Last-Mile Network Design for Urban Commodity Distribution in Latin America

by

Esteban Ezequiel Mascarino

Industrial Engineer
Universidad Nacional de Rosario, Santa Fe, Argentina (2012)

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN TRANSPORTATION

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2018

Signature redacted

Author ....
Department of Civil and Environmental Engineering
August 10, 2018

Certified by.............. Signature redacted ....
Matthias Winkenbach
Director, Megacity Logistics Lab
Thesis Supervisor

Signature redacted
Accepted by ....
Heidi Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

[This page was intentionally left blank]

Last-Mile Network Design for Urban Commodity Distribution in Latin America

by

Esteban Ezequiel Mascarino

Submitted to the Department of Civil and Environmental Engineering
on August 10, 2018, in partial fulfillment of the
requirements for the degree of
MASTER OF SCIENCE IN TRANSPORTATION

Abstract

Transportation, as the carrier of freight and passengers, is undeniably one of the fundamental components required for both economic growth and development. In an urban environment, freight movements support most city-based activities, while detrimentally impacting the quality of life through negative externalities (e.g., congestion, noise and air pollution, etc.). Specifically, last-mile delivery is regarded as an important yet highly expensive section within every supply chain. This is partially caused by inherent inefficiencies such as prolonged delays in traffic and unproductive idle periods at customers locations, among others. Consequently, there is a need for methodologies addressing the design of improved last-mile delivery networks. In this context, the optimal design of distribution systems requires an integrated view of strategic, tactical, and operational decisions. This work contributes with a mathematical framework that provides such an integrated view while leveraging both customer-generated waiting time inefficiencies and existing network infrastructure to serve additional clients. It also provides computationally feasible algorithms to obtain solutions for realistic situations.

First, we formulate a single-echelon, multi-depot, capacitated routing problem. Employing a brownfield approach, this model optimizes the fleet composition as well as the delivery schedule and allocation to distribution facilities of medium- and high-dropsize clients, hereafter 'big-box' customers. This Routing Problem (RP) is modeled as a special case of a Bin Packing Problem (BPP) combined with a customer clustering approach. However, given its high combinatorial complexity, two alternative methodologies, a two-step approach and Benders Decomposition (BD), are tested to reduce computational times. Second, we develop a two-echelon extension, which builds on the previous model, to evaluate the economic impact of including a large number of low-dropsize customers, also known as 'nanostores', into the original distribution footprint. Those newly added customers will be served through the second echelon using a subset of the original big-box customer locations as transshipment points. To solve this Location-Routing Problem (LRP), a three-step iterative optimization approach is developed and tested.

Both models are applied to a real-world consumer goods distribution case study in Latin America. Results suggest that a systematic and properly framed optimization approach, which makes efficient use of available resources, can significantly reduce the total distribution cost. Further, we show that the case study company, leveraging its existing assets and addressing inherent network inefficiencies, can efficiently expand its distribution footprint towards nanostores.

Thesis Supervisor: Matthias Winkenbach
Title: Director, Megacity Logistics Lab

[This page was intentionally left blank]

"Science, technology, and research are the basis of health, well-being, wealth, power, and independence of modern peoples. There are those who believe that scientific research is merely a luxury or an interesting yet dispensable amusement. Serious mistake; it is an urgent, immediate, and inescapable need in order to get ahead. The dilemma is then clear; either science, technology, and research are cultivated so the country gets prosperous, powerful, and moves forward; or they are not properly practiced so the country stagnates and retreats, living in poverty and mediocrity. Rich countries are so because they dedicate resources to scientific and technological development. Poor countries remain as such if they do not invest in those areas. Science is not expensive, what is expensive is ignorance." (translated from Spanish)

Dr. Bernardo Houssay

Nobel laureate in Medicine - 1947, Argentinean

Acknowledgments

My first thanks go to my advisor, Dr. Matthias Winkenbach, for his constant support, motivation, patience, and kindness during these past years. I highly regard his professional guidance, inspiration, and encouragement, which have certainly transmitted to me his enthusiasm for last-mile logistics. It has not only been a great pleasure, but also a huge honor for me to work under the supervision of such a respected scholar. I look forward to continue collaborating with him in the near future.

I am also profoundly thankful to Dr. Daniel Merchán, Dr. Sergio Caballero, and André Snoeck, my first coworkers at the Megacity Logistics Lab and now certainly my friends, for their mentoring and guidance inside and outside academia. Their experience and knowledge have always been enlightening since the three of them are role-models for me.

In addition, I thank Dr. Mohammad Moshref Javadi, Dr. Milena Janjevic, Maggie Wilson, and Xavier Lavenir. It has been a memorable experience joining this talented research team packed with promising young scientists.

Moreover, I would like to express my gratitude to Professor Yossi Sheffi, Dr. Chris Caplice, Jim Rice, Dr. Josué Velázquez Martínez, Katie Date, and Dr. Christopher Mejía for providing me with useful advice and various academic opportunities at CTL.

None of this would have been possible without the love and support of my family. Marcela, Sergio, Matías, María del Carmen, and Miguel Ángel have definitely been the cornerstone of this achievement, teaching me that dedicated work is the main path to fulfill any goal in life.

Furthermore, I express my appreciation to the professors at my alma mater who have always encouraged me to dream big and pursue this endeavor abroad.

Thanks to my friends in Argentina and to my new friends in Boston for being present in relevant moments throughout this transformative challenge.

Last and but not least, I would like to thank Fulbright-BecAr, CTL, and MIT for funding me and my research during these two years.

In loving memory of Nélida María, Miguel Ángel, and Carolina

Esteban E. Mascarino

Biographical Note

Esteban E. Mascarino is an Industrial Engineer from the National University of Rosario (UNR), Santa Fe, Argentina. After his graduation in 2012, he received several honors for his academic performance including the Engineering School's gold medal —the highest award for graduating students—, and a recognition from the National Engineering Academy, among others. Before joining MIT in 2016, both as a Master of Science in Transportation (MST) candidate and as a Research Assistant at the Megacity Logistics Lab (MLL), Esteban led the design and implementation of various agro-mechanical manufacturing projects in Industrias John Deere Argentina. In 2015, he was awarded the Fulbright-BecAr Fellowship to pursue his Master's degree in the United States.

Contents

List of Acronyms

The following list includes, in alphabetical order, all the acronyms used throughout this work.

API Application Programming Interface

BD Benders Decomposition

BIP Binary Integer Program

BPP Bin Packing Problem

DC Distribution Center

IP Integer Program

LAP Location-Allocation Problem

LP Linear Program

LRP Location-Routing Problem

MILP Mixed-Integer Linear Program

MP Master Problem

OR Operations Research

PapComp The Paper Company

RLE Route Length Estimation

RP Routing Problem

SKU Stock Keeping Unit

SP Subproblem

TP Transshipment Point

TSP Traveling Salesman Problem

UGT Urban Goods Transportation

ULN Urban Logistics Network

UN United Nations

VRP Vehicle Routing Problem

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

1.1 Outline

This thesis consists of eight chapters, addressing specific aspects of the decision problem that forms the basis of our analysis. Chapter 1 outlines the motivation, research objectives and contributions of the present work. Chapter 2 broadly reviews the existing literature on routing models, Location-Allocation Problems (LAPs), and LRPs. Chapter 3 provides a brief description of The Paper Company (PapComp), a manufacturer of paper tissue products in Santiago de Chile that serves as a real-world case study. Moreover, this chapter positions the problems being addressed by each of the optimization models formulated in this work in light of their real-world application. Next, Chapter 5 introduces a deterministic single-echelon distribution network design model to meet the demand of large- and medium-dropsize customers, referred to as 'big-box' customers in this work. A two-echelon extension of this model that incorporates, to the existing urban distribution footprint, a set of highly fragmented low-dropsize traditional retail establishments, known as 'nanostores', is addressed in Chapter 6. In Chapter 7, we present the results after applying the developed models to specific problem instances, and we provide insights at this respect. Finally, Chapter 8 concludes by summarizing our key findings while exploring potential extensions and future research avenues.

1.2 Motivation

On average, the global population living in urban areas is expected to keep growing by 65 million people annually until 2050, according to a study conducted by the United Nations (UN) Department of Social and Economic Affairs (United Nations Population Division 2014). This rapid growth in urban population is particularly pronounced in emerging economies such as Latin America. In 2012, the urban population accounted for 79% of the so-called emerging markets' inhabitants, and it has been projected to reach approximately 85% by 2030 (United Nations Population Division 2014).

Increasing urbanization and population growth directly translate to a rising demand for goods and services, and their supporting logistics activities (Taniguchi and Thompson 2014). This increased demand leads to a rise in urban traffic load from commercial vehicles with its associated externalities, such as congestion and pollutant emissions (Taniguchi 2014). Nevertheless, in a globalized world, where economic opportunities have been increasingly related to the mobility of people, freight, and information, the transportation sector is still the irreplaceable piece that makes economic and social development possible (Rodrigue and Notteboom 2017, Greene and Plotkin 2011). Therefore, under the previously described scenario, models and tools aiming at optimizing the utilization of transportation-related resources play an important role for both private and public sectors.

Within the transportation ecosystem, the field of Urban Goods Transportation (UGT) is particularly complex to manage. UGT focuses on the last mile of the supply chain to fulfill customer requests in light of increasing service level expectations within urban environments (Ehmke 2012, Bektas et al. 2015). UGT is regarded as one of the most expensive yet typically least efficient parts of every supply chain (Gevaers et al. 2011). Therefore, an effective Urban Logistics Network (ULN) that is properly designed to serve demand while coping with externalities such as congestion is paramount for efficient and reliable distribution operations in every urban market (Crainic et al. 2004, Snoeck et al. 2017). The previous is especially accentuated in emerging markets where the nanostore channel accounts for around 50% of the market share in many emerging-market megacities (Fransoo and Blanco 2013).

According to Fransoo et al. (2017), these nanostores can be defined as small, independent businesses that have a minimum headcount, a reduced amount of commercial volume, and generally function with a single showroom. These traditional retailers are usually family-run businesses that operate in a low entry barrier market. Moreover, since they serve a relatively low number of consumers in their immediate neighborhood, they know their customers and can then grant informal credit to them, reason why nanostore operators are usually short of cash (Boulaksil 2012). Thus, one of the main challenges in the logistics planning and execution is to schedule deliveries such that these stores have enough cash available to pay for their merchandise (Fransoo and Blanco 2013). Furthermore, due to the nanostores constrained space, the product assortment depth is limited.

Consequently, their suppliers go through large efforts in trying to secure part of this shelf space. Additionally, there is a high concentration of nanostores in an urban environment which is generally influenced by population density. The aforementioned features of the traditional channel are, in almost all aspects, opposite to the modern channel (i.e., the channel that groups big-box customers) leading to different distribution network designs.

Hence, to serve a specific market, companies can opt from four relevant distribution channel policies: modern-channel-only, traditional-channel-only, first-modern-then-traditional-channel, and first-traditional-then-modern-channel (Ge 2017). For the purpose of this research, we start considering a modern-channel-only policy in which a company exclusively serves big-box customers. We then transition to a combined first-modern-then-traditional-channel policy to integrate nanostore deliveries. Both commercial and operational reasons motivate this second approach. For instance, when integrating nanostores into the network's footprint, a company can better control the distribution channel bridging the gap with its end consumers. Hence, this company can better segment the market and target specific products to well-defined niches directly influencing the final price. Moreover, it can leverage idle times at big-box customer locations to perform new value-added activities. Furthermore, this policy provides a direct source of revenue with potentially higher relative margins than the ones obtained from the modern channel.

Therefore, in the design of efficient ULNs, applied Operations Research (OR) can deliver a valuable contribution. OR, also referred to as management science or decision science, is a systematic approach aiming to model complex real-world managerial problems. Employing techniques from other mathematical sciences, such as mathematical modeling, statistical analysis, and mathematical optimization, OR arrives at optimal or near-optimal solutions to complex decision-making problems (Salimifard and Shahbandarzadeh 2012). Thus, OR models can lead to a more efficient use of resources, while helping us to identify the trade-offs between different dimensions of a particular problem (Dekker et al. 2012). OR, in particular applied optimization, is an important tool to streamline last-mile distribution operations, which are often considered to be one of the cornerstones of every supply chain, representing the face of the company to the consumers (Goodman 2005).

1.3 Research Objective and Contributions

This work contributes to the existing research on ULN design integrating nanostore deliveries into distribution networks that are big-box customer-oriented by leveraging operational inefficiencies of the current network setup. To achieve this, we first develop a methodology to optimize the existing distribution network by simultaneously accounting for big-box customer allocation to facilities and scheduling decisions. Second, we extend the previous methodology by integrating nanostore deliveries into our distribution network leveraging the available infrastructure as well as inefficiencies derived from waiting times at big-box customers.

Therefore, we propose an analytical framework to address both the capacitated RP and the capacitated LRP (i.e., serving only big-box customers, and serving both big-box customers as well as neighboring nanostores, respectively) embedded with customer scheduling and a capacity-constrained mixed vehicle fleet by implementing a variation of the well-known BPP (cf., Martello and Toth (1990) for a detailed description of this problem, some of its variations, and solution strategies).

We further suggest a 'divide-and-conquer' approach to efficiently solve the routing part of both problems, by geographically clustering customers before implicitly routing vehicles within these customer clusters. By considering only intra-cluster routes within our models, we ensure their tractability even for large-scale problem instances. Hence, we significantly reduce their combinatorial complexity and, as a direct consequence, their computational burden.

Given the computational complexity of our proposed models, which are generally known to be NP-hard (Korte and Vygen 2018), we explore different solution strategies and compare their results in terms of computational effort (runtime), resulting total delivery cost (objective function), and network design characteristics (decision variables).

We also evaluate the economic impact of an optimization-based network re-design using a stylized example of an actual industry case study. We further examine the impact of integrating highly fragmented nanostores in their currently existing distribution network.

Chapter 2

Literature Review

The subsequent sections of this chapter broadly explore the available literature on mathematical models and solution techniques for RPs, LAPs, and LRPs. We conclude by stating the intended contributions of our work to the existing literature.

2.1 Routing Models

According to Kumar and Panneerselvam (2012), the goal of the Vehicle Routing Problem (VRP) is to design cost optimal delivery routes from a point of origin to a group of geographically dispersed locations, e.g., customers, subject to a set of constraints. Starting with the seminal papers of Dantzig and Ramser (1959) and Clarke and Wright (1964), the VRP has been extensively studied in the specialized literature. Several variants were formulated depending on the nature of the transported goods, the required service level, and the customer- and vehicle-specific characteristics, among other criteria. For instance, four relevant variants are the Capacitated VRP, the Multi-depot VRP, the VRP with Time Windows, and the Dynamic VRP.

To address this problem and its variations, several exact solution approaches have been proposed over the years including branch-and-bound, branch-and-cut, and branch-and-price algorithms. A branch-and-cut-and-price strategy was suggested by Ropke et al. (2007) for the Capacitated VRP. Similarly, Desaulniers et al. (2014) proposed a branch-and-price algorithm to solve a VRP with Time Windows. Moreover, Azi et al. (2010) and Qureshi et al. (2009) solved the VRP with scheduling and time windows using a column generation approach and Dantzig-Wolfe decomposition, respectively. Other works, such as the ones of Balinski and Quandt (1964), Agarwal et al. (1989), and Alvarenga et al. (2007), use a variation of the Set Partitioning formulation to solve VRP instances. However, being the VRP an NP-hard type of problem, currently available exact solution methodologies are limited to instances in the order of 50 to 100 customers.

Consequently, current research focuses on approximate algorithms (dividing them into construction heuristics, improvement heuristics, metaheuristics, and hybrid methods) that are capable of finding high quality solutions within a limited runtime. The previous renders the approximate methodologies suitable for real-life problem instances which are characterized by large vehicle fleet sizes and significant number of customers. Among the most relevant heuristics we find the cluster-first-route-second approach and seed-based approximations described in Fisher and Jaikumar (1981), neighborhood search heuristics proposed by Gendreau et al. (2006), and scatter search algorithms used in Belfiore and Yoshida Yoshizaki (2009). Regarding the most prominent metaheuristics we can mention Tabu Search (cf., Potvin et al. (1996)), Genetic Algorithms (see Blanton and Wainright (1993) and Salhi and Gamal (2003)), Evolutionary Algorithms (cf., Garcia-Najera and Bullinaria (2011)), and Ant Colony Optimization algorithms (see Montemanni et al. (2005)). Finally, hybrid methodologies use a combination of exact, heuristic, or metaheuristic procedures to solve the problem. For example, De Backer and Furnon (1997) have proposed a hybrid method that combines constraint programming with metaheuristics. In this context, VRP literature offers an ample variety of heuristic, metaheuristic, and hybrid approaches which are surveyed in, for instance, Laporte (1992), Cordeau et al. (2005), and Gendreau et al. (2002).

For comprehensive literature reviews about VRPs and their variations see Toth and Vigo (2014), Eksioglu et al. (2009), and Montoya-Torres et al. (2015).

2.2 Location-Allocation Problems

The LAP's objective is to determine the optimal location for one or more facilities that will serve demand from a given set of customers within a specific area (Azarmand and Jami 2009). Even though LAPs are not being developed in this work, it is worth mentioning them since they represent a formulation, in which routing decisions are ignored, that precedes LRPs.

Starting with the seminal works of Cooper (1963) and Hakimi (1965), Badri (1999) provided an extensive review of LAPs. Scaparra and Scutellà (2001) then proposed a unified framework for characterizing the different components of these problems, namely facilities, demand and customer features, and locations.

Since LAPs are NP-hard models, numerous algorithms have been designed, involving branch-and-bound (cf., Kuenne and Soland (1972)), Simulated Annealing (see Murray and Church (1996)), Tabu Search (cf., Brimberg and Mladenovic (1996) and Ohlemüller (1997)), and P-median (see Hansen et al. (1998)). Some hybrid approaches have also been suggested, such as the ones based on Simulated Annealing and Random Descent methods (e.g., Ernst and Krishnamoorthy (1999)) as well as those applying Lagrange relaxation and Genetic Algorithms (e.g., Gong et al. (1997)). Finally, Brimberg et al. (2000) improved some solution methodologies with variable neighborhood search, which proved to obtain better results when the number of facilities to locate is large.

2.3 Location-Routing Problems

LRPs combine two basic planning tasks in logistics. For these problems, decisions on the location of arbitrary facilities (e.g., production plants, Distribution Centers (DCs), Transshipment Points (TPs), etc.) are jointly taken with vehicle routing decisions (Drexl and Schneider 2015). It is well-known that making these types of decisions independently of one another may lead to highly suboptimal planning results (Salhi and Rand 1989). It is important to highlight that the allocation of customers to delivery facilities is defined within the routing decisions of the problem. However, the computational complexity of the VRP has heavily limited the possibility of simultaneously addressing these decisions (Perl and Daskin 1985).

The literature on LRP is fairly extensive and mostly focused on single-echelon distribution systems (Merchan et al. 2015). Several surveys on LRP developments were released over time including the works by Laporte et al. (1988), Nagy and Salhi (2007), and, more recently, by Prodhon and Prins (2014) and Drexl and Schneider (2015).

Regarding multi-echelon LRPs, several models have been proposed over the past years in the context of ULNs. For example, Ambrosino and Scutellà (2005) developed mathematical programming formulations for the three-echelon LRP. Moreover, Gonzalez-Feliu (2012) generalized the LRP modeling to include n-echelon networks. The previous work presented a Mixed-Integer Linear Program (MILP) formulation based on partitioned sets. It also elaborated on its potential applications including, for example, ULNs. In this context, Boccia et al. (2011) explored extensions for the ULN systems described in Crainic et al. (2004).

The use of metaheuristic solution procedures has been studied over the past years. For instance, a Tabu Search heuristic based on an iterative-nested approach has been proposed both by Boccia et al. (2010) and Crainic et al. (2011a). Furthermore, Contardo et al. (2012) developed a branch-and-bound procedure for mid-size instances and an adaptive large neighbor metaheuristic for bigger problems. Nevertheless, none of the previously mentioned works have included real-world applications. In this context, Winkenbach et al. (2016) have addressed a real industry case-study problem by proposing a MILP formulation for a capacitated two-echelon distribution network, along with a two-stage solution heuristic and a closed form route cost approximation.

2.4 Intended Contributions

Based on the existing literature and given the applied nature of the problems being solved, this work proposes the following specific contributions:

1. Generate an analytical framework to address both a single-echelon and a two-echelon distribution network design optimization that simultaneously accounts for customer allocation and scheduling decisions as well as a capacity-constrained mixed vehicle fleet definition.

2. Leverage network inherent time inefficiencies as well as the available distribution infrastructure to extend the network's footprint and serve new customers with a limited cost increase.

3. Develop alternative solutions strategies that can efficiently address larger problem instances using the proposed analytical framework.

Chapter 3

Case Study

3.1 The Company

PapComp is a Chilean pulp and paper company focused on the integrated forest industry, with a strong footprint in the pulp and paper industry, as well as the sawmill industry. PapComp consists of three independently operating business units: Cellulosic Products, Paper and Paper Products, and Tissue. The company has branches all over Latin America.

In the following, we apply our research to the operation of the Tissue division within Chile's Metropolitan Region, composed of Santiago de Chile and 51 surrounding municipalities. The Tissue division represents about 71% of the whole Chilean tissue market (i.e., toilet paper, paper towels, paper napkins, and facial tissues), as well as 29% of the country's sanitary product market (i.e., baby and adult diapers, feminine care, and wipes). Similar market share figures also hold for the Metropolitan Region's niche.

Currently, PapComp's distribution network operates with two DCs and one satellite facility. The DCs, located in the municipalities of Puente Alto and Talagante, also function as production facilities, manufacturing roughly 20% and 80% of the Stock Keeping Units (SKUs), respectively. A similar split also holds volume-wise. The satellite facility, situated in the municipality of Pudahuel, is mainly used for storage purposes and rarely selected as a delivery origin. Figure 3.1 presents a map with the facilities' geographical locations.

From the two DCs (i.e., Talagante and Puente Alto), PapComp delivers to more than 500 big-box customers within the Metropolitan Region, including supermarkets, pharmacy chains, wholesalers, and distributors. Figure 3.2(a) summarizes the composition of PapComp's client base per customer type. Figure 3.2(b) illustrates the fraction of PapComp's annual demand volume (in [m^3]) associated

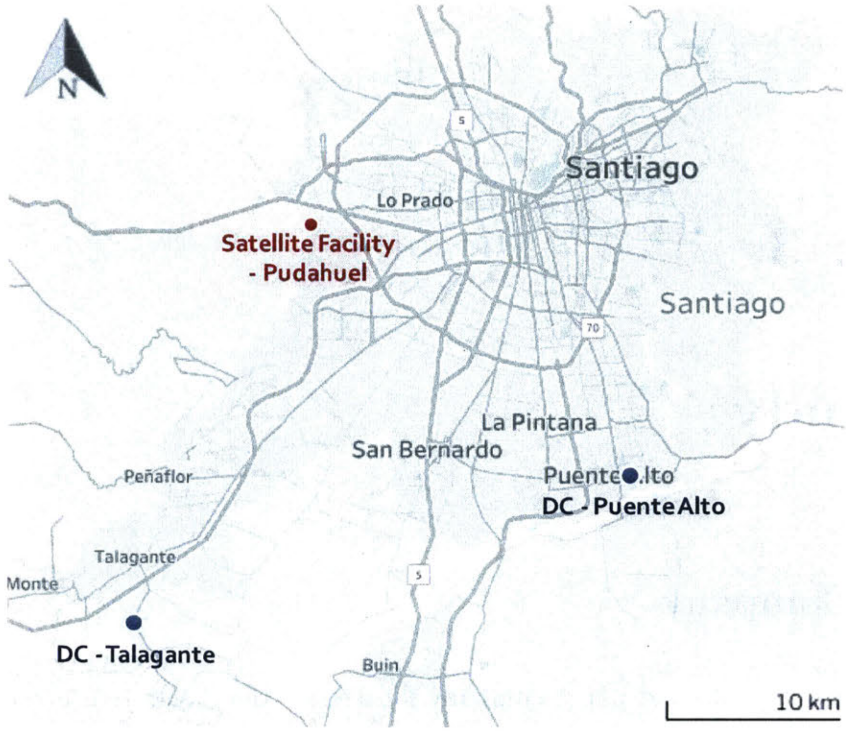
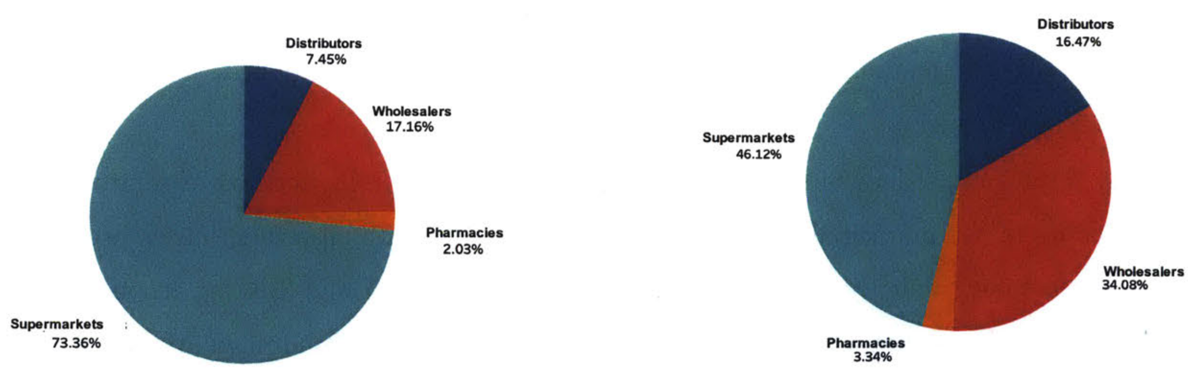


Figure 3.1: Geographical locations of facilities

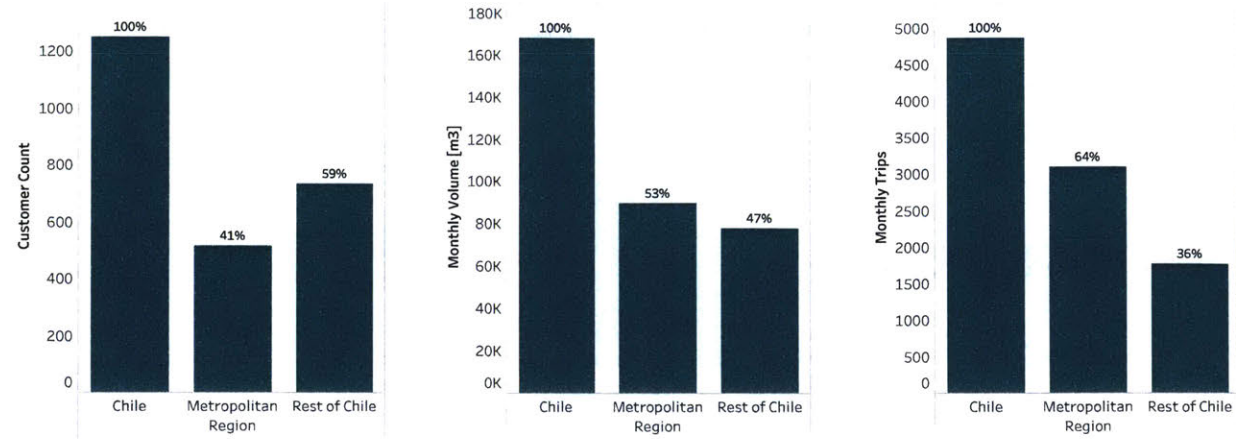
with each of these customer groups. Supermarkets, wholesalers, and distributors comprise almost 97% of the Metropolitan Region's demand in volume. Moreover, it is worth highlighting that distributors and wholesalers generate more than 50% of the demand even though they represent less than 25% of PapComp's customer base.



