network ($IM-Hubs$); total flows that use intermodal shipments between hubs ($IM-Flows$); total flows that use single hub shipments ($S-Flows$) and the location of the chosen hubs. Additionally, the waiting times at unloading and loading processes at each hub and the total waiting times of shipments were also recorded.

RESULTS: TOTAL NETWORK COSTS

The results indicate that the total network costs of a logistics network is affected by hub resource levels (μ), inter-hub discount factor (α) and service time requirements (TW), as shown in Figure 19. The first plot in Figure 19(a) shows the effect of discount factor on total network costs. It can be seen that when the transportation costs of inter-hub transfers are discounted at a higher level ($\alpha = 0.5$), the total network costs decrease. The higher inter-hub discount also attracts more inter-hub shipments which contributes to reducing the total network costs.

The second plot in Figure 19(a) shows that the total network costs increase when the hub resource level (μ) is low. The logistics network compensates for insufficient hub resources by using more hubs. The increase in the number of logistics hubs results in higher fixed operating and intermodal service costs which consequently also increase the total network costs. Furthermore, the negative effect of lower hub resource levels on shipment waiting times also divert flows from inter-hub shipments to single hub shipments, which have shorter total shipment times as there are no consolidation and break-bulk activities.

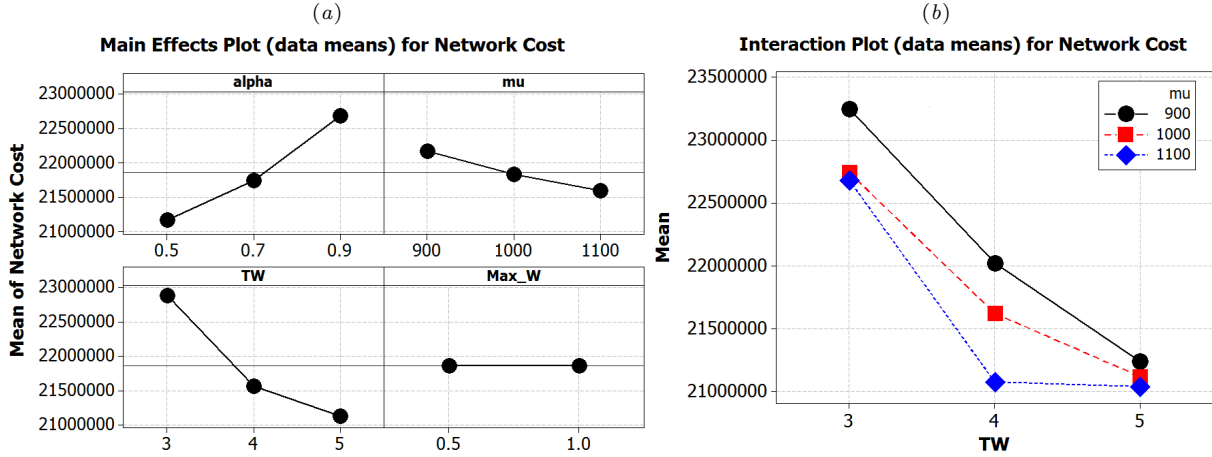


Figure 19: Total Network Costs

The third plot in Figure 19(a) shows that a shorter service time requirement results in higher total network costs. This increase occurs due to the change from inter-hub routes to single hub routes which use road transportation over links with shorter transit times. Such shipments however have higher transportation costs due to the use of a more expensive mode of transportation (i.e., road) and at transportation costs which are not discounted. Finally, the fourth plot in Figure 19(a) shows that the maximum allowed waiting time (Max_Wait) did not effect the results over the problem instances tested. This factor is dropped from further analysis.

The results also indicate an interaction effect between the service time requirements and the hub resource levels, as shown in Figure 19(b). The marginal increase (slope of the cost lines) in network costs is more significant for lower hub resource levels. This trend identifies an increasing effect of smaller hub resource levels on the total network costs. In general, smaller service times require the use of faster transportation and smaller waiting times in hub(s). It is seen in Figure 19(b) that the change in total network costs due to different levels of μ is small when service requirements are not restrictive (see the case of $TW=5$). This means that even larger waiting times which result from lower hub resource levels (μ) does not significantly affect the choice of shipments (single-hub and inter-hub). However these choices are affected more so when service time requirements are shorter, resulting in a higher use of faster shipments which have higher transportation costs.

