Congestion Reduction Through Efficient Empty Container Movement Under Stochastic Demand

June 2020

A Research Report from the National Center for Sustainable Transportation

Maged Dessouky, University of Southern California
Santiago Carvajal, University of Southern California
Siyuan Yao, University of Southern California

National Center for Sustainable Transportation

METRANS Transportation Center
USC | CSULB

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. NCST-USC-RR-20-19	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A

4. Title and Subtitle
Congestion Reduction Through Efficient Empty Container Movement Under Stochastic Demand

5. Report Date
June 2020

6. Performing Organization Code
N/A

7. Author(s)
Maged Dessouky, PhD, https://orcid.org/0000-0002-9630-6201
Santiago Carvajal, PhD
Siyuan Yao

8. Performing Organization Report No.
N/A

9. Performing Organization Name and Address
METRANS Transportation Center
University of Southern California
University Park Campus
VKC 367 MC:0626
Los Angeles, California 90089-0626

10. Work Unit No.
N/A

11. Contract or Grant No.
Caltrans 65A0686 Task Order 018
USDOT Grant 69A3551747114

12. Sponsoring Agency Name and Address
California Department of Transportation
Division of Research, Innovation and System Information, MS-83
1727 30th Street, Sacramento, CA 95816

13. Type of Report and Period Covered
Final Report (December 2018 – April 2020)

14. Sponsoring Agency Code
USDOT OST-R

15. Supplementary Notes
DOI: https://doi.org/10.7922/G2W09464
Dataset DOI: https://doi.org/10.5061/dryad.mcvdncjxf

16. Abstract
In today's world, there is a significant amount of investigation regarding how to efficiently distribute loaded containers from the ports to the consignees. However, to fully maximize the process and become more environmentally friendly, one should also study how to allocate the empty containers created by these consignees. This is an essential part in the study of container movement since it balances out the load flow at each location. The problem of coordinating the container movement to reuse empty containers and lower truck miles is called the "Empty Container Problem". In this work we develop a scheduling assignment for loaded and empty containers that builds on earlier models but incorporates stochastic (random) future demand. Since this problem is meant to be solved daily and the solution implemented today affects tomorrow's starting state, incorporating future demand is an important aspect. This report shows that the truck miles needed to satisfy the demand at all locations is reduced by about 4-7% when considering future stochastic demand as opposed to only considering today's demand. Thus, leading to a cleaner and greener solution, creating less congestion and lowering the impact of freight movement on the environment.

17. Key Words Container Movement, Vehicle Routing, Stochastic Demand, Double Container Trucks	**18. Distribution Statement** No restrictions.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	50	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

About the National Center for Sustainable Transportation

The National Center for Sustainable Transportation is a consortium of leading universities committed to advancing an environmentally sustainable transportation system through cutting-edge research, direct policy engagement, and education of our future leaders. Consortium members include: University of California, Davis; University of California, Riverside; University of Southern California; California State University, Long Beach; Georgia Institute of Technology; and University of Vermont. More information can be found at: ncst.ucdavis.edu.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the accuracy of the data and information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, the California Department of Transportation and the METRANS Transportation Center in the interest of information exchange. The U.S. Government, the California Department of Transportation, and the University of Southern California assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views or policies of the State of California, USC, or the Department of Transportation. This report does not constitute a standard, specification, or regulation.

Acknowledgments

This study was funded, partially or entirely, by a grant from the National Center for Sustainable Transportation (NCST), supported by the U.S. Department of Transportation (USDOT) and the California Department of Transportation (Caltrans) through the University Transportation Centers program. The authors would like to thank the NCST, the USDOT, and Caltrans for their support of university-based research in transportation, and especially for the funding provided in support of this project.

Congestion Reduction Through Efficient Empty Container Movement Under Stochastic Demand

A National Center for Sustainable Transportation Research Report

June 2020

Maged Dessouky, Department of Industrial and Systems Engineering, University of Southern California

Santiago Carvajal, Department of Industrial and Systems Engineering, University of Southern California

Siyuan Yao, Department of Industrial and Systems Engineering, University of Southern California

NCST

[page intentionally left blank]

TABLE OF CONTENTS

EXECUTIVE SUMMARY v
1. Introduction 1
1.1 Background 1
1.2 Motivation 3
1.3 Problem Description 4
1.4 Structure of the Report 6
2. Literature Review 7
3. Container Assignment Problem 11
3.1 Mathematical Formulation 11
4. Vehicle Routing Problem 19
5. Experimental Analysis 24
5.1 Randomized Experiments 24
5.2 Ports of Los Angeles and Long Beach Experiments 27
6. Implementation 29
7. Conclusion 30
8. References 31
9. Data Management 34
10. Appendix 35

List of Tables

Table 1. Parameter settings for randomized experiments 25
Table 2. Average of the ratios for randomized experiments 26
Table 3. Standard Deviation of the ratios for randomized experiments 26
Table 4. Parameter settings for LA & LB Port experiments 27
Table 5. Average of the ratios for LA & LB Port experiments 28
Table 6. Standard Deviation of the ratios for LA & LB Port experiments 28
Table A1. Transitional probability distribution 1 35
Table A2. Transitional probability distribution 2 36
Table A3. Transitional probability distribution 3 37
Table A4. Demand Distributions 1 38
Table A5. Demand Distributions 2 38
Table A6. Demand Distributions 3 39
Table A7. Demand Distributions 4 39
Table A8. Demand Distributions 5 40
Table A9. Demand Distributions 1 for the San Pedro Ports Scenarios 40
Table A10. Demand Distributions 2 for the San Pedro Ports Scenarios 40
Table A11. Demand Distributions 3 for the San Pedro Ports Scenarios 40

List of Figures

Figure 1. San Pedro Bay Container Trade Imbalance [4] 2
Figure 2. Current container flow 5
Figure 3. Proposed single container flow 5
Figure 4. Proposed double container flow 5
Figure 5. Truck Job Construction 20
Figure 6. Truck Job Network 20

List of Acronyms and Abbreviations

ALNS	Adapted Large Neighborhood Search
DCAM	Double Container Assignment Model
SDCAM	Stochastic Double Container Assignment Model
I-110	California Interstate-110
I-710	California Interstate-710
NP-hard	Non-deterministic Polynomial-time Hardness
OD	Origin-Destination
SR47/SR	California State Route 47
TEU	Twenty-Foot Equivalent Unit
VRP	Vehicle Routing Problem

Congestion Reduction Through Efficient Empty Container Movement Under Stochastic Demand

EXECUTIVE SUMMARY

In today's world, there is a significant amount of investigation regarding how to efficiently distribute loaded containers from the ports to the consignees. However, to fully maximize the process and become more environmentally friendly, one should also study how to allocate the empty containers created by these consignees. This is an essential part in the study of container movement since it balances out the load flow at each location.

The problem of coordinating the container movement to reuse empty containers and lower truck miles is called the "Empty Container Problem". In this work, we develop a scheduling assignment for loaded and empty containers that builds on earlier models but incorporates stochastic (random) future demand. It is worth mentioning that in the previous research [5], the empty container problem was divided into two subproblems, including an assignment problem and a vehicle routing problem (VRP).

The previous research only considered the problem as a one-day horizon. But in reality, the container movements are not only to fulfill today's demand at each location but also prepare for the next day's delivery. Thus, incorporating future demand is an essential aspect of the problem.

By considering the future demand, a better solution can be constructed compared to solving the problem as a one-day horizon problem. This report shows that the truck miles needed to satisfy the demand at all locations is reduced by about 4-7% when considering future stochastic demand as opposed to only considering today's demand, thus, leading to a cleaner and greener solution, creating less congestion and lowering the impact of freight movement on the environment.

1. Introduction

1.1 Background

The increasing of international economic activities has caused rapid growth in international transportation of commodities. Compared with 2012, in 2013, the total volume of global containerized trade increased from 153 million to 160 million Twenty-foot Equivalent Units (TEU); and the Ports of Los Angeles and Long Beach, as the largest US marine terminals, had a 0.2 million TEUs increment in 2015 from 2014 [1]. In the United States, 1.5 trillion dollars were spent on goods transportation and services in 2017 [2]. A large portion of the international products is shipped by sea because of the low transportation cost and reliability. In 2018, there are approximately 9 million TEUs shipped in the Port of Los Angeles, which is 20 percent more than in 2010 [3].

Currently, there is a container size difference between the cargo ship and the United States convention. The cargo ships usually carry 20- and 40-foot containers, while the United States mainly uses 52-foot containers. To repack these containers into the corresponding sized container, the warehouses near the Ports of Los Angeles and Long Beach are built to satisfy the requirement. These repacking locations are called transloading stations.

As the port is playing an increasingly more important role in cargo transportation, it also raises the traffic congestion and environmental pollution problems near the port area. A large portion of the trailers are carrying empty containers. A large number of empty containers are generated due to the severe trade imbalance. Figure 1 indicates that there is an increasing trend of trade imbalance in the next few years at the San Pedro Ports [4]. Additionally, taking the Port of Los Angeles as an example, the data shows that in May 2019, about 28% of the 0.83 million TEU were empty containers [3], indicating a significant amount of unnecessary empty container movement. With a modest assumption of 3% annual growth, in 2045, there will be nearly 30 million TEUs traffic volume near the Ports of Los Angeles and Long Beach which is double its current volume. About 75% of the containers are carried by trucks for at least one segment of their trip. Most containerized traffic is using the I-110, I-710, and SR47/SR freeways as the primary transportation corridors, which exacerbates traffic congestion in the Los Angeles area.

Presently, there is almost no container exchange between inland locations at the Ports of Los Angeles and Long Beach, only about 2% of the empty containers are reused. Specifically, the current situation mostly has movements of loaded containers from exporters to the port, or from the port to the importers, while the reverse movements see mostly empty containers. If the reuse rate could be increased to 5%, there would be an annual 350 thousand truck trips saving or an average of 953 trips daily [4]. In this study, we propose a model that allows a "street exchange" which means empty containers can go directly from the importers to the exporters without returning to the port. In this way, redundant empty container movement can be reduced so that the traffic situation could be relieved.