(a) Customer type composition (b) Volume split per customer type

Figure 3.2: Customer types and volume split within Metropolitan Region

As shown in Figure 3.3, customers located within the Metropolitan Region (i.e., slightly more than 40% of PapComp's customer base) account for more than 50% of the company's total volume as well as 60% of the trips departing from the three available facilities.



(a) Geographical customer split (b) Geographical volume split (c) Geographical trips split

Figure 3.3: General comparison between Chile, Metropolitan Region, and the other regions

Currently, the company has completely outsourced its last-mile delivery activities to more than 12 independent carriers. These transportation firms offer a combined capacity of around 210 vehicles, which are classified in six different categories based on their loading capacity and crew size. Table 3.1 summarizes the characteristics associated with each vehicle type.

Table 3.1: Main characteristics per vehicle type

Vehicle type	**Capacity range** [m^3]	**Median capacity** [m^3]	**Crew size**	**Maximum availability** [vehicles]
C1	< 7.0	5.8	1	30
C2	7.0 – 13.6	9.3	1	35
C3	13.6 – 22.5	13.3	1	30
C4	22.5 – 30.0	26.8	2	15
C5	30.0 – 66.0	39.5	2	50
C6	> 66.0	103.0	3	50

The payment scheme that PapComp negotiated with its carriers can be described as a flat fare per vehicle trip. This fare is a function of the vehicle's category, the facility of origin, and the municipality in which the farthest customer to be visited within a trip is located. Hence, there is no explicit limit to the number of stops per vehicle trip. Instead, this number is bounded by the total delivery time available, which is limited to 10 hours a day by law, and/or the vehicle's maximum payload. This flat-fare scheme is common practice for outsourced delivery operations.

It provides a clear accrual methodology for both parties by simplifying the billing process while avoiding potential misinterpretations. Besides, a flat-fare payment scheme streamlines the daily optimization of routing decisions.

3.2 Problem Setting

As mentioned in Section 3.1, PapComp's Tissue division serves over 500 big-box customers in Chile's Metropolitan Region aggregating, on average, 18,000 m^3 of delivered products per week. The vast majority of these customers have a specific weekly delivery frequency and a schedule associated to them (e.g., visited three times a week, namely on Mondays, Wednesdays, and Fridays). PapComp must honor these schedules every time a customer places an order.

Generally speaking, PapComp has to manage its customers' orders to assure on-time delivery; manage its stock levels in order to guarantee the availability of each SKU within each facility; define a daily delivery route for each vehicle; pick and prepare the orders; load the vehicles at their facilities; manage payments, returns, complaints, etc.

The carriers must then transport the cargo from the origin facility to each customer included in a delivery tour, fill the associated paper work at the customer's location, unload the products requested by each customer, and report the delivery status (e.g., complete delivery, partial delivery, failed delivery) back to PapComp.

The goal of this research is to develop two large-scale network optimization models for the urban distribution of paper products and to apply them to PapComp's last-mile distribution operations in Chile's Metropolitan Region. In particular, the network design models we develop will address the following questions related to the company's last-mile operational footprint:

1. Distribution facilities
    - How many distribution facilities of which kind (i.e., DCs and/or satellite facility) should be used to serve the urban market under consideration?
2. Service areas and route territories
    - Which customer should be served from which distribution facility?
    - On which day(s) of the week should each customer be served?
    - Which customers should be assigned to which delivery route?

The models to be developed will be optimizing PapComp's operational footprint for cost (cost minimization objective), while respecting certain constraints and minimum requirements with regards to service level (see Chapters 5 and 6 for further details).

3.3 Scenarios of Analysis

Starting off with PapComp's current customer footprint, each model presented in this work will be used to address one of the following scenarios:

1. Serving PapComp's existing customer base (*optimizing the status quo*): This scenario will suggest potential changes to the distribution footprint in order to make current operations more efficient. In particular, this analysis will help to validate
   i) whether the company is operating the right number and type of facilities,
   ii) whether the company is deploying the right vehicle types to the right customers from the right facilities, and
   iii) whether all customers are scheduled on the optimal delivery days.

   In summary, we aim to determine the optimal operational footprint if PapComp kept serving only big-box customers.

2. Adding traditional retail to the customer base (*integrating nanostores*): This scenario serves to quantify the potential impact of future changes to PapComp's distribution strategy on the required underlying network infrastructure and fleet. Specifically, we are considering a future scenario in which a subset of nanostores is served strictly by PapComp instead of indirectly through wholesalers and distributors. The demand from the integrated nanostores will be cannibalized from the existing wholesalers and distributors. Analyzing this scenario will help
   i) to determine the optimal operational footprint under this new instance and
   ii) to quantify the cost effects associated with this change in customer footprint.

Chapter 4

Data Availability and Processing

This chapter describes the different sources of data gathered and queried to obtain useful and accurate information to feed the optimization models. These sources are a combination of company-specific tables (see Section 4.1) and external databases (see Section 4.2).

4.1 Company-specific Data

These data provide information related to PapComp's customers, their associated orders, and additional details about every delivery visit. Moreover, they present relevant details about facilities and vehicles. Figure 4.1 provides a simplified relational map that schematizes the connection between company-specific data sources. Algorithm 4.1 conceptualizes the customer-related data processing strategy by combining information from the customers', deliveries', and shipping report's tables. Appendixes B and C partially compiles this information and provides further data (e.g., delivery fares, vehicle-specific waiting times, etc.) for the stylized problem instances that are solved in this work.

4.1.1 Customers

The customers' database is of central relevance for any distribution network design model since it supplies vital information about the points of delivery to be visited. In this particular work, the database provided us with the following key pieces of information in which each record represented one specific point of delivery to be considered by our model:

- customer identification number, name, and type (i.e., supermarket, wholesaler, distributor, or pharmacy)

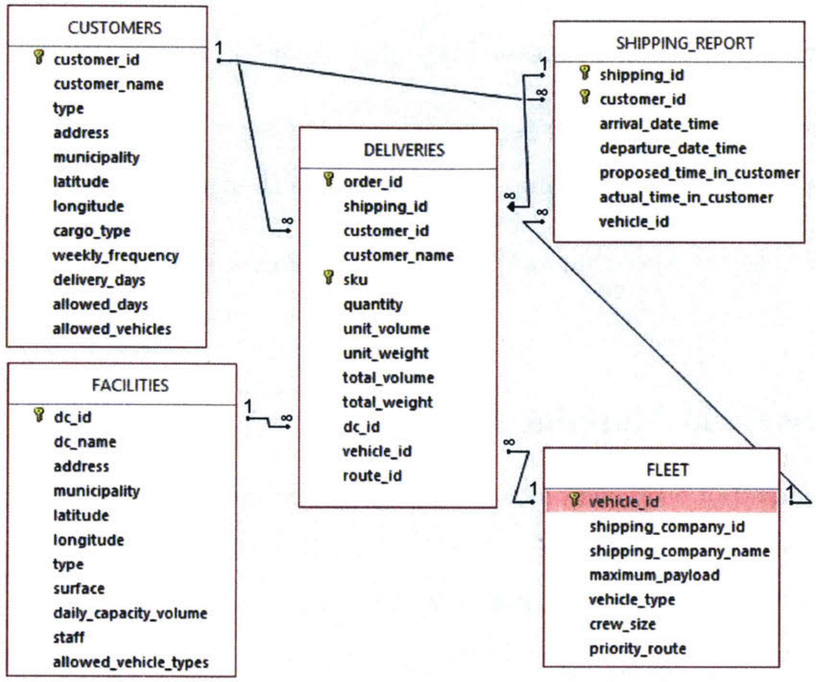


Figure 4.1: Schematic relational database structure

Algorithm 4.1: Customer data processing steps

Data:

- *customers* table
- *deliveries* table
- *shipping report* table

Result: List of *customers* with average demand per delivery day as well as average waiting time

remove duplicated *customers*;
filter *deliveries* based on the analyzed time horizon (e.g., one-year period);
filter those *customers* that did not receive any *deliveries* within the predefined time horizon;
infer weekly delivery frequency (number of deliveries/number of active weeks) and visit days (based on the estimated frequency) for those *customers* for which these fields have not been specified;
calculate the average demand per delivery day for each *customer* (total volume delivered in each delivery day/total number of active weeks);
calculate the average waiting time for each *customer* (sum of waiting times on each delivery/number of deliveries) based on the waiting time provided by the *shipping report*;
return updated *customers* table.

- address, geolocalization (i.e., latitude and longitude), and the associated municipality
- cargo-type requested by the customer (i.e., either 'bulk' or 'palletized')
- weekly delivery frequency, agreed delivery days, and the set of allowed delivery days on which each customer can be visited, if the delivery schedule changes
- set of allowed vehicles types that can be used to visit each customer based on its vehicle access restrictions

4.1.2 Deliveries and Shipping Report

To estimate the demand of each customer on the associated delivery days, historic data on orders and deliveries are required. Moreover, for a deeper understanding of the customer visit dynamics, additional information on each delivery is provided by the shipping report.

The deliveries' table, contains a record for each SKU being delivered to a particular customer. It then provides the following relevant data fields:

- order and shipping identification numbers
- customer name and identification number
- delivered SKU, associated quantity (in [bulks]), volume (in [m^3]), and weight (in [kg])
- facility of origin
- truck and route identification numbers

For each combination of shipment (defined by its identification number) and customer, the shipping report provides the arrival and departure date-time stamps, the delivery vehicle, and the planned and actual duration of the visit (i.e., the difference between the departure and the arrival times).

4.1.3 Facilities

Information about the relevant delivery facilities (i.e., DCs and satellite facilities) is also required. Their geolocations (i.e., latitude and longitude) as well as their daily capacities are key inputs for our models. Moreover, information about vehicle access restrictions, defining which vehicle types are allowed to operate in each facility, is also relevant. Finally, details about the type of facility (i.e., either DC or satellite facility), headcount, and footprint are not needed but offer complementary information.

4.1.4 Vehicle Fleet

Finally, data about the available fleet are necessary to derive a correct vehicle profile. For this reason, we mostly require the maximum payload (in [m^3]) that each vehicle can carry together with its crew size and the associated vehicle type (i.e., C1 to C6). That information has been previously summarized in Table 3.1.

4.2 External Data

Further sources of data are necessary to compute additional parameters required by our models. In this section, we describe the use of two external databases provided by Google® which were queried through their Application Programming Interfaces (APIs) accessible from Python 2.7.

4.2.1 Travel Distances

The developed models require inter-customer driving distances to approximate the delivery distance of each vehicle trip. Moreover, the driving distance from each facility to the centroid of every municipality is necessary to approximate the line-haul on each delivery route.

For this reason, we have generated a driving distance matrix containing the previously mentioned elements (i.e., inter-customer driving distance and driving distance from each delivery facility to each municipality centroid) through repetitive queries to the Google Distance Matrix® service. The Distance Matrix API is a service that provides travel distance and time for a matrix of origins and destinations, based on the recommended route between start and end points[1].

Chapters 5 and 6 illustrate how these distances (or times) are integrated within our models.

4.2.2 Commercial Establishments

If we model the integration of nanostores in the existing delivery network, the geographic location of these potentially new customers must be estimated when disaggregated economic census data or Chamber of Commerce's data is unavailable. For this reason, we employ the Google Places®

[1]https://developers.google.com/maps/documentation/distance-matrix/start

API, which is a service that returns information about places using HTTP requests. Those places are defined within this API as either establishments, geographic locations, or prominent points of interest[2]. For our purposes, we consider only those establishments classified as either convenience stores, grocery stores, or supermarkets. We then filter out any other commercial classification (e.g., pharmacies, gas stations, etc.) as well as specific commercial establishments (i.e., those medium- and big-size stores which are currently being served by PapComp) that are irrelevant for the scope of our research. As a result, a total of approximately 17,000 nanostores within Chile's Metropolitan Region are deemed relevant for PapComp's delivery purposes.

Although this methodology provides us with highly granular data, it encompasses clear limitations. First, since nanostores are generally characterized by their informality (Fransoo and Blanco 2013), a certain number of them might not be considered because they are simply not uploaded in Google's database. Moreover, no information about the establishment's size or its associated sales level is provided by this service. Finally, and closely connected with the previous point, no conclusions about its delivery frequency can be inferred from the available data.

[2]https://developers.google.com/places/web-service/intro

Chapter 5

Single-echelon Routing

This chapter introduces a mathematical formulation for the capacitated RP with customer scheduling and a mixed vehicle fleet as a variation of the BPP. Besides a one-step exact solution approach, we present two alternative approaches that aim at reducing the initial model's runtime: an approximate method, using a two-step solution approach, and the original program's transformation applying BD. In line with Section 3.2, these MILPs determine the cost-optimal network configuration for last-mile-delivery to big-box customers by defining

i) the optimal allocation of existing big-box customers to delivery facilities,

ii) a delivery day schedule for each customer that is compatible with its weekly delivery frequency, and

iii) the optimal fleet composition to serve existing customers.

In the following, we will first outline the set of assumptions that apply to each of the optimization models presented in this chapter. Second, the initial one-step formulation of the optimization problem is developed. Next, we present a two-step relaxation of the initial model. Finally, the application of BD to the original formulation is described. A comparison of the results obtained from all three modeling and solution techniques for various problem instances is presented in Section 7.1.

5.1 Model Assumptions

The following set of assumptions, associated with the network design, is considered within the formulation of the problem. For clarity purposes, we will divide them across the different components of the model.

I. Demand

The information concerning the demanded volume is assumed to be static and deterministic for a one-week time horizon. Moreover, all demands must be satisfied (i.e., a 100% service level is assumed), eliminating the need to include costs of lost sales.

II. Customers

All customers must be scheduled according to their predefined weekly delivery frequency and to one of their allowed delivery schedules. Therefore, they are visited on as many days in a week as defined by their weekly delivery frequency. Moreover, each customer can be allocated to only one distribution facility, and can be categorized as either 'bulk' or 'palletized' based on the containerization type of the delivered products.

Furthermore, customers are clustered based on the municipality in which they are located as well as their associated cargo-type (i.e., 'bulk' or 'palletized').

Regarding the proposed delivery schedules, the associated delivery days (within a delivery schedule with multiple delivery days) must be as spread-out as possible. Hence, the model considers the following discrete choices of schedules to pick from:

- *Frequency 1*: {M, T, W, Th, F, S}
- *Frequency 2*: {M-Th, T-F, W-S}
- *Frequency 3*: {M-W-F, T-Th-S}
- *Frequency 4*: {M-T-Th-F, T-W-F-S, M-W-Th-F}
- *Frequency 5*: {M-T-W-Th-F, M-T-W-Th-S, M-T-W-F-S, M-T-Th-F-S, M-W-Th-F-S, T-W-Th-F-S}
- *Frequency 6*: {M-T-W-Th-F-S}

where M, T, W, Th, F, and S denotes Monday to Saturday, respectively.

On each delivery day, the number of visits to a customer must be minimized. In other words, split deliveries are allowed only if the daily demand of a customer exceeds the maximum payload of the largest allowable vehicle which can visit that customer. Should delivery splitting be required, the orders can be partitioned always minimizing the number of visits to each customer in a day.

Some customers present vehicle restrictions, i.e., they can be visited only by specific vehicle types (e.g., some customers cannot be visited with C5 or C6 trucks due to limited parking space). Moreover, some customers do exhibit delivery day restrictions. Hence, not all delivery schedules associated with the customer's weekly delivery frequency are allowed (e.g., some customers might not be open on Saturdays, rendering all schedules that contain Saturday deliveries infeasible for them).

III. Facilities

All facilities (i.e., both DCs and the satellite facility) are open every day of the week, and no activation cost is considered. Furthermore, since the product transfer between DCs, which are also production centers in our problem, is simple to execute, we consider that all SKUs are available at each facility. The previous triggers the inclusion of transfer costs within the model. These costs are then assumed to be proportional to the volume delivered from each facility. Finally, not all vehicle types are allowed to operate at each facility (e.g., only C6 vehicles can be dispatched from the satellite facility in Pudahuel).

IV. Fleet

In this model, vehicles are not required to return to their facility of origin after delivering to the last customer on a route. Therefore, the line-haul distance to reach the centroid of the municipality of destination will be considered only once per vehicle trip. The previous is motivated by the fact that the whole delivery service is being outsourced to carrier companies; hence, vehicles do not need to return to their point of origin.

Moreover, a vehicle can serve customers only from a single municipality in each trip to avoid paying an extra fee. Hence, the cost of a vehicle trip depends only on the vehicle type, the facility of origin, and the municipality of destination.

Furthermore, cargo-types (i.e., 'bulk' and 'palletized') cannot be mixed within a vehicle trip. Besides, not all vehicles can carry all cargo-types (e.g., C1, C2, C3, and C4 trucks can take only 'bulk' loads).

In addition, there is no explicit limit to the number of customers that can be visited during each vehicle trip. Last but not least, vehicles can be flexibly allocated to any facility on different weekdays.

5.2 Single-step Formulation

5.2.1 Network Definition

Let $\mathcal{I}$ be a set of customers. Each customer i is located in a specific municipality, m, from a set of municipalities, $\mathcal{M}$, and demanding only one cargo type, q, from a set of cargo types, $\mathcal{Q}$. Further, let $\mathcal{F}$ be a set of facilities from which to serve the previously mentioned customers with a set of vehicles $\mathcal{K}$, each belonging to a vehicle type v, from a set of vehicle types $\mathcal{V}$. The model must then allocate each customer i to a single facility f and to a single delivery schedule s from the set $\mathcal{S}_i$ of feasible delivery schedules for customer i. The assigned delivery schedule s is associated with a set of weekdays, $\mathcal{J}_s$, in which customer i must be visited. On every visit day $j \in \mathcal{J}_s$, customer i must be served from facility f with a specific vehicle k belonging to the set $\mathcal{K}_i \cap \mathcal{K}_f \cap \mathcal{K}_q$ of vehicles that are able to serve customer i, operate from facility f and accept cargo-type q. In this vehicle trip, vehicle k could also visit other customers in the set $\mathcal{I}_{mq}$ of customers that share the same municipality m and require the same cargo-type q as customer i. Table 5.1 summarizes the previously described notation as well as other relevant sets and subsets for this particular model.

Table 5.1: General sets and subsets

$\mathcal{I}$	set of customers to be served; $i \in \mathcal{I}$
$\mathcal{M}$	set of municipalities where customers are located; $m \in \mathcal{M}$
$\mathcal{Q}$	set of cargo-types being delivered; $q \in \mathcal{Q}$
$\mathcal{I}_{mq}$	subset of customers located within municipality m to be served cargo-type q products; $\mathcal{I}_{mq} \subseteq \mathcal{I}$
$\mathcal{F}$	set of delivery facilities (includes DCs and satellite facilities); $f \in \mathcal{F}$
$\mathcal{V}$	set of vehicle types; $v \in \mathcal{V}$
$\mathcal{K}$	set of vehicles; $k \in \mathcal{K}$
$\mathcal{K}_v$	subset of vehicles belonging to a particular type v; $\mathcal{K}_v \subseteq \mathcal{K}$
$\mathcal{K}_f$	subset of vehicles allowed to operate in facility f; $\mathcal{K}_f \subseteq \mathcal{K}$
$\mathcal{K}_i$	subset of vehicles allowed to serve customer i; $\mathcal{K}_i \subseteq \mathcal{K}$
$\mathcal{K}_q$	subset of vehicles that can carry cargo-type q; $\mathcal{K}_q \subseteq \mathcal{K}$
$\mathcal{Q}_k$	subset of cargo-types that can be carried by vehicle k ; $\mathcal{Q}_k \subseteq \mathcal{Q}$
$\mathcal{J}$	set of weekdays; $j \in \mathcal{J}$
$\mathcal{S}$	set of delivery schedules; $s \in \mathcal{S}$
$\mathcal{J}_s$	subset of weekdays included in delivery schedule s; $\mathcal{J}_s \subseteq \mathcal{J}$
$\mathcal{S}_i$	subset of all feasible delivery schedules allowed for customer i; $\mathcal{S}_i \subseteq \mathcal{S}$
$\mathcal{P}$	set of normalized days (indexed weekdays for a delivery schedule s); $p \in \mathcal{P} \equiv \{1, 2, ..., \lvert\mathcal{J}_s\rvert\}$

5.2.2 Decision Variables

The following decisions are addressed within the model we formulate in Section 5.2.3:

- Determine the best delivery schedule for each customer: given a weekly visit frequency, define on which day(s) $\mathcal{J}_s$ of the week each customer i should be served given the allowed delivery schedules $\mathcal{S}_i$.
- Determine the best facility allocation for each customer: from which facility f each customer i should be served.
- Define the vehicle fleet mix: number of vehicles $|\mathcal{K}_v|$ required per vehicle type v to serve the current demand on every day j of the week.

Table 5.2 summarizes the decision variables included in the model.

Table 5.2: Decision variables

X_{ifksj}	binary variable that represents whether customer i is served from facility f with vehicle k on weekday j allowed by delivery schedule s that is feasible for that customer
$\tilde{X}_{ifksj}$	continuous variable that represents the demand proportion of customer i that is served from facility f with vehicle k on weekday j allowed by delivery schedule s that is feasible for that customer
Y_{isf}	binary variable that represents whether customer i is served according to a feasible delivery schedule s from facility f
Z_{fkjmq}	binary variable that represents if vehicle k is routed from facility f to municipality m on weekday j taking cargo-type q

5.2.3 Mathematical Formulation

Before we cover the optimization model, both the customer clustering approach as well as the implicit route approximation strategy will be described in detail. Moreover, a summary of the mathematical notation used in the program formulation is presented. After this, the MILP formulation in itself appears. Finally, a thorough description of each model component is provided.

Customer Clustering

It is well-known that routing problems and their variations are of central interest for both practitioners and scholars due to their practical relevance as well as the associated difficulty in solving them to optimality. They are considered NP-hard, so that the task of finding the best set of vehicle tours by solving an exact MILP model is computationally prohibitively expensive for real-world applications. As a result, different types of heuristic solution methodologies are usually applied (Dondo and Cerdá 2007, Crainic et al. 2010). For instance, the CLUST heuristic has been developed on the assumption that if the customers are distributed in clusters, a solution procedure which locates and routes accordingly should be more efficient (Srivastava 1993).

Bowerman et al. (1994), based on the taxonomic classification of Bodin and Golden (1981), divided the heuristic approaches to the VRP into five different groups:

1. cluster-first/route-second,
2. route-first/cluster-second,
3. savings/insertion,
4. improvement/exchange, and
5. simpler mathematical programming representations by relaxing some constraints.

From the two clustering procedures, the cluster-first/route-second yields more effective results (Dondo and Cerdá 2007). This approach includes an initial preprocessing phase in which a heuristic-based clustering algorithm is applied to group the customers into a small number of clusters (Crainic et al. 2010, 2011b). In this way, the general model can be written in terms of clusters rather than customers to generate a reduced problem formulation. After this phase, the routing problem is locally solved by considering only customers within each particular cluster (i.e., no extra-cluster trips are considered). Other applications of the clustering approach have also been proposed by Ambrosino et al. (2009), and Salhi and Sari (1997).

In our particular case study application, it is natural for the model to leverage the existing municipality-based payment scheme to geographically cluster customers based on their municipalities. Despite yielding a negligible cost increase, this approach demands considerably lower runtimes. The benefits of this strategy are then twofold.

First, it simplifies the mathematical formulation required to calculate the delivery cost consistently reducing runtime. These costs are now constant coefficients for the model's objective function which are determined by the facility of origin, the vehicle type, and the municipality of destination. Otherwise, if customers from different municipalities are included in the same trip, these costs would have been dependent on the subset of municipalities associated with those customers included in each vehicle trip. This, consequently increases the formulation's complexity.

Second, it yields geographically constrained delivery routes. Since municipalities are well-defined territories, the optimal routes to serve a subset of customers located within them are fairly compact. As stated by Mourgaya and Vanderbeck (2006), in tactical planning, the objective of regionalizing routes reflects a desire to specialize them to restricted geographical areas well-known by the truck drivers. Figure 5.1 shows an example, for our case study, of the customer geographical distribution per municipality. Each dot represents a customer and the different colors are associated with different municipalities. For a clearer distinction, the municipalities' political boundaries are provided in black.

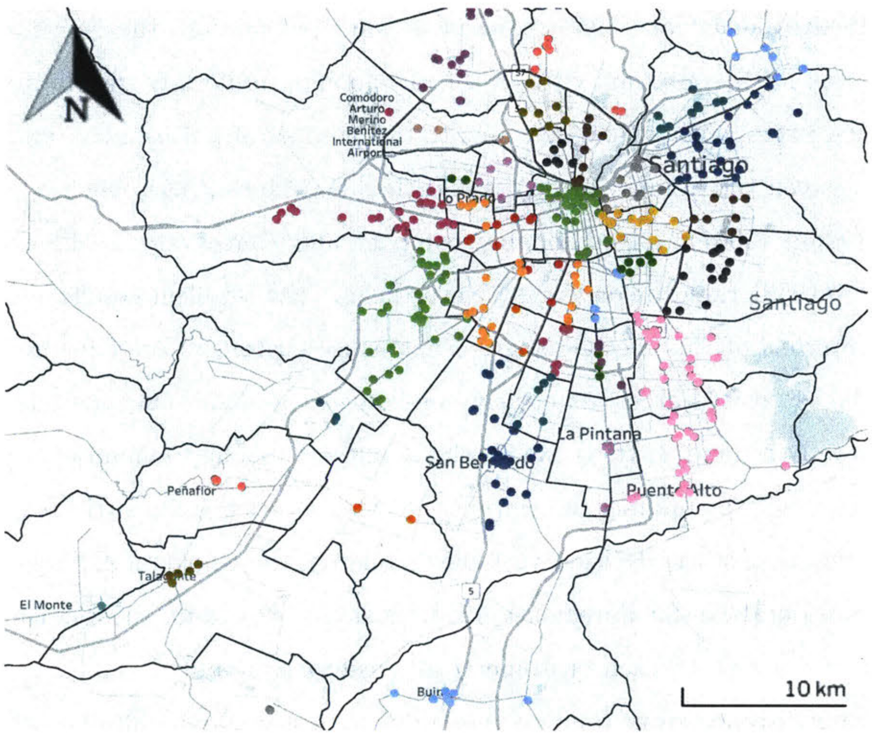


Figure 5.1: Geographical location of customers clustered by municipalities

Implicit Route Approximation

In many logistics problems, it is necessary to estimate the distance that a fleet of vehicles travel to meet a set of customer demands. Traveled distance is not only an important element of carriers' variable costs but it is also a key input in tactical and strategic models that solve, for instance, network design and fleet sizing problems (Figliozzi 2008). As proposed by Figliozzi (2009), routing approximations are intended for strategic and tactical planning analyses of such problems, in which the number and location of customers as well as their demands vary daily and are, a priori, unknown. Following the previous rationale, a wide set of heuristics as well as closed form solutions for Route Length Estimations (RLEs) have been proposed in the literature.