NUMBER AND LOCATION OF HUBS

In this section, the effect of hub resource levels (μ) and service time requirements (TW) on the network structure i.e., number and location of hubs is analyzed. For each test problem in this research study, the best number of hubs and the location of these hubs were recorded. Table 34 reports the selection frequency of different cities as hub locations. The results show that New York, Chicago and Atlanta were the most suitable cities for locating logistics hubs in this research study. The results also indicate that depending on the modal network infrastructure and the corresponding transportation costs, some cities can be used as alternative sites for hub locations. For example, Houston and Dallas served as alternate hub cities in almost all the test problems. Similarly, Chicago which was selected as a hub city in 85% of the test problems, was replaced with Indianapolis in some test problems. A review of the results indicates that Indianapolis was preferred over Chicago in test problems with service time requirements set at TW=3 and a very small fraction of flows used intermodal rail. The best locations for logistics hubs along with their alternate sites are shown graphically in Figure 20.

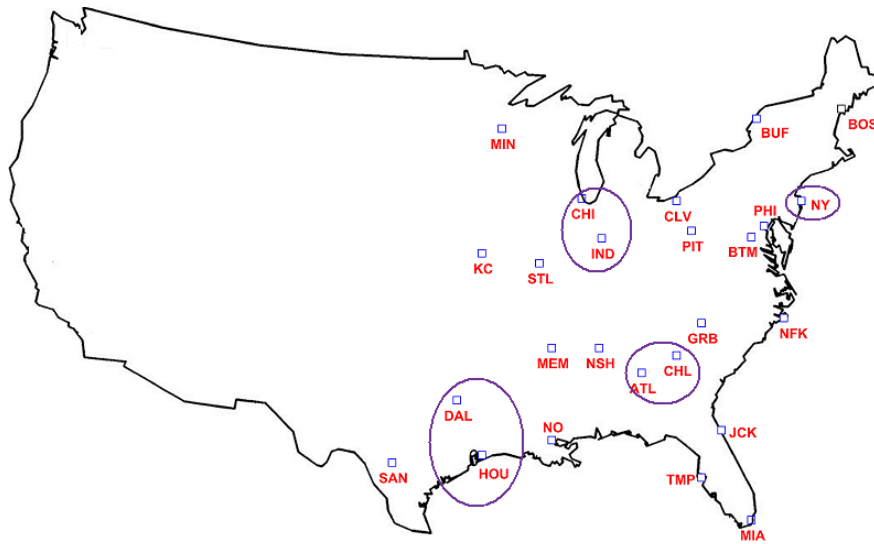


Figure 20: Best Locations for Logistics Hub

The results also showed a relationship between the level of hub resources and the number of hubs in a logistics network under service time requirements. As discussed in an earlier section, the aggregation of flows at hubs in a hub-and-spoke logistics network affects the shipment

Table 34: Selection Frequency of Hub Cities

<i>Cities</i>	<i>Frequency</i>
New York, NY	90%
Chicago, IL	85%
Atlanta, GA	74%
Dallas, TX	48%
Houston, TX	48%
Baltimore, MD	26%
Charlotte, NC	26%
Philadelphia, PA	15%
Indianapolis, IN	11%

Table 35: Effects of TW and μ on Number of Hubs

TW	$\mu = 900$	$\mu = 1000$	$\mu = 1100$
3	4 hubs in 100% cases	4 hubs in 100% cases	4 hubs in 100% cases
4	5 hubs in 100% cases	5 hubs in 100% cases	4 hubs in 100% cases
5	5 hubs in 100% cases	4 hubs in 60% cases 5 hubs in 40% cases	4 hubs in 100% cases