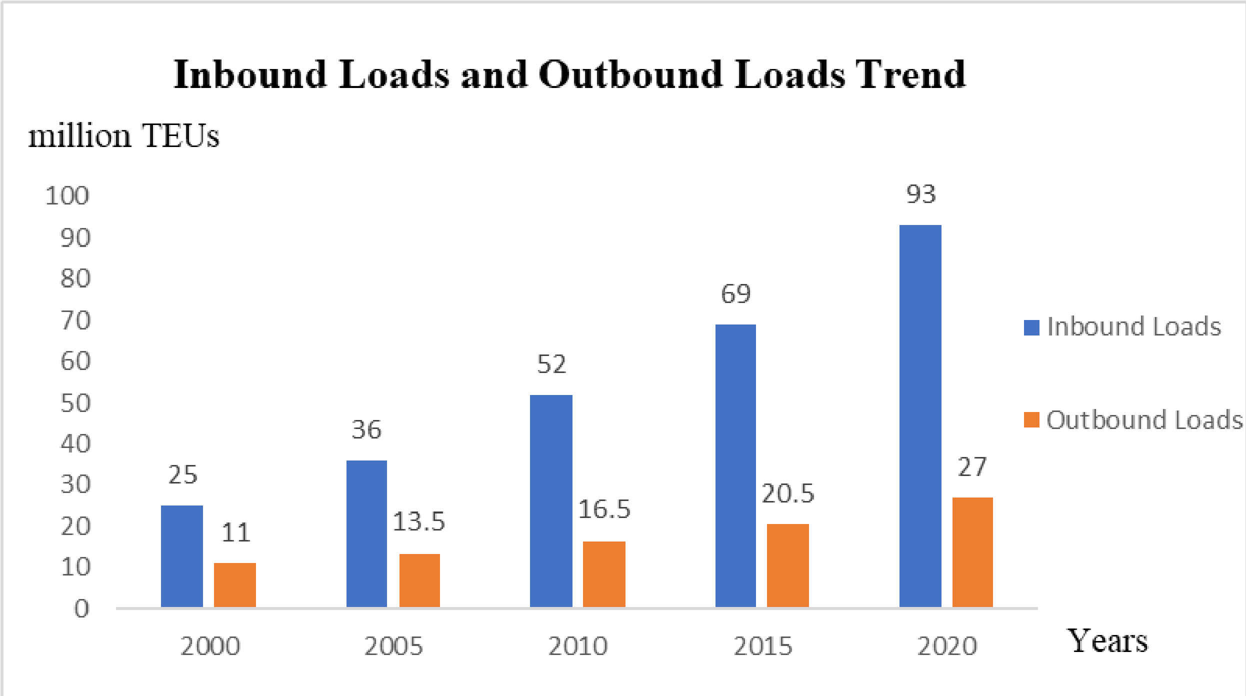

Figure 1. San Pedro Bay Container Trade Imbalance [4]

Unlike land freight that has the uncertainty in size and amount, sea transportation has standardized cargo with a relatively predictable schedule. The ship arrival and dispatch records can provide information about the demand and supply for the container movement for the next day, though some delay may happen because of the weather or uncertainty in port operations. For example, if ships are arriving in three consecutive days, the next day there would likely be a ship dispatching. Instead of focusing on just one day's container assignment and vehicle routing problem, the stochastic demand requires researchers to consider consecutive days and each day's final state is the initial state of the next day.

At the beginning of the day, it is reasonable to know if a container will become available for pickup and drop off. Thus, transloading stations around the Ports of Los Angeles and Long Beach make demand requests one day at a time. At this point, trucking companies can make their daily schedule to fulfill such requests for that day. However, it is possible to study past data to derive predictions of what tomorrow's demand could be based on different scenarios, such as a ship arriving or not. This information is unexploited if not incorporated to today's movement of containers. Specifically, this information can be used to position the containers at the end of the day to be better equipped to handle tomorrow's stochastic demand.

Container movement is limited by the carrier capacity. Nowadays, multiple-trailer trucks (also called Road Train) are employed in many countries such as Argentina, Australia, Mexico, the United States, and Canada, to move freight efficiently. Particularly, in the United States, trucks on public roads are allowed to connect at most two trailers with one dolly connection. The availability of multiple-trailer trucks adds research potential to the empty container transportation problem. A truck with multiple trailers, which are loaded or empty, can meet the demands of customers for goods and containers in a single operation at a lower cost and time consumption. Therefore, we could employ multiple-trailer trucks to our vehicle routing framework to construct a more intelligent truck routing algorithm to reuse empty containers, whereby reducing total truck movements in the road network.

1.2 Motivation

The problem with street exchanges is known as the "Empty Container Reuse Problem". This study builds on the earlier work of Dessouky and Carvajal [5] which only considered today's demand in developing the container reuse plan by considering future days in demand. As previously mentioned, double container trailers are used worldwide, such as Mexico, Canada, China, etc. However, the double container trailers are prohibited in the Ports of Los Angeles and Long Beach. This study shows the importance and possible benefits of using double container trailers in the operation of the port which can reduce the container truck miles and redundant empty container movement.

Focusing only on a single day does not optimize the system since it does not consider future demand. However, future demand is not known and depends on several factors, like whether a ship is arriving at the port on a given day. Therefore, most models only incorporate today's demand and optimize one day at a time. Future demand can be assumed to follow some underlying probability distribution. We can learn something about this distribution by studying past data. This information can be used in a stochastic optimization model that minimizes the expected cost under several possible operating scenarios such as normal, ship arrival, ship departure, etc. More specifically how today's scheduling affects tomorrow's scheduling is not that only unfulfilled demand must be met tomorrow, but the system's state at the end of today, will be the system's state at the beginning of the day tomorrow. Moreover, since the demand is stochastic, we propose a Markov Chain to predict the demand in the future. By classifying the scenarios of the port based on the cargo ship schedule, it is possible to find out the transitional probabilities from the historical data. After finding the transitional probabilities, we are able to model the problem by optimizing a container reuse problem over two days that allows us to account for the uncertain future demand. To take advantage of the knowledge of future demand, we use a solution approach for vehicle routing that meets today's demand, and incorporates tomorrow's stochastic demand. The approach is to have a model that can be solved at the beginning of the day and yield a feasible daily schedule for that day that considers customer time windows and the availability of containers. The schedule will also take into account tomorrow's stochastic demand, such that today's container movement is done in such a way to better handle tomorrow's stochastic demand. The model will then be solved again tomorrow, yielding tomorrow's container schedule movement, and taking into account the demand of the next day as well. This process is repeated until the planning horizon is reached.

1.3 Problem Description

In this study, we assume today's container demand at each location is deterministic at the beginning of the day and tomorrow's demand is stochastic. For each port scenario, the transitional probabilities from one state to another is known. In this way, the model is able to optimize the system over the long run. We consider two days in one run since, at the end of the today, the container allocation is the offset of the next day. The decision variables correspond to the number of containers sent from location *i* to *j* at time *t*.

The following figures show the current and proposed single container truck movement with empty or full containers (Figure 2 and Figure 3). In Figure 2, the current container movements are inefficient because a truck moves back and forth between the port and an exporter/importer. There is no street exchange in this scenario. Figure 3 shows the proposed single container flow with street exchanges so that an empty container can be shipped directly from an importer to an exporter. In order to extend the number of possible routes between the transloading stations, we proposed to use double container trailers in the model. Figure 4 indicates the container flow with double container trailers with the assumption that a double container truck has to drop off both loads before another pickup. At the same time, depots need to be considered in the model to make street exchanges easier to occur. Thus, all the locations in the network include importers, exporters, depots, and the port. For each location except the port, it has a demand for either loaded containers or empty containers; and it also yields empty or loaded containers or both. For example, an exporter has the demand for empty containers and can turn the empty container into loaded containers that can satisfy the port's demand. Because of the assumption we made, the possible supply-demand pairs are constrained in our model. For example, importers can only receive the truck from the port, etc.

As shown in Figure 4, there are three types of double container trailers, including the one with two empty containers (dash line), the one with two loads (solid line), and the one with one empty and one loaded container (dash line with dots). For example, the dashed-dotted line can only start from the port since only the port yields both empty and loaded containers. Clearly, the using of double container trailers increases the possible truck routes compared to the other two systems, though there are limits in the pickup and drop off. One major strength of the model is that as long as an OD pair is determined, the container operation is also calculated; and the second container delivery route is limited in Figure 3. For example, a truck with both an empty container and a loaded container arrives at an exporter; it can only drop off the empty container at the location and then go to an importer.