If we consider routing heuristics, they do perform well for routing-focused problems. However, in this particular application, in which we are making scheduling decisions for a large set of customers constrained by vehicle fleet availability, those heuristics render incapable of handling that level of complexity (in terms of decision types and problem size). Therefore, we decided to replace the explicit vehicle routing step by a route distance approximation embedded within the problem formulation.

Following the previous rationale, the first RLE of a shortest tour through a set of points was established by Beardwood et al. (1959). Based on that publication, the length of the Traveling Salesman Problem (TSP) tour that visits a set of n points, uniformly distributed in an area A, asymptotically converges, with probability one, to the product of a constant κ and the square root of the product between the number of points and the area (i.e.: $\kappa\sqrt{nA}$), when $n \to \infty$. According to Larson and Odoni (2007), for reasonably compact and convex areas, the limit provided by Beardwood et al. (1959) rapidly converges. Nevertheless, our problem conflicts with some of the underlying assumptions of this methodology. First, since customers are clustered based on their municipalities, the associated areas are clearly non-convex. Second, being the number of stops per vehicle trip (generally around 3) fairly small, the quality of this approximation tends to decrease. Last and foremost, since the number of customers to visit on each vehicle trip is unknown ex-ante, n will then be a function of the decision variables rendering the whole model non-linear.

Taking into consideration the aforementioned limitations, we opt for the following RLE strategy. In this approach, we assign to each customer i an easily precomputed value r_{imq} that represents the average driving distance from (and to) customer i to (and from) all the other customers in municipality m that demand the same cargo-type q. This value is a proxy for the average additional driving distance incurred within a delivery tour if customer i is visited. Therefore, a trip's total driving distance can be approximated as the sum of the line-haul distance and the delivery tour distance being:

- Line-haul distance: driving distance from the facility of origin f to the centroid of the municipality of destination m (r_{fm}).
- Delivery tour distance: approximate driving distance to serve all customers in a particular delivery tour once the vehicle is positioned in the municipality of destination. This is the sum of the average distances r_{imq} to reach each customer i that is served within the delivery tour.

Figure 5.2 presents a schematic version of the described RLE strategy. According to this strategy, a truck departs from a facility (e.g., DC), drives the line-haul to the centroid of the destination municipality and starts its delivery tour visiting the assigned customers. Once the delivery tour is completed, the truck does not return to the facility of origin.

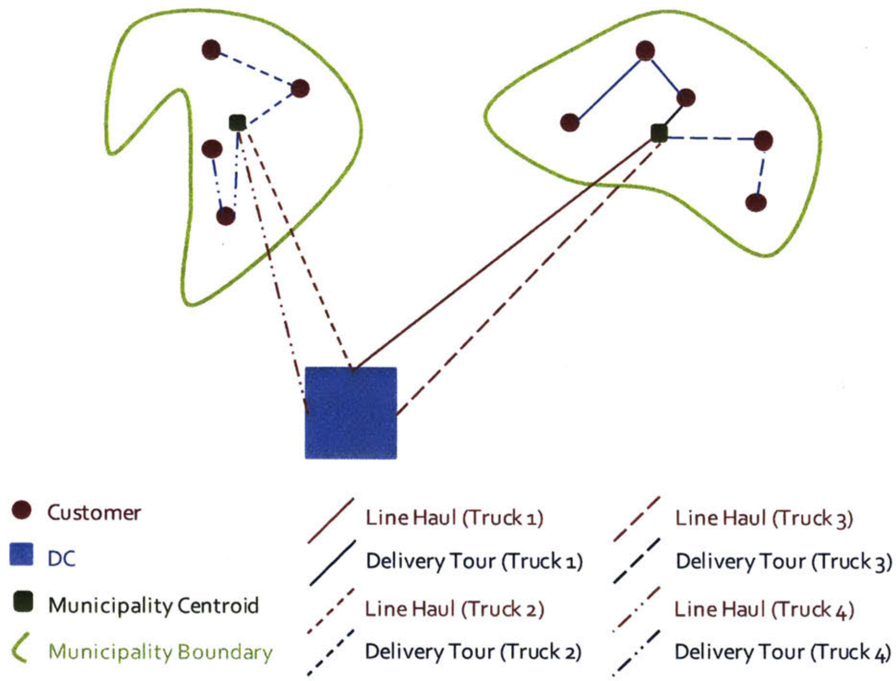


Figure 5.2: Route composition proposed for the single-echelon distribution network model

Equation (5.1) presents the calculation of r_{imq}. Let r_{ig} be the driving distance (in $[km]$) to go from customer i to customer g, both located in municipality m and requiring cargo-type q. Similarly, let r_{gi} be the driving distance (in $[km]$) to go from customer g to customer i (in particular, the distance between a customer and itself will always be zero, mathematically $r_{ii} = r_{gg} = 0$). Finally, be $|\mathcal{I}_{mq}|$ the number of customers in municipality m demanding cargo-type q; then r_{imq} can be computed ex-ante the optimization as follows

$$r_{imq} = \begin{cases} \sum_{g \in \mathcal{I}_{mq}} \frac{\frac{r_{ig}+r_{gi}}{2}}{|\mathcal{I}_{mq}|-1} & \text{if } |\mathcal{I}_{mq}| > 1 \\ 0 & \text{if } |\mathcal{I}_{mq}| = 1 \end{cases}, \qquad i \in \mathcal{I}_{mq}, \forall m, \forall q. \tag{5.1}$$

This strategy clearly penalizes remotely located customers by assigning higher r_{imq} values since their associated distances, r_{ig} and r_{gi}, will be higher. Conversely, for those customers concentrated within a limited area their associated r_{imq} values tend to be smaller. Moreover, if the customer is the only one in municipality m demanding cargo-type q the total trip distance will be approximated by the line-haul distance, r_{fm}, as $r_{imq} = 0$.

Simplicity may certainly come at the expense of accuracy. However, given that the driving time represents a small fraction of the total available time (as the number of customers being visited per trip is extremely limited) and that the model's goal is not the generation of explicit routes, the previously proposed approximation serves our purpose while enormously simplifying the model's combinatorial complexity.

MILP formulation

As metioned before, we can mathematically formulate this problem as a BPP with embedded implicit routing and customer scheduling and allocation. Tables 5.3–5.7 summarize the notation used for the model components and parameters.

Table 5.3: Time, distance, and speed parameters

T_{max}	allowed service time per day [h]
T_i^w	average fixed waiting time at customer i [h]
T_{fkq}^l	average fixed loading time for vehicle k taking cargo-type q at facility f [h]
t_{fkq}^l	average variable loading time for vehicle k taking cargo-type q at facility f [h/m^3]
t_{kq}^u	average variable unloading time for vehicle k taking cargo-type q [h/m^3]
V_k^d	average inter-stop (delivery) speed for vehicle k [km/h]
V_k^{lh}	average line-haul speed for vehicle k [km/h]
r_{fm}	driving distance from facility f to the centroid of municipality m [km]
r_{imq}	average delivery driving distance required to serve customer i, receiving cargo-type q within municipality m, in a delivery tour [km]

Table 5.4: Capacity parameters

ξ_f	daily delivery capacity available for facility f [m^3]
η_{kq}	delivery capacity for vehicle k carrying cargo-type q [m^3]
η_i^{max}	maximum delivery capacity of those vehicles allowed to visit customer i [m^3]
a_v	number of available vehicles of type v [vehicles]

Table 5.5: Cost parameters

c_{fkm}	cost of moving a vehicle k from facility f to municipality m [\$/vehicle trip]
c_f	variable transfer cost to move products to facility f [\$/m^3]

Table 5.6: Assignment function and general parameter

$P(s,j)=p$	function that maps a weekday j to a normalized day p based on a delivery schedule s (e.g., $P(\{M,W,F\},W)=2$ or $P(\{T,F\},T)=1$)
μ	minimum fraction of a customer order to be delivered per visit (i.e., $0<\mu\leq 1$)

Table 5.7: Customer parameters

d_{ip}	demand volume (in [m^3]) to be delivered to customer i in normalized day p, where p is determined as a function of the delivery schedule s and the specific weekday j in that delivery schedule, i.e.: $p = P(s, j)$
q_i	cargo-type associated to customer i

Here, the objective is to minimize the total transportation cost presented in Equation (5.2). That cost is composed of the last-mile-delivery cost, in the first term, and the associated transfer cost, in the second term. Regarding the last-mile-delivery cost, the parameter c_{fkm} denotes the flat delivery fare accrued when a vehicle k is routed from facility f to deliver to a subset of customers in municipality m. The transfer cost is assumed to be proportional to the total volume delivered from each facility, as detailed in Section 5.1. The parameter c_f then represents the associated cost per volume that is incurred when transferring products from other facilities to make them available at facility f. That specific value can be estimated based on historic data. The model can, therefore, be mathematically expressed as:

$$\min \left\{ \sum_{f \in \mathcal{F}} \sum_{k \in \mathcal{K}_f} \sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}} \sum_{q \in \mathcal{Q}_k} c_{fkm} Z_{fkjmq} + \sum_{i \in \mathcal{I}} \sum_{f \in \mathcal{F}} \sum_{k \in \mathcal{K}_i \cap \mathcal{K}_f} \sum_{s \in \mathcal{S}_i} \sum_{j \in \mathcal{J}_s} c_f d_{ip} \tilde{X}_{ifksj} \right\} \tag{5.2}$$

subject to

$$\sum_{s \in \mathcal{S}_i} \sum_{f \in \mathcal{F}} Y_{isf} = 1, \qquad \forall i, \tag{5.3}$$

$$\sum_{k \in \mathcal{K}_i \cap \mathcal{K}_f} X_{ifksj} = \left\lceil \frac{d_{ip}}{\eta_i^{max}} \right\rceil Y_{isf}, \qquad \forall i, s \in \mathcal{S}_i, j \in \mathcal{J}_s, \forall f, \tag{5.4}$$

$$\sum_{k \in \mathcal{K}_i \cap \mathcal{K}_f} \tilde{X}_{ifksj} = Y_{isf}, \qquad \forall i, s \in \mathcal{S}_i, j \in \mathcal{J}_s, \forall f, \tag{5.5}$$

$$\sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}_i \cap \mathcal{K}_f} \sum_{s \in \mathcal{S}_i | j \in \mathcal{J}_s} d_{ip} \tilde{X}_{ifksj} \leq \xi_f, \qquad \forall f, \forall j, \tag{5.6}$$

$$\sum_{f \in \mathcal{F} | k \in \mathcal{K}_f} \sum_{m \in \mathcal{M}} \sum_{q \in \mathcal{Q}_k} Z_{fkjmq} \leq 1, \qquad \forall k, \forall j, \tag{5.7}$$

$$\sum_{f \in \mathcal{F}} \sum_{k \in \mathcal{K}_f \cap \mathcal{K}_v} \sum_{m \in \mathcal{M}} \sum_{q \in \mathcal{Q}_k} Z_{fkjmq} \leq a_v, \qquad \forall j, \forall v, \tag{5.8}$$

$$\sum_{i \in \mathcal{I}_{mq} | k \in \mathcal{K}_i} \sum_{s \in \mathcal{S}_i | j \in \mathcal{J}_s} d_{ip} \tilde{X}_{ifksj} \leq \eta_{kq} Z_{fkjmq}, \qquad \forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k, \tag{5.9}$$

$$\sum_{i \in \mathcal{I}_{mq} | k \in \mathcal{K}_i} \sum_{s \in \mathcal{S}_i | j \in \mathcal{J}_s} \left[\left(\frac{r_{imq}}{V_k^d} + T_i^w \right) X_{ifksj} + \left(t_{kq}^u + t_{fkq}^l \right) d_{ip} \tilde{X}_{ifksj} \right] \leq \left(T_{max} - \frac{r_{fm}}{V_k^{lh}} - T_{fkq}^l \right) Z_{fkjmq}, \quad \forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k, \tag{5.10}$$

$$X_{ifksj} \geq \tilde{X}_{ifksj}, \quad \forall i, \forall f, k \in \mathcal{K}_i \cap \mathcal{K}_f, s \in \mathcal{S}_i, j \in \mathcal{J}_s, \tag{5.11}$$

$$\mu X_{ifksj} \leq \tilde{X}_{ifksj}, \quad \forall i, \forall f, k \in \mathcal{K}_i \cap \mathcal{K}_f, s \in \mathcal{S}_i, j \in \mathcal{J}_s, \tag{5.12}$$

$$X_{ifksj} \in \{0, 1\}, \quad \forall i, \forall f, k \in \mathcal{K}_i \cap \mathcal{K}_f, s \in \mathcal{S}_i, j \in \mathcal{J}_s, \tag{5.13}$$

$$\tilde{X}_{ifksj} \in [0, 1], \quad \forall i, \forall f, k \in \mathcal{K}_i \cap \mathcal{K}_f, s \in \mathcal{S}_i, j \in \mathcal{J}_s, \tag{5.14}$$

$$Y_{isf} \in \{0, 1\}, \quad \forall i, s \in \mathcal{S}_i, \forall f, \tag{5.15}$$

$$Z_{fkjmq} \in \{0, 1\}, \quad \forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k. \tag{5.16}$$

Constraints (5.3) guarantee that each customer can only be served according to one feasible delivery schedule s and from only one facility f. Equations (5.4) set the required number of daily visits to each customer avoiding unnecessary cargo splits. This is essentially a function of the customer's demand d_{ip} and the capacity of the largest truck that can be used to deliver it η_i^{max}. Constraints (5.5) force the model to deliver the whole volume demanded by the customer on each scheduled day. Equations (5.6) prevent the model from overloading available facilities given their maximum daily capacity ξ_f measured in [m^3]. Constraints (5.7) avoid multiple uses of the same vehicle k on a particular day j. Equations (5.8) guarantee that the model will only assign, on each weekday j, vehicles of category v based on the associated available fleet a_v. Constraints (5.9) enforce that, given a specific vehicle k and a cargo type q, the vehicle's maximum payload η_{kq} cannot be exceeded. Equations (5.10) limit the total time that a vehicle k can operate per day. There, T^{max} denotes the maximum allowed service time per day. Several time components are also considered in these constraints including: a fixed waiting time at a facility, T_{fkq}^l (which varies per combination of vehicle, facility and cargo-type); a volume dependent loading time, t_{fkq}^l (also varying per combination of vehicle, facility and cargo-type); a line-haul driving time that is calculated as the division of the line-haul distance, r_{fm} (depending on the facility of origin and the municipality of destination), and the line-haul driving speed, V_k^{lh} (determined by the vehicle itself); a delivery tour driving time, computed as the division between the average distance to reach the customer, r_{imq}, and the delivery driving speed, V_k^d (determined by the vehicle itself); a fixed waiting time at each customer, T_i^w; and a volume dependent unloading time, t_{kq}^u (defined by the combination of vehicle and cargo-type). Constraints (5.11) and (5.12) link the values that the customer demand allocation binary variables and the associated demand partition variables can take. They guarantees that, for any positive value that $\tilde{X}_{ifksj}$ takes, X_{ifksj} will be forced to one. In Equations (5.12), μ represents the minimum fraction of an order that can be delivered per customer visit. For instance, $\mu = 1$ will prevent order partitioning what renders the whole model infeasible if only one customer demands a volume that exceeds the largest vehicle payload η_i^{max} allowed to serve that client. $\mu = 0.25$ implies that at least one quarter of the ordered volume should be delivered on each visit. Similarly, $\mu = 0.01$ forces the model to deliver at least 1% of the demanded volume on each visit. Therefore, the lower the value of μ, the higher the model flexibility for reaching the optimal partition and allocation of large orders. Finally, Equations (5.13) through (5.16) define the allowable domains of all decision variables.

5.3 Two-step Hierarchical Formulation

As mentioned in Section 1.3, the model developed in Section 5.2.3 is intrinsically NP-hard. This makes even simple problem instances difficult to solve due to its hard combinatorial features. According to Koch et al. (2012), there are several reasons why MILPs are hard to solve using a sequential algorithm. The most related ones to this work are:

- The presence of symmetry in the solution space, i.e., many equivalent solutions of similar cost requiring substantial amounts of enumeration.
- Large enumeration trees that simply take too long to sequentially explore them. In such cases, the ability to evaluate more branch and bound nodes clearly helps, as long as the tree is balanced enough to effectively divide the computation.

In general, heuristics can approximately solve problems of larger sizes in reduced computational time. Nevertheless, they usually lack robustness and their performance is problem dependent. Instead, optimization algorithms offer the best promise in terms of robustness (Fisher 1995). However, given the enormous complexity of our model, it does not seem realistic to apply a pure optimization method in a single-step approach. Instead, we can focus on hybrid solution algorithms that can be as robust as optimization methods. These hybrid methods are capable of providing reasonably good solutions for large problem instances within acceptable computational times.

In this work, we develop a hierarchical solution approach that partitions the original model into two optimization subproblems which are sequentially solved. Many methods decompose the original problem into subproblems motivated by overlapping decision types (i.e., strategic, tactical, or operational) that must be simultaneously made (Ambrosino et al. 2009, Crainic et al. 2010). In our case, we model a combination of tactical and operational decisions. Therefore, the ultimate goal is to construct a model that can be easily understood and that provides high quality solutions in a reasonable amount of computing time.

Hence, we proceed in two steps by applying a separation strategy based on the decision types to be made (i.e., tactical and operational). This splits the initial problem into two subproblems. In the first step, the associated model defines and fixes the scheduling and facility allocation for each customer while approximating the required fleet size and mix to serve them. The individual vehicles are replaced by their associated vehicle types what greatly reduces the model's complexity. This stage only provides an approximation of the transportation cost. In the second step, and taking as a given the scheduling and allocation decisions of the previous step, the new submodel implicitly routes customers (considering individual vehicles), consequently, refining the fleet size and mix calculation. This step then provides a better estimation for the final transportation cost incurred by the operation.

In summary, the first step focuses on tactical decisions (i.e., customer scheduling and allocation, and approximation of fleet size and mix). The second step takes the operational decisions (i.e., implicit vehicle routing) and refines some of the tactical ones (i.e., definition of the fleet size and mix) while yielding a better approximation for the final transportation cost. Figure 5.3 depicts the previously described two-step hierarchical strategy.

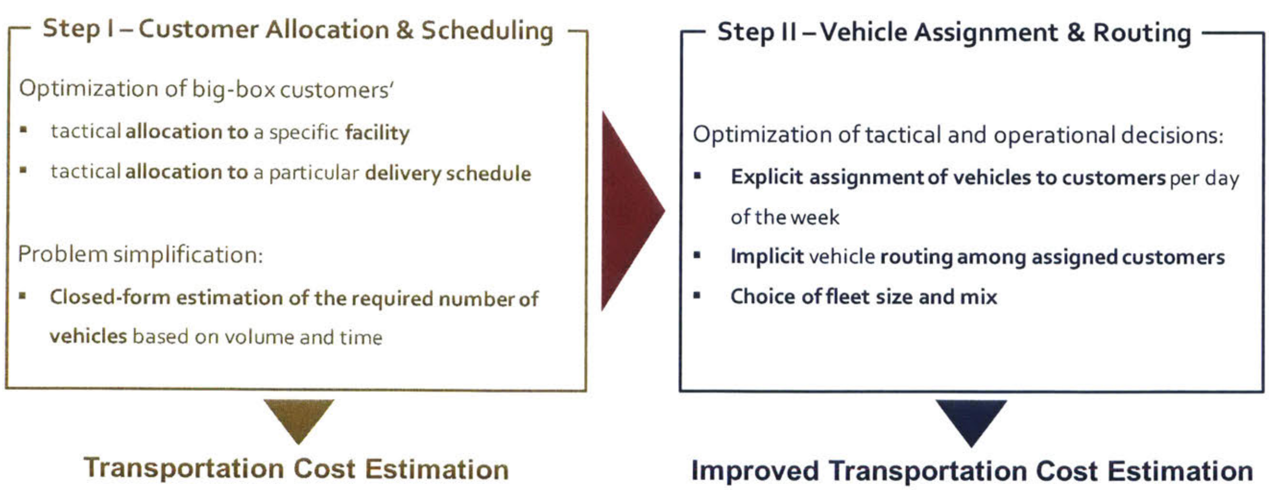


Figure 5.3: Two-step hierarchical optimization strategy

In the following, both steps are properly formalized and mathematically modeled.

5.3.1 Step I – Customer Allocation and Scheduling

Before introducing the submodel's mathematical formulation, Tables 5.8 to 5.10 present the new variables as well as additional sets and parameters that are relevant for this model.

Table 5.8: Step I – notation of submodel decision variables

X_{ifvsj}	binary variable that represents whether customer i is served from facility f with vehicle type v on weekday j allowed by delivery schedule s feasible for that customer
Y_{isf}	binary variable that represents whether customer i is served according to delivery schedule s from facility f
L_{fvjmq}	continuous variable that represents how many vehicles of type v are used from facility f on weekday j to municipality m taking cargo-type q

Table 5.9: Step I – notation of additional sets

$\mathcal{V}_f$	subset of vehicle types allowed to operate in facility f; $\mathcal{V}_f \subseteq \mathcal{V}$
$\mathcal{V}_i$	subset of vehicle types allowed to delivery to customer i; $\mathcal{V}_i \subseteq \mathcal{V}$
$\mathcal{Q}_v$	subset of cargo types that can be carried by vehicle type v; $\mathcal{Q}_v \subseteq \mathcal{Q}$

Table 5.10: Step I – notation of parameters

T^l_{fvq}	average fixed loading time for vehicle type v taking cargo-type q at facility f [h]
t^l_{fvq}	average variable loading time for vehicle type v taking cargo-type q at facility f [h/m^3]
t^u_{vq}	average variable unloading time for vehicle type v taking cargo-type q [h/m^3]
η_{vq}	delivery capacity for vehicle type v carrying cargo-type q [m^3]
V^d_v	average inter-stop (delivery) speed for vehicle type v [km/h]
V^{lh}_v	average line-haul speed for vehicle type v [km/h]
c_{fvm}	cost of moving a vehicle of type v to municipality m from facility f [\$/vehicle trip]

The mathematical formulation follows.

$$\min\left\{\sum_{f\in\mathcal{F}}\sum_{v\in\mathcal{V}_f}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}}\sum_{q\in\mathcal{Q}_v} c_{fvm}L_{fvjmq} + \sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}}\sum_{v\in\mathcal{V}_i\cap\mathcal{V}_f}\sum_{s\in\mathcal{S}_i}\sum_{j\in\mathcal{J}_s} c_f d_{ip} X_{ifvsj}\right\} \tag{5.17}$$

subject to

$$\sum_{s\in\mathcal{S}_i}\sum_{f\in\mathcal{F}} Y_{isf} = 1, \qquad \forall i, \tag{5.18}$$

$$\sum_{v\in\mathcal{V}_i\cap\mathcal{V}_f} X_{ifvsj} = Y_{isf}, \qquad \forall i, s\in\mathcal{S}_i, j\in\mathcal{J}_s, \forall f, \tag{5.19}$$

$$\sum_{i\in\mathcal{I}}\sum_{v\in\mathcal{V}_i\cap\mathcal{V}_f}\sum_{s\in\mathcal{S}_i|j\in\mathcal{J}_s} d_{ip}X_{ifvsj} \leq \xi_f, \qquad \forall f,\forall j, \tag{5.20}$$

$$\sum_{f\in\mathcal{F}}\sum_{m\in\mathcal{M}}\sum_{q\in\mathcal{Q}_v} L_{fvjmq} \leq a_v, \qquad \forall j,\forall v, \tag{5.21}$$

$$\sum_{i\in\mathcal{I}_{mq}|v\in\mathcal{V}_i}\sum_{s\in\mathcal{S}_i|j\in\mathcal{J}_s} d_{ip}X_{ifvsj} \leq \eta_{vq}L_{fvjmq}, \qquad \forall f, v\in\mathcal{V}_f,\forall j,\forall m, q\in\mathcal{Q}_v, \tag{5.22}$$

$$\sum_{i\in\mathcal{I}_{mq}|v\in\mathcal{V}_i}\sum_{s\in\mathcal{S}_i|j\in\mathcal{J}_s} \left[\frac{r_{im}}{V_v^d} + T_i^w + \left(t_{vq}^u + t_{fvq}^l\right)d_{ip}\right]X_{ifvsj} \leq \left(T_{max} - \frac{r_{fm}}{V_v^{lh}} - T_{fvq}^l\right)L_{fvjmq}, \qquad \forall f, v\in\mathcal{V}_f,\forall j,\forall m, q\in\mathcal{Q}_v, \tag{5.23}$$

$$X_{ifvsj} \in \{0,1\}, \qquad \forall i,\forall f, v\in\mathcal{V}_i\cap\mathcal{V}_f, s\in\mathcal{S}_i, j\in\mathcal{J}_s, \tag{5.24}$$

$$Y_{isf} \in \{0,1\}, \qquad \forall i, s\in\mathcal{S}_i,\forall f, \tag{5.25}$$

$$L_{fvjmq} \in \mathbb{R}_0^+, \qquad \forall f, v\in\mathcal{V}_f,\forall j,\forall m, q\in\mathcal{Q}_v. \tag{5.26}$$

The objective function (5.17) minimizes an approximation of the total transportation cost since the final number of vehicles and their destination is not determined in this step. Similar to Equation (5.2), the total transportation cost aggregates both the delivery and the transfer cost components.

Constraints (5.18) guarantee that each customer can only be served according to one feasible delivery schedule s and from only one facility f. Equations (5.19) force customers to be delivered according to the delivery schedules defined in Restrictions (5.18). Constraints (5.20) prevent the model from overloading available facilities given their maximum daily capacity ξ_f. Equations (5.21) guarantee that the model will only assign, on each weekday j, vehicles of category v based on the associated available fleet a_v. Constraints (5.22) enforce that, given a vehicle type v and a cargo-type q, the aggregated payload that is allocated is not exceeded. Equations (5.23) work in a similar way as Constraints (5.10). In this case, the aggregated available time of a specific vehicle type should not be exceeded. Finally, Equations (5.24) through (5.26) define the allowable domains of all decision variables.

After solving this model, the allocation and scheduling results for each customer are fed into Step II.

5.3.2 Step II – Vehicle Assignment and Implicit Routing

Once again, before presenting the mathematical formulation, Tables 5.11 and 5.12 introduce the relevant subsets and parameters as well as the new decision variables for Step II.