waiting times in hub(s). The lack of sufficient hub resources to match the flow volume results in long waiting times which impact the service performance of the network. Table 35 summarizes the percentage of test problems in this research study which resulted in network designs with a specific number of hubs. These results show that when sufficient hub resources were available (see case of $\mu=1100$, in the last column), best number of hubs which minimized the total network costs and satisfied the service time requirements was 4 hubs. On the other hand, when hub resource levels were not sufficient, i.e., $\mu = \{900,1000\}$, the volume of flows caused congestion at the hubs resulting in long waiting times. In order to mitigate the effects of congestion, either more hubs or additional hub resources are needed. The former case is shown in Table 35 (see first and second column), where the best number of hubs for $\mu=900$ ($TW=\{4,5\}$) and $\mu=1000$ ($TW=\{4,5\}$) was 5 hubs. Opening one more hub, re-distributed the flows serviced by each hub to a level which matched available hub resources.

Additionally, consider the case of $\mu=900$. The best number of hubs for test problems with $TW=\{4,5\}$ was 5 hubs. When the service time requirement was decreased to 3 days, the logistics networks used only 4 hubs. This effect occurs because at shorter service time requirements the use of single hub shipments increases. The single hub shipments spend less time in hubs because there are no time delays in consolidation/break-bulk activities. In situations where a

high number of flows use single hub shipments, fewer hubs will suffice. However note that this tradeoff depends on the availability of sufficient hub resources to manage the increase in flows at a hub which may result due to fewer hubs in the network. For example, in this research study, $\mu=900$ was found to be sufficient to satisfy $TW=3$ service requirement with 4 hubs.

MODAL FLOWS

In this research, shipments can move between origins and destinations using either a single hub (referred to as $S-F$) or a hub-pair using road transportation (referred to as $R-F$) or intermodal rail transportation (referred to as $Ra-F$). The use of inter-hub shipments results in discounted transportation costs over the inter-hub link. The downside of inter-hub travel is that it requires shipments to wait extra time at hubs for consolidation and break-bulk activities. Furthermore, in order to utilize the inter-hub discount, such shipments may travel over links which are not the quickest route to their destination. The effects of inter-hub discount factor (α), hub resource level (μ) and service time requirements (TW) on modal flows are shown in Figure 21.

Since road and intermodal rail offer two alternate means of sending inter-hub shipments, a complementary effect in the plots of Figure 21(b) and (c) can be observed, except for the effect of service time requirements, which impacts the intermodal rail shipments more significantly than road transportation. A higher discount on inter-hub road costs attracts more hub shipments to use road transportation. Figure 21(b) shows that road flows increased from 5% to 16% when α (discount factor) is varied from 0.9 to 0.5, while intermodal rail flow decreased from 23% to 18%. However, note that a rather large change in inter-hub discount did not result in attracting significantly more flows to use inter-hub routes, as shown in Figure 21(a). This plot shows that single hub shipments fell from 72% to 65%, only a 6% drop.