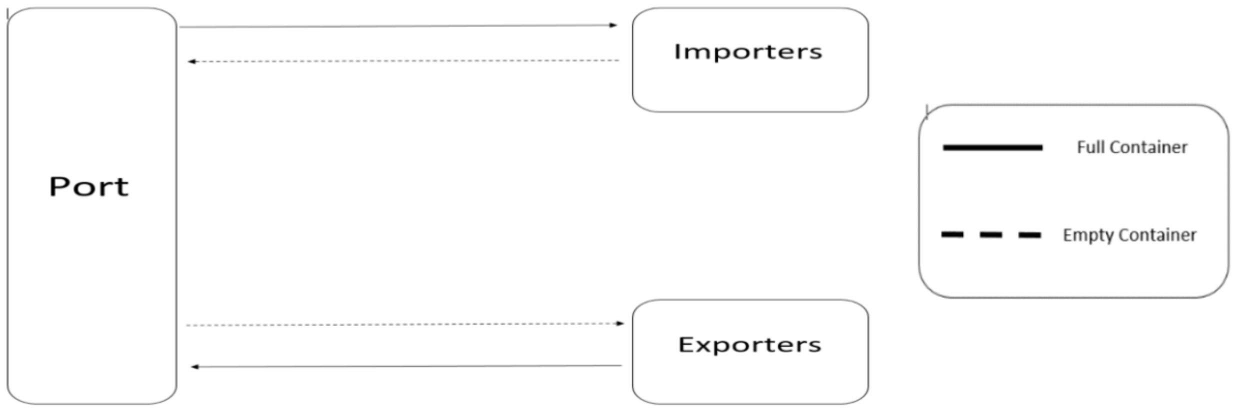

Figure 2. Current container flow

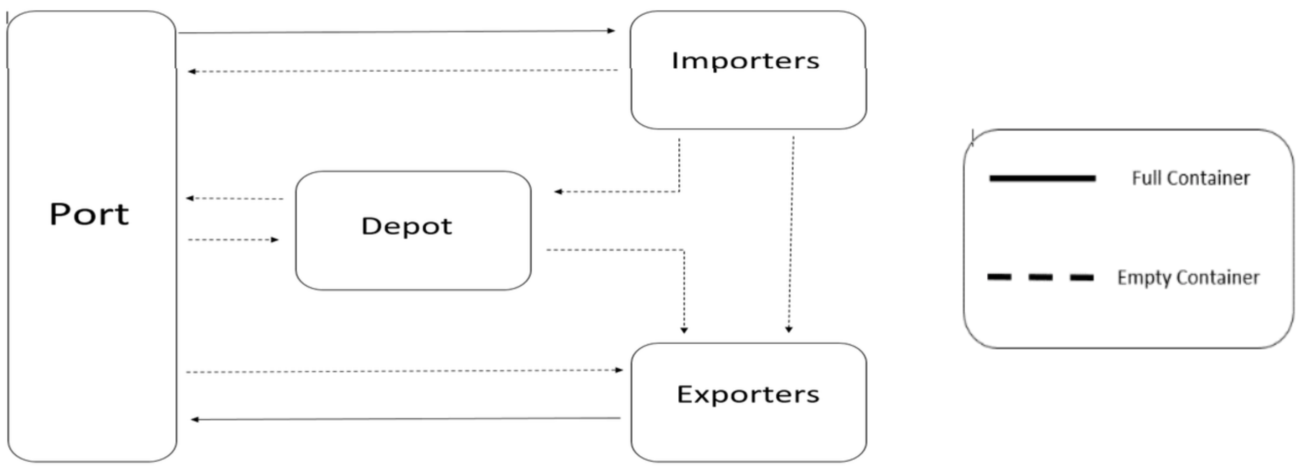

Figure 3. Proposed single container flow

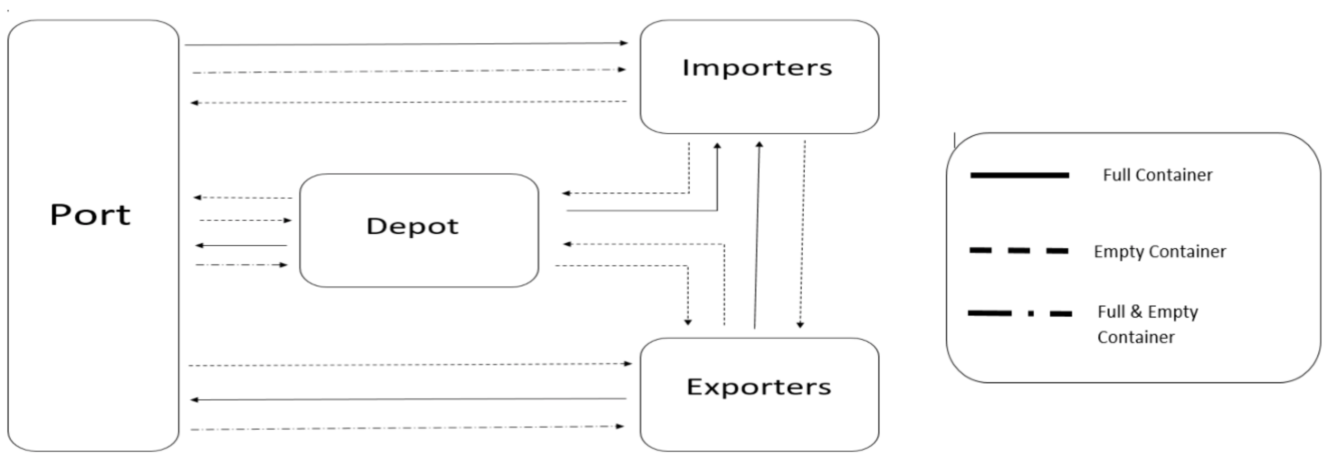

Figure 4. Proposed double container flow

In order to track the number of containers at each location, we introduce two groups of new variables. The first group of variables is used to keep track of the number of containers received at each location at a certain period. The second group of variables is used to record the number of containers dispatched at each location at a certain period. Similar to the previous study [5], the problem is solved by dividing the problem into the assignment problem and the VRP problem. First, we solve the problem as a container allocation assignment problem with time discretization. At this state, we assume the truck resources are unlimited since the port area has a sufficiently large number of trailers. Then, we create vehicle routes to find out the minimum number of trucks that can satisfy our assignment problem to reduce the total number of trucks in the system, which can reduce the traffic congestion in the port area.

1.4 Structure of the Report

The rest of this report is organized as follows. In Section 2, a literature review of relevant problems is presented. Section 3 presents the incorporation of the future stochastic demand to the container assignment problem. Section 4 presents an algorithm to build truck routes based on the container assignments found in Section 3. In Section 5, experimental results of our approach for randomly generated data sets and data that is representative of the Ports of Los Angeles and Long Beach are presented. In Section 6, we discuss the implementation and applicability of our work. Finally, in Section 7, some conclusions are drawn.

2. Literature Review

There has been some prior research on the Empty Container Reuse Problem due to the fact that container repositioning has become increasingly more expensive over the years. Historically, the problem has been subdivided into two sub-problems. The first problem focuses on empty container reuse in inland destinations. The second sub-problem focuses on the movement of containers that are near the port areas, usually no more than 20 miles from the port. It is this second problem that is the focus of this research.

In the literature, most people point to the paper by Dejax and Crainic [6] as the one that first identified the problem that empty containers were not being properly routed and empty container movement should be incorporated into container movement planning. In this paper, they describe how loaded and empty containers need to be managed together to fully optimize the container movement process. They proposed successive research with new ideas such as adding a depot center and integrating empty and loaded container movements at an industry level. They wrote several papers afterward developing different models for different situations and assumptions. For example, Crainic et al. [7] added the fact that maritime demand was stochastic and thus developed an integer model for stochastic demand, with a time window. The model was too complex for solving, but their objective was to present formulations for the problem and introduce the "Empty Container Problem" to the literature.

Later on, researchers started developing models and actually solving them for reasonably sized problems. Among these, Shen and Khoong [8] studied the problem from the perspective of a single company. In their paper, they develop a decision model that yielded when and how to move a container, as well as when to lease a container. They then performed some constraint relaxations on the model that allows the model to react quickly to supply and demand changes. Likewise, Li et al [9] studied the problem at a more global view. They built a model that maximized profit for the shipping company. Their model was deterministic and operated on a rolling horizon basis. They then tested their model on a real-life example using some ports from the east coast of China and showed that not only is their approach more profitable but also provides a greener solution. Song and Dong [10] also studied the problem by taking a global view of the problem. In their paper, they considered two ports and the intermodal locations around the ports. They assume that there is a trade imbalance between the two ports, to simulate real-life conditions such as a port from Asia and a port in the USA. They then derive a unique solution. In this approach, they model each port separately as integer programs, but then they use an inventory control based simulation model to decide when to ship containers between the two ports.

There has also been research done on how other factors affect the solving of the problem. For example, Bourbeau et al. [11] developed a mixed integer model that could be subdivided into semi-independent problems. They were then able to use parallel programming to solve the subproblems faster. They used their model to test how the location of the depots affects the problem and came up with an algorithm to find the optimal number and location of the depots. Another example would be how Choong at al. [12] studied the effect of the length of the planning horizon with regard to the empty container problem. In their paper, they use a model developed in Choong [13] and other empty container models to find the effect of the length of the planning horizon and how far in advance does planning need to start.

Francesco et el. [14] studied the problem of repositioning containers under stochastic demand, with uncertain parameters for the probability distribution functions. They built a multi-scenario multi-step model that was solved sequentially as events happened. They used a modified benders decomposition to solve their multi-step model. Later, Long et al. [15] enhanced this research by allowing demand to be non independent and identically distributed on each day.

Braekers et al [16] tackled the dynamic empty container problem, in which origins and destinations are not known beforehand. They constructed a network flow model to optimize the movement from importers, exporters, depots, and the port. They used a sequential approach and an integrated approach to solve the model. This yielded a sub-optimal result, but decreased the complexity of the model, thus reducing the solving time. They tested their solving methods using a small example that they created, as well as other examples from other papers for comparison.

Zhang et al. [17] developed a tabu search algorithm for the multi-depot truck transportation problem. In their paper, they developed two algorithms to solve this NP-hard problem. The first algorithm used clusters while the second one used a reactive tabu search. They then compared their algorithms to the optimal solution that could be found for small size instances. They determined that the cluster algorithm performed well, but the reactive tabu algorithm performed extremely well, although the cluster algorithm is more scalable than the tabu algorithm. Funke and Kopfer [18] also studied the VRP problem for the Empty Container Problem. In their paper, they assumed that loaded containers had a known origin and destination, but empty containers only had a known origin or destination, and the other was unknown and needed to be solved. To solve their problem, they build a time and space graph with the arcs representing possible truck movements from one location to another. They are then able to build an integer model for the graph and solve it to optimality using branch and bound techniques. However, they are only able to do so for a small size problem and mention that heuristics would be needed to make the model scalable.

Vidovic et al. [19] studied the problem of solving a VRP for multi-size containers. In their problem, they had multiple locations with multiple requests for either 20 foot or 40 foot containers, and a depot. They studied the problem in which the demand for the containers was known, and each location had a known initial number of containers. They then developed a container assignment and vehicle routes where all demands were met. To solve their problem, they implemented a modified version of Popovic et al's [20] variable neighborhood search heuristic, which was in turn inspired by Mladenovic and Hansen's [21] metaheuristic. The idea for the heuristic is to find an initial solution, and then do a neighborhood search to find a better solution. However, what makes their search interesting and successful is that they change the size of their search based on the polar angle between two routes. The idea being that if two adjacent routes has a big gap between them, then they are less likely to share a job between them and one can eliminate any route exchanges between those two routes.

Tan et al. [22] also studied the Empty Container Problem, but at a company level. They developed a hybrid model that could be used by transportation companies to determine their truck schedules for the day. The model incorporates local search heuristics and several specialized genetic algorithms, with the objective being to minimize truck miles, and the number of trucks used. The model also determines when it is more efficient for the company to outsource a certain job. Sterzik and Kopfer [23] further advanced this research by introducing a robust model that different companies could incorporate. They solved their algorithm using a tabu search with several heuristics. They tested their algorithm against the algorithm in Zhang [24] outperforming it in almost every instance. In another paper by the same authors [25], they augment their previous work by building a mixed integer programming model that considers both vehicle routing and empty container repositioning simultaneously. Their model minimizes trucking operating time which includes travel, service, and waiting times. They solve the model by first building an initial solution using a modified version of Clark and Wright's [26] savings algorithm and then searching the solution space using their previous tabu search algorithm. In their paper, they specifically run scenarios in which container leasing companies are allowed to exchange containers and scenarios in which they are not, and show the benefits of such exchanges.

Probably the most extensive research of container movement in the Ports of Long Beach and Los Angeles was done by the Tioga Group [4]. They did extensive research on container movement in and out of the Port of Long Beach. After compiling extensive data, they suggested a concept of how empty container reuse could be increased in this area. Their work has served as a foundation to various other empty container models that use the Ports of Long Beach and Los Angeles as their research scenario, especially when using their data. For example, Jula et al. [27] built a dynamic model that used the Tioga report data to come up with a feasible solution of how to allocate containers on a daily basis. Taking into account that on any single day all demand is deterministic, but the demand for the next day is stochastic, they use dynamic programming to find the best match of a bipartite transportation network. In that way, they meet all the daily demand and to optimize the containers for future days as well. Chang et al. [28] studied when and where containers should be substituted with another type. They proposed a heuristic method that divided the problem into dependent and independent parts. They were then able to apply a branch and bound algorithm to arrive at an integer solution relatively fast. They tested their procedure for the Ports of Los Angeles and Long Beach using data from the Tioga report, and on randomized scenarios, comparing it to other commercial mixed integer programming solvers. Similarly, Lam et al [29] demonstrated how dynamic programming can lead to a competitive approximate solution that improves efficiency. They first built an empty container model and then used linear approximations to simplify the model. They then used dynamic programming to arrive at an approximation of the optimal solution. They tested their approach in a simple model, and for a real-life example. They then compared their solution to other heuristics used in the industry.