Table 5.11: Step II – notation of relevant subsets and indexes

$\mathcal{J}_i$	set of optimal weekdays in which customer i is delivered; $j \in \mathcal{J}_i$. This results from the scheduling decision of Step I.
f_i	facility from which customer i is delivered based on the allocation decision of Step I; $f_i \in \mathcal{F}$

Table 5.12: Step II – notation of decision variables

X_{if_ikj}	binary variable that represents whether customer i, served from facility f_i on weekday j (determined on Step I), is delivered with vehicle k
$\tilde{X}_{if_ikj}$	continuous variable that represents the demand proportion of customer i, served from facility f_i on weekday j (determined on Step I), that is delivered with vehicle k
Z_{fkjmq}	binary variable that represents whether vehicle k is routed from facility f to municipality m on weekday j taking cargo-type q

Considering the previously defined elements, this subproblem can be formulated as

$$\min\Big\{\sum_{f\in\mathcal{F}}\sum_{k\in\mathcal{K}_f}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}}\sum_{q\in\mathcal{Q}_k} c_{fkm}Z_{fkjmq} + \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}_i\cap\mathcal{K}_{f_i}}\sum_{j\in\mathcal{J}_i} c_f d_{ip}\tilde{X}_{if_ikj}\Big\} \tag{5.27}$$

subject to

$$\sum_{k\in\mathcal{K}_i\cap\mathcal{K}_{f_i}} X_{if_ikj} = \left\lceil \frac{d_{ip}}{\eta_i^{max}} \right\rceil, \qquad \forall i, j \in \mathcal{J}_i, \tag{5.28}$$

$$\sum_{k\in\mathcal{K}_i\cap\mathcal{K}_{f_i}} \tilde{X}_{if_ikj} = 1, \qquad \forall i, j \in \mathcal{J}_i, \tag{5.29}$$

$$\sum_{f\in\mathcal{F}|k\in\mathcal{K}_f}\sum_{m\in\mathcal{M}}\sum_{q\in\mathcal{Q}_k} Z_{fkjmq} \leq 1, \qquad \forall k, \forall j, \tag{5.30}$$

$$\sum_{f\in\mathcal{F}}\sum_{k\in\mathcal{K}_f\cap\mathcal{K}_v}\sum_{m\in\mathcal{M}}\sum_{q\in\mathcal{Q}_k} Z_{fkjmq} \leq a_v, \qquad \forall j, \forall v, \tag{5.31}$$

$$\sum_{i\in\mathcal{I}_{mq}|f=f_i,k\in\mathcal{K}_i,j\in\mathcal{J}_i} d_{ip}\tilde{X}_{if_ikj} \leq \eta_{kq}Z_{fkjmq}, \qquad \forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k, \tag{5.32}$$

$$\sum_{i \in \mathcal{I}_{mq} | k \in \mathcal{K}_i \& j \in \mathcal{J}_i} \left[X_{if_ikj} \left(\frac{r_{im}}{V_k^d} + T_i^w \right) + \tilde{X}_{if_ikj} \left(t_{kq}^u + t_{fkq}^l \right) d_{ip} \right] \leq$$
$$\left(T_{max} - \frac{r_{fm}}{V_k^{lh}} - T_{fkq}^l \right) Z_{fkjmq},$$
$$\forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k, \tag{5.33}$$

$$X_{if_ikj} \geq \tilde{X}_{if_ikj}, \quad \forall i, k \in \mathcal{K}_i \cap \mathcal{K}_{f_i}, j \in \mathcal{J}_i, \tag{5.34}$$
$$\mu X_{if_ikj} \leq \tilde{X}_{if_ikj}, \quad \forall i, k \in \mathcal{K}_i \cap \mathcal{K}_{f_i}, j \in \mathcal{J}_i, \tag{5.35}$$
$$X_{if_ikj} \in \{0, 1\}, \quad \forall i, k \in \mathcal{K}_i \cap \mathcal{K}_{f_i}, j \in \mathcal{J}_i, \tag{5.36}$$
$$\tilde{X}_{if_ikj} \in [0, 1], \quad \forall i, k \in \mathcal{K}_i \cap \mathcal{K}_{f_i}, j \in \mathcal{J}_i, \tag{5.37}$$
$$Z_{fkjmq} \in \{0, 1\}, \quad \forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k. \tag{5.38}$$

In this case, the objective function (5.27) minimizes the total transportation cost based on the optimized vehicle fleet. Moreover, since the customer allocation to facilities is not modified in Step II, the second term of Equation (5.27) remains constant when compared with Step I.

Equations (5.28) minimizes the number of visits per day to each customer avoiding unnecessary cargo splits. Moreover, Constraints (5.29) force the model to deliver the whole volume demanded by each customer on each scheduled day. Restrictions (5.30) avoid multiple uses of the same vehicle k on a particular day j. Equations (5.31) guarantee that the model will only assign, on each weekday j, vehicles of category v based on the associated available fleet a_v. Constraints (5.32) enforce that, given a specific vehicle k and a cargo type q, the vehicle's maximum payload η_{kq} cannot be exceeded. Similar to Equations (5.10) (i.e., using the same coefficients and formulation), Restrictions (5.33) limit the total time that a vehicle k can operate per day. Constraints (5.34) and (5.35) have the same interpretation of Equations (5.11) and (5.12). The coefficient μ has already been explained in Section 5.2.3. Finally, Equations (5.36) through (5.38) define the allowable domains of all decision variables.

In this submodel, there is no need for a set of constraints that limits each facility's daily capacity. This has already been considered in Step I where we defined the customers allocations.

5.4 Benders Decomposition

MILPs are usually difficult to solve, even though state-of-the-art mixed integer programming solvers are, in many cases, remarkably effective and have radically improved in the past years. As stated in Vanderbeck and Wolsey (2009), these solvers typically use branch-and-cut methods to obtain improved linear programming bounds and branching to carry out an implicit enumeration of possible solutions. However, these systems essentially ignore the inherent problem structure. Fortunately, there are numerous ways in which, given an initial MILP formulation, its underlying structure can be leveraged to obtain improved problem formulations as well as more effective solution approaches. Klotz et al. (2013) provide and extensive portfolio of guidelines to carefully (re)formulate MILPs. This can vastly improve the performance of solution algorithms.

One common way to obtain reformulations is by adding valid inequalities (also known as cutting planes) in the original variables. A valid inequality for a MILP is any constraint that does not eliminate any feasible integer solutions (i.e., a cut that eliminates part of the feasible region which does not contain any integer solution in the Linear Program (LP) relaxation's). Hence, the general motivation is to obtain a reformulation for which the optimal value of the LP relaxation is closer to the optimal value of the reformulated MILP. In other words, we try to approximate the boundaries in the MILP's solution space to the problem's convex hull (i.e., the smallest feasible region that contains all of the integer solutions). However, the convex hull for the BPP, in particular, for our variation (i.e., with embedded customer scheduling and allocation, and implicit routing), is impossible to find (Bertsimas and Tsitsiklis 1997).

Other alternatives rely on a decomposition of the problem. BD, introduced by Benders (1962), is an example of this strategy that is used to solve certain large-scale optimization problems. In this case, instead of simultaneously considering all decision variables and constraints in a problem, BD partitions the original formulation into multiple smaller problems (Master Problem (MP) and Subproblem (SP)). These subproblems are either a pure Integer Program (IP) or a pure LP that could each be iteratively solved to arrive at the solution of the original problem. The main advantage of this scheme is that the two subproblems are easier to solve than the original problem and, although they might need to be executed several times to arrive at the solution, this is still likely to be quicker than trying to solve the original large problem (Murphy 2013). In other words, since computational difficulty of optimization problems significantly increases with the number of variables and constraints, iteratively solving these smaller subproblems can be more efficient than dealing with the originally large formulation (Taşkin 2011).

The BD algorithm has been successfully applied to a wide range of complex optimization problems in the context of production planning (see, e.g., Behnamian (2014), Adulyasak et al. (2015)), scheduling (see, e.g., Canto (2008), Cordeau et al. (2001)), transportation and network design (see, e.g., Boland et al. (2016), Cordeau et al. (2006), Corréa et al. (2007), Fortz and Poss (2009), Gelareh et al. (2015), Jiang et al. (2009), Pishvaee et al. (2014)), and other research areas. Appendix A presents an extensive mathematical explanation of this methodology together with an iterative relaxation strategy which further reduces computation time.

In this work, BD has been applied to the single-step problem formulation presented in Section 5.2. The main objective is arriving at near-optimal solutions in reduced computational time. In BD, the original formulation can be decomposed based on the binary (X_{ifksj}, Y_{isf}, and Z_{fkjmq}) and continuous ($\tilde{X}_{ifksj}$) decision variables, respectively. A comparison of the results obtained with BD versus the original model formulation and the two-step approach is presented in Chapter 7.

Chapter 6

Two-echelon Location Routing

This chapter develops a mathematical formulation to solve the two-echelon multi-depot capacitated LRP. The model includes big-box customer scheduling, a capacity-constrained mixed vehicle fleet, and integrates nanostore deliveries within the network's footprint. The volume of nanostore demand that must be satisfied is a strategic decision exogenous to our model. Hence, it is captured as an input parameter. Given the inherently high combinatorial complexity that a single formulation would yield this problem has thus been partitioned in three subproblems combined within a three-step iterative approach that improves its mathematical tractability. As mentioned in Sections 3.2 and 3.3, this program determines the cost-optimal last-mile-delivery network configuration to serve existing big-box customers (i.e., tier I) and a subset of strategically selected nanostores (i.e., tier II) by defining

i) the optimal allocation of existing big-box customers to delivery facilities,

ii) the optimal activation of big-box customers as TPs to serve neighboring nanostores,

iii) the strategic definition of which nanostores to serve and their optimal allocation to active TPs,

iv) a delivery day schedule for each customer (i.e., big-box customers and served nanostores) that is compatible with its weekly delivery frequency, and

v) the optimal fleet composition to serve both existing big-box customers as well as the newly incorporated nanostores.

Similar to Chapter 5, we will first elaborate on the assumptions relevant to this scenario of analysis. Second, we present the associated notation for the updated network definition. Finally, we introduce a three-step iterative solution approach and we individually formulate each optimization model. The results obtained for a stylized problem instance are presented in Section 7.2.

6.1 Model Assumptions

The following assumptions are considered within the problem formulation. This set complements those assumptions in Section 5.1.

I. Nanostores

Nanostore demanded volume is assumed to be static and deterministic for a one-week time horizon. Moreover, given a specific nanostore, its dropsize remains constant on each delivery day. Furthermore, a predefined volume of the traditional channel total demand must be satisfied. This volume, provided as an input to the model, is uniformly cannibalized from those big-box customers that typically serve nanostores (i.e., distributors and wholesalers). This cannibalization is accounted by reducing the demand, d_{ip}, of those big-box customers. Besides, no sales effort cost is accrued to serve nanostores (in other words, we are ignoring the sales methodology that the company might use, e.g., presales, vansales, etc.).

In addition, the weekly delivery frequency of each nanostore is known and must be honored when that nanostore is served by our network. At this respect, nanostores can be visited either once, twice, or three times a week on different weekdays.

II. Big-box Customers

If a big-box customer is activated as a TP, it will be allowed to serve only those neighboring nanostores whose delivery frequencies are less than or equal to its own. Besides, if activated, a big-box customer can be used as a TP every day that is visited. Furthermore, since demand varies during daily operation, only one visit to a TP per delivery day can be leveraged to serve nanostores. Finally, no activation cost is considered whenever a big-box customers is chosen as a TP.

III. Nanostore Delivery Strategy

Only the average waiting time at each big-box customer can be used to serve nanostores. In other words, when products are being served to a big-box customer, all crew members are engaged in this activity. Moreover, only one crew member is allowed to serve nanostores. This person will use a fixed-capacity handcart and will walk to the allocated nanostores. Depending on the available time and the nanostore demands, this person could complete one or several tours per delivery day.

Besides, each served nanostore will always be delivered from the same TP (i.e., big-box customer). That TP is the closest —in terms of euclidean distance— active one. Finally, nanostores always demand 'bulk' cargo, hence, only 'bulk'-type big-box customers are potential TPs.

IV. Fleet

The associated delivery fares and unloading times for each vehicle type will be updated according to their new crew sizes.

At this point, it is important to clarify that each candidate TP is always associated with a specific big-box customer. However, not all big-box customers can be candidate TPs (e.g., 'palletized'-type customers fall in this category). Therefore, when we mention a TP we are indistinctly making reference to the associated big-box customer that hosts it.

6.2 Network Definition

This network definition complements the one presented in Section 5.2.1. Within this context, let $\mathcal{N}$ be a set of nanostores. Each nanostore, n, has an associated weekly delivery frequency, ϕ_n, that must be honored if it is served. Moreover, let $\mathcal{A}_i$ be the subset of activation states that a customer i is allowed to take as a TP (i.e., either 'active' or 'inactive'). In particular, let $\mathcal{I}_{active}$ be the subset of big-box customers that are candidate TPs and $\mathcal{I}_{inactive}$, the subset of big-box customers that cannot be considered candidate TPs (i.e., $\mathcal{I}_{inactive} \cup \mathcal{I}_{active} \equiv \mathcal{I}$ and $\mathcal{I}_{inactive} \cap \mathcal{I}_{active} \equiv \emptyset$). This model must also decide which customers, i, to activate as TPs in order to deliver a minimum weekly volume, Ψ (defined as an external input), to the selected nanostores. For this reason, the model should define which nanostores, n, to serve from which active TP, i, on which normalized weekdays, p, from the subset $\mathcal{P}_i$ of feasible normalized weekdays for customer i. Equation (6.1) presents a mathematical definition of $\mathcal{P}_i$ using the function $P(s,j)$ displayed in Table 5.6:

$$\mathcal{P}_i \equiv \Big\{p|p = P\big(s,j\big), s \in \mathcal{S}_i, j \in \mathcal{J}_s\Big\}, \qquad \forall i. \tag{6.1}$$

Table 6.1 formalizes the additional sets described beforehand. Finally, table 6.2 presents the additional parameters relevant to this model.

Table 6.1: Notation of additional sets and subsets for the multi-echelon LRP

$\mathcal{N}$	set of nanostores to be considered; $n \in \mathcal{N}$
$\mathcal{N}_i$	subset of nanostores allocated to big-box customer i; $\mathcal{N}_i \subseteq \mathcal{N}$
$\mathcal{A}$	set of activation status in a potential TP (i.e., *'active'* or *'inactive'*); $a \in \mathcal{A}$
$\mathcal{A}_i$	subset of allowed TP activation status for big-box customer i; $\mathcal{A}_i \subseteq \mathcal{A}$
$\mathcal{P}_i$	subset of normalized delivery days for big-box customer i; $\mathcal{P}_i \subseteq \mathcal{P}$, (e.g., if customer i has weekly frequency three, then $\mathcal{P}_i = \{1, 2, 3\}$)
$\mathcal{I}_{active}$	subset of big-box customers that can be activated as TPs; $\mathcal{I}_{active} \subseteq \mathcal{I}$
$\mathcal{I}_{inactive}$	subset of big-box customers that cannot be activated as TPs; $\mathcal{I}_{inactive} \subseteq \mathcal{I}$

Table 6.2: Notation of additional parameters for the multi-echelon LRP

d_n	volume demanded by nanostore n per visit day [m^3]
ϕ_n	weekly delivery frequency requested by nanostore n [-]; $1 \leq \phi_n \leq 3$
ϕ_i	weekly delivery frequency requested by big-box customer i [-]; $1 \leq \phi_i \leq 6$
$r_{in}^{\Vert 2 \Vert}$	euclidean distance from big-box customer i to nanostore n [km]
ζ_i	average circuity factor for each walking trip departing from big-box customer i [-]
T_n^w	average fixed waiting time at nanostore n [h]
$T^{l,H}$	average fixed setup time for the handcart [h]
$t^{l,H}$	average variable loading time for the handcart [h/m^3]
$t^{u,H}$	average variable unloading time for the handcart [h/m^3]
V^H	average walking speed while delivering with a handcart [km/h]
η^H	maximum loading capacity of the handcart [m^3]
Ψ	weekly volume to deliver to the nanostore channel [m^3]

6.3 Three-step Iterative Approach

For similar reasons to the ones expressed in Section 5.3, the formulation of this problem as a single MILP renders any solving effort fruitless given its high combinatorial complexity. Even the application of BD is incapable of yielding close-to-optimal solutions within reasonable runtimes for real-life problem instances. Therefore, a 'divide-and-conquer' approach provides a feasible roundabout to sort this difficulty. Perl and Daskin (1985) and Wu et al. (2002), among others, elaborate on this type of strategies introducing multi-step iterative approaches to solve, for instance, different variations of the multi-depot LRP.

Therefore, to design the multi-echelon distribution network described in Section 6.2 we propose a multi-step iterative solution approach. We divide the problem in three steps combined with a cost-improvement loop that iterates between the first two steps. For this reason, we apply a separation strategy that is based not only on the decision types (i.e., strategic, tactical and operational) to be made but also on the network Tier (i.e., Tier I for big-box customers or Tier II for nanostores) being approached. In the next paragraphs, we briefly describe each step and the associated cost-improvement loop.

Step 0, addresses the distribution to nanostores from big-box customers (i.e., Tier II strategic and tactical decisions). This step is further divided in two substeps. First, each nanostore is preallocated to the closest —in terms of Euclidean distance— candidate TP whose weekly delivery frequency is greater than or equal to the nanostore's one. Second, given the preallocation of the first substep, an optimization model is executed for each big-box customer that is a candidate TP. This model selects which of the nanosotres, allocated to the candidate TP, should be served within the associated big-box customer's waiting time. Furthermore, the model schedules the visits of the selected nanosotres across the associated big-box customer's delivery days. The objective of this model is to maximize the total weekly volume delivered to nanostores for each candidate TP.

Step I then decides which candidate TPs to activate, defines and fixes the scheduling and facility allocation for each big-box customer, and approximates the required fleet size and mix to serve them (i.e., Tier I strategic and tactical decisions). As described in Section 5.3, individual vehicles are replaced by their associated vehicle types to greatly reduce the computational complexity. This step only provides an approximate transportation cost.

With the main aim of minimizing the associated delivery cost (leveraging economies of scale), serving a specific volume quota of the nanostore channel, and reducing the number of active TPs (exploiting their potential to reach nanostores), we propose and implement an iterative strategy that links steps 0 and I. It consists of an inner and an outer loop. The inner loop cycles between steps 0 and I reducing the number of active TPs until the activation decisions do not change (i.e., the active TPs are the same in two consecutive iterations). It is important to highlight that the convergence of the inner loop is guaranteed since consecutive iterations yield reduced subsets of candidate TPs (i.e., $\mathcal{I}_{active}^{\text{iteration N}} \equiv \mathcal{I}_{active}^{\text{iteration (N-1)}} \subset \mathcal{I}_{active}^{\text{iteration (N-2)}} \subset \cdots \subset \mathcal{I}_{active}^{\text{iteration 2}} \subset \mathcal{I}_{active}^{\text{iteration 1}}$) that can serve the required volume quota of the nanostore channel. At this point, the configuration that have yielded the lowest estimated cost is chosen. The outer loop prevents the algorithm from getting trapped in local optima. For this reason, the outer loop randomly inserts a set of B deactivated TPs (i.e., belonging to $\mathcal{I}_{inactive}$) into Step 0 triggering back the inner loop. The number of outer iterations H is a user-defined parameter.

Step II is run after leaving the outer loop. It takes as inputs the scheduling, allocation to facilities, and TP activation decisions of the network design that has yielded the lowest approximate cost. The model in Step II implicitly routes customers (i.e., Tier I operational decisions) while refining the fleet size and mix calculation (i.e., Tier I tactical decisions). Hence, this step provides a better estimation for the final transportation cost incurred by the whole operation (i.e., reaching both big-box customers and nanostores).

In brief, Step 0 focuses on Tier II strategic and tactical decisions (nanostore selection, allocation to a TP, and scheduling). Step I concentrates on Tier I strategic and tactical decisions (TP activation, big-box customer scheduling and allocation, and approximation of fleet size and mix) providing an approximation for the transportation cost. Finally, Step II takes Tier I operational decisions (implicit vehicle routing) and refines some of the tactical ones (definition of the fleet size and mix) while yielding a better approximation for the final transportation cost. Figure 6.1 schematizes the previously described solution strategy and Algorithm 6.1 formalizes the iterative approach.

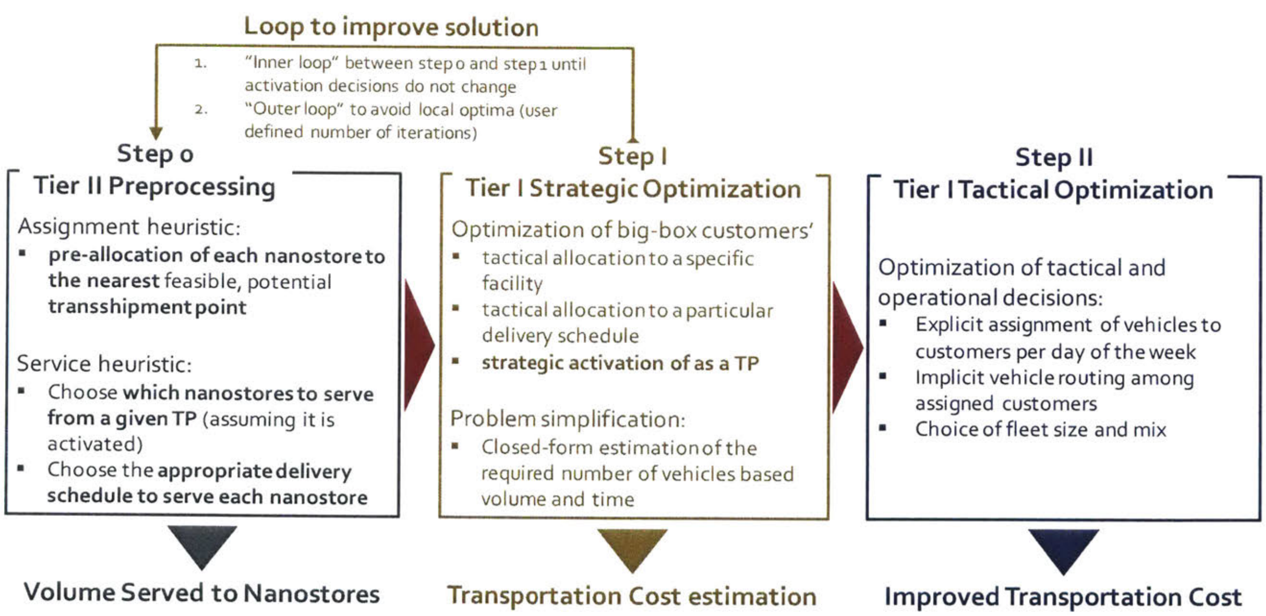


Figure 6.1: Three-step iterative approach framework

6.3.1 Step 0 – Strategic and Tactical Preprocessing of Second-tier Transportation

Step 0 is divided in two sequential substeps. First, we preallocate each nanostore n to the closest candidate TP —measured in Euclidean distance— that has a weekly delivery frequency greater than or equal to the nanostore's one. In this context, any nanostore can be assigned to only one

Algorithm 6.1: Three-step iterative solution approach

Data:

- updated *customers* table
- *nanostores* table
- *optimization parameters*
- H and B

Result:

- scheduling, allocation to facilities, and activation of big-box customers as TPs
- vehicle fleet size and mix definition
- selection, allocation to TPs, and scheduling of nanostores to serve

```
initialize I_active ≡ {i|'active' ∈ A_i, ∀i} and I_inactive ≡ {i|'active' ∉ A_i, ∀i};
run Step 0 considering all candidate TPs in I_active;
run Step I given the nanostores' selection, allocation to a TP, and scheduling from Step 0;
record the solution's approximated cost and the updated subsets I_active and I_inactive given the
  results from Step I;
for h ← 1 to H do
    while I_active has changed do
        run Step 0 with the updated subset I_active;
        run Step I given the nanostores' selection, allocation to a TP, and scheduling from
          Step 0;
        record the solution's approximated cost and the updated subsets I_active and I_inactive
          given the results from Step I;
    end
    select the network configuration that has yielded the lowest estimated cost;
    randomly pick, from the selected network design, B big-box customers from I_inactive such
      that 'active' ∈ A_i and place them in I_active as candidate TPs;
end
select the network configuration that has yielded the lowest estimated cost;
run Step II given the big-box customer allocation to facilities, scheduling, and activation as
  TPs defined in Step I;
return results from Step II.
```

candidate TP in each iteration. Equation (6.2) formalizes this idea:

$$n \in \mathcal{N}_i \Leftrightarrow \begin{cases} r_{in}^{||2||} & = \min\limits_{g \in \mathcal{I}_{active}} r_{gn}^{||2||} \\ \phi_n & \leq \phi_i \end{cases}, \qquad \forall n, i \in \mathcal{I}_{active}. \tag{6.2}$$

Second, an optimization model is executed for each candidate TP (i.e., for each big-box customer $i \in \mathcal{I}_{active}$) based on the nanostore preallocation of the the first substep. This model maximizes the volume that each candidate TP can potentially deliver to nanostores by leveraging the associated big-box customer's average waiting time. It decides which nanostores to serve and how to schedule them based on their weekly delivery frequency. It is important to highlight that no nanostore scheduling restrictions are considered in this model. In other words, each nanostore n, preallocated to a big-box customer i, can be scheduled in any combination of normalized days in $\mathcal{P}_i$ as long as the nanostore weekly delivery frequency is honored. The separation between consecutive nanostore visit days is thus given by the associated big-box customer delivery schedule.

To select a feasible subset of nanostores to serve by each candidate TP within its available average waiting time, the model needs to approximate the feasible delivery tours for each normalized day. To compute the amount of time that a delivery tour takes, we need to know, among other components, the associated traveled distance. Therefore, to approximate the walking distance when a particular nanostore n is served from a candidate TP i, we apply a RLE technique, proposed by Christofides and Eilon (1969), which is based on the problem's Median Relaxation. These authors have shown that the expected length of vehicle routes are monotonically related to the sum of the radial distances of the customers from a center (Christofides and Beasley 1984). In our case, those customers are the selected nanostores and the center is the candidate TP itself. Since each radial (euclidean) distance, $r_{in}^{||2||}$, is known ex-ante, this relaxation yields an associated program that is considerably easier to solve. To improve the accuracy of this estimation, each radial distance must be affected by the average walking circuity factor, ζ_i, associated to the TP. Merchan and Winkenbach (2017) present a data-driven methodology to estimate this factor. This is particularly relevant in our case since the average inter-stop distance is relatively short and, consequently, the effect of network —defined by curbsides and other walkable paths— circuity amplifies.

In the following, we present the mathematical formulation of this problem as a Binary Integer Program (BIP) preceded by Table 6.3 that summarizes the model's decision variables. The relevant additional parameters were already described in Table 6.2.