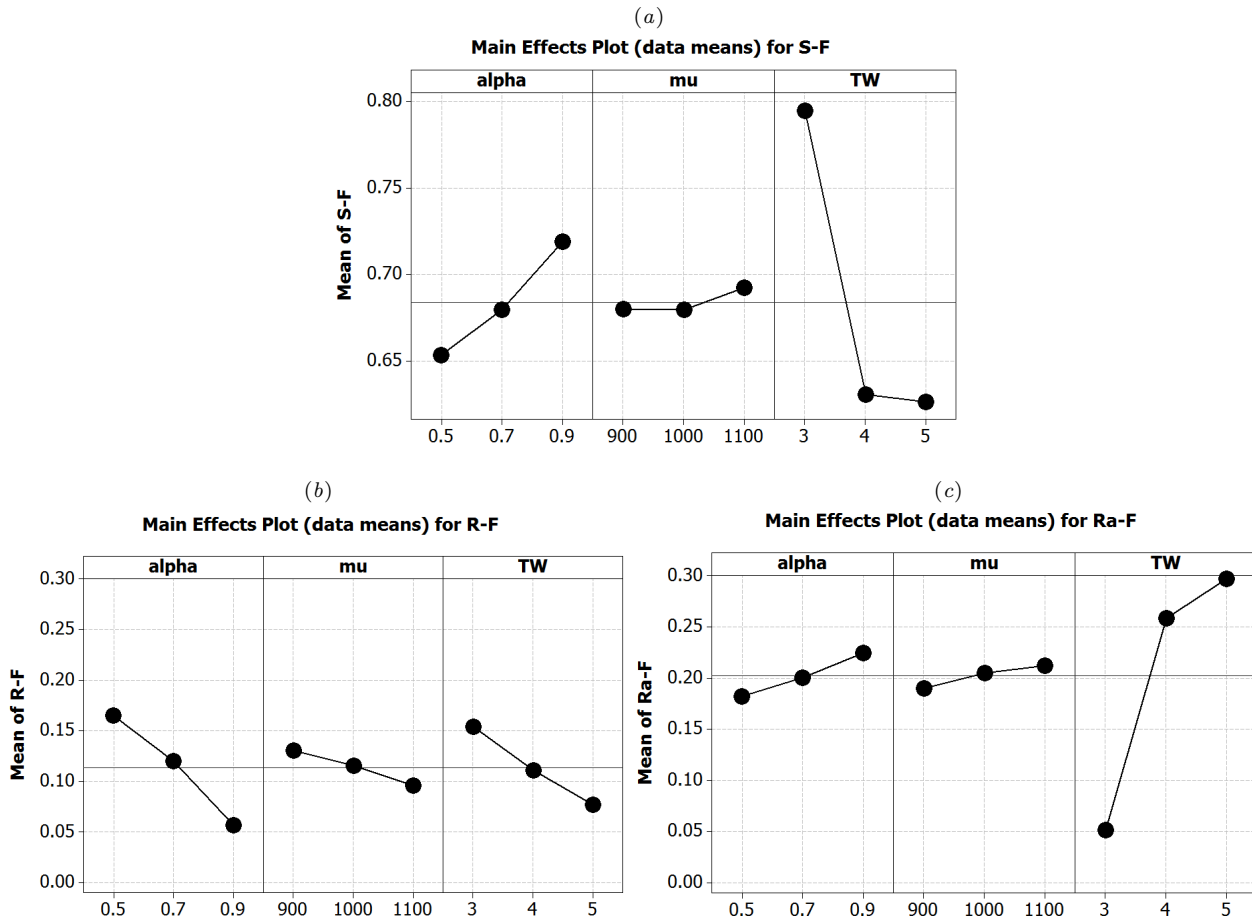


Figure 21: Modal Flows

Figure 21 also shows the effect of service time requirement on modal flows. These results show that service requirements affect the choice of mode, with more flows moving from inter-modal rail to single hub shipments for a shorter service time requirement. The plots in Figure 21 show that intermodal rail flows fell from 30% to 5% when service time requirement was changed from 5 to 3 days, while road and single hub flows increased from 7% to 15% and from 62% to 80%, respectively.

SHIPMENT WAITING TIMES

The total waiting time for a shipment passing through hub(s) consists of the waiting times at unloading, consolidation, break-bulk and loading processes. Figure 22 shows the average waiting times for inter-hub road shipments (W_{H_R}), inter-hub intermodal rail shipments (W_{H_Ra}) and single hub shipments (W_S). These plots show the changes in average waiting times for

different levels of hub resources (μ) and service time requirements (TW). The latter affects the waiting times indirectly, because of its influence on the number of hubs used in the network and the consequent effect on the flow volume which passes through each hub.

Figure 22(a) shows that the average waiting times (W_S) of single hub shipments were not affected by the service time requirements and hub resource levels as much as the inter-hub shipments. The average waiting times for single hub shipments is small because it consists only of the waiting times at the loading and unloading process, which were small compared to the consolidation and break-bulk times which affect the inter-hub shipments.