Dam Le [30] has also assessed the problem from the perspective of the logistics involved to make container reuse possible in Southern California. She conducted several interviews with field experts to make recommendations on where depots would make the most sense according to expected demand from the different importers and exporters. The problem has also been studied in different ports around the world. Islam et al. [31] conducted an extensive study of the Port of Auckland. They studied how containers were moved and how truck congestion changed throughout the day. They, in particular, studied when empty containers should be relocated such that the effect of the empty container movement is minimized, especially during peak hours.

The empty container problem traditionally focuses on meeting the demand for one day. This is done because daily demand is deterministic. However, a natural extension to this problem is how can we accommodate future stochastic demand and reposition containers such that we are better able to handle future demand. Bandeira et al [32] developed a rolling horizon model to coordinate different customer demands as to minimize costs. Their model is solved in two steps. First, it meets all the demand for that time period. Then it adjusts the solution to allocate containers to minimize costs. Erera et al. [33] built a robust optimization framework for container allocation. For their solution framework, they first solve an integer model to acquire an initial solution. They then use heuristics to manage container inventory at the locations and then update their solution as time goes on. This allowed them to find an approximate optimal solution and then change their solution to adapt to a stochastic world.

3. Container Assignment Problem

This research enhances the model and methodology of Dessouky and Carvajal [5] by incorporating stochastic future demand into the container assignments. This incorporation is beneficial since the model is not meant to be solved in a vacuum, but instead it is meant to be solved considering future demand. Thus, the solution that is implemented today will have future repercussions on tomorrow's solution. More specifically, the final state of today will be the starting state of tomorrow. It is for this reason that if we want to minimize the cost and truck miles in the long run, we should take tomorrow's demand into consideration when solving the container assignment today.

The main challenge in including tomorrow's demand into our model is that, unlike today's demand which is deterministic, tomorrow's demand is stochastic. However, it is possible to study historical data and derive a stochastic model for tomorrow's demand. In this case, one way to develop a stochastic model is twofold. First, there are different scenarios that could take place. For example, an importer demand would look rather different depending if a ship arrives on time or is delayed. Thus, there are different scenarios like this one that would affect how the demand for containers for the different locations would be. It is then possible to estimate the transitional probabilities of the scenarios from historical data.

3.1 Mathematical Formulation

We next present the formulation for the Stochastic Double Container Assignment Model (SDCAM) given the different scenarios for tomorrow's demand.

Parameters:

$I = Total\ number\ of\ importers$

$E = Total\ number\ of\ exporters$

$D = Total\ number\ of\ Depots$

$T = Number\ of\ time\ discretizations$

$l_{i,j,t} = time\ it\ takes\ to\ go\ from\ location\ i\ to\ location\ j\ leaving\ at\ time\ t$

$o_{i,j,t} = time\ it\ takes\ to\ go\ from\ location\ i\ to\ location\ j\ arriving\ at\ time\ t$

$r_i = Container\ turnover\ time\ at\ location\ i$

$p_i = Number\ of\ containers\ available\ at\ the\ beginning\ of\ the\ day\ at\ location\ i$

$d_{i,t} = Number\ of\ containers\ demanded\ at\ location\ i\ by\ time\ t$

$c_i = Capacity\ of\ location\ i$

$e_{i,j,t} = Cost\ of\ first\ leg\ of\ a\ two\ container\ route\ going\ from\ location\ i\ to\ location\ j\ starting\ at\ time\ t$

$f_{i,j,t} = Cost\ of\ second\ leg\ of\ a\ two\ container\ route\ going\ from\ location\ i\ to\ location\ j\ starting\ at\ time\ t$

$g_{i,j,t} = Cost\ of\ a\ one\ container\ route\ from\ location\ i\ to\ location\ j\ starting\ at\ time\ t$

$\bar{T} = Number\ of\ time\ discretizations\ for\ tomorrow$

$S = Number\ of\ different\ scenarios\ for\ tomorrow$

$\theta_s = Probability\ for\ scenario\ s$

$\bar{d}_{i,t,s} = Mean\ number\ of\ containers\ demanded\ at\ location\ i\ by\ time\ t\ for\ scenario\ s$

$\varphi = Penalty\ for\ not\ fullfilying\ one\ unit\ of\ demand$

$\mu_s = Mean\ number\ of\ containers\ that\ arrive\ to\ the\ port\ in\ scenarios$

Sets:

$SI = \{1, \dots, I\}\ (locations\ of\ all\ importers)$

$SE = \{\ I+1, \dots, I+E\}\ (locations\ of\ all\ exporters)$

$SD = \{I+E+1, \dots, I+E+D\}\ (locations\ of\ all\ depots)$

$SP = \{\ I+E+D+1\ \}\ (location\ of\ the\ port)$

$SA = \{SI \cup SE \cup SD \cup SP\}(all\ locations)$

$ST = \{1, \dots, T\}\ (times\ of\ the\ day)$

$S\bar{T} = \{T+1, \dots, \bar{T}\ (times\ of\ tomorrow's\ day)$

$SN = \{1, \dots I, \dots I+E, \dots I+E+D\}\ (all\ non-port\ locations)$

$S\theta = \{1, \dots, S\}\ \ \ (set\ of\ all\ scenario)$

Decision Variables:

$x_{i,j,t} = Number\ of\ first\ leg\ two\ container\ trucks\ going\ from\ location\ i\ to\ location\ j\ at\ time\ t$

$y_{i,j,t} = Number\ of\ second\ leg\ two\ container\ trucks\ going\ from\ location\ i\ to\ location\ j\ at\ time\ t$

$z_{i,j,t} = Number\ of\ single\ container\ trucks\ going\ from\ location\ i\ to\ j\ at\ time\ t$

$m_{i,t} = Number\ of\ containers\ supplied\ by\ location\ i\ at\ time\ t$

$n_{i,t} = Number\ of\ containers\ delivered\ to\ location\ i\ at\ time\ t$

$a_{i,t} = Number\ of\ containers\ that\ have\ been\ supplied\ by\ location\ i\ by\ time\ t$

$b_{i,t} = Number\ of\ containers\ that\ have\ been\ delivered\ to\ location\ i\ by\ time\ t$

$\bar{x}_{i,j,t} = Number\ of\ first\ leg\ two\ container\ trucks\ going\ from\ location\ i\ to\ location\ j\ at\ time\ t\ for\ tomorrow\ of\ a\ truck$

$\bar{y}_{i,j,t} = Number\ of\ second\ leg\ two\ container\ trucks\ going\ from\ location\ i\ to\ location\ j\ at\ time\ t\ for\ tomorrow\ of\ a\ truck$

$\bar{z}_{i,j,t} = Number\ of\ single\ container\ trucks\ going\ from\ location\ i\ to\ j\ at\ time\ t\ for\ tomorrow$

$\bar{m}_{i,t} = Number\ of\ containers\ supplied\ by\ location\ i\ at\ time\ t\ for\ tomorrow$

$\bar{n}_{i,t} = Number\ of\ containers\ delivered\ to\ location\ i\ at\ time\ t\ for\ tomorrow$

$\bar{a}_{i,t} = Number\ of\ containers\ that\ have\ been\ supplied\ by\ location\ i\ by\ time\ t\ for\ tomrrow$

$\bar{b}_{i,t} = Number\ of\ containers\ that\ have\ been\ delivered\ to\ location\ i\ by\ time\ t\ for\ tomrrow$

$ƨ_{i,t,s} = Number\ of\ unmet\ demand\ for\ location\ i\ at\ time\ t\ for\ scenario\ s$

Objective:

$$\min \sum_{t\in ST} \sum_{i\in SA} \sum_{j\in SA} (e_{i,j,t} * x_{i,j,t} + f_{i,j,t} * y_{i,j,t} + g_{i,j,t} * z_{i,j,t}) + \sum_{t\in S\bar{T}} \sum_{i\in SA} \sum_{j\in SA} (e_{i,j,t} * \bar{x}_{i,j,t} + f_{i,j,t} * \bar{y}_{i,j,t} + g_{i,j,t} * \bar{z}_{i,j,t}) \ldots$$

$$\ldots + \textstyle\sum_{s\in S\theta} \sum_{t\in S\bar{T}} \sum_{i\in SA} \varphi * \theta_s * ƨ_{i,t,s}$$

s.t.

Today's Constraints:

Containers provided at time t:

$$2 \sum_{\in SE \cup SD \cup SP} x_{i,j,t} + \sum_{j\in SE \cup SD \cup SP} z_{i,j,t} = m_{i,t} \quad \forall i \in SI \quad \forall t \in ST \quad (Importers) \tag{1}$$

$$2 \sum_{j\in SP} x_{i,j,t} + \sum_{j\in SP} z_{i,j,t} = m_{i,t} \quad \forall i \in SE \qquad \forall t \in ST \quad (Exporters) \tag{2}$$

$$2 \sum_{j\in SE \cup SD \cup SP} x_{i,j,t} + \sum_{j\in SE \cup SD \cup SP} z_{i,j,t} = m_{i,t} \quad \forall i \in SD \quad \forall t \in ST \quad (Depots) \tag{3}$$

$$2 \sum_{j\in SI \cup SE \cup SD} x_{i,j,t} + \sum_{j\in SI \cup SE \cup SD} z_{i,j,t} = m_{i,t} \quad \forall i \in SP \quad \forall t \in ST \quad (Port) \tag{4}$$

Containers received at time t:

$$\sum_{i\in SP} x_{i,j,t-o_{i,j,t}} + \sum_{i\in SI \cup SE \cup SD} y_{i,j,t-o_{i,j,t}} + \sum_{i\in SP} z_{i,j,t-o_{i,j,t}} = n_{j,t} \quad \forall j \in SI \quad \forall t \in ST \quad (Importers) \tag{5}$$

$$\sum_{i\in SI \cup SD \cup SP} x_{i,j,t-o_{i,j,t}} + \sum_{i\in SI \cup SE \cup SD} y_{i,j,t-o_{i,j,t}} + \sum_{i\in SI \cup SD \cup SP} z_{i,j,t-o_{i,j,t}} = n_{j,t} \quad \forall j \in SE \quad \forall t \in ST \ (Exporters) \tag{6}$$

$$\sum_{i\in SI \cup SD \cup SP} x_{i,j,t-o_{i,j,t}} + \sum_{i\in SI \cup SE \cup SD} y_{i,j,t-o_{i,j,t}} + \sum_{i\in SI \cup SD \cup SP} z_{i,j,t-o_{i,j,t}} = n_{j,t} \quad \forall j \in SD \quad \forall \in ST \quad (Depots) \tag{7}$$

$$\sum_{i\in SI \cup SE \cup SD} x_{i,j,t-o_{i,j,t}} + \sum_{i\in SA} y_{i,j,t-o_{i,j,t}} + \sum_{i\in SI \cup SE \cup SD} z_{i,j,t-o_{i,j,t}} = n_{j,t} \quad \forall j \in SP \quad \forall t \in ST \quad (Port) \tag{8}$$