Table 6.3: Step 0 – notation of decision variables

U_n	binary variable that represents whether nanostore n is being served
O_{np}	binary variable that represents whether nanostore n is being served on normalized day p
G_p	integer variable that represents the number of pedestrian delivery trips to serve those nanostores scheduled on normalized day p

The model is then expressed, for each candidate TP (i.e., $i \in \mathcal{I}_{active}$), as:

$$\max = \sum_{n \in \mathcal{N}_i} \sum_{p \in \mathcal{P}_i} d_n O_{np} \tag{6.3}$$

subject to

$$\sum_{p \in \mathcal{P}_i} O_{np} = \phi_n U_n, \qquad n \in \mathcal{N}_i, \tag{6.4}$$

$$\sum_{n \in \mathcal{N}_i} d_n O_{np} \leq \eta^H G_p, \qquad p \in \mathcal{P}_i, \tag{6.5}$$

$$\sum_{n \in \mathcal{N}_i} \left[d_n \left(t^{l,H} + t^{u,H} \right) + \frac{\zeta_i r_{in}^{||2||}}{V^H} + T_n^w \right] O_{np} \leq T_i^w - T^{l,H} G_p, \qquad p \in \mathcal{P}_i, \tag{6.6}$$

$$U_n \in \{0,1\}, \qquad n \in \mathcal{N}_i, \tag{6.7}$$

$$O_{np} \in \{0,1\}, \qquad n \in \mathcal{N}_i, p \in \mathcal{P}_i, \tag{6.8}$$

$$G_p \in \mathbb{Z}_0^+, \qquad p \in \mathcal{P}_i. \tag{6.9}$$

The objective function (6.3) maximizes, for each candidate TP i, the weekly volume that it can deliver to neighboring nanostores. Constraints (6.4) guarantee that each visited nanostore, n, is served according to its weekly delivery frequency, ϕ_n. Equations (6.5) enforce that the handcart's maximum payload η^H (which is assumed to be constant for all trips) cannot be exceeded. Restrictions (6.6) limit the total time that can be spent, on each normalized day p, delivering to nanostores. There, T_i^w denotes, for each candidate TP i, the maximum allowed time per visit day that can be allocated to serve nanostores. Several time components are considered in these constraints including: a fixed setup time for each delivery trip, $T^{l,H}$ (which is assumed equal for all delivery tours); volume dependent loading and unloading times, $t^{l,H}$ and $t^{u,H}$, respectively; a delivery tour walking time, computed as the division between the expected walking distance to reach the served nanostore, $\zeta_i r_{in}^{||2||}$, and the walking speed, V^H (assumed to be the average human walking speed); and a fixed waiting time at each nanostore, T_n^w (which is assumed to be the same for all nanostores). Finally, Equations (6.7) through (6.9) define the allowable domains of all decision variables.

6.3.2 Step I – Strategic and Tactical Optimization of First-tier Transportation

The optimization model formulated in Step I is sequentially executed after Step 0. This program presents obvious similarities with the one in Section 5.3.1 in terms of decision variables, objective function, and mathematical formulation of the associated constraints. However, this program also decides which big-box customers, that are candidate TPs (i.e., $i \in \mathcal{I}_{active}$), must be activated to serve a predefined volume quota of the nanostore channel while reducing the estimated transportation cost. Therefore, this problem considers, for each candidate TP, the optimized volumes (derived from Step 0) that a big-box customer can deliver to nanostores on each normalized day if activated as a TP. Let d_{ip}^N be the optimized volume (in [m^3]) that a candidate TP, i, can deliver to the nanostore channel on a normalized day, p (i.e., $d_{ip}^N = \sum_{n\in\mathcal{N}_i} d_n O_{np}$ if $i \in \mathcal{I}_{active}$ —as optimized in Step 0— and $d_{ip}^N = 0$ if $i \in \mathcal{I}_{inactive}$). We can thus redefine the volume, d_{ipa}, to deliver to each big-box customer, i, on a normalized day, p, given an allowed TP activation status, a, according to equation (6.10):

$$d_{ipa} = \begin{cases} d_{ip}, & \text{if customer } i \text{ is not activated as a TP,} \\ d_{ip} + d_{ip}^N, & \text{if customer } i \text{ is activated as a TP,} \end{cases} \qquad \forall i, p \in \mathcal{P}_i, a \in \mathcal{A}_i. \tag{6.10}$$

Table 6.4 introduces the decision variables and is followed by the MILP formulation for this problem.

Table 6.4: Step I – notation of decision variables

X_{ifvsja}	binary variable that represents whether customer i is served from facility f with vehicle type v on weekday j allowed by delivery schedule s using TP activation status a feasible for that customer
Y_{isfa}	binary variable that represents whether customer i is served according to delivery schedule s from facility f using TP activation status a feasible for that customer
L_{fvjmq}	continuous variable that represents how many vehicles of type v are used from facility f on weekday j to municipality m taking cargo-type q

The model can be formulated as:

$$\min\left\{\sum_{f\in\mathcal{F}}\sum_{v\in\mathcal{V}_f}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{M}}\sum_{q\in\mathcal{Q}_v} c_{fvm}L_{fvjmq} + \sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}}\sum_{v\in\mathcal{V}_i\cap\mathcal{V}_f}\sum_{s\in\mathcal{S}_i}\sum_{j\in\mathcal{J}_s}\sum_{a\in\mathcal{A}_i} c_f d_{ipa} X_{ifvsja}\right\} \tag{6.11}$$

subject to

$$\sum_{s\in\mathcal{S}_i}\sum_{f\in\mathcal{F}}\sum_{a\in\mathcal{A}_i} Y_{isfa} = 1, \qquad \forall i, \tag{6.12}$$

$$\sum_{v\in\mathcal{V}_i\cap\mathcal{V}_f} X_{ifvsja} = Y_{isfa}, \qquad \forall i, s \in \mathcal{S}_i, j \in \mathcal{J}_s, \forall f, a \in \mathcal{A}_i, \tag{6.13}$$

$$\sum_{i\in\mathcal{I}}\sum_{v\in\mathcal{V}_i\cap\mathcal{V}_f}\sum_{s\in\mathcal{S}_i|j\in\mathcal{J}_s}\sum_{a\in\mathcal{A}_i} d_{ipa}X_{ifvsja} \leq \xi_f, \qquad \forall f,\forall j, \tag{6.14}$$

$$\sum_{f\in\mathcal{F}}\sum_{m\in\mathcal{M}}\sum_{q\in\mathcal{Q}_v} L_{fvjmq} \leq a_v, \qquad \forall j,\forall v, \tag{6.15}$$

$$\sum_{i\in\mathcal{I}_{mq}|v\in\mathcal{V}_i}\sum_{s\in\mathcal{S}_i|j\in\mathcal{J}_s}\sum_{a\in\mathcal{A}_i} d_{ipa}X_{ifvsja} \leq \eta_{vq}L_{fvjmq}, \qquad \forall f,v\in\mathcal{V}_f,\forall j,\forall m,q\in\mathcal{Q}_v, \tag{6.16}$$

$$\sum_{i\in\mathcal{I}_{mq}|v\in\mathcal{V}_i}\sum_{s\in\mathcal{S}_i|j\in\mathcal{J}_s}\sum_{a\in\mathcal{A}_i}\left[\frac{r_{im}}{V_v^d}+T_i^w+t_{fvq}^l d_{ipa}+t_{vq}^u d_{ip}\right]X_{ifvsja} \leq \left(T_{max}-\frac{r_{fm}}{V_v^{lh}}-T_{fvq}^l\right)L_{fvjmq}, \qquad \forall f,v\in\mathcal{V}_f,\forall j,\forall m,q\in\mathcal{Q}_v, \tag{6.17}$$

$$\sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{F}}\sum_{v\in\mathcal{V}_i\cap\mathcal{V}_f}\sum_{s\in\mathcal{S}_i}\sum_{j\in\mathcal{J}_s}\sum_{a\in\mathcal{A}_i}(d_{ipa}-d_{ip})X_{ifvsja} \geq \Psi, \tag{6.18}$$

$$X_{ifvsja}\in\{0,1\}, \qquad \forall i,\forall f,v\in\mathcal{V}_i\cap\mathcal{V}_f,s\in\mathcal{S}_i,j\in\mathcal{J}_s,a\in\mathcal{A}_i, \tag{6.19}$$

$$Y_{isfa}\in\{0,1\}, \qquad \forall i,s\in\mathcal{S}_i,\forall f,a\in\mathcal{A}_i, \tag{6.20}$$

$$L_{fvjmq}\in\mathbb{R}_0^+, \qquad \forall f,v\in\mathcal{V}_f,\forall j,\forall m,q\in\mathcal{Q}_v. \tag{6.21}$$

The objective function (6.11) minimizes an approximation of the total transportation cost aggregating both the delivery and the inter-facility transfer cost components. The demand of each big-box customer now depends on its TP activation status (i.e., whether or not it is used to serve the nanostore channel).

Constraints (6.12) guarantee that each customer can only be served according to one feasible delivery schedule s, from only one facility f, and using TP activation status a. Equations (6.13) then force customers to be delivered according to the delivery schedules defined in the previous set of restrictions. Constraints (6.14) prevent the model from overloading available facilities given their maximum daily capacity ξ_f. Equations (6.15) guarantee that the model will only assign, on each weekday j, vehicles of category v based on the associated available fleet a_v. Constraints (6.16) enforce that, given a vehicle type v and a cargo type q, the aggregated payload is not exceeded. Equations (6.17) ensure that the aggregated available time of a specific vehicle type v should not be exceeded. It is important to highlight that, the variable loading time is calculated using the total volume carried by the trucks (i.e., the volume delivered to big-box customers as well as the volume destined to nanostores). However, only the volume that is served to big-box customers affects the variable unloading time. The variable unloading time associated with nanostores has already been considered in Restrictions (6.6) of Step 0 when the handcart is being loaded. Therefore, the nanostore-related variable unloading time is contained within the associated big-box customer's average waiting time T_i^w. The new Constraint (6.18) forces the model to activate as many TPs as necessary to deliver to the nanostore channel the predefined minimum weekly volume Ψ. This equation only considers the volume that big-box customers send to nanostores when activated as TPs. Finally, Equations (6.19) through (6.21) define the allowable domains of all decision variables.

6.3.3 Step II – Tactical and Operational Optimization of First-tier Transportation

After solving the previous model and exiting the cost improvement loop, the allocation, scheduling, and TP activation results for each big-box customer, determined by the lowest cost solution, are fed into Step II. This particular program is analogous to the one in Section 5.3.2. Specifically, we redefine the big-box customer demands based on their TP activation status. Here, d'_{ip} represents the aggregated demand that must be sent to big-box customer i on its normalized day p. It groups the customer-specific demand, d_{ip}, and the volume delivered to neighboring nanostores, d'^{N}_{ip}, if i is an active TP (otherwise $d'^{N}_{ip} = 0$).

Considering these modifications and the variables defined in Table 5.12, the mathematical model for this optimization can be formulated as:

$$\min \left\{ \sum_{f\in\mathcal{F}} \sum_{k\in\mathcal{K}_f} \sum_{j\in\mathcal{J}} \sum_{m\in\mathcal{M}} \sum_{q\in\mathcal{Q}_k} c_{fkm} Z_{fkjmq} + \sum_{i\in\mathcal{I}} \sum_{k\in\mathcal{K}_i\cap\mathcal{K}_{f_i}} \sum_{j\in\mathcal{J}_i} c_f d'_{ip} \tilde{X}_{if_ikj} \right\} \tag{6.22}$$

subject to

$$\sum_{k\in\mathcal{K}_i\cap\mathcal{K}_{f_i}} X_{if_ikj} = \left\lceil \frac{d'_{ip}}{\eta_i^{max}} \right\rceil, \qquad \forall i, j \in \mathcal{J}_i, \tag{6.23}$$

$$\sum_{k\in\mathcal{K}_i\cap\mathcal{K}_{f_i}} \tilde{X}_{if_ikj} = 1, \qquad \forall i, j \in \mathcal{J}_i, \tag{6.24}$$

$$\sum_{f\in\mathcal{F}|k\in\mathcal{K}_f} \sum_{m\in\mathcal{M}} \sum_{q\in\mathcal{Q}_k} Z_{fkjmq} \le 1, \qquad \forall k, \forall j \tag{6.25}$$

$$\sum_{f\in\mathcal{F}} \sum_{k\in\mathcal{K}_f\cap\mathcal{K}_v} \sum_{m\in\mathcal{M}} \sum_{q\in\mathcal{Q}_k} Z_{fkjmq} \le a_v, \qquad \forall j, \forall v \tag{6.26}$$

$$\sum_{i\in\mathcal{I}_{mq}|f=f_i,k\in\mathcal{K}_i,j\in\mathcal{J}_i} d'_{ip} \tilde{X}_{if_ikj} \le \eta_{kq} Z_{fkjmq}, \qquad \forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k, \tag{6.27}$$

$$\sum_{i \in \mathcal{I}_{mq} | k \in \mathcal{K}_i \& j \in \mathcal{J}_i} \left[X_{if_ikj} \left(\frac{r_{im}}{V_k^d} + T_i^w \right) + t_{fkq}^l d_{ip}' \tilde{X}_{if_ikj} + t_{kq}^u d_{ip} \tilde{X}_{if_ikj} \right] \leq \left(T_{max} - \frac{r_{fm}}{V_k^{lh}} - T_{fkq}^l \right) Z_{fkjmq}, \quad \forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k, \tag{6.28}$$

$$X_{if_ikj} \geq \tilde{X}_{if_ikj}, \quad \forall i, k \in \mathcal{K}_i \cap \mathcal{K}_{f_i}, j \in \mathcal{J}_i, \tag{6.29}$$

$$\mu X_{if_ikj} \leq \tilde{X}_{if_ikj}, \quad \forall i, k \in \mathcal{K}_i \cap \mathcal{K}_{f_i}, j \in \mathcal{J}_i, \tag{6.30}$$

$$X_{if_ikj} \in \{0, 1\}, \quad \forall i, k \in \mathcal{K}_i \cap \mathcal{K}_{f_i}, j \in \mathcal{J}_i, \tag{6.31}$$

$$\tilde{X}_{if_ikj} \in [0, 1], \quad \forall i, k \in \mathcal{K}_i \cap \mathcal{K}_{f_i}, j \in \mathcal{J}_i, \tag{6.32}$$

$$Z_{fkjmq} \in \{0, 1\}, \quad \forall f, k \in \mathcal{K}_f, \forall j, \forall m, q \in \mathcal{Q}_k. \tag{6.33}$$

The description of Equations (5.27) through (5.38) also holds for Equations (6.22) through (6.33). A similar explanation of Constraints (6.17) applies for Restrictions (6.28). The execution of this model yields the final and 'optimal' configuration for this two-echelon distribution network that integrates both big-box customers and nanostores.

Chapter 7

Experimentation and Analysis

7.1 Single-echelon Network Design

The goal of this experimentation is twofold. First, to compare the performances of the proposed solution approaches (i.e., single-step, BD, and two-step formulations) in terms of total delivery cost and computational time (i.e., algorithmic efficiency) for different problem sizes. Second, to contrast the network designs yielded by each solution methodology regarding customer (and volume) allocation to facilities as well as fleet size and mix definition for a representative problem instance.

Experiments were performed on several stylized instances each containing an increasing number of customers (11, 20, 30, 50, 72, and 109) from different municipalities and with specific demand characteristics (in terms of weekly frequency, demanded volume, and cargo-type, among others). All these instances were addressed using the three solution strategies proposed in Sections 5.2 through 5.4. The input parameters as well as some complementary results are summarized in Appendix B.

7.1.1 Network Cost

Table 7.1 contrasts the different network costs, classified by their main components (i.e., delivery, transfer, and total cost), that every solution strategy yielded for each problem instance.

Several insights can be derived from these results. First and foremost, both the single-step and the BD formulations could not retrieve cost-optimal solutions for the 72- and 109-customer problem instances since they failed to reach the set optimality gap (0.5%) within the maximum allowed run-time (84 [h]). Second, it is worth noting that several network configurations produced near-optimal solutions as presented, for example, by the 20- and 30-customer problem instances. In these cases, while both the one-step and BD approaches produced cost-optimal solutions, the two-step strategy

Table 7.1: Cost comparison (in [USD]) per solution approach

Instance	Single-step			BD				Two-step			
# Customers	Delivery Cost	Transfer Cost	Total Cost	Delivery Cost	Transfer Cost	Total Cost	% Difference	Delivery Cost	Transfer Cost	Total Cost	% Difference
11 customers	1,668	93	1,761	1,668	93	1,761	**0.00**	1,668	93	1,761	**0.00**
20 customers	3,484	244	3,728	3,484	244	3,728	**0.00**	3,508	246	3,754	**0.70**
30 customers	4,213	272	4,485	4,213	272	4,485	**0.00**	4,216	272	4,488	**0.01**
50 customers	6,269	322	6,591	6,297	322	6,619	**0.40**	6,112	480	6,592	**0.02**
72 customers	-	-	-	-	-	-	-	8,556	629	9,185	-
109 customers	-	-	-	-	-	-	-	12,030	1,182	13,212	-

yielded near-optimal ones. In other words, even though the network configurations proposed by the two-step approach were different than those returned by the other two methodologies, the total transportation costs differed by less than 1%. This unveils a highly symmetrical solution space, as described in Section 5.3, that presents a multiplicity of (near-)optimal network configurations. This topology of the solution space highly increases the computational effort required to solve this type of problems.

Regarding BD, this approach arrived at optimal network designs except for the 50-customer problem instance. In that particular case, the methodology provided a slightly more expensive solution (still within the 0.5% optimality gap) given by a close-to-optimal fleet mix definition, scheduling, and routing as expressed by a higher delivery cost. However, the customer allocation seems to be optimal since the transfer cost components are equal for both one-step and BD approaches.

Similar conclusions can be drawn for the two-step approach. This methodology yielded close-to-optimal results for every comparable problem instance. It was also well-behaved for larger problems (i.e., with 72 and 109 customers) returning more economical network configurations when compared with the sub-optimal ones produced by both the single-step and BD approaches. Finally, no obvious correlation can be appreciated between the relative cost difference (compared to the one-step formulation) yielded by the two-step approach and the problem size. However, these high quality solutions render the two-step approach as an attractive methodology to address this type of problems with a neglegible increase in cost.

7.1.2 Network Design

In this section, we take the 50-customer problem instance (i.e., the largest instance that could be solved to 'optimality' by the three methodologies within the allowed runtime) and compare the network designs (in terms of customer allocation to facilities, volume delivered per day and per facility, and vehicle fleet mix and size) produced by each solution approach. We have selected the largest comparable problem instance since it provides the most interesting variations in terms of network design.

Figure 7.1 depicts the daily volume being delivered per facility (i.e., Puente Alto and Talagante) for each solution strategy. In terms of customer allocation, both the single-step and BD approaches served all customers through Talagante, which was generally the most convenient alternative given the combination of delivery and transfer costs (see Tables B.3 and B.7 for further information). The two-step approach; however, partially used Puente Alto to serve two high dropsize big-box customers (see Figure B.2(c)). Besides, the different volumes being served on each weekday varies per solution approach. In addition, none of the facilities were ever saturated in terms of capacity regardless the solution methodology. The previous is based on efficient customer scheduling decisions.

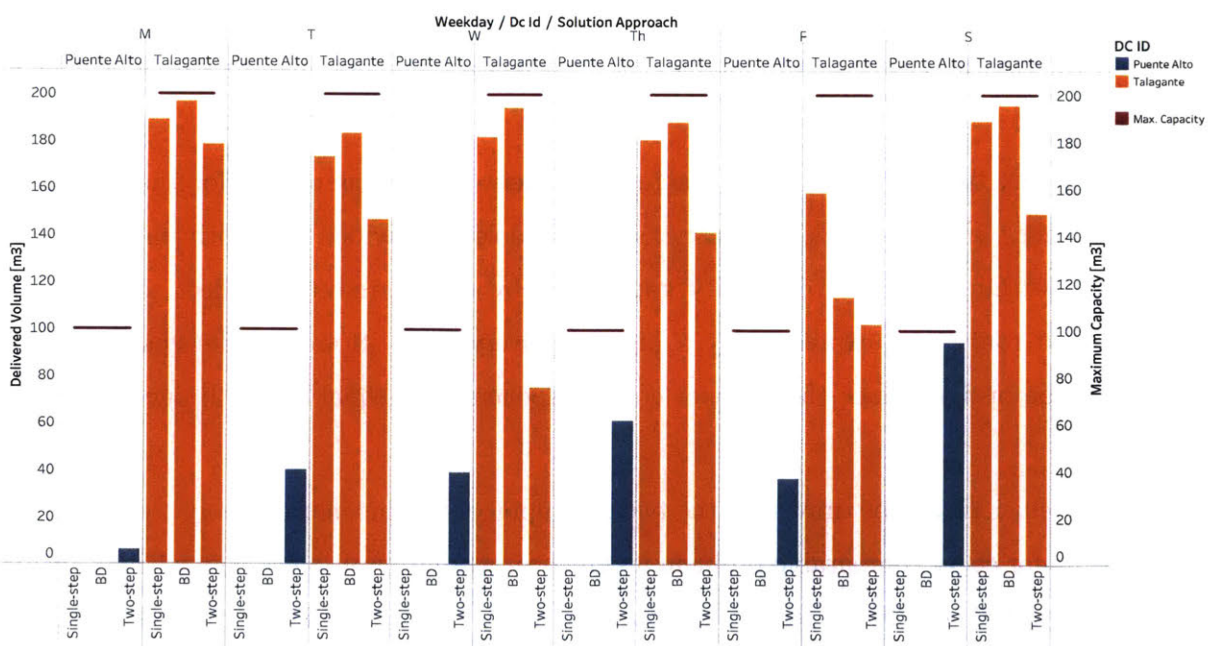
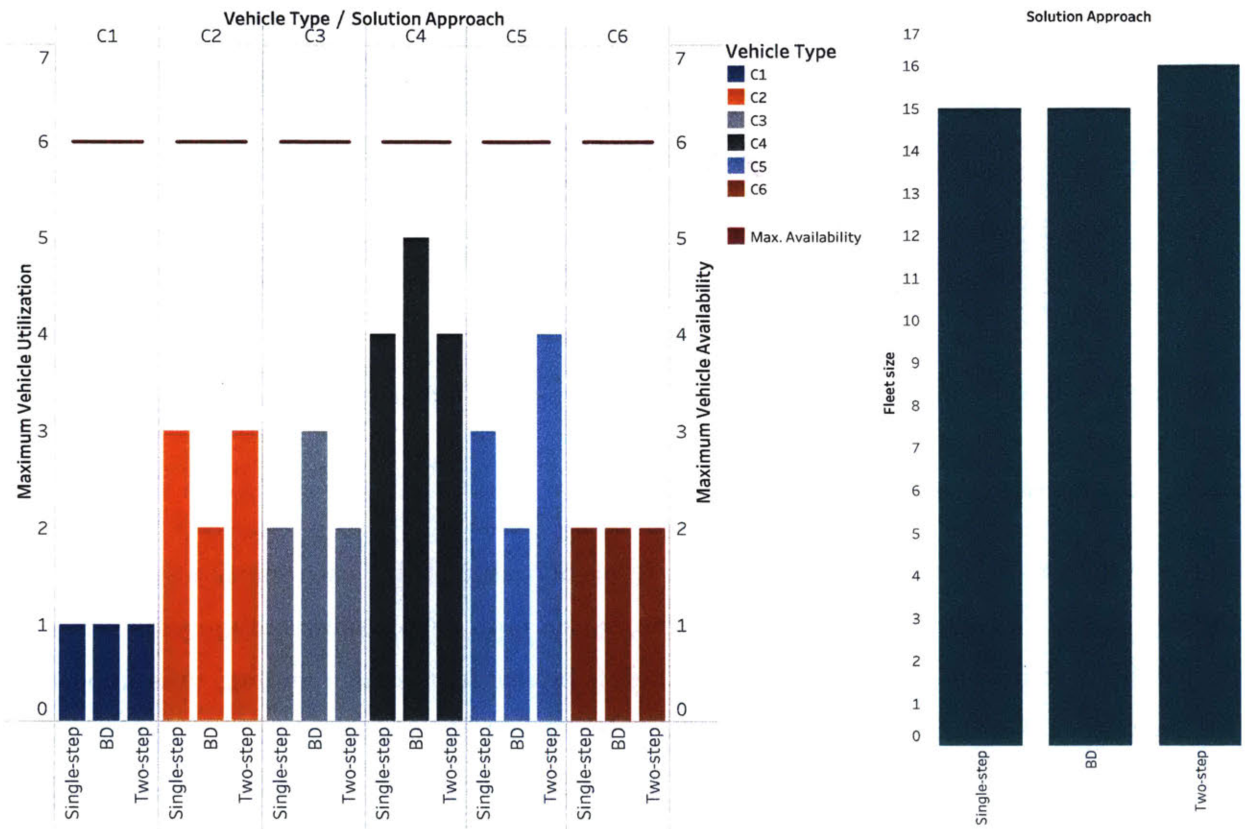


Figure 7.1: Volume distribution and DC capacity (in [m^3]) per weekday and solution strategy – 50-customer instance

Regarding the fleet mix, BD, on the one hand, used higher capacity, more expensive vehicles (i.e., C3 and C4) returning a near-optimal solution. On the other hand, the two-step approach suggested a similar fleet mix to the one yielded by the one-step methodology but combined with a different customer allocation. Figure 7.2(a) illustrates the previous by depicting the maximum vehicle utilization per vehicle-type and solution approach. According to this, none of the vehicle categories is, in any case, a limiting resource. Moreover, according to Figure 7.2(b), the total number of vehicles required by each solution strategy is similar. The single-step approach and BD required 15 vehicles while the two-step strategy, 16.



(a) Vehicle fleet composition (b) Vehicle fleet total size

Figure 7.2: Vehicle fleet mix and size per solution strategy – 50-customer instance

Finally, Figure 7.3 presents, based on the two-step approach results, the average relative split per vehicle- and cargo-type spent on each relevant time component modeled in this scenario (i.e., fixed waiting time at DCs and customers, variable loading and unloading time, and line-haul and delivery driving time). Between 30% and 50% of the total trip time is spent, on average, waiting at customer locations. The total amount of time waiting at customers is invariant for the other solution approaches since the average fixed waiting time per customer, T_i^w, is an input for our models. The previous motivates the use of this system-inherent inefficiencies to serve neighboring nanostores. This alternative is analyzed in the next section for a larger problem instance.