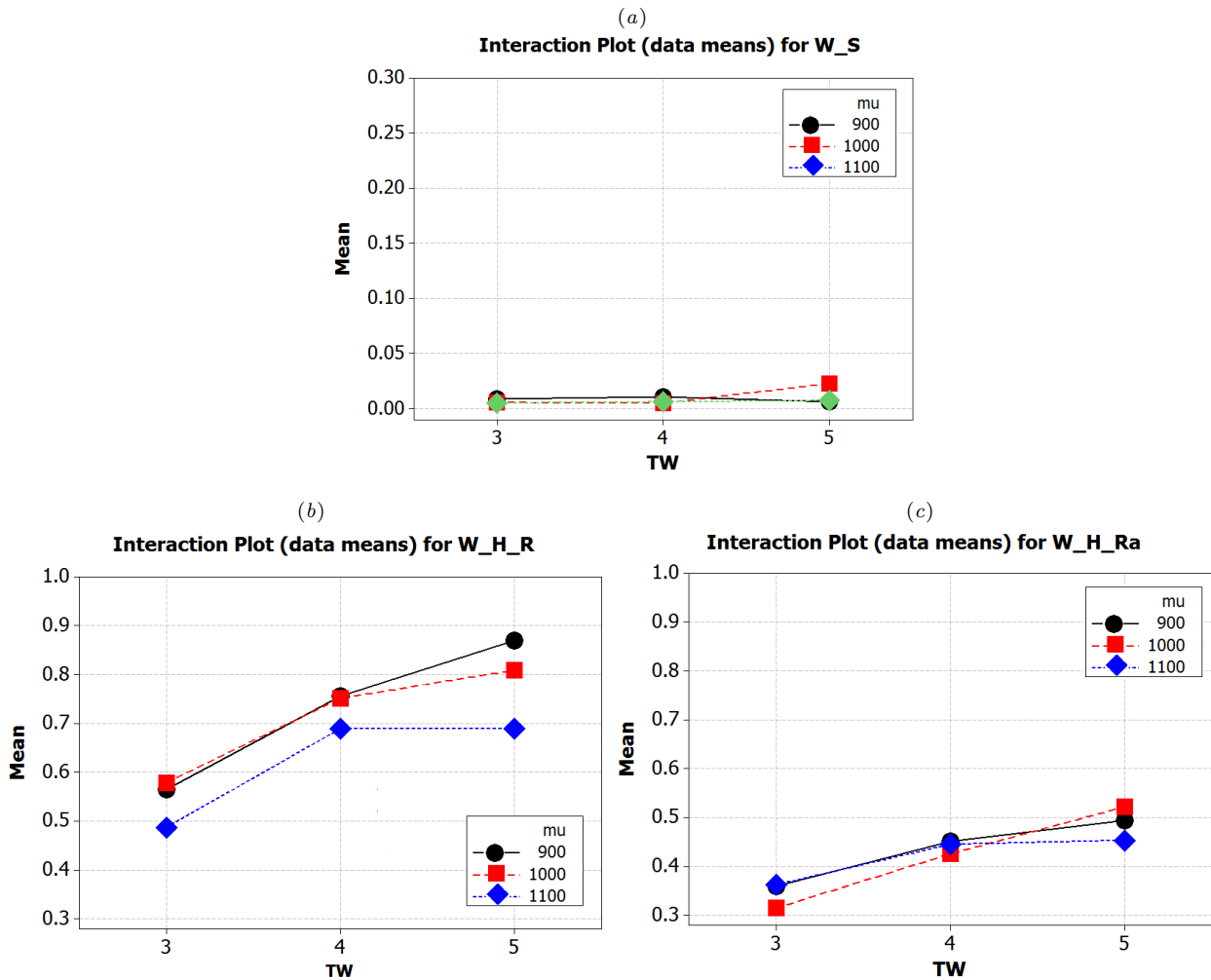


Figure 22: Shipment Waiting Times

Figure 22(b) and (c) show that the waiting times for both inter-hub road and inter-hub intermodal rail tend to decrease for smaller service times. This effect represents the tradeoff between the use of fewer hubs and lower volume of inter-hub shipments for shorter service time requirements. As was seen previously in Figure 21, the average intermodal rail flows ($Ra-F$) are higher than road flows ($R-F$) which results in smaller average waiting times for intermodal rail.

CONCLUSIONS

This research deals with the design of an intermodal logistics network with consideration for flow congestion at hubs. A modeling framework for the operations of an intermodal logistics hub was developed using a queuing system approach. This framework describes a job classification scheme for individual shipments which are routed through the unloading, consolidation, break-bulk and loading processes of hub operations. By integrating this queuing model with a p-hub median approach for interacting hub network design, a mathematical model was developed. Building on prior research in open serial queuing systems with product classes and deterministic routings, traffic rate and variability equations were developed. These equations were used to compute average shipment waiting times in the logistics network.