Demand and Feasibility constraints:

$$a_{i,t} = \sum_{q=1}^{t} m_{i,q} \qquad \forall i \in SA \qquad \forall t \in ST \qquad (Number\ provided\ at\ i\ by\ time\ t) \tag{9}$$

$$b_{i,t} = \sum_{q=1}^{t} n_{i,q} \qquad \forall i \in SA \qquad \forall t \in ST \qquad (Number\ received\ at\ i\ by\ time\ t) \tag{10}$$

$$b_{i,t-r_i} + p_i - a_{i,t} \geq 0 \qquad \forall i \in SA \qquad \forall t \in ST \qquad (non-negativity\ constriant\ ) \tag{11}$$

$$b_{i,t} \geq d_{i,t} \qquad \forall i \in SN \qquad \forall t \in ST \quad (Demand\ at\ location\ i\ must\ be\ met\ by\ time\ t\ for\ non-port) \tag{12a}$$

$$b_{i,t} - a_{i,t} \geq d_{i,t} \qquad \forall i \in SP \quad \forall t \in ST \quad (Demand\ at\ location\ i\ must\ be\ met\ by\ time\ t\ for\ the\ port) \tag{12b}$$

$$b_{i,t} + p_i - a_{i,t} \leq c_i \qquad \forall i \in SA \qquad \forall t \in ST \quad (Capacity\ at\ i\ cannot\ be\ exceeded) \tag{13}$$

$$\sum\nolimits_{i \in SA} x_{i,j,t} = \sum_{k \in SA} y_{j,k,t+l_{i,j,t}} \qquad \forall j \in SA \qquad \forall t \in ST \quad (Two\ container\ trucks\ must\ provide\ two\ containers) \tag{14}$$

$$x_{i,j,t}, y_{i,j,t}, z_{i,j,t} \geq 0 \qquad \forall i \in SA \quad \forall j \in SA \quad \forall \in ST \quad (Non-negative\ Constraint) \tag{15}$$

$$x_{i,j,t}, y_{i,j,t}, z_{i,j,t} \in \mathbb{Z} \qquad \forall i \in SA \quad \forall j \in SA \quad \forall \in ST \quad (Integer\ Constraint) \tag{16}$$

Tomorrow's Constraints:

$$2 \sum_{j \in SE \cup SD \cup SP} \bar{x}_{i,j,t} + \sum_{j \in SE \cup SD \cup SP} \bar{z}_{i,j,t} = \bar{m}_{i,t} \qquad \forall i \in SI \qquad \forall t \in S\bar{T} \qquad (Importers) \tag{17}$$

$$2 \sum_{j \in SP} \bar{x}_{i,j,t} + \sum_{j \in SP} \bar{z}_{i,j,t} = \bar{m}_{i,t} \qquad \forall i \in SE \qquad \forall t \in S\bar{T} \qquad (Exporters) \tag{18}$$

$$2 \sum_{j \in SE \cup SD \cup SP} \bar{x}_{i,j,t} + \sum_{j \in SE \cup SD \cup SP} \bar{z}_{i,j,t} = \bar{m}_{i,t} \qquad \forall i \in SD \qquad \forall t \in S\bar{T} \qquad (Depots) \tag{19}$$

$$2 \sum_{j \in SI \cup SE \cup SD} \bar{x}_{i,j,t} + \sum_{j \in SI \cup SE \cup SD} \bar{z}_{i,j,t} = \bar{m}_{i,t} \qquad \forall i \in SP \qquad \forall t \in S\bar{T} \qquad (Port) \tag{20}$$

Containers received at time t for tomorrow:

$$\sum_{i \in SP} \bar{x}_{i,j,t-o_{i,j,t}} + \sum_{i \in SI \cup SE \cup SD} \bar{y}_{i,j,t-o_{i,j,t}} + \sum_{i \in SP} \bar{z}_{i,j,t-o_{i,j,t}} = \bar{n}_{j,t} \qquad \forall j \in SI \qquad \forall t \in S\bar{T} \qquad (Importers) \tag{21}$$

$$\sum_{i \in SI \cup SD \cup SP} \bar{x}_{i,j,t-o_{i,j,t}} + \sum_{i \in SI \cup SE \cup SD} \bar{y}_{i,j,t-o_{i,j,t}} + \sum_{i \in SI \cup SD \cup SP} \bar{z}_{i,j,t-o_{i,j,t}} = \bar{n}_{j,t} \qquad \forall j \in SE \qquad \forall t \in S\bar{T} \qquad (Exporters) \tag{22}$$

$$\sum_{i \in SI \cup SD \cup SP} \bar{x}_{i,j,t-o_{i,j,t}} + \sum_{i \in SI \cup SE \cup SD} \bar{y}_{i,j,t-o_{i,j,t}} + \sum_{i \in SI \cup SD \cup SP} \bar{z}_{i,j,t-o_{i,j,t}} = n_{j,t} \qquad \forall j \in SD \qquad \forall \in S\bar{T} \qquad (Depots) \tag{23}$$

$$\sum_{i \in SI \cup SE \cup SD} \bar{x}_{i,j,t-o_{i,j,t}} + \sum_{i \in SA} \bar{y}_{i,j,t-o_{i,j,t}} + \sum_{i \in SI \cup SE \cup SD} \bar{z}_{i,j,t-o_{i,j,t}} = \bar{n}_{j,t} \qquad \forall j \in SP \qquad \forall t \in S\bar{T} \qquad (Port) \tag{24}$$

Demand and Feasibility constraints:

$$\bar{a}_{i,t} = \sum_{q=T+1}^{t} \bar{m}_{i,q} \qquad \forall i \in SA \qquad \forall t \in S\bar{T} \qquad (Number\ provided\ at\ i\ by\ time\ t) \tag{25}$$

$$\bar{b}_{i,t} = \sum_{q=T+1}^{t} \bar{n}_{i,q} \qquad \forall i \in SA \qquad \forall t \in S\bar{T} \qquad (Number\ received\ at\ i\ by\ time\ t) \tag{26}$$

$$(b_{i,T} + p_i - a_{i,T}) - \bar{a}_{i,t} \geq 0\ \forall\{i,t | i \in SN, t \in ST, t - r_i < T+1\}\ (Non-negative\ constriant\ for\ non-port) \tag{27a}$$

$$\bar{b}_{i,t-r_i} - \bar{a}_{i,t} + (b_{i,T} + p_i - a_{i,T} - d_{i,T}) + \sum_{s \in S\theta} \theta_s * \mu_s \geq 0 \qquad \forall i \in SP \qquad \forall t \in S\bar{T} \qquad (Non-negative\ constraint\ for\ Port) \tag{27b}$$

$$-\bar{a}_{i,t} + (b_{i,T} + p_i - a_{i,T} - d_{i,T}) + \sum_{s \in S\theta} \theta_s * \mu_s \geq 0 \qquad \forall i \in SP \qquad \forall t \in S\bar{T}, t - r_i < T+1 \qquad (Non-negative\ constraint\ for\ Port) \tag{27c}$$

$$\bar{b}_{i,t} - \bar{d}_{i,t,s} + z_{i,t,s} \geq 0 \qquad \forall i \in SN \qquad \forall t \in ST \quad \forall s \in S\theta \quad (Demand\ at\ location\ i\ must\ be\ met\ by\ time\ t) \tag{28a}$$

$$\bar{b}_{i,t} + p_j - \bar{n}_{j,t} - \bar{d}_{i,t,s} + z_{i,t,s} \geq 0 \qquad \forall j \in SP \quad \forall t \in S\bar{T} \quad \forall s \in S\theta \quad (Demand\ for\ port\ location\ must\ be\ met\ by\ time\ t) \tag{28b}$$

$$\bar{b}_{i,t} + (p_i + b_{i,T} - a_{i,T}) - \bar{a}_{i,t} \leq c_i \qquad \forall i \in SN \qquad \forall t \in S\bar{T} \qquad (Capacity\ at\ i\ cannot\ be\ exceeded\ for\ non-port) \tag{29a}$$

$$\bar{b}_{i,t-r_i} - \bar{a}_{i,t} + (b_{i,T} + p_i - a_{i,T} - d_{i,T}) + \sum_{s \in S\theta} \theta_s * \mu_s \leq c_i \qquad \forall i \in SP \qquad \forall t \in S\bar{T} \qquad (Capacity\ at\ i\ cannot\ be\ exceeded\ for\ the\ port) \tag{29b}$$

$$\sum_{i \in SA} \bar{x}_{i,j,t} = \sum_{k \in SA} \bar{y}_{j,k,t+l_{i,j,t}} \qquad \forall j \in SA \qquad \forall t \in S\bar{T} \qquad (Two\ container\ trucks\ must\ provide\ two\ containers) \tag{30}$$

$$\bar{x}1_{i,j,t}, \bar{y}1_{i,j,t}, \bar{z}_{i,j,t} \geq 0 \qquad \forall i \in SA \qquad \forall j \in SA \qquad \forall t \in S\bar{T} \qquad (Non-negative\ Constraint) \tag{31}$$

$$\bar{x}1_{i,j,t}, \bar{y}1_{i,j,t}, \bar{z}_{i,j,t} \in \mathbb{Z} \qquad \forall i \in SA \qquad \forall j \in SA \qquad \forall \in S\bar{T} \qquad (Integer\ Constraint) \tag{32}$$

The objective of the SDCAM is to minimize the transportation costs needed to meet all of today's deterministic demand and all of tomorrow's expected demand for all the possible scenarios. The model is separated into two main parts. The first part deals with container movements made today while the second part deals with container movements made tomorrow. Both parts are similar with the second part having a few modifications to account for the container movements the model performs today. We will next explain the first part of the model and then explain the second part of the model.

The model has three main integer variables. The $x_{i,j,t}$ variables correspond to the first leg of a double-container truck going from location *i* to location *j* at time *t* to drop off the first container. The $y_{i,j,t}$ variables correspond to the second leg of a double-container truck going from location *i* to location *j* at time *t* to drop off the second container. Finally, the *z* variables represent a single-container truck going from location *i* to location *j* at time *t*. Note that *i* and *j* cannot be the same for any *x* or *z* variable since it does not make sense that a location can provide itself with containers; however the *y* variables can have *i* and *j* be the same since that means the second container is being dropped off at the same location as the first container. The rest of the variables only serve to record the total number of received and delivered containers at each location for each time period, and are determined by specific summations of the main three variables.