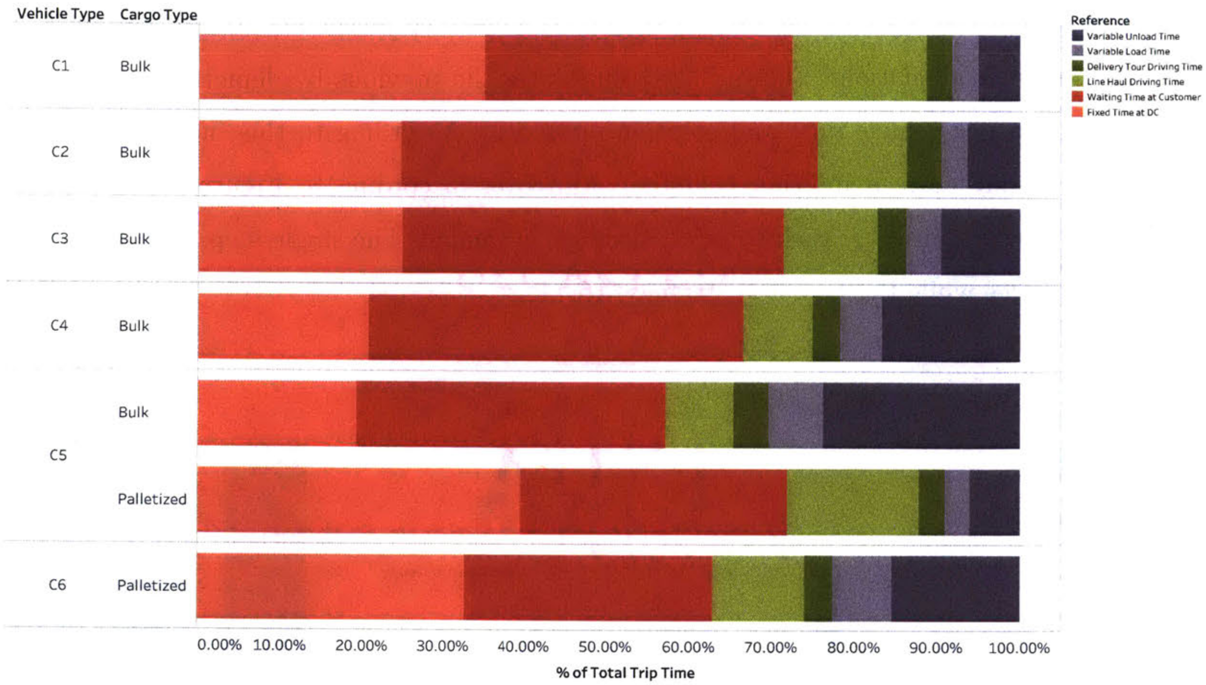


Figure 7.3: Total trip time split yielded by the two-step approach – 50-customer instance

7.1.3 Computational Time

Table 7.2 compares the computational effort (in terms of runtime), that each solution strategy took to produce optimal results.[1] For our experiments, we set both a 0.5% optimality gap threshold as well as an 84-hour maximum allowed runtime. Before presenting the main findings, it is important to highlight that, since the problem instances with 72 and 109 customers could not be solved to optimality (within the allowed runtime limit) using the single-step and BD approaches, their associated times are not presented. In addition, since these types of problems are NP-hard, runtimes exponentially increase with instance size.

The first key point is related with the clear advantage in computational time that the two-step method provides in comparison with the other two approaches. At this respect, one-order-of-magnitude runtime differences can be detected for instances with more than 20 customers. The previous, combined with the high quality solutions yielded by this methodology, suggests its suitability for approaching larger problem instances. Moreover, if we analyze the the two-step approach runtime composition, it is clear that the first step (i.e., the allocation and scheduling decisions) accounts for the vast majority of the total computational effort. Once those decisions have been made,

[1]The calculations were carried out on a 3.1 GHz Intel Core i7 MacBook Pro with 16.0 GB of installed RAM running OS Sierra version 10.12.6. The models were implemented in Python 2.7 and used Gurobi Optimizer version 7.5.2 for the optimization.

Table 7.2: Runtime (in [sec]) for each single-echelon distribution network solution approach

Instance	**Single-step**	**BD**		**Two-step**			
# Customers	Runtime [sec]	Runtime [sec]	% Difference	Step I - Runtime [sec]	Step II - Runtime [sec]	Total Runtime [sec]	% Difference
11 customers	17	23	**+35**	10	1	11	**-35**
20 customers	2,512	2,305	**-8**	202	1	203	**-92**
30 customers	42,026	40,989	**-3**	424	1	425	**-99**
50 customers	186,403	179,264	**-4**	20,368	156	20,524	**-89**
72 customers	-	-	**-**	66,423	129	66,552	**-**
109 customers	-	-	**-**	196,735	793	197,528	**-**

the fleet definition and the implicit routing take a neglegible fraction of the total computational time regardless the problem size.

When we consider both the single-step formulation and BD, the former proved faster for the smallest problem instance (i.e., with 11 customers). However, for larger cases, BD was marginally quicker (e.g., 1,000 seconds for the 30-customer instance). That meager improvement renders this solution alternative ill-favoured due to its intrinsically high coding complexity. Since, modern solvers such as Gurobi already implement runtime enhancement techniques (e.g., different types of cutting plane strategies and heuristics) on their default settings, the potential computational gains are marginal.

7.2 Two-echelon Network Design

Experiments were performed on two stylized instances each simulating increasing cannibalization percentages of nanostore demand. In this section, we will compare, in terms of cost and network design, an instance with no cannibalization against two instances with 10.0% and 16.2% cannibalization rates, respectively. The latter instance simulates the maximum feasible volume share that can be served to nanostores by leveraging only the idle time spent at big-box customers. All cannibalization instances, involving 105 potential TPs (i.e., 'bulk'-type customers) out of 109 big-box customers and 5,064 nanostores that could be potentially served, were solved using the methodology introduced in Chapter 6. Both new and updated parameters as well as additional network design results are presented in Appendix C.

7.2.1 Network Cost

Table 7.3 summarizes the main information concerning delivered volumes, active TPs, nanostores served, and different cost components for each cannibalization rate instance. Based on these results, 58% of the potential TPs (i.e., 61 out of 105) were activated to serve 10% of the nanostore channel volume (i.e., 122.2 m^3) reaching 8.6% of the nanostores (i.e., 433 out of 5,064). Nanostores low drop-sizes as well as big-box customers constrained waiting times demanded a considerable proportion of TPs to cannibalize the targeted market share with a 7.1% increase in the total transportation cost. Similarly, for the maximum possible cannibalization rate instance (i.e., leveraging the total available waiting time at potential TPs as intensively as possible), 96% of the potential TPs (i.e., 101 out of 105) were activated to serve 16.2% of the nanostore channel volume (i.e., 197.1 m^3) yielding a 12.4% increase in the total transportation cost.

Table 7.3: Volume (in [m^3]) and cost (in [USD]) comparison per cannibalization rate

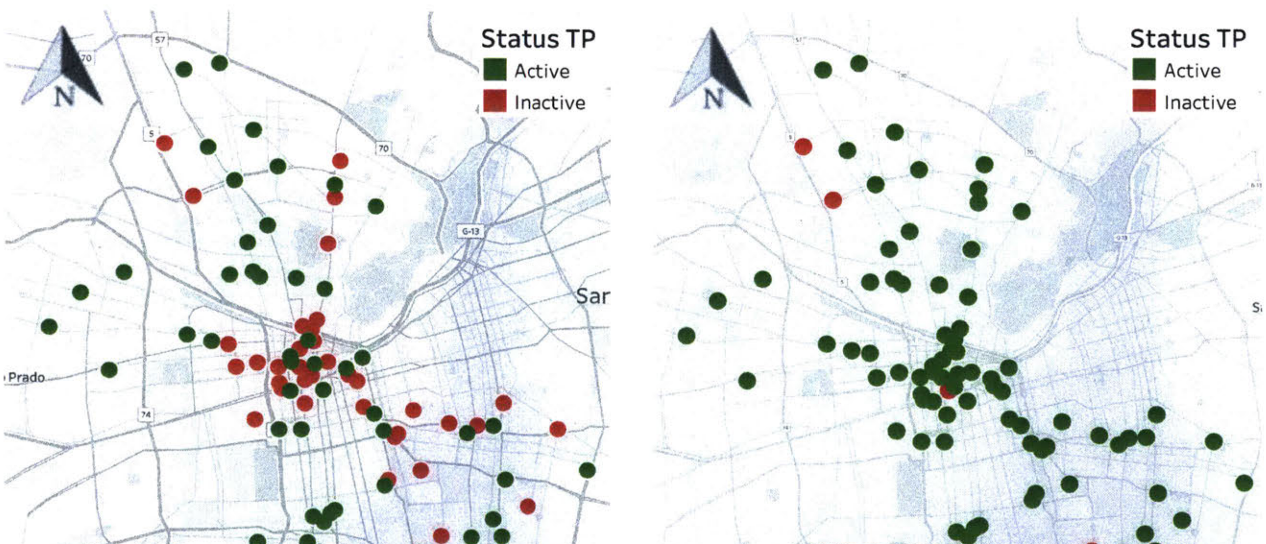
Instance	**Volume**			**Network**		**Cost component**				
% Cannibalization	to Big-box	to Nanostores	Total	# Active TPs	# Nanostores served	Delivery	Extra crew member	Transfer	Total	% Difference
0.0	2,469.4	-	2,469.4	-	-	12,030	-	1,182	13,212	-
10.0	2,347.2	122.2	2,469.4	61	433	12,348	725	1,073	14,146	**7.1**
Maximum (16.2)	2,272.3	197.1	2,469.4	101	741	12,958	800	1,087	14,845	**12.4**

7.2.2 Network Design

In this section, Figures 7.4 through 7.6 present a summary of the network design results yielded for both the 10% and the 16.2% cannibalization instances in terms of TP activation (Figures 7.4(a) and 7.4(b)), number of nanostores and volume served from each active TP (Figures 7.5(a) and

7.5(b)), and the walking distance distribution from active TPs to served nanostores (Figures 7.6(a) and 7.6(b)). Further information about big-box customer allocation to facilities, the geographical distribution of the served nanostores, and the fleet mix and size can be found, for each problem instance, in Appendix C.

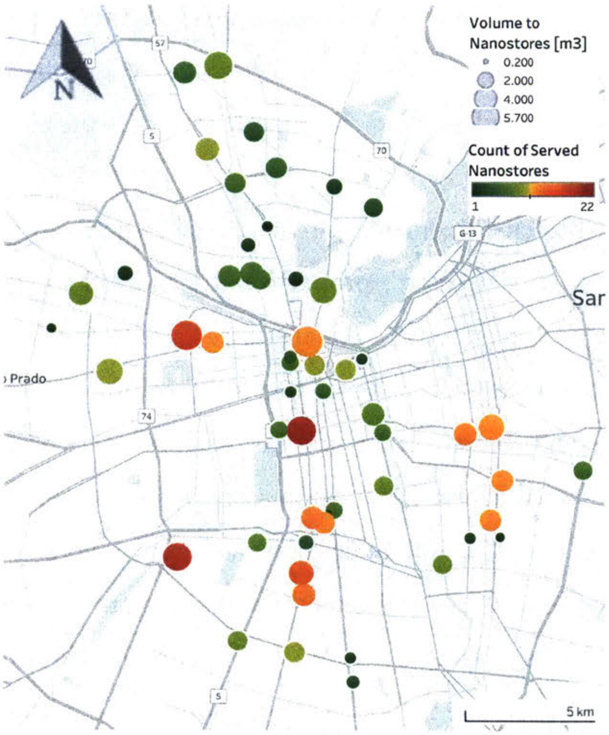
According to Figures 7.4 and 7.5, neighboring big-box customers interfere with each other leading to slightly diminishing returns in cannibalization rates. Therefore, proportionally more TPs are needed to serve a given volume increment in the nanostore channel. For instance, a 66% increase in the number of active TPs (i.e., 101 over 61) leads to a 61% raise in the total volume served to nanostores (i.e., 197.1 m^3 over 122.1 m^3). Moreover, based on Figure 7.6, as more TPs were activated to capture a higher proportion of the traditional channel, nanostores are naturally served by closer TPs. The previous is coherent with the nanostore preallocation strategy defined by Equation (6.2). In the 10%-cannibalization-rate instance, 80% of the served nanostores are located within 650 meters (with a 395-meter mean) of their TPs. However, in the 16.2%-cannibalization-rate instance, 80% of the nanostores are now located within 650 meters (with a 350-meter mean) of their sources, since more TPs were activated. Finally, it is worth noting that, as shown in Figure C.5, none of the problem instances required drastically different fleet sizes.



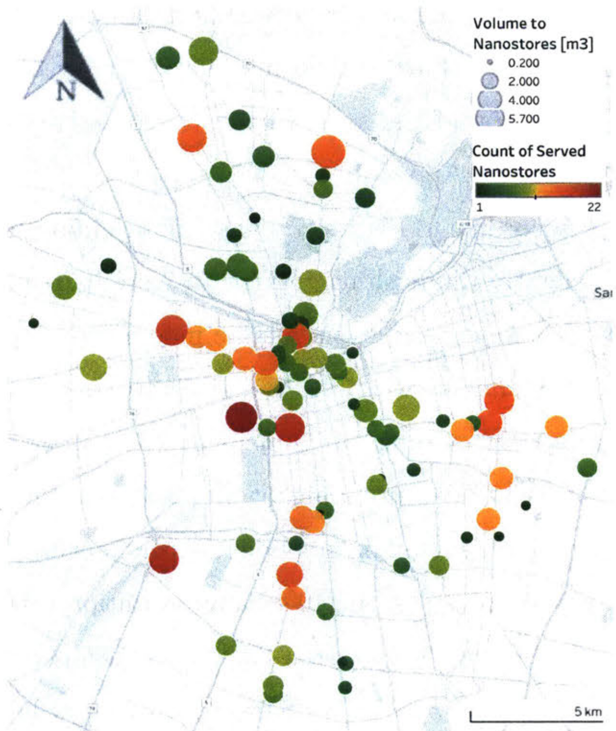
(a) 10.0% cannibalization rate

(b) Maximum (16.2%) cannibalization rate

Figure 7.4: TP status per cannibalization instance

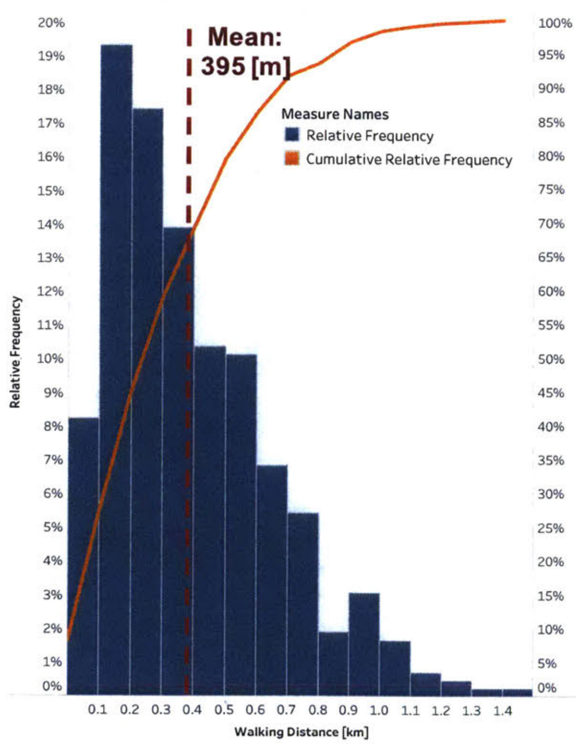


(a) 10.0% cannibalization rate

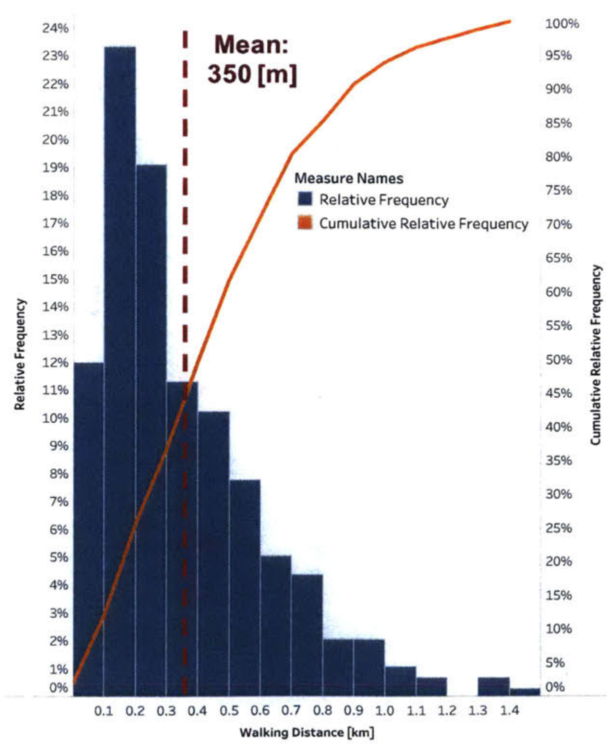


(b) Maximum (16.2%) cannibalization rate

Figure 7.5: Volume (in [m^3]) and number of nanostores served per active TP



(a) 10.0% cannibalization rate



(b) Maximum (16.2%) cannibalization rate

Figure 7.6: Distribution of walking distance (in [km]) from active TPs to served nanostores

Chapter 8

Conclusions and Future Research

The transportation of goods constitutes both an extremely important and an increasingly disturbing activity taking place in urban areas. Freight movements support most city-based activities, while negatively impacting the quality of life in cities through significant contributions to the levels of congestion, noise, and pollution (Crainic et al. 2004). From the private perspective, transportation, in particular last-mile delivery, is a vital yet highly inefficient link within every supply chain. Hence, there is an imperative need for methodologies and tools addressing the planning of optimized delivery-related activities. In this context, as stated by Lin and Lei (2009), the optimal design of distribution systems requires an integrated view of strategic (e.g., facility investment, which specific customers to serve, among others), tactical (e.g., customers' scheduling and their allocation to facilities, vehicle mix definition, etc.), and operational decisions (e.g., how to dispatch available vehicles in delivery routes to satisfy customers' demand). This work has thus developed mathematical models that provide such an integrated view as well as computationally feasible algorithms for obtaining solutions in realistic situations. The proposed framework allows the comprehensive modeling of distribution strategies, deemed relevant in practice, while simultaneously capturing the complexity of urban-last-mile network design decisions. The application of our modeling framework makes it possible to infer how choices on a conceptual level impact the outcomes of the network design in terms of the strategic, tactical, and operational decision levels. Moreover, computational experiments enable the replication of common trade-offs that were identified through the case study under analysis.

In the first part of this thesis, we formulated a single-echelon capacitated RP integrating customer scheduling decisions as well as a capacity-constrained mixed vehicle fleet to optimize the delivery to big-box customers within the urban environment. This problem was modeled as a special case of a BPP combining it with a customer clustering approach. However, given the inherently high numerical complexity of this model, two alternative solution methodologies were tested to reduce computational times. On the one hand, a two-step approach has been designed, implemented, and tested on a set of reduced problem instances, obtaining near-optimal solutions in a limited amount of time. Its effects are found to be remarkably important as only a fraction of the original formulation's computing time is needed, and this economy is achieved with a negligible loss in quality. The previous fosters the scalability of this methodology for real-life problem instances providing near-optimal solutions in considerably reduced runtimes. On the other hand, the decomposition of the original formulation using BD was implemented and analyzed. Despite yielding similar total costs and a slight improvement in computational effort, when compared with the time required to solve our original formulation, BD does not bring noticeable benefits as those obtained with our two-step approach. Although both the single-step and BD strategies produce optimal solutions, they render computationally prohibitive for larger problem instances

The second part of this work built upon the first one to integrate a number of highly fragmented nanostores within the distribution network's footprint and its existing infrastructure. Here, the main motivation was to leverage existing idle times at the original points of delivery (i.e., waiting time at each big-box customer) to cannibalize part of the traditional channel's market share that will potentially generate a more than proportional increase in revenue. Hence, a three-step iterative optimization model has been developed —reusing techniques and formulations from the first part— and tested yielding promising results. The main takeaway lies in the opportunity of leveraging system inefficiencies (improving assets utilization and optimizing the use of idle times) to grow the network market reach without a considerable increase in the associated delivery cost. However, to render this alternative profitable, the previous cost increase must be outbalanced by the additional revenue generated with the cannibalized market share.

The research problems presented in this thesis can be extended in multiple ways. By no means is the following intended to be an absolute or infallible description of limitations and potential improvements of the developed models.

First of all, in distribution network design problems, customer demand and the number of orders (Snoeck et al. 2017, Klibi et al. 2010, Baldi et al. 2012), travel times (Ghaffari-Nasab et al. 2013, Zarandi et al. 2011), and time windows (Zarandi et al. 2013) may be affected by uncertainties. These uncertainties can be modeled as random variables whose probability distributions are, for instance, estimated using historical data. Those characteristics can be handled using various techniques such as simulation, stochastic programming, and recourse methods. For instance, Snoeck and Winkenbach (2018) study the value of including flexibility in the design phase of last-mile distribution networks to better handle future exogenous variability that will arise in their operations.

Moreover, results and insights derived from the two-echelon network design are based on a known, deterministic, and constant aggregated demand, which is certainly not very realistic. A natural way to relax this assumption is to assume that the integration of nanostores might spark a growth in demand, which can be either deterministic or stochastic. Ge (2017) presents interesting conclusions on how to model demand increases when companies decide to integrate the traditional retail segment into their modern-channel-oriented distribution networks.

Furthermore, this work refers to no sales methodology employed to take nanostore orders. With pre-sales and van-sales the more wide-spread alternatives, a thorough study is required to select the most efficient alternative to add its associated cost into the network design model. Boulaksil and Belkora (2017) compare both strategies for a case study in Casablanca, arriving at interesting results about the attractiveness of each methodology under various demand scenarios and urban landscapes.

Finally, the two-echelon model, having a cost minimization objective function combined with a restriction on the minimum volume to be served to nanostores, does not consider any revenue originated from this delivery activity. The key characteristic of the problems with integrated profits is that the set of customers to serve is not given. Therefore, two different decisions have to be simultaneously taken: which customers to serve and how to group them into one or several routes. In general, a profit is associated with each client that makes such customer more or less attractive. Thus, any route or set of routes can be measured both in terms of length (i.e., either time or distance) and profit. The two measures may either be combined in a single objective function, or one of them may be bounded in a constraint (Archetti et al. 2014). These problems, in which it is impossible to serve all customers, are common in the service industry, for instance, in maintenance and repair activities, and in disaster logistics.

In conclusion, OR models can play a role and provide their positive contribution to this challenge by enabling an optimized use of a network's available assets and infrastructure. However, when assessing the implementation of these models, it is crucial to consider the various objectives of all stakeholders so as to avoid choosing shortsighted solutions.

Bibliography

Adulyasak, Y., Cordeau, J.-F., and Jans, R. (2015). Benders Decomposition for Production Routing Under Demand Uncertainty. *Operations Research*, 63(4):851–867.

Agarwal, Y., Mathur, K., and Salkin, H. M. (1989). A set-partitioning-based exact algorithm for the vehicle routing problem. *Networks*, 19(7):731–749.

Alvarenga, G., Mateus, G., and de Tomi, G. (2007). A genetic and set partitioning two-phase approach for the vehicle routing problem with time windows. *Computers & Operations Research*, 34(6):1561–1584.

Ambrosino, D., Sciomachen, A., and Scutellà, M. G. (2009). A heuristic based on multi-exchange techniques for a regional fleet assignment location-routing problem. *Computers and Operations Research*, 36(2):442–460.

Ambrosino, D. and Scutellà, M. G. (2005). Distribution network design: New problems and related models. *European Journal of Operational Research*, 165(3):610–624.

Archetti, C., Speranza, M., and Vigo, D. (2014). Vehicle Routing Problems with Profits. *Working Paper Department of Economics and Management - University of Brescia*, pages 273–298.

Azarmand, Z. and Jami, E. N. (2009). Facility Location. In *Facility Location: Concepts, Models, Algorithms and Case Studies*, number 2001, chapter 6, pages 93–109.

Azi, N., Gendreau, M., and Potvin, J.-Y. (2010). An exact algorithm for a vehicle routing problem with time windows and multiple use of vehicles. *European Journal of Operational Research*, 202(3):756–763.

Badri, M. A. (1999). Combining the analytic hierarchy process and goal programming for global facility location-allocation problem. *International Journal of Production Economics*, 62(3):237–248.

Baldi, M. M., Ghirardi, M., Perboli, G., and Tadei, R. (2012). The Capacitated Transshipment Location Problem Under Uncertainty: A Computational Study. *Procedia - Social and Behavioral Sciences*, 39:424–436.

Balinski, M. L. and Quandt, R. E. (1964). On an Integer Program for a Delivery Problem. *Operations Research*, 12(2):300–304.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest math through many points. In *Proceedings of the Cambridge Philosophical Society*, number 55, pages 299–327.

Behnamian, J. (2014). Decomposition based hybrid VNS-TS algorithm for distributed parallel factories scheduling with virtual corporation. *Computers and Operations Research*, 52:181–191.

Bektas, T., Crainic, T., and van Woensel, T. (2015). From managing urban freight networks to smart city logistics networks. Technical report, CIRRELT.

Belfiore, P. and Yoshida Yoshizaki, H. T. (2009). Scatter search for a real-life heterogeneous fleet vehicle routing problem with time windows and split deliveries in Brazil. *European Journal of Operational Research*, 199(3):750–758.

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252.

Bertsekas, D. P., Nedic, A., and Ozdaglar, A. E. (2003). *Convex Analysis and Optimization.*

Bertsimas, D. and Tsitsiklis, J. N. (1997). *Introduction to Linear Optimization.*

Blanton, J. L. and Wainright, R. L. (1993). Multiple Vehicle Routing with Time and Capacity Constraints using Genetic Algorithms. In *Proceedings of the 5th International Conference on Genetic Algorithms*, pages 452–459.

Boccia, M., Crainic, T., Sforza, A., and Sterle, C. (2011). Location-Routing Models for Two-Echelon Freight Distribution System Design Location-Routing Models for Two-Echelon Freight. *CIRRELT Working Paper*, page 28.

Boccia, M., Crainic, T. G., Sforza, A., and Sterle, C. (2010). A Metaheuristic for a Two Echelon Location-Routing Problem. In Festa, P., editor, *Experimental Algorithms*, pages 288–301, Berlin, Heidelberg. Springer Berlin Heidelberg.

Bodin, L. and Golden, B. (1981). Classification in vehicle routing and scheduling. In *Proceedings of the International Workshop on Current and Future Direction in the Routing and Scheduling of Vehicles and Crews*, volume 11, pages 97–108.

Boland, N., Fischetti, M., Monaci, M., and Savelsbergh, M. (2016). Proximity Benders: a decomposition heuristic for stochastic programs. *Journal of Heuristics*, 22(2):181–198.

Boulaksil, Y. (2012). The operational performance of small retailers in developing countries – the case of Morocco. In *Proceedings of the 4th P&OM World Conference.*

Boulaksil, Y. and Belkora, M. J. (2017). Distribution Strategies Toward Nanostores in Emerging Markets: The Valencia Case. *Interfaces*, 47(6):505–517.

Bowerman, R. L., Calamai, P. H., and Brent Hall, G. (1994). The spacefilling curve with optimal partitioning heuristic for the vehicle routing problem. *European Journal of Operational Research*, 76(1):128–142.

Brimberg, J., Hansen, P., Mladenovic, N., and Taillard, E. (2000). Improvements and Comparison of Heuristics for Solving the Uncapacitated Multisource Weber Problem. *Operations Research*, 48(3):444–460.

Brimberg, J. and Mladenovic, N. (1996). Solving the continuous Location-Allocation problem with tabu search. Technical report.

Canto, S. P. (2008). Application of Benders' decomposition to power plant preventive maintenance scheduling. *European Journal of Operational Research*, 184(2):759–777.

Chaisiri, S., Lee, B., and Niyato, D. (2011). Optimization of Resource Provisioning Cost in Cloud Computing. *IEEE Transactions on Services Computing*, 5:164–177.