In a research study based on real world data from the US Department of Transportation and actual networks of interstate highway and intermodal rail, a large size (25 city) intermodal logistics network was designed. This network was used to study the impact of limited processing resources (unloading, loading, material handling and storage) on total network cost, number and location of hubs, modal flows, choice of shipments and shipment waiting times, under service time requirements. The results showed that the availability of sufficient hub resources is a big determinant of the network structure. A lack of sufficient hub resources can considerably increase the dwell time of hub shipments, consequently discouraging the use of inter-hub shipments. This situation reduces the impact of the inherent benefit of discounted transportation costs for inter-hub transfers. The results also show that a logistics network compensates for insufficient hub resources by using more hubs. Opening extra hub(s), re-distributes the flows

serviced by each hub to a level which matches the available hub resources. However, the increase in the number of logistics hubs results in higher fixed operating and intermodal service cost which increase the total network costs. The results also show that increasing the inter-hub shipment discount factor did not result in any significant increase in hub flows due to limited hub resources.

This research can be extended by considering the option of direct shipments in addition to single-hub and inter-hub shipments. Furthermore, the impact of modal variability and batch size (truck-load/container loads) on shipment waiting times need to be investigated. Another aspect which is of further interest is that of service design in which customers are offered different delivery times, such as regular or premium. The effects of such service options on the design and management of logistics networks can be evaluated using a similar approach as the one presented in this paper.

REFERENCES

- Alumur, S., Kara, B. Y., 2008. Network hub location problems: The state of the art. *European Journal of Operational Research* 190 (1), 1–21.
- Aykin, T., 1994. Lagrangian relaxation based approaches to capacitated hub-and-spoke network design problem. *European Journal of Operational Research* 79 (3), 501–523.
- Aykin, T., 1995. Networking policies for Hub-and-Spoke systems with applications to the air transportation systems. *Transportation Science* 29 (3), 201–221.
- Bazaraa, M. S., Sherali, H. D., C.M.Shetty, 1979. *Nonlinear Programming*. John Wiley & Sons.
- Bitran, G. R., Morabito, R., 1996. Open queueing networks: Optimization and performance evaluation models for discrete manufacturing systems. *Production and Operations Management* 5 (2), 163–193.
- Bitran, G. R., Tirupati, D., 1988. Multiproduct queueing networks with deterministic routing: Decomposition approach and the notion of interference. *Management Science* 34 (1), 75–100.
- Bonami, P., Biegler, L. T., Conn, A. R., Cornujols, G., Grossmann, I. E., Laird, C. D., Lee, J., Lodi, A., Margot, F., Sawaya, N., Wachter, A., 2008. An algorithmic framework for convex mixed integer nonlinear programs. *Discrete Optimization* 5 (2), 186 – 204.
- Campbell, J. F., 1994. Integer programming formulations of discrete hub location problems. *European Journal of Operational Research* 72 (2), 387–405.