Constraints (1)-(4) are the definition of the $m_{i,t}$ variables and thus sum all the containers provided by a specific location at a specific point in time. This is done for all locations and all times. Notice that a double-container truck trip uses 2 containers and thus there is a 2 multiplying the *x* variable, while a single-container truck trip only uses 1 container. Each set of variables is for different type of locations. For example, constraint (1) sums up containers provided by the importers which can provide empty containers that can go to the exporters, depots, or the port, but they cannot go to the other importers.

Similarly, constraints (5)-(8) are the definition of the $n_{i,t}$ variables and thus sum all the containers received by a specific location at a specific point in time. This is done for all locations and all times. In these constraints each variable represents one container being dropped off at a location and thus there is no multiplication of 2. For example, constraint (5) sums up all the containers received by an importer. An importer only needs loaded containers, and thus can only receive containers from the port. For this reason, the *x* and *z* variables can only originate from the port. However, notice that the *y* variable cannot originate from the port, and instead must originate from an importer, exporter, or depot. This is because logistically it does not make sense to go into the port to drop off a container and then come out to drop another container at an importer.

Constraints (9) and (10) are the definitions of the $a_{i,t}$ and $b_{i,t}$ variables respectively, which sum up all the containers provided/received by location *i* by time *t* by aggregating all the $n_{i,t}$ or $m_{i,t}$ variables up to time t. The $a_{i,t}$ and $b_{i,t}$ variables will then become the variables used in the latter constraints to ensure that feasibility of the solution and that demand is met at all locations and at all points in time.

Constraint (11) is the feasibility constraints that the total number of containers received at a location, plus the number of containers at the start of the day at that location, minus the number of containers provided by that location cannot be negative. Notice that $b_{i,t}$ is offset by r_i, which is the turnover time at location *i.* This is because after a container arrives at location *i* it must be processed (either unloaded or loaded) before it can be moved again (r_i may be zero for locations like the depot where containers do not need to be processed). Constraint (12) ensures that demand is met, and notice that the $b_{i,t}$ variable is aggregated, and thus demand is also aggregated. Constraint (13) is the capacity constraint and is similar to constraint (11), by ensuring that capacity is not exceeded for all locations.

Finally, constraint (14) makes sure that a double container truck delivers two containers. The *x* variables represent a truck going from location *i* to location *j* at time *t* after some delay, given by the parameter $l_{i,j,t}$. This truck must go to another location (this can be the same location) to deliver the second container. This is represented by the *y* variable. This constraint states that all the *x* variables that arrive at a certain location by time *t* must have a corresponding *y* variable associated with them. Then constraints (15) and (16) are the variable integrality and non-negativity constraints.

These are all the constraints for the first part of the model representing today's container movements. As stated before these constraints are repeated in the second part of the model representing tomorrow's container movements. The constraints are the same, apart from three; the non-negativity constraints (27a-c)), the demand constraints (28a-b), and the capacity (constraints (29a-b)). For tomorrow's set of constraints however, they are divided into port locations and non-port locations. For the non-port locations the constraint states that the sum of the total number of container received tomorrow at a location, minus the total number of containers provided tomorrow by that location, plus the number of containers at the end of today, must be non-negative and less than the location's capacity. Note that the end of today's state is the beginning of tomorrow, thus any containers left at the end of the day at a given location is the number of containers that location will have at the beginning of the day tomorrow. The port however functions a little different. The containers demanded by the port are containers that will be loaded into ships to be transported overseas. Thus, we also need to subtract all the containers that the port demanded for today, thus removing those containers from the system. However, the port also provides new containers to the system when ships arrive at the port with new containers. The number of new containers available depends on what scenario tomorrow is. Therefore, we also add the mean number of containers that will become available multiplied by the probability for that scenario.

The other constraint that changes is the demand constraint (28a-b). As previously explained, the demand for each location is based on a probability density function that depends on which scenario is occurring. Thus, instead of having a single demand constraint for each location, we have *S* constraints, where *S* is the number of scenarios that can occur, thus having one constraint for each scenario. We also introduce $ƨ_{i,t,s}$ which acts as a slack variable, allowing the demand constraint to be violated if $ƨ_{i,t,s}$ is positive for location *i* for time *t* under scenario *s*. However, we then bring this slack variable to the objective and penalize it by the parameter φ for each unit of demand not met under that scenario. This creates a penalty function for violations of the demand constraint. This allows the model to hedge container movements based on the probability of certain scenarios of occurring or not.

It is worth reiterating here that the model assumes that tomorrow's stochastic demand is modeled in two steps. First, there are several scenarios that can occur, and each scenario has a probability of occurring θ_s. Furthermore, it is assumed that no other scenarios exist outside of these. Therefore $\sum_{s=1}^{S} \theta_s = 1$. Additionally, each scenario has a known probability distribution that determines tomorrow's demand. Hence, the demand constraint can be thought as being subdivided into *S* different constraints, where each constraint is the realization of each possible scenario. Each constraint takes the difference of the total number of containers received by time *t* of tomorrow minus the mean aggregated demand for scenario *s* by time *t*. To strictly enforce this demand this difference should be positive. However, the demand is not known and we do not strictly enforce it. We achieve this by introducing a penalty variable ɛ for each of these constraints, which yields *S* penalty variables. Each of these variables determines how many containers we are not fulfilling under scenario *s*. We can then take these penalty variables and bring them to the objective. We then penalize each container that we are not fulfilling in this scenario, and weight that by the probability of that scenario happening.

If solved the model will yield container movements for today and tomorrow. However, the container movements for tomorrow are never actually made. Instead, the model works on a rolling horizon in which the model is solved at the beginning of each day. The container assignments for today are then fulfilled, then at the beginning of tomorrow, once the random variables have become realized, the model is solved again, and the new solution is then implemented. This is done iteratively ad infinitum solving two days at a time.

The advantages of the SDCAM compared to the deterministic model (DCAM) presented in [5] is that the SDCAM takes into consideration future container movements that need to be made in later days, and can take advantage of situations were moving a container preemptively might reduce truck miles. For example, suppose a container is needed today and it can be taken from either location *i* or location *j*. In the DCAM model the container would be taken from the location which minimizes the distance, let's say location *i.* However, there may be a good probability that location *i* will need more containers than location *j* tomorrow. Therefore, in order to reduce the cost and truck miles in the long run over all the days one should move the container from location *j* instead of location *i* and this would be considered in the SDCAM, but not in the DCAM. It is also for this reason that if SDCAM is run for only one day and compared to the DCAM, the DCAM will have less cost and trucking miles. This is because the DCAM is set to minimize the trucking cost and trucking miles for only one day, while the SDCAM is set to optimize the cost and trucking miles in the long run, and as such it will sometimes sacrifice local minimization, for a global minimization, as seen in the example.

Another situation where considering tomorrow's demand leads to a reduction of cost and truck miles is when one location has a large probability of needing a lot of containers. The DCAM optimizing for today, will simply not take this information into account by performing container movements that are only necessary to meet today's demand. However, the SDCAM will move some containers at the end of the day to either the location that needs the containers, or to a warehouse close by. Therefore, tomorrow's container movement will be reduced. This also has the incidental advantage that the SDCAM will proactively move containers more at the end of the day, probably at a time with less congestion in the streets, thus reducing congestion at peak hour times.

4. Vehicle Routing Problem

After solving the Container Assignment model, a truck schedule that supports such assignments needs to be found. Each container assignment is treated as a job that needs to be satisfied by an empty truck. Each job is defined by a starting location, an ending location, and a time window. This allows the conversion of the container assignments (both double container trips and single container trips) to be converted into jobs. For example, a container assignment $x_{i,j,t}$ and $y_{j,k,t+l_{i,j,t}}$ can be transformed into a job starting at time *t* that begins at location *i*, goes to location *j*, and ends at location *k*. This can be done for all container assignments to convert them to jobs. These jobs can then be converted to a time and space graph and an edge can be added from one job to another, if it is possible for a truck to service job *v* and then be able to move to the starting location for another job *w* before its start time. The cost of this edge is equal to the distance the truck must move to get from the ending location of job *v* to the starting location of job *w*. This can be done for all jobs resulting in a directed time and space graph with non-Euclidean distances. Solving a Vehicle Routing Problem (VRP) on this graph would yield a truck schedule supporting the container assignments that were previously found, thus finding a solution to the Empty Container Problem. The specific VRP represented by this graph is a VRP with tight time windows with non-Euclidean distances.

It is known that the VRP is an NP-hard problem, making it difficult to solve optimally for problems of practical size. Thus, to get a good solution for real sized problems, this research uses a modified version of Ropke and Pisinger's Adapted Large Neighborhood Search (ALNS) [34]. This algorithm is a tabu search algorithm that has no theoretical guarantees but has been shown to perform well on VRP problems with tight time windows. Its power lies on the ability to search large neighborhoods of variable size and move fast towards a solution, while being able to become infeasible at some iterations to escape local minima.

There are two main steps to solve the VRP using ALNS. The first step is finding an initial feasible solution, and the second is exploring the search space iteratively for good solutions until convergence or the max number of iterations is reached. Next, we present our algorithm for finding an initial feasible solution and our modified ALNS for our specific problem.

The first step needed to find an initial solution is to transform the container assignments into jobs, such that we can build a graph that will represent the problem. To do this we assume that the Container Assignment Problem has been solved and that the container movement variables ($x_{i,j,t}$, $y_{i,j,t}$, or $z_{i,j,t}$) are available. We then introduce our truck job nodes as $w_{i,j,k,t}$ in Figure 5.