Christofides, N. and Beasley, J. E. (1984). The period routing problem. *Networks*, 14(2):237–256.

Christofides, N. and Eilon, S. (1969). Expected Distances in Distribution Problems. *Journal of the Operational Research Society*, 20(4):437–443.

Clarke, G. and Wright, J. W. (1964). Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. *Operations Research*, 12(4):568–581.

Contardo, C., Hemmelmayr, V., and Crainic, T. G. (2012). Lower and upper bounds for the two-echelon capacitated location-routing problem. *Computers & Operations Research*, 39(12):3185–3199.

Cooper, L. (1963). Location-Allocation Problems. *Operations Research*, 11(3):331–343.

Cordeau, J.-F., Gendreau, M., Hertz, A., Laporte, G., and Sormany, J.-S. (2005). New Heuristics for the Vehicle Routing Problem. In Langevin, A. and Riopel, D., editors, *Logistics Systems: Design and Optimization*, pages 279–297. Springer US, Boston, MA.

Cordeau, J. F., Pasin, F., and Solomon, M. M. (2006). An integrated model for logistics network design. *Annals of Operations Research*, 144(1):59–82.

Cordeau, J.-F., Stojković, G., Soumis, F., and Desrosiers, J. (2001). Benders Decomposition for Simultaneous Aircraft Routing and Crew Scheduling. *Transportation Science*, 35(4):375–388.

Corréa, A. I., Langevin, A., and Rousseau, L. M. (2007). Scheduling and routing of automated guided vehicles: A hybrid approach. *Computers and Operations Research*, 34(6 SPEC. ISS.):1688–1707.

Crainic, T., Sforza, A., and Sterle, A. (2011a). Tabu Search Heuristic for a Two- Echelon Location-Routing Problem.

Crainic, T. G., Mancini, S., Perboli, G., and Tadei, R. (2010). *Clustering-Based Heuristics for the Two-Echelon Vehicle Routing Problem.* Number November.

Crainic, T. G., Mancini, S., Perboli, G., and Tadei, R. (2011b). Multi-start Heuristics for the Two-Echelon Vehicle Routing Problem. In Hutchison, D. and Mitchell, J. C., editors, *Evolutionary Computation in Combinatorial Optimization*, pages 179–190.

Crainic, T. G., Ricciardi, N., and Storchi, G. (2004). Advanced freight transportation systems for congested urban areas. *Transportation Research Part C: Emerging Technologies*, 12(2):119–137.

Dantzig, G. B. and Ramser, J. H. (1959). The Truck Dispatching Problem. *Management Science*, 6(1):80–91.

De Backer, B. and Furnon, V. (1997). Meta-heuristics in Constraint Programming Experiments with Tabu Search on the Vehicle Routing Problem. In *Proceedings of the 2nd International Conference on Meta-Heuristics*, Sophia Antipolis, France.

Dekker, R., Bloemhof, J., and Mallidis, I. (2012). Operations Research for green logistics - An overview of aspects, issues, contributions and challenges. *European Journal of Operational Research*, 219(3):671–679.

Desaulniers, G., Madsen, O. B. G., and Ropke, S. (2014). The vehicle routing problem with time windows. In *Vehicle routing: Problems, Methods, And Applications*, pages 119–159.

Dondo, R. and Cerdá, J. (2007). A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows. *European Journal of Operational Research*, 176(3):1478–1507.

Drexl, M. and Schneider, M. (2015). A survey of variants and extensions of the location-routing problem.

Ehmke, J. (2012). *Integration of information and optimization models for routing in city logistics.* Springer Science & Business Media.

Eksioglu, B., Vural, A. V., and Reisman, A. (2009). The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering*, 57(4):1472–1483.

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

Figliozzi, M. (2008). Planning Approximations to the Average Length of Vehicle Routing Problems with Varying Customer Demands and Routing Constraints. *Transportation Research Record: Journal of the Transportation Research Board*, 2089:1–8.

Figliozzi, M. A. (2009). Planning approximations to the average length of vehicle routing problems with time window constraints. *Transportation Research Part B: Methodological*, 43(4):438–447.

Fisher, M. (1995). Vehicle routing. *Handbooks in Operations Research and Management Science*, 8(C):1–33.

Fisher, M. L. and Jaikumar, R. (1981). A generalized assignment heuristic for vehicle routing. *Networks*, 11(2):109–124.

Fortz, B. and Poss, M. (2009). An improved Benders decomposition applied to a multi-layer network design problem. *Operations Research Letters*, 37(5):359–364.

Fransoo, J. and Blanco, E. (2013). Reaching 50 million nanostores: Retail distribution in emerging megacities.

Fransoo, J., Blanco, E., and Mejía Argueta, C. (2017). *Reaching 50 million nanostores: retail distribution in emerging megacities.* CreateSpace Independent Publishing Platform, Cambridge.

Garcia-Najera, A. and Bullinaria, J. A. (2011). An improved multi-objective evolutionary algorithm for the vehicle routing problem with time windows. *Computers & Operations Research*, 38(1):287–300.

Ge, J. (2017). *Traditional retail distribution in megacities.* PhD thesis, Technische Universiteit Eindhoven.

Gelareh, S., Neamatian Monemi, R., and Nickel, S. (2015). Multi-period hub location problems in transportation. *Transportation Research Part E: Logistics and Transportation Review*, 75:67–94.

Gendreau, M., Guertin, F., Potvin, J.-Y., and Séguin, R. (2006). Neighborhood search heuristics for a dynamic vehicle dispatching problem with pick-ups and deliveries. *Transportation Research Part C: Emerging Technologies*, 14(3):157–174.

Gendreau, M., Laporte, G., and Potvin, J.-Y. (2002). Metaheuristics for the Capacitated VRP. In *The Vehicle Routing Problem*, chapter 6, pages 129–154. SIAM, Philadelphia.

Geoffrion, A. M. (1970a). Elements of Large-Scale Mathematical Programming Part I: Concepts. *Management Science*, 16(11):652–675.

Geoffrion, A. M. (1970b). Elements of Large Scale Mathematical Programming Part II: Synthesis of Algorithms and Bibliography. *Management Science*, 16(11):676–691.

Geoffrion, A. M. (1972). Generalized Benders Decomposition. *Journal of Optimization Theory and Applications*, 10(4):238–260.

Gevaers, R., Van de Voorde, E., and Vanelslander, T. (2011). Characteristics and typology of last-mile logistics from an innovation perspective in an urban context. *City Distribution and Urban Freight Transport: Multiple Perspectives, Edward Elgar Publishing*, 56-71.

Ghaffari-Nasab, N., Jabalameli, M. S., Aryanezhad, M. B., and Makui, A. (2013). Modeling and solving the bi-objective capacitated location-routing problem with probabilistic travel times. *International Journal of Advanced Manufacturing Technology*, 67(9-12):2007–2019.

Gong, D., Gen, M., Yamazaki, G., and Xu, W. (1997). Hybrid evolutionary method for capacitated location-allocation problem. *Computers & Industrial Engineering*, 33(3-4):577–580.

Gonzalez-Feliu, J. (2012). Cross-Docking : N-Echelon Location Routing Problem. *Promet - Traffic&Transportation*, 24(2):143–149.

Goodman, R. W. (2005). Whatever You Call It, Just Don't Think of Last-mile Logistics, Last. *Global Logistics and Supply Chain Strategies*, (December):1–6.

Greene, D. and Plotkin, S. (2011). Reducing greenhouse gas emission from US transportation. *Arlington: Pew Center on Global Climate Change.*

Hakimi, S. L. (1965). Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems. *Operations Research*, 13(3):462–475.

Hansen, P., Mladenovic, N., and Taillard, E. (1998). Heuristic solution of the multisource Weber problem as a p-median problem. *Operations Research Letters*, 22(2-3):55–62.

Jiang, W., Tang, L., and Xue, S. (2009). A hybrid algorithm of tabu search and benders decomposition for multi-product production distribution network design. *2009 IEEE International Conference on Automation and Logistics*, (70728001):79–84.

Klibi, W., Lasalle, F., Martel, A., and Ichoua, S. (2010). The Stochastic Multiperiod Location Transportation Problem. *Transportation Science*, 44(2):221–237.

Klotz, E., Newman, A. M., and Tyson, M. (2013). Practical Guidelines for Solving Difficult Mixed Integer Linear Programs. *Surveys in Operations Research and Management Science*, 18(1-2):18–32.

Koch, T., Ralphs, T., and Shinano, Y. (2012). Could we use a million cores to solve an integer program? *Mathematical Methods of Operations Research*, 76(1):67–93.

Korte, B. and Vygen, J. (2018). Bin-Packing. In *Combinatorial Optimization: Theory and Algorithms*, pages 489–507. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kuenne, R. E. and Soland, R. M. (1972). Exact and approximate solutions to the multisource weber problem. *Mathematical Programming*, 3(1):193–209.

Kumar, S. N. and Panneerselvam, R. (2012). A Survey on the Vehicle Routing Problem and Its Variants. *Intelligent Information Management*, 04(03):66–74.

Laporte, G. (1992). The vehicle routing problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, 59(3):345–358.

Laporte, G., Nobert, Y., and Taillefer, S. (1988). Solving a Family of Multi-Depot Vehicle Routing and Location-Routing Problems. *Transportation Science*, 22(3):161–172.

Larson, R. C. and Odoni, A. R. (2007). *Urban Operations Research.* Dynamic Ideas, Belmont, Massachusetts, 2nd edition.

Lin, J. R. and Lei, H. C. (2009). Distribution systems design with two-level routing considerations. *Annals of Operations Research*, 172(1):329–347.

Martello, S. and Toth, P. (1990). Bin-Packing problem.

Merchan, D., Blanco, E. E., and Winkenbach, M. (2015). Transshipment Networks for Last-Mile Delivery in Congested Urban Areas. In *Information Systems, Logistics and Supply Chain: Sixth International Conference*, number June, page 61.

Merchan, D. and Winkenbach, M. (2017). High-Resolution Last-Mile Network Design. In Taniguchi, E. and Thompson, R. G., editors, *City Logistics 3: Towards Sustainable and Liveable Cities*, pages 201–214, Phuket, Thailand. ISTE - John Wiley and Sons.

Montemanni, R., Gambardella, L. M., Rizzoli, A. E., and Donati, A. V. (2005). Ant Colony System for a Dynamic Vehicle Routing Problem. *Journal of Combinatorial Optimization*, 10(4):327–343.

Montoya-Torres, J. R., López Franco, J., Nieto Isaza, S., Felizzola Jiménez, H., and Herazo-Padilla, N. (2015). A literature review on the vehicle routing problem with multiple depots. *Computers & Industrial Engineering*, 79:115–129.

Mourgaya, M. and Vanderbeck, F. (2006). Problème de Tournées de Véhicules Multipériodiques: Classification et Heuristique pour la Planification Tactique. *RAIRO Operations Research*, (40):169–194.

Murphy, J. (2013). Benders, Nested Benders and Stochastic Programming: An Intuitive Introduction. Technical Report December, Cambridge University Engineering Department, Cambridge.

Murray, A. T. and Church, R. L. (1996). Applying simulated annealing to location-planning models. *Journal of Heuristics*, 2(1):31–53.

Nagy, G. and Salhi, S. (2007). Location-routing: Issues, models and methods. *European Journal of Operational Research*, 177(2):649–672.

Nemhauser, G. L. and Wolsey, L. A. (1999). *Integer and Combinatorial Optimization.*

Ohlemüller, M. (1997). Tabu search for large location–allocation problems. *Journal of the Operational Research Society*, 48(7):745–750.

Perl, J. and Daskin, M. S. (1985). A warehouse location-routing problem. *Transportation Research Part B*, 19(5):381–396.

Pishvaee, M. S., Razmi, J., and Torabi, S. A. (2014). An accelerated Benders decomposition algorithm for sustainable supply chain network design under uncertainty: A case study of medical needle and syringe supply chain. *Transportation Research Part E: Logistics and Transportation Review*, 67:14–38.

Potvin, J.-Y., Kervahut, T., Garcia, B.-L., and Rousseau, J.-M. (1996). The Vehicle Routing Problem with Time Windows Part I: Tabu Search. *INFORMS Journal on Computing*, 8(2):158–164.

Prodhon, C. and Prins, C. (2014). A survey of recent research on location-routing problems. *European Journal of Operational Research*, 238(1):1–17.

Qureshi, A. G., Taniguchi, E., and Yamada, T. (2009). An exact solution approach for vehicle routing and scheduling problems with soft time windows. *Transportation Research Part E: Logistics and Transportation Review*, 45(6):960–977.

Rahmaniani, R., Crainic, T. G., Gendreau, M., and Rei, W. (2016). The Benders Decomposition Algorithm: A Literature Review. *European Journal of Operational Research*, 259(3):801 – 817.

Rodrigue, J.-P. and Notteboom, T. (2017). Transportation, Economy and Society. In Rodrigue, J.-P., Comtois, C., and Slack, B., editors, *The Geography of Transport Systems*, chapter 7, page 440. Routledge, New York, fourth edition.

Ropke, S., Cordeau, J. F., Iori, M., and Vigo, D. (2007). Branch-and-cut-and-price for the capacitated vehicle routing problem with two-dimensional loading constraints. In *Proceedings of ROUTE*, Jekyll Island.

Salhi, S. and Gamal, M. D. H. (2003). A Genetic Algorithm Based Approach for the Uncapacitated Continuous Location–Allocation Problem. *Annals of Operations Research*, 123(1):203–222.

Salhi, S. and Rand, G. K. (1989). The effect of ignoring routes when locating depots. *European Journal of Operational Research*, 39(2):150–156.

Salhi, S. and Sari, M. (1997). A multi-level composite heuristic for the multi-depot vehicle fleet mix problem. *European Journal of Operational Research*, 103(1):95–112.

Salimifard, K. and Shahbandarzadeh, H & Raeesi, R. (2012). Green Transportation and the Role of Operation Research. *Proceedings of 2012 International Conference on Traffic and Transportation Engineering*, 26(May 2014):1–6.

Scaparra, M. P. and Scutellà, M. G. (2001). Facilities, Locations, Customers: Building Blocks of Location Models. A Survey. Technical report, Università di Pisa, Pisa, Italy.

Snoeck, A. and Winkenbach, M. (2018). The Value of Flexibility in Urban Distribution Networks Under Demand Uncertainty. *CTL Working Paper Series*, pages 1–34.

Snoeck, A., Winkenbach, M., and Mascarino, E. E. (2017). Establishing a Robust Urban Logistics Network at FEMSA through Stochastic Multi-Echelon Location Routing. In Taniguchi, E. and Thompson, R. G., editors, *City Logistics 2: Modeling and Planning Initiatives*, pages 59–78, Phuket, Thailand.

Srivastava, R. (1993). Alternate Solution Procedures For The Location Routing Problem. *Omega-International Journal Of Management Science*, 21(4):497–506.

Taniguchi, E. (2014). Concepts of city logistics for sustainable and liveable cities. *Procedia-Social and Behavioral Sciences*, 151:310–317.

Taniguchi, E. and Thompson, R. G. (2014). *City Logistics: Mapping The Future.* CRC Press.

Taşkin, Z. C. (2011). Benders' Decomposition. *Wiley Encyclopedia of Operations Research and Management Science.*

Toth, P. and Vigo, D. (2014). *Vehicle Routing: Problem, Methods, and Applications.* Society for Industrial and Applied Mathematics, second edi edition.

United Nations Population Division (2014). World Urbanization Prospects, the 2014 Revision. Technical report, United Nations.

Vanderbeck, F. and Wolsey, L. A. (2009). Reformulation and decomposition of integer programs. Technical report, CORE.

Winkenbach, M., Roset, A., and Spinler, S. (2016). Strategic Redesign of Urban Mail and Parcel Networks at La Poste. *Interfaces*, 46(5).

Wu, T. H., Low, C., and Bai, J. W. (2002). Heuristic solutions to multi-depot location-routing problems. *Computers and Operations Research*, 29(10):1393–1415.

Zarandi, M. H. F., Hemmati, A., and Davari, S. (2011). The multi-depot capacitated location-routing problem with fuzzy travel times. *Expert Systems with Applications*, 38(8):10075–10084.

Zarandi, M. H. F., Hemmati, A., Davari, S., and Burhan Turksen, I. (2013). Capacitated location-routing problem with time windows under uncertainty. *Knowledge-Based Systems*, 37:480–489.

Appendix A

Benders Decomposition

The original version of the BD algorithm along with its iterative relaxed implementation, is broadly described in this appendix. Both explanation and formulation are taken from Rahmaniani et al. (2016).

This decomposition methodology is based on a sequence of projections, outer linearizations, and relaxations (Geoffrion 1970a,b). The original model is first projected onto the subspace defined by the set of integer variables, also referred to as 'complicating variables'. Here, binary variables can be regarded as a special case of integer variables. After this projection, the dual model of the resulting formulation is generated. The extreme rays and points of this dual model respectively define the feasibility requirements (feasibility cuts) and the projected costs (optimality cuts) of those 'complicating variables' (see Bertsekas et al. (2003), and Bertsimas and Tsitsiklis (1997) for further details about extreme points and rays). Thus, the dual formulation can be built by enumerating all the extreme points and rays. However, this brute force enumeration is generally computationally intensive. Hence, one solves the equivalent model by applying a relaxation strategy to the feasibility and optimality cuts, yielding a MP and a SP. These two problems are iteratively solved in order to guide the search process and generate the violated cuts (Rahmaniani et al. 2016).

BD starts by considering a MILP of the form

$$\min \mathbf{f}^T\mathbf{y} + \mathbf{c}^T\mathbf{x} \tag{A.1}$$

subject to

$$\mathbf{A}\mathbf{y} = \mathbf{b}, \tag{A.2}$$

$$\mathbf{B}\mathbf{y} + \mathbf{D}\mathbf{x} = \mathbf{d}, \tag{A.3}$$

$$\mathbf{x} \geq \mathbf{0}, \tag{A.4}$$

$$\mathbf{y} \in \mathbb{Z}_0^{+n_1}. \tag{A.5}$$

there, the 'complicating variables' $\mathbf{y} \in \mathbb{Z}_0^{+n_1}$ must satisfy constraints (A.2), where $\mathbf{A} \in \mathbb{R}^{m_1 \times n_1}$ is a known matrix and $\mathbf{b} \in \mathbb{R}^{m_1}$ is a given vector. The continuous variables $\mathbf{x} \in \mathbb{R}^{n_2}$, together with $\mathbf{y}$, must satisfy the set of linking constraints (A.3), with $\mathbf{B} \in \mathbb{R}^{m_2 \times n_1}$, $\mathbf{D} \in \mathbb{R}^{m_2 \times n_2}$, and $\mathbf{d} \in \mathbb{R}^{m_2}$. The objective function in Equation (A.1) minimizes total cost with the cost vectors $\mathbf{f} \in \mathbb{R}^{n_1}$ and $\mathbf{c} \in \mathbb{R}^{n_2}$.

The model given by Equations (A.1) through (A.5) can be reformulated as

$$\min_{\bar{\mathbf{y}} \in \mathcal{Y}} \left\{ \mathbf{f}^T \bar{\mathbf{y}} + \min_{\mathbf{x} \geq \mathbf{0}} \left\{ \mathbf{c}^T \mathbf{x} : \mathbf{D}\mathbf{x} = \mathbf{d} - \mathbf{B}\bar{\mathbf{y}} \right\} \right\}, \tag{A.6}$$

where $\bar{\mathbf{y}}$ is a given value for the integer variables, which belongs to the set $\mathcal{Y} = \{\mathbf{y} | \mathbf{A}\mathbf{y} = \mathbf{b}, \mathbf{y} \in \mathbb{Z}_0^{+n_1}\}$. The inner minimization is a continuous linear problem that can be dualized by means of dual variables $\boldsymbol{\pi}$ associated with the set of constraints $\mathbf{D}\mathbf{x} = \mathbf{d} - \mathbf{B}\bar{\mathbf{y}}$:

$$\min_{\boldsymbol{\pi} \in \mathbb{R}^{m_2}} \left\{ \boldsymbol{\pi}^T (\mathbf{d} - \mathbf{B}\bar{\mathbf{y}}) : \boldsymbol{\pi}^T \mathbf{D} \leq \mathbf{c} \right\} \tag{A.7}$$

Based on duality theory, the primal and dual formulations can be interchanged to obtain the following equivalent formulation:

$$\min_{\bar{\mathbf{y}} \in \mathcal{Y}} \left\{ \mathbf{f}^T \bar{\mathbf{y}} + \max_{\boldsymbol{\pi} \in \mathbb{R}^{m_2}} \left\{ \boldsymbol{\pi}^T (\mathbf{d} - \mathbf{B}\bar{\mathbf{y}}) : \boldsymbol{\pi}^T \mathbf{D} \leq \mathbf{c} \right\} \right\} \tag{A.8}$$

The feasible space of the inner maximization, i.e., $\mathcal{U} = \{\boldsymbol{\pi} | \boldsymbol{\pi}^T \mathbf{D} \leq \mathbf{c}\}$, is independent of the choice of $\bar{\mathbf{y}}$. Thus, if $\mathcal{U}$ is not empty, the inner problem can be either unbounded or feasible for any arbitrary choice of $\bar{\mathbf{y}}$. In the former case, given the set of extreme rays $\mathcal{L}$ of the feasible space $\mathcal{U}$, there is a direction of unboundedness $\mathbf{r}_l$, with $l \in \mathcal{L}$, for which $\mathbf{r}_l^T (\mathbf{d} - \mathbf{B}\bar{\mathbf{y}}) > 0$; this must be avoided because it indicates the infeasibility of the solution $\bar{\mathbf{y}}$. We then add a cut

$$\mathbf{r}_l^T(\mathbf{d} - \mathbf{B}\bar{\mathbf{y}}) \leq 0, \qquad \forall l \in \mathcal{L}, \tag{A.9}$$

to the problem to restrict movement in this direction. In the latter case, the solution of the inner maximization is one of the extreme points $\boldsymbol{\pi}_e$, $e \in \mathcal{E}$, where $\mathcal{E}$ is the set of extreme points of the feasible space $\mathcal{U}$. If we add all the cuts of the form presented by Equation (A.9) to the outer minimization problem, the value of the inner problem will be one of its extreme points. Consequently, the problem given by Equation (A.8) can be reformulated as

$$\min_{\bar{\mathbf{y}} \in \mathcal{Y}} \mathbf{f}^T \bar{\mathbf{y}} + \max_{e \in E} \left\{ \boldsymbol{\pi}_e^T (\mathbf{d} - \mathbf{B}\bar{\mathbf{y}}) \right\} \tag{A.10}$$

subject to

$$\mathbf{r}_l^T(\mathbf{d} - \mathbf{B}\bar{\mathbf{y}}) \leq 0, \qquad \forall l \in \mathcal{L}. \tag{A.11}$$

This problem can easily be linearized via a continuous variable $\eta \in \mathbb{R}^1$ to give the following equivalent formulation to the initial problem, which is referred to as the Benders MP:

$$\min_{\mathbf{y},\eta} \mathbf{f}^T \mathbf{y} + \eta \tag{A.12}$$

subject to

$$\mathbf{A}\mathbf{y} = \mathbf{b}, \tag{A.13}$$

$$\eta \geq \boldsymbol{\pi}_e^T(\mathbf{d} - \mathbf{B}\mathbf{y}), \qquad \forall e \in \mathcal{E}, \tag{A.14}$$

$$0 \geq \mathbf{r}_l^T(\mathbf{d} - \mathbf{B}\mathbf{y}), \qquad \forall l \in \mathcal{L}, \tag{A.15}$$

$$\mathbf{y} \in \mathbb{Z}_0^{+n_1}. \tag{A.16}$$

Constraints (A.14) and (A.15) are referred to as optimality and feasibility cuts, respectively. As mentioned before, the complete enumeration of these cuts is generally not practical. Therefore, Benders (1962) proposed a relaxation of the feasibility and optimality cuts blended with an iterative approach. Thus, the algorithm repeatedly solves the MP, which includes only a subset of constraints (A.14) and (A.15), to obtain a trial value, $\bar{\mathbf{y}}$, for the variables in $\mathbf{y}$. It then solves SP (A.7) with $\bar{\mathbf{y}}$. If the subproblem is feasible and bounded, a cut of type (A.14) is produced. Otherwise, if the SP is unbounded, a cut of type (A.15) is generated. If the cuts are violated by the current solution, they are inserted into the current MP and the whole process starts over again.

To confirm the convergence of the obtained solution, the optimality gap can be calculated at each iteration. The objective function of the MP gives a valid lower bound on the optimal cost because it is a relaxation of the equivalent Benders reformulation. On the other hand, if solution $\bar{\mathbf{y}}$ yields a feasible SP, then the sum of both $\mathbf{f}^T\bar{\mathbf{y}}$ and the objective value associated to the subproblem provides a valid upper bound for the original problem. The described iterative approach is schematized in Figure A.1.

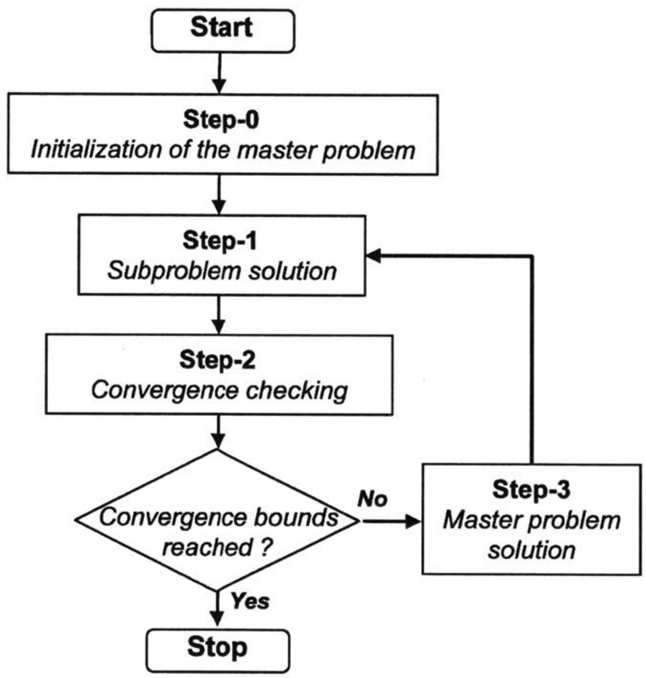


Figure A.1: Flow chart of BD's iterative approach
Source: Chaisiri et al. (2011)

For an extensive review and a thorough explanation of the BD methodology, see, e.g., Geoffrion (1972), Nemhauser and Wolsey (1999), and Bertsimas and Tsitsiklis (1997).