- Campbell, J. F., 2009. Hub location for time definite transportation. *Computers & Operations Research* 36 (12), 3107–3116.
- Colliers Research Services, 2009. A survey of commercial real estate market. Tech. rep.
- de Camargo, R., Jr., G. M., Luna, H., 2008. Benders decomposition for uncapacitated multiple allocation hub location problem. *Computers & Operations Research* 35 (4), 1047–1064.
- 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.
- Elhedhli, S., Hu, F. X., 2005. Hub-and-spoke network design with congestion. *Computers & Operations Research* 32 (6), 1615–1632.
- Ernst, A. T., Krishnamoorthy, M., 1998. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *European Journal of Operational Research* 104 (1), 100–112.
- Ernst, A. T., Krishnamoorthy, M., 1999. Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research* 86, 141–159.
- Floudas, C., 1995. *Nonlinear and mixed-integer programming-fundamentals and applications*. Oxford University Press 4, 249–281.
- FMCSA, 2008. Hours of service of drivers. *Federal Register* 73 (224), 69567–69586.
- Glover, F., 1989. Tabu search - part I. *ORSA Journal on Computing* 1 (3), 190–206.
- Grove, G., O’Kelly, M., 1986. Hub networks and simulated schedule delay. *Papers of the Regional Science Association* 59, 103119.
- Limbourg, S., Jourquin, B., 2009. Optimal rail-road container terminal locations on the european network. *Transportation Research (E)* 45 (4), 551–563.
- Luenberger, D. G., 1989. *Linear and Nonlinear Programming*. Addison-Wesley Publishing Company Inc.
- Marianov, V., Serra, D., 2003. Location models for airline hubs behaving as M/D/c queues. *Computers & Operations Research* 30 (7), 983–1003.
- O’Kelly, M. E., 1986. The location of interacting hub facilities. *Transportation Science* 20 (2), 92–106.
- Pirkul, H., Schilling, D. A., 1998. An efficient procedure for designing single allocation hub and spoke systems. *Management Science* 44 (12(2)), 235–242.
- Rardin, R. L., Uzsoy, R., 2001. Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics* 7, 261–304.
- Rodriguez, V., Alvarez, M. J., Barcos, L., 2007. Hub location under capacity constraints. *Transportation Research Part E: Logistics and Transportation Review* 43, 495–505.

- Sevcik, K. C., Levy, A., Tripathi, S., Zahorjan, J., 1977. Improving approximations of aggregated queuing network subsystems. In: International Symposium on Computer Performance Modelling, Measurement, and Evaluation. Yorktown Heights.
- UPS, 2008. Facts sheet: UPS operations. <http://www.pressroom.ups.com/Fact+Sheets>.
- Waller, D. L., 2003. Operations Management: A Supply Chain Approach, 2nd Edition. Cengage Learning Business Press, p. 1088.
- Warsing, D. P., Souza, G. C., Greis, N. P., 2001. Determining the value of dedicated multimodal cargo facilities in a multi-region distribution network. *European Journal of Operational Research* 133 (1), 81–93.
- Whitt, W., 1983. The queueing network analyzer. *The Bell System Technical Journal* 62 (9), 2779–2815.
- Whitt, W., 1984. Approximations for departure processes and queues in series. *Naval Research Logistics* 31, 499–521.
- Whitt, W., 1988. A Light-Traffic approximation for Single-Class departure processes from Multi-Class queues. *Management Science* 34 (11), 1333–1346.

OVERALL CONCLUSIONS

This research explored the impact of using intermodal shipments on the logistics strategy of a third-party logistics service provider. The contributions of this research lie in the development of different modeling frameworks, fast and accurate solution approaches and managerial insights gained through research studies. The first article developed a modeling framework which incorporates the fixed cost of operating a hub, the cost of providing intermodal services and service time requirements in a road-rail intermodal network. The modeling framework was used to develop a mixed integer linear program, which was solved using a meta-heuristic (tabu search) solution approach. The performance of this solution approach was benchmarked with lower bounds developed through a Lagrangian relaxation approach. The computational study demonstrated that the meta-heuristic solution approach finds high quality solutions in a reasonable length of time. This research also showed that the Lagrangian lower bounds for this problem are very tight. Large-size test problems (up to 100 nodes) were used to demonstrate the performance of the tabu search metaheuristic. The solutions obtained by tabu search were compared with the corresponding lower bounds, and the results showed that for a large number (84%) of these problems, the solution approach yielded solutions that were within 5% of the lower bound.

The model developed in the first article was used with Civil Aeronautics Board (CAB) data to gain relevant managerial insights into the intermodal logistics networks. The results showed that the modal connectivity cost affects both the number of intermodal hubs and the level of use of intermodal shipments in a logistics network. The results identified a threshold level for the modal connectivity costs beyond which the cost benefits of intermodal shipments are outweighed by the costs of providing intermodal services at the hubs. The results also showed that the difference in road and rail transportation rates, called the cost ratio, affects the structure of the intermodal network. The cost ratio impacts the interplay between fixed

costs, hub discount factor and service time requirements, which in turn affect the resulting network structure. The results showed that under shorter service time requirements, some of the lowest cost network paths are infeasible. If the service time requirements are too small, the cost benefit of intermodal transportation cannot be realized and the network structure shifts to an over-the-road network.