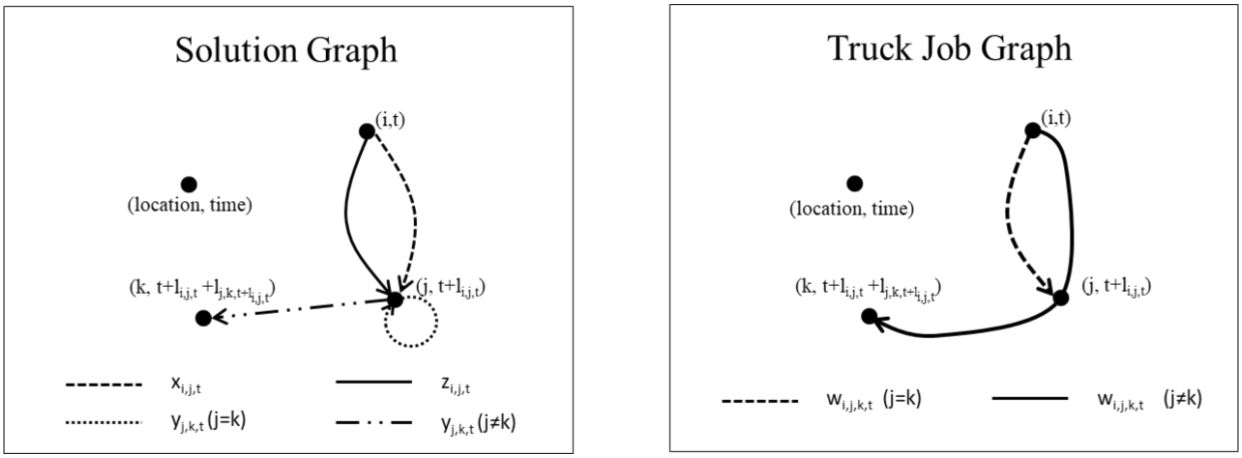

Figure 5. Truck Job Construction

These nodes refer to both single and double container truck trips, with the formal definition found below.

$$\textbf{\textit{Truck Job.}}\ w_{i,j,k,t} = \begin{cases} \frac{x_{i,j,t}+y_{j,k,t}}{2} + z_{i,j,t} & if\ j = k \\ \frac{x_{i,j,t}+y_{j,k,t+l_{i,j,t}}}{2} & otherwise \end{cases}$$

We can then graph these nodes and connect two jobs with a directed edge from node $w_{i,j,k,t}$ to node $w_{p,q,r,s}$ if $t + l_{i,j,t} + l_{j,k,t+l_{i,j,t}} + l_{k,p,t+l_{j,k,t+l_{i,j,t}}+l_{i,j,t}} \leq s$. That is, a single truck has enough time to serve job $w_{i,j,k,t}$ at time t and then have enough time to move from location k to location p before the start of job $w_{p,q,r,s}$ at time s. Figure 6 gives a simple demonstration of the network created by the truck job construction.

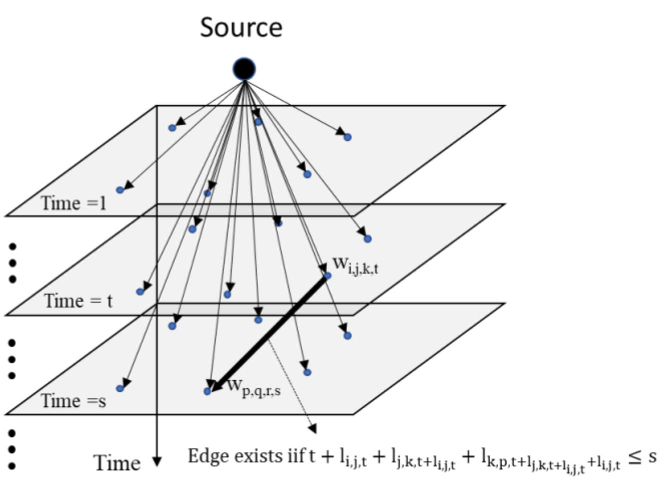

Figure 6. Truck Job Network

Before we formally present the algorithm to find an initial feasible solution we introduce some parameters. Each truck schedule is held in ordered sets L_p. Each ordered set L_p holds the truck schedule for truck *p*. The elements of each set are the container jobs $w_{i,j,k,t}$, with truck *p* serving the first element in the ordered set, moving to the second element, then serving the second element and so on until all elements have been served. For example, suppose L_p only has two elements with the first element being $w_{i,j,k,t}$ and the second element being $w_{\alpha,\beta,\gamma,s}$ (we will assume both jobs are double container jobs). Then the truck route for *p* would be to start its route at location *i* at time *t*. Pickup two containers and move to location *j* to drop the first container. Then drop the second container at location *k*. Then the empty truck *p* moves from location *k* to location α. Pickup two containers at time *s*, drop the first container at location β, and end its route at location γ. From construction there will be enough time for the truck to move from location *k* to location α, such that it arrives before time *s*. The next parameter σ_p represents the ending time of the last job in truck route *p*. That is, $\sigma_p = t + l_{i,j,t} + l_{j,k,t+l_{i,j,t}}$ were all subscripts come from the last element $\{i,\ j,\ k,\ t\}$ in the ordered set L_p. If the ordered set L_p is empty, then σ_p is set to 0. Finally, we let $\delta_{\alpha,i,t}$ be the travel time from the last location of the last job in the ordered set L_p (α) to location *i* arriving at time *t*. To be more specific if the last element in the ordered set L_p is $\{\beta, \gamma, \alpha,\ s\}$ and we are considering adding the tuple $\{i,\ j,\ k,\ t\}$, then $\delta_{\alpha,i,t}$= $o_{\alpha,i,t}$. We now present our algorithm to get an initial feasible solution below. We call this algorithm "VRP Initial Solution Construction".

VRP Initial Solution Construction

1. Solve the empty container problem to get the job variables ($x_{i,j,t}$, $y_{i,j,t}$, or $z_{i,j,t}$).
2. Set all $w_{i,j,k,t}$ such that

$$w_{i,j,k,t} = \begin{cases} \dfrac{x_{i,j,t} + y_{j,k,t}}{2} + z_{i,j,t} & \text{if } j = k \\ \dfrac{x_{i,j,t} + y_{j,k,t+l_{i,j,t}}}{2} & \text{otherwise} \end{cases}$$

3. Set $p = 1$
4. Set $\alpha = 0$
5. Choose a positive $w_{i,j,k,t}$ with the smallest t subscript such that:
   $t - \delta_{\alpha,i,t} \geq \sigma_p$. Break ties based on the smallest distance between α and i. Suppose we choose $w_{i',j',k',t'}$
6. Add the tuple $\{i', j', k', t'\}$ to the ordered set L_p.
7. Set $w_{i',j',k',t'} = w_{i',j',k',t'} - 1$
8. Set $\alpha = i'$
9. Repeat Steps 5 to 8 until no more $w_{i,j,k,t}$ can be chosen in Step 5.
10. If there is at least one positive $w_{i,j,k,t}$ set $p = p + 1$ and go back to Step 4. Otherwise STOP.

The algorithm above is a greedy algorithm that tries to minimize truck idle time. The idea being that if idle time is minimized then the number of trucks required to service all truck jobs is decreased. It does this by first adding a new truck to the system. It assigns the earliest possible job to this truck. It then calculates the ending time of the job and eliminates all truck jobs that cannot be serviced by the truck due to time window constraints. It then adds the earliest job the truck can service breaking ties by the necessary distance travelled to get to the job. The algorithm does this until no more jobs can be added to the truck. At this point, another truck is introduced, and the process is repeated until all jobs are serviced. This algorithm yields a feasible solution since at termination the number of jobs are finite and by construction each truck route is feasible. The algorithm (excluding the solving of the empty container problem) can be implanted in $O(n^2)$ time were *n* is the number of truck jobs.

After obtaining a feasible solution we can apply a modified version of ALNS. Before we present the algorithm we introduce some parameters. The first parameter ζ determines how many single job truck routes each iteration will try to eliminate. The second parameter Δ determines how many jobs will be removed and reinserted at every iteration. Notice that $\Delta \geq \zeta$ since removing one truck means that one job is also removed. The third parameter Ψ determines how many iterations of the heuristics will be performed. Conversely, the variable ψ gives the current iteration number. Furthermore, let p_{max} represent the maximum number of trucks that are currently being used. We introduce a new set *G* which will hold the removed jobs that later will need to be reinserted back to some route in order to preserve feasibility. Next, let ƨ represent the minimum cost of adding a job. Finally, ƕ holds the place where the minimum cost of inserting a job appears. We now introduce our modified ALNS.

Modified ALNS

1. Set $\psi = 1$
2. Set $\mathrm{p} = 1$
3. If ordered set $\mathrm{L_p}$ contains only one tuple. Remove it from the ordered set $\mathrm{L_p}$ and add the tuple to set G. Then set $\mathrm{p_{max}} = \mathrm{p_{max}} - 1$
4. If set G has ζ elements go to Step 6. Otherwise, CONTINUE.
5. If $\mathrm{p} = \mathrm{p_{max}}$, CONTINUE. Otherwise, set $\mathrm{p} = \mathrm{p} + 1$ and go back to Step 3.
6. Randomly remove any tuple from a random truck route ($\mathrm{L_p}$) and add it to set G.
7. If G has less than Δ elements go back to Step 6. Otherwise, CONTINUE.
8. Sort the tuples in G based on their starting time (t).
9. Remove the first tuple $\{\mathrm{i'}, \mathrm{j'}, \mathrm{k'}, \mathrm{t'}\}$ from G.
10. Set $\mathrm{p} = 1$
11. Set ƨ $= \infty$ and ƕ $= \{0\}$
12. If tuple $\{\mathrm{i'}, \mathrm{j'}, \mathrm{k'}, \mathrm{t'}\}$ can be inserted on truck route $\mathrm{L_p}$. Calculate the additional cost of inserting the job on route $\mathrm{L_p}$. If this cost is less than ƨ. Set ƕ $= \mathrm{p}$. Otherwise, CONTINUE.
13. If $\mathrm{p} = \mathrm{p_{max}}$, CONTINUE. Otherwise, set $\mathrm{p} = \mathrm{p} + 1$ and go back to Step 12.

14. If $\text{ƨ} < \infty$ insert tuple $\{i', j', k', t'\}$ to truck route ƕ. Otherwise, set $p_{max} = p_{max} + 1$ and add tuple $\{i', j', k', t'\}$ to truck route $L_{p_{max}}$.
15. If G is empty, CONTINUE. Otherwise, go back to Step 9.
16. If $\psi = \Psi$, STOP. Otherwise, set $\psi = \psi + 1$ and go back to Step 2.

As mentioned before the ALNS is a tabu search algorithm, that at each iteration it removes Δ truck jobs at random and then reintroduces them based on starting time and cost. Additionally, because more trucks are more expensive than any distance reduction, the algorithm prioritizes removing jobs from trucks that have only one truck job. The power of the algorithm is that it searches the solution space quickly, and can escape from a local minima. It achieves this by allowing the cost function to increase on some iterations. As shown in the paper by Ropke and Pisinger the algorithm in practice does tend to perform very well compared to other well-known algorithms.

5. Experimental Analysis

In the previous research [5], the benefits of using DCAM have been shown compared with the current container movement operations. In this study, these experiments are designed to compare the model performance under stochastic demand between DCAM and SDCAM. In this section we compare the two models on randomized data sets as well as on a data set representative demand near the port areas in Los Angeles County.