Appendix B

Single-echelon Distribution Network – Model Inputs and Additional Results

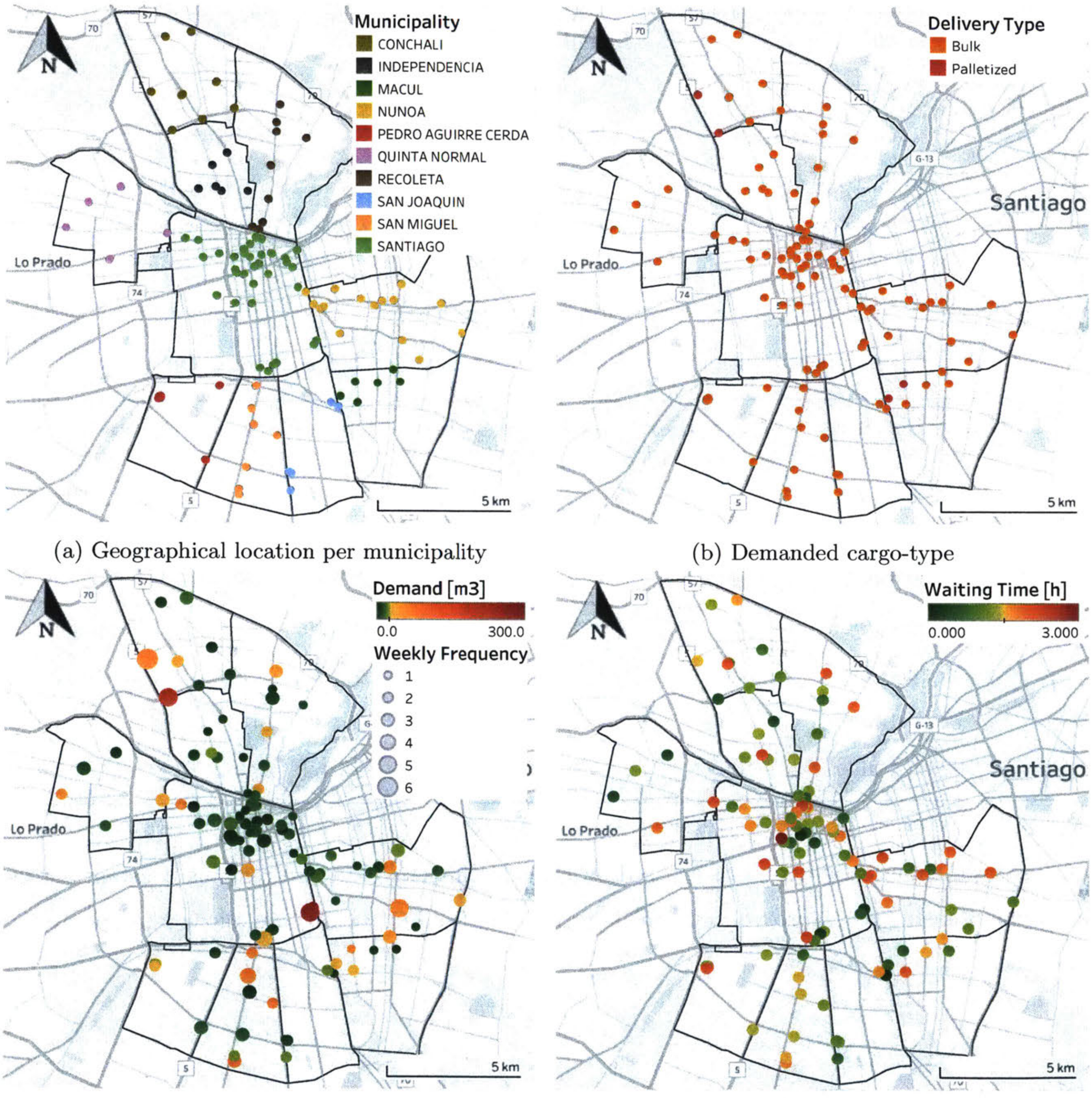
In this appendix, we present the data and parameters for those problem instances analyzed in the single-echelon distribution network scenario (i.e., to serve only big-box customers). For clarity purposes, we will group them by the specific model component with which they are associated.

B.1 Customers

Tables B.1 and B.2 summarize the general information regarding big-box customers for each problem instance. Besides, Figure B.1(a) depicts the geographical distribution per municipality of all 109 customers that were considered in this work. Moreover, Figure B.1(b) classifies each customer according to the cargo-type that it requires. Furthermore, Figure B.1(c) presents a heatmap showing the demand distribution (in [m^3/week]) per customer as well as their associated weekly delivery frequencies. Finally, Figure B.1(d) portrays the average waiting time (in [h]) for each customer. For simplicity, we do not present any information related to the feasible delivery days as well as vehicle-specific access restrictions per customer.

B.2 Facilities

In these problem instances, we have only considered the DCs of Puente Alto (i.e., 1101) and Talagante (i.e., 1102) as our delivery facilities. Table B.3 summarizes the relevant parameters associated with them on each problem instance. It is important to highlight that the transfer cost per volume,



(a) Geographical location per municipality

(b) Demanded cargo-type

(c) Weekly demand (in $[m^3]$) and delivery frequency

(d) Average waiting time (in [h])

Figure B.1: General characteristics of the selected subset of customers

Table B.1: Customer demand information for the single-echelon distribution network

Problem instance	**Municipalities**	**Weekly volume** [m^3]
11 customers	Macul, Quinta Normal	257.8
20 customers	+ Conchalí	624.3
30 customers	+ Pedro Aguirre Cerdá	760.3
50 customers	+ Ñuñoa, San Joaquín	1072.0
72 customers	+ Independencia, Recoleta, San Miguel	1573.0
109 customers	+ Santiago	2469.4

Table B.2: Customer additional information for the single-echelon distribution network

Problem Instance	**Customers per Cargo Type**		**Customers per Frequency**					
	Bulk	Palletized	1	2	3	4	5	6
11 customers	9	2	3	5	3	-	-	-
20 customers	16	4	3	11	4	-	1	1
30 customers	26	4	7	16	5	-	1	1
50 customers	46	4	12	26	9	-	2	1
72 customers	68	4	19	36	14	1	1	1
109 customers	105	4	28	52	23	2	3	1

c_f, has been empirically determined based on available historic data and remains constant across all problem instances. Moreover, all vehicle types (i.e., from C1 to C6) are allowed to operate at both facilities.

B.3 Vehicle Fleet

We now present the values associated with the vehicle-specific parameters. Table B.4 summarizes the vehicle fleet availability for each combination of problem instance and vehicle type. Table B.5 details the capacity and cargo-types allowed for each vehicle type. Finally, Table B.6 condenses the vehicle dependent time parameters.

Table B.3: Facility information for the single-echelon distribution network

Problem Instance	Daily Capacity [m^3]		Transfer Cost [\$/m^3]	
	1101	1102	1101	1102
11 customers	40	60	0.88	0.30
20 customers	50	100	0.88	0.30
30 customers	65	130	0.88	0.30
50 customers	100	200	0.88	0.30
72 customers	125	250	0.88	0.30
109 customers	150	300	0.88	0.30

Table B.4: Vehicle type availability per problem instance

Problem instance	Vehicle Type Availability					
	C1	C2	C3	C4	C5	C6
11 customers	2	2	2	2	2	2
20 customers	3	3	3	3	3	3
30 customers	4	4	4	4	4	4
50 customers	6	6	6	6	6	6
72 customers	7	7	7	7	7	7
109 customers	12	12	12	12	12	12

Table B.5: Vehicle capacity (in [m^3]) and allowed cargo-types per vehicle type

Vehicle type	Payload [m^3]		Allowed Cargo Types	
	Bulk	Palletized	Bulk	Palletized
C1	5.8	-	✓	✗
C2	9.3	-	✓	✗
C3	13.3	-	✓	✗
C4	26.8	-	✓	✗
C5	39.5	26.9	✓	✓
C6	103.0	70.0	✓	✓

Table B.6: Vehicle-specific time parameters (in [h] and [h/m^3])

Vehicle Type	**Setup [h]**				**Load [h/m^3]**				**Unload [h/m^3]**	
	Bulk		Palletized		Bulk		Bulk		Bulk	Palletized
	1101	1102	1101	1102	1101	1102	1101	1102		
C1	2.2	1.6	-	-	0.04	0.03	-	-	0.05	-
C2	2.4	1.7	-	-	0.03	0.03	-	-	0.05	-
C3	2.4	1.7	-	-	0.02	0.02	-	-	0.05	-
C4	2.2	1.7	-	-	0.02	0.02	-	-	0.06	-
C5	2.1	1.6	2.0	1.2	0.02	0.02	0.01	0.01	0.06	0.02
C6	1.5	1.9	2.0	1.2	0.01	0.01	0.01	0.01	0.05	0.02

B.4 General Parameters and Vehicle Fares

For all six problem instances we considered a maximum service time, T_{max}, of 10 hours per vehicle trip. Moreover, the associated line-haul travel times, r_{fm}/V_k^{lh}, and delivery travel times, r_{fm}/V_k^{lh}, were calculated from a massive query to the Google Distance Matrix® service.

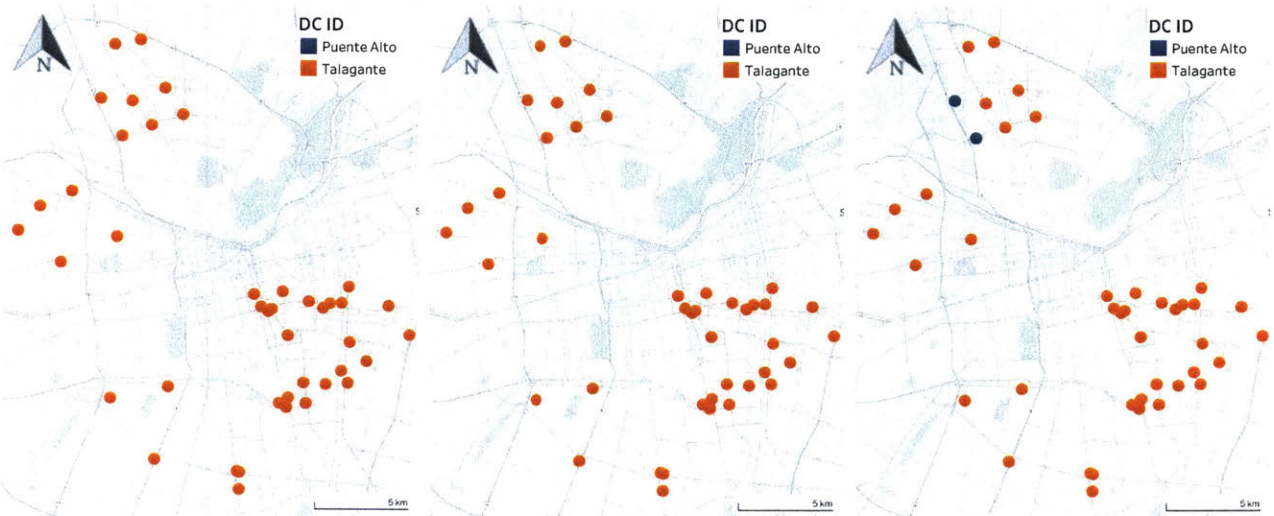
Finally, table B.7 summarizes the associated fares (applicable to all problem instances) based on the facility of origin, the vehicle type, and the municipality of destination.

Table B.7: Delivery fares (in [USD]) per facility, vehicle type, and municipality

Municipality of destination	**Facility of origin: 1101**						**Facility of origin: 1102**					
	C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6
Conchalí	71.11	83.65	102.47	117.64	146.20	189.60	72.38	83.65	107.46	121.58	171.97	214.40
Independencia	71.11	83.65	102.47	117.64	146.20	189.60	72.38	83.65	107.46	121.58	171.97	214.40
Quinta Normal	71.11	83.65	102.47	117.64	146.20	189.60	72.38	83.65	107.46	121.58	171.97	214.40
Recoleta	71.11	83.65	102.47	117.64	146.20	189.60	72.38	83.65	107.46	121.58	171.97	214.40
Macul	69.34	83.97	108.97	135.88	165.68	189.60	70.61	81.59	109.61	125.24	177.15	214.40
Ñuñoa	69.34	83.97	108.97	135.88	165.68	189.60	70.61	81.59	109.61	125.24	177.15	214.40
Pedro Aguirre Cerda	67.64	86.97	108.97	135.88	165.68	189.60	68.60	79.27	106.43	122.81	171.87	214.40
San Joaqín	67.64	86.97	108.97	135.88	165.68	189.60	68.60	79.27	106.43	122.81	171.87	214.40
San Miguel	67.64	86.97	108.97	135.88	165.68	189.60	68.60	79.27	106.43	122.81	171.87	214.40
Santiago	67.64	86.97	108.97	135.88	165.68	189.60	68.60	79.27	106.43	122.81	171.87	214.40

B.5 Additional Network Design Results

Figure B.2 presents the allocation of big-box customers to DCs yielded by each solution approach for the 50-customer problem instance.



(a) One-step approach (b) BD (c) Two-step approach

Figure B.2: Customer allocation to facilities for each solution approach – 50-customer instance

Appendix C

Two-echelon Distribution Network – Model Inputs and Additional Results

In this appendix, we present the additional data and parameters for those problem instances analyzed in the two-echelon distribution network scenario (i.e., to serve both big-box customers and nanostores).

C.1 Customer Data

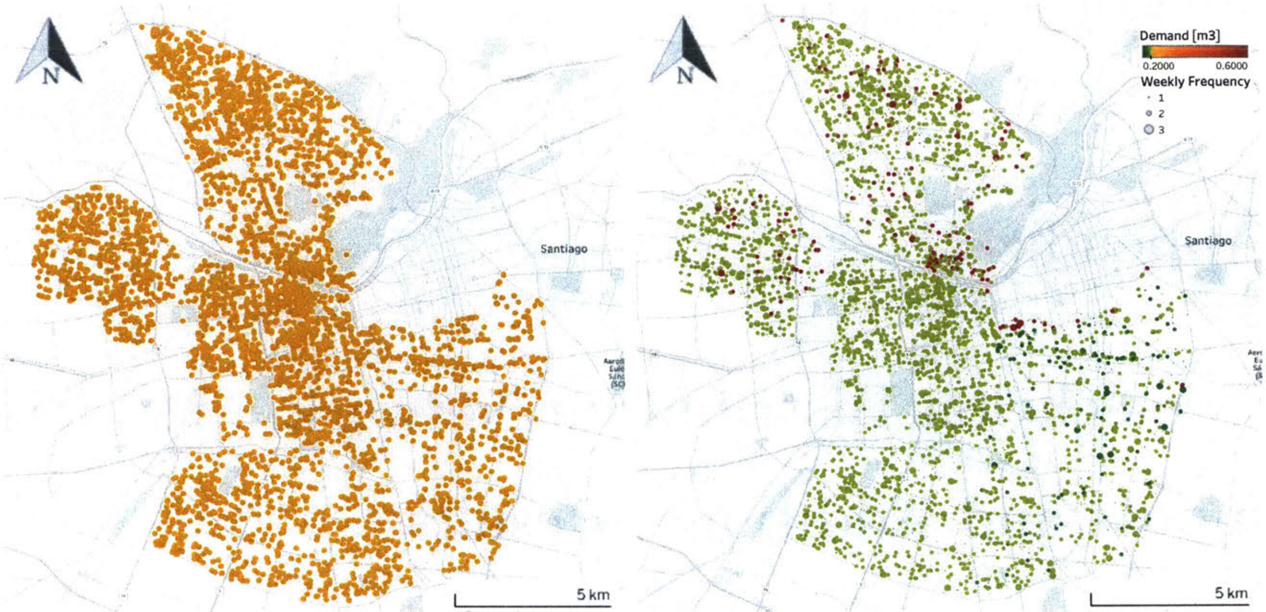
For clarity, we further divide this information across nanostores and big-box customers.

C.1.1 Nanostores

Figure C.1(a) depicts the geographical distribution of all 5,064 nanostores considered in this work. Furthermore, Figure C.1(b) presents a heatmap showing the demand distribution (in [m^3/week]) as well as the associated weekly delivery frequency per nanostore.

C.1.2 Big-box Customers

Our two-echelon network model requires the average walking circuity factor, ζ_i, specific to each big-box customer. Figure C.2 summarizes this information in a heatmap. These factors vary between 1.28 and 2.21 (with a mean of 1.42) greatly influencing our estimation for the expected walking distance to each served nanostore.



(a) Geographical location (b) Weekly demand (in [m^3]) and delivery frequency

Figure C.1: General characteristics of the selected subset of nanostores

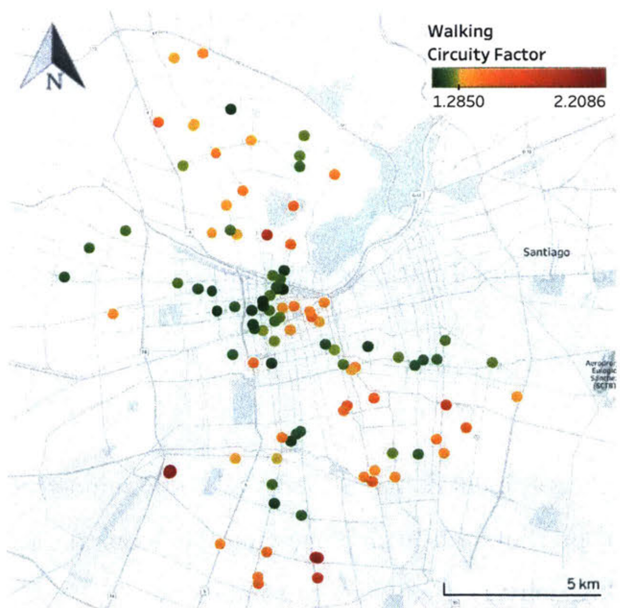


Figure C.2: Average walking circuity factor associated with each big-box customer

C.2 Additional Parameters

In addition to those parameters described in Appendix B.4, we present the following new ones. The capacity of the handcart, η^H, was assumed, based on past experience, constant and equal to 0.3 [m^3]. Moreover, the walking speed, V^H, was set at 3 [km/h] (i.e., the human average walking speed).

Furthermore, the handcart's loading and unloading variable times, $t^{l,H}$ and $t^{u,H}$, were estimated as 0.05 [h/m^3]. Finally, both the average waiting time at each nanostore, T_n^w, and the fixed set-up time for each handcart trip, $T^{l,H}$, were taken as 0.083 [h]. Regarding the cost improvement loop, we set five iterations for the outer loop (i.e., $H = 5$) and we randomly picked three eliminated TPs to reinclude them as feasible transshipment options on each outer iteration (i.e., $B = 3$).

C.3 Updated Vehicle-specific Times and Fares

To serve the nanostore channel, an additional crew member was allocated to each C1-, C2-, and C3-type vehicle. Since that additional person would also be helping with big-box customer deliveries, the variable unloading times for these vehicle types were then adjusted to account for the new crew size. Table C.1 presents the updated vehicle-type-specific time parameters.

Table C.1: Updated vehicle-type-specific time parameters (in [h] and [h/m^3])

Vehicle Type	**Setup [h]**				**Load [h/m^3]**				**Unload [h/m^3]**	
	Bulk		Palletized		Bulk		Bulk		Bulk	Palletized
	1101	1102	1101	1102	1101	1102	1101	1102		
C1	2.2	1.6	-	-	0.04	0.03	-	-	0.025	-
C2	2.4	1.7	-	-	0.03	0.03	-	-	0.025	-
C3	2.4	1.7	-	-	0.02	0.02	-	-	0.025	-
C4	2.2	1.7	-	-	0.02	0.02	-	-	0.06	-
C5	2.1	1.6	2.0	1.2	0.02	0.02	0.01	0.01	0.06	0.02
C6	1.5	1.9	2.0	1.2	0.01	0.01	0.01	0.01	0.05	0.02

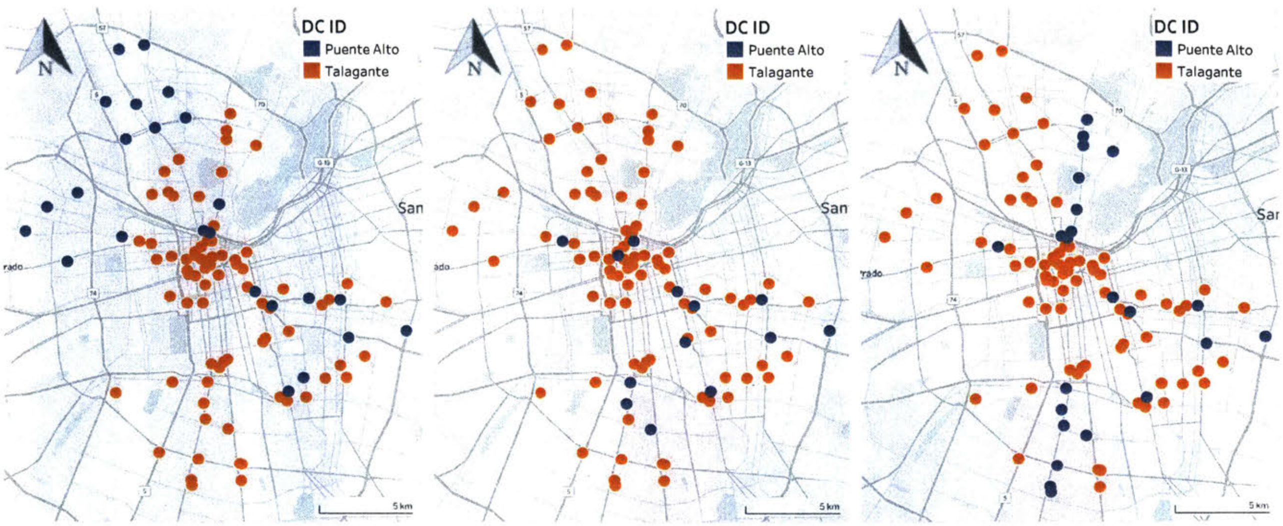
Moreover, table C.2 summarizes the vehicle updated fares considering the additional crew member required for vehicle categories C1, C2, and C3.

Table C.2: Delivery fares (in [USD]) per facility, vehicle type, and municipality

Municipality of destination	Facility of origin: 1101						Facility of origin: 1102					
	C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6
Conchalí	96.11	108.65	127.47	117.64	146.20	189.60	97.38	108.65	132.46	121.58	171.97	214.40
Independencia	96.11	108.65	127.47	117.64	146.20	189.60	97.38	108.65	132.46	121.58	171.97	214.40
Quinta Normal	96.11	108.65	127.47	117.64	146.20	189.60	97.38	108.65	132.46	121.58	171.97	214.40
Recoleta	96.11	108.65	127.47	117.64	146.20	189.60	97.38	108.65	132.46	121.58	171.97	214.40
Macul	94.34	108.97	133.97	135.88	165.68	189.60	95.61	106.59	134.61	125.24	177.15	214.40
Ñuñoa	94.34	108.97	133.97	135.88	165.68	189.60	95.61	106.59	134.61	125.24	177.15	214.40
Pedro Aguirre Cerda	92.64	111.97	133.97	135.88	165.68	189.60	93.60	104.27	131.43	122.81	171.87	214.40
San Joaqín	92.64	111.97	133.97	135.88	165.68	189.60	93.60	104.27	131.43	122.81	171.87	214.40
San Miguel	92.64	111.97	133.97	135.88	165.68	189.60	93.60	104.27	131.43	122.81	171.87	214.40
Santiago	92.64	111.97	133.97	135.88	165.68	189.60	93.60	104.27	131.43	122.81	171.87	214.40

C.4 Additional Network Design Results

Figure C.3 presents the allocation of big-box customers to DCs for each problem instance.



(a) 0% cannibalization rate (b) 10% cannibalization rate (c) Maximum cannibalization rate

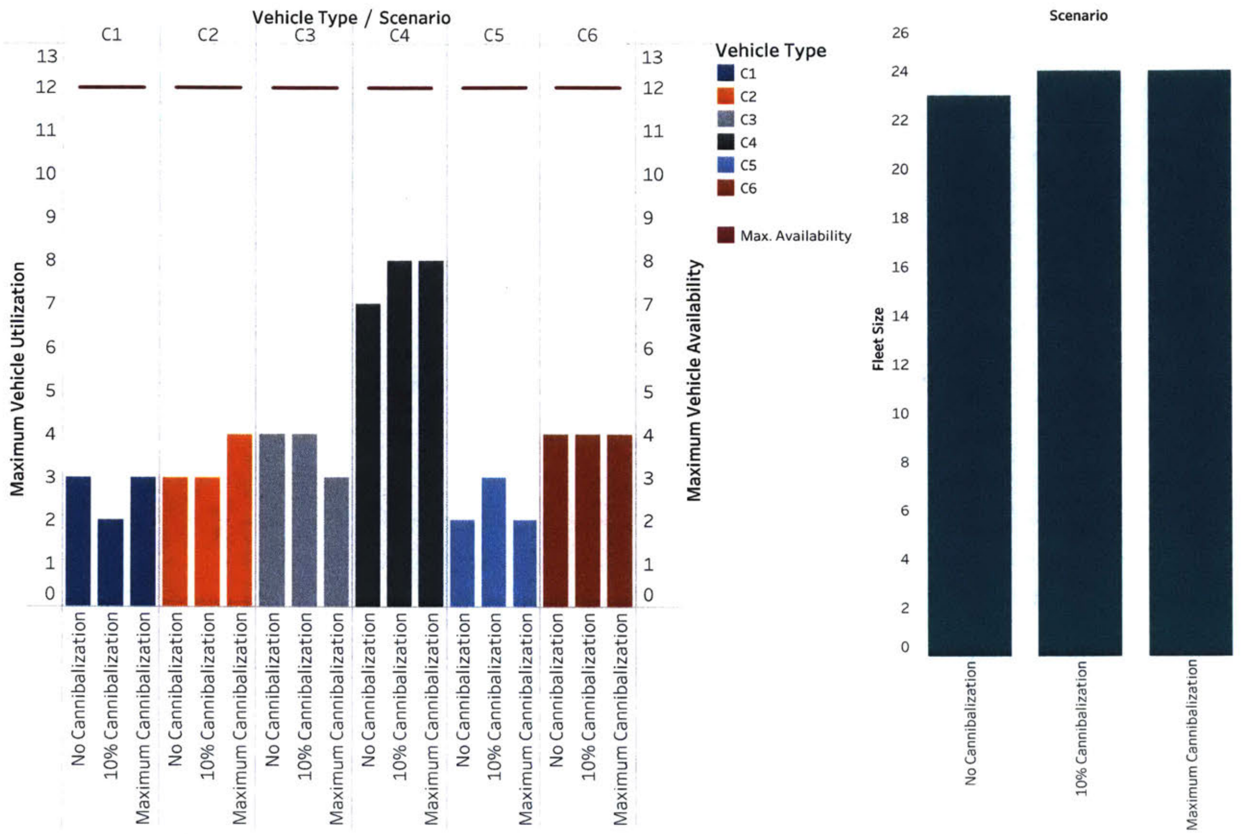
Figure C.3: Big-box customer allocation to facilities per problem instance

Figure C.4 portrays the set of nanostores being served on each cannibalization instance.

(a) 10.0% cannibalization rate (b) Maximum (16.2%) cannibalization rate

Figure C.4: Geographical distribution of served nanostores and active TPs

Finally, Figure C.5 compares the vehicle fleet mix as well as the fleet size for each problem instance.



(a) Vehicle fleet composition

(b) Vehicle fleet total size

Figure C.5: Vehicle fleet mix per cannibalization instance