The insights gained in the first article led to a research study in the second article which explored the interplay between the financial, operational and service aspects of an intermodal logistics hub network. In this article an enhanced modeling framework was developed that captures the realities of an actual intermodal logistics hub network. This framework and the resulting mathematical model incorporate different types of shipments (direct, inter-hub and single hub), different modes of transportation (road, rail and air), fixed hub operating costs, modal connectivity costs, economies-of-scale and service time requirements. The solution approach devised for this model is based on the work in the first article. New additions and enhancements were made which allowed for handling the additional complexities introduced in the second article. The performance of the new meta-heuristic solution approach was evaluated in a computational study by benchmarking it with the corresponding optimal solutions.

The second article provided relevant managerial insights into the differences in the structure of over-the-road and intermodal logistics networks. This study found that the location of hubs in the over-the-road and intermodal networks differed greatly. The choice of best hub locations in the intermodal networks was found to be affected by the network parameters, whereas the hub locations in over-the-road networks were very robust. The results identified cities which have a strategic geographical location and adequate modal network connectivity that make them suitable candidates for hub locations. The results also showed that while the use of intermodal shipments reduced the transportation costs, the modal connectivity costs and service time requirements limited the benefits of those intermodal shipments. It was shown that the location of hubs and the design of service regions in the intermodal networks allowed for more hub-based shipments than the over-the-road networks. The study also found that changes in service time requirements required reorganizing the intermodal hub network by

moving hubs and using different types of shipments and transportation modes. Under short service time requirements, the modal usage transitioned from intermodal rail to road to single hub shipments to direct shipments and air shipments. The study also found large variations in the inbound and outbound flows at the hubs with changing service requirements. The results showed that efficiency gains in over-the-road networks can reduce the cost benefit of intermodal shipments, however it would require a prohibitive level of effort.

The third article is focused on the design and management of an intermodal logistics network under hub congestion considerations. A modeling framework for the hub operations of an intermodal logistics hub was developed using a queuing system approach. In the hub operations queuing system, each shipment is classified as a job that is routed through unloading, consolidation, break-bulk and loading processes of hub operations. The integration of this queuing model with the interacting hub network design model of the first article resulted in a new mathematical model. Building on prior research in open serial queuing systems with product classes and deterministic routings, traffic rate and variability, and shipment waiting time estimates were developed.

A research study was conducted based on real world data from US Department of Transportation and actual networks of interstate highway and intermodal rail. A large size intermodal logistics network was designed. This network was used to study the impact of limited processing resources (unloading, loading, material handling and storage) on total network cost, number and location of hubs, modal flows, choice of shipments and shipment waiting times, under service time requirements. The results showed that lack of sufficient hub resources can considerably increase the dwell time of hub shipments, consequently discouraging the use of inter-hub shipments. This situation reduces the impact of inherent benefit of discounted transportation costs for inter-hub transfers. The results also showed that a logistics network compensates for insufficient hub resources by using more hubs. Opening extra hub(s) redistributes the flows serviced by each hub to a level which matches the available hub resources.

FUTURE WORK

The research work in this dissertation can be extended in a number of ways:

- The models developed in this research used a single product or a generic shipment. One of the ways to extend this research is to include the use of multiple commodities. This will provide an opportunity to incorporate different requirements (weight and volume) of different commodities.
- Another extension is in the area of service design. Instead of using a network-wide service time requirement, different types of service requirements can be modeled. For example, a transportation mode can be differentiated in terms of capacity such as vans, small trucks, 20-foot and 48-foot trailers and in terms of speed such as regular service and premium service. This will enrich the analysis from the standpoint of developing a service based network design.
- The considerations of link capacities can also be incorporated, which is based on the limited availability of transportation assets required to move freight from one point to the other. This case can be further extended to the use of multiple modes along each lane instead of the all-or-nothing approach used in this research.
- The research on network design under hub congestion can be extended by considering the option of direct shipments, impact of modal variability and batch size (truck-load/container loads) on shipment waiting times.
- Finally, the use of ocean freight can be included in the model and the data, which will extend the scope of this research to the domain of global logistics.