5.1 Randomized Experiments

To test the effect of including future stochastic demand when solving the Empty Container Problem, we perform experiments where we change both the demand distribution for each location, and the probability distribution for the different scenarios. In our experiments we represent three main events (scenarios) that could impact demand at the port: (1) no ship arrival, (2) a ship arrival, and (3) a ship departure. The transitional probabilities then give the probability of which event will occur tomorrow given the events (scenario) of yesterday and today. Each scenario is represented by a set of transitional probabilities and the demand distribution. We also assumed that the cost of using an additional truck is much greater than any mileage cost. For each experimental trial, we model ten days and it is assumed that day zero and the first day are scenario one (i.e., no ship arrival). We simulate a grid size of 25x25 with all locations being uniformly randomly located in the grid, except for the port which is located at the bottom center of the grid. We have a location capacity of 24 containers at the importers and exporters, a depot capacity of 36 containers, and a port capacity of 1500 containers. We also have an unloading and loading rate of containers of 1 hour at all locations. Additionally, the time it takes a truck to enter and leave the port is 2 hours. We ran the modified ALNS for 1000 iterations each day, removing 15 jobs and trying to eliminate a total of 3 trucks per iteration. We simulate a 12 hour day. These parameters are summarized in Table 1 below

Table 1. Parameter settings for randomized experiments

Parameter name	Parameter value
# of importers (I)	7
# of exporters (E)	5
# of depots (D)	2
Loading and unloading time of containers	1 hour
Location of port	bottom center
Truck turnover time at port	2 hours
Daily time horizon	12 hours
Time discretization size	15 mins
Grid size	25x25
Location capacity	24
Time horizon	10 days
Number of scenarios	3
Number of ALNS iterations (Ψ)	1000
Number of jobs to remove at each iteration (Δ)	15
Number of trucks to be removed at each iteration (ζ)	3

We ran three different transitional probability distributions representing the probability of tomorrow's scenario (i.e., no ship arrival, ship arrival, or ship departure) depending on the scenario of yesterday and today. These three different transitional probability distributions are shown in the Appendix. As seen in the tables, the transitional probabilities represented in Table A1 have the least variability while the ones in Table A3 have the most variability. We then ran five different uniform distribution cases for tomorrow's demand. For each case, the demand distribution also depends on the scenario for tomorrow. Similarly, the demand distributions in Table A4 have the least variability and these increase for each case until the 5th case (Table A8) which has the most variability. Finally, for each experiment we ran 10 replications.

We next test how including future stochastic demand affects our methodology for solving the Empty Container Problem. To do so in each replication we ran two models. In the first model we used DCAM to obtain the container assignments for each of the 10 days in our time horizon, and the modified ALNS model to obtain the truck schedule. In the second model, we used the SDCAM to obtain the container assignments for each of the 10 days in our time horizon, and then ran the modified ALNS model. Note that the SDCAM explicitly accounts for the uncertain future demand while the DCAM model does not. We compare both of these models against a solution knowing perfect information which is computed by assuming that the entire demand for the 10 days is known and the container assignments are solved collectively for these 10 days. Afterwards, these assignments are used as the jobs for the modified ALNS. Since in reality the demand is not known for the 10 days, the solution for this model will serve as a point for comparison to both the DCAM and the SDCAM results. We compare the results by using the ratio of either the SDCAM or the DCAM result over the perfect information model result. Below we show the average of the ratios in Table 2 and standard deviation of the ratios in Table 3 for our experiments.

Table 2. Average of the ratios for randomized experiments

		Transition Probability Distribution 1		Transitional Probability Distribution 2		Transitional Probability Distribution 3	
		SDCAM	DCAM	SDCAM	DCAM	SDCAM	DCAM
Demand Distribution	1	1.07	1.14	1.10	1.15	1.15	1.17
	2	1.12	1.18	1.16	1.20	1.22	1.24
	3	1.13	1.23	1.16	1.22	1.17	1.18
	4	1.16	1.25	1.19	1.25	1.18	1.22
	5	1.19	1.25	1.20	1.27	1.23	1.27

Table 3. Standard Deviation of the ratios for randomized experiments

		Transitional Probability Distribution 1		Transitional Probability Distribution 2		Transitional Probability Distribution 3	
		SDCAM	DCAM	SDCAM	DCAM	SDCAM	DCAM
Demand Distribution	1	0.025	0.026	0.038	0.044	0.038	0.047
	2	0.035	0.027	0.042	0.033	0.059	0.048
	3	0.041	0.048	0.045	0.030	0.050	0.048
	4	0.053	0.056	0.044	0.028	0.077	0.086
	5	0.057	0.041	0.055	0.033	0.073	0.061

As seen in Table 2, the SDCAM performs better than the DCAM; however, Table 3 shows that for the most part the SDCAM has a slightly higher standard deviation than the DCAM. SDCAM has a higher standard deviation because the DCAM will try to minimize today's demand and thus it will not make any unnecessary moves trying to minimize tomorrow's demand. It will also move containers in such a way to minimize the cost for today, lowering its standard deviation. However, the SDCAM will make some preemptive movements to try and minimize tomorrow's movement. These movements will sometimes pay off, but sometimes it will not. This will increase the variability of the model, but in the long run these movements lower the average cost for the system. It is for this reason that the SDCAM has a lower average cost, but a slightly higher standard deviation.

Another result that can be observed in Table 3 is that the ratios increases both as the demand and transitional probability variability increases. Both models perform worse when there is a higher variability. DCAM does not incorporate the transitional probabilities and instead only focuses only on the today's demand. Meanwhile, the SDCAM relies on the transitional probabilities to make a prediction on tomorrow's demand. If the variability for these probabilities starts to grow to the point where each scenario is equally likely then the model has no useful information on what tomorrow's scenario is going to be, and thus has less information about the demand. Meaning that the SDCAM would perform worse, because it will make unnecessary container movements more often.

In summary, SDCAM leads to solutions that use around 4% to 7% less truck miles to solve the Empty Container Problem as compared to the DCAM. SDCAM hedges container movements depending on what is the probabilities of tomorrow's demand and the distance that a container must move. This leads to SDCAM to preemptively move some containers, or to move containers from further away locations. In the long run this leads to a reduction in miles, at the expense of increasing the standard deviation, compared to a solution that will always move the containers the same way regardless of future demand.

5.2 Ports of Los Angeles and Long Beach Experiments

We next test how the DSCAM compares to the DCAM using data reflective of the Ports of Long Beach. We use the same Los Angeles and Long Beach scenario used in [5], with the same locations, distance matrix, and parameters. However, the Los Angeles and Long Beach port data available only has the mean demand for each location, but in order to test DSCAM different scenarios and demand distributions are needed. Thus, we use the same transition probabilities as in A1-A3 but the new demand distributions probabilities A9-A11 can be found in the Appendix. We also ran each trial for 10 days and each experiment was run for 10 trials. Table 4 below shows the parameters used in the experiments.

Table 4. Parameter settings for LA & LB Port experiments

Parameter name	Parameter value
# of importers (I)	5
# of exporters (E)	3
# of depots (D)	2
Loading and unloading time of containers	1 hour
Truck turnover time at port	2 hours
Daily time horizon	12 hours
Time discretization size	1 hour
Location capacity	10
Time horizon	10 days
Number of scenarios	3
Number of ALNS iterations (Ψ)	700
Number of jobs to remove at each iteration (Δ)	10
Number of trucks to be removed at each iteration (ζ)	2

For each experimental trial we ran three different models. In the first model we used the DCAM to obtain the container assignments for each of the 10 days in our time horizon, and the modified ALNS model. In the second model, we used the SDCAM to obtain the container assignments for each of the 10 days in our time horizon, and then ran the modified ALNS model. We compare both of these models against a solution knowing perfect information which is computed by assuming that the entire demand for the 10 days is known and the container assignments are solved collectively for these 10 days. Afterwards, these assignments are used as the jobs for the modified ALNS. Since in reality the demand is not known for 10 days, the solution for this model will serve as a point for comparison to both the DCAM and the SDCAM results. We compare the results by using the ratio of either the SDCAM or the DCAM result over the perfect information model result. Below we show the average of the ratios in Table 5 and standard deviation of the ratios in Table 6 for our experiments.

Table 5. Average of the ratios for LA & LB Port experiments

		Transition Probability Distribution 1		**Transitional Probability Distribution 2**		**Transitional Probability Distribution 3**	
		SDCAM	DCAM	SDCAM	DCAM	SDCAM	DCAM
Demand Distribution	1	1.04	1.10	1.05	1.11	1.07	1.13
	2	1.07	1.14	1.09	1.16	1.11	1.15
	3	1.10	1.16	1.12	1.18	1.14	1.19

Table 6. Standard Deviation of the ratios for LA & LB Port experiments

		Transitional Probability Distribution 1		**Transitional Probability Distribution 2**		**Transitional Probability Distribution 3**	
		SDCAM	DCAM	SDCAM	DCAM	SDCAM	DCAM
Demand Distribution	1	0.015	0.019	0.023	0.024	0.019	0.025
	2	0.017	0.011	0.045	0.030	0.036	0.019
	3	0.025	0.024	0.065	0.036	0.029	0.045

The results for the Ports of Los Angeles and Long Beach data sets are similar to the randomized experiments. Table 5 shows that the SDCAM performs around 4% to 6% better than the DCAM because it considers important information about future demand that the DCAM does not consider. Also, as predictions for tomorrow's state and demand locations improve the SDCAM solution will also improve.

6. Implementation

This problem addresses how to efficiently move empty containers to reduce the number of total trucks and truck miles that are required to meet demand. As more and more containers pass through ports every year, it becomes increasingly more important to efficiently move these containers. As shown in this report empty container reuse helps improve the container movement and reduce congestion at the port. Furthermore, it has been shown that if laws and infrastructure were to be modified to allow double container trucks to operate, there would be a lot of efficiency gained. We ran experiments, both on randomized data sets and using data from the Ports of Los Angeles and Long Beach to show that these gains can be significant. Additionally, the approach that this paper developed can be implemented to yield truck routes for both loaded and empty container movements. The implementation of our approach will require a programming language, such as Julia, and an optimization solver, such as Gurobi. Additionally, historical data needs to be collected to obtain the transitional probabilities between scenarios and probability density functions for the demands at the different locations.

7. Conclusion

In this report we build on the model proposed in [5] by incorporating future stochastic demand leading to a better solution in the long run. We test the model on randomized data sets and on scenarios based on data from the Port of Los Angeles and Long Beach. The results show that the stochastic model performs about 4-7% better in terms of total miles traveled than the model that does not take into account future demand. One future research direction could account for unforeseen events throughout the day that might delay some trucks or container movements. This may cause the current solution to become inefficient or even infeasible.

8. References

[1] J. Rogers, Ed., Review of Maritime Transport 2014. United Nations Publication, 2015.

[2] Bureau of Transportation Statistics. (2017). Transportation Statistics Annual Report 2017. *Bureau of Transportation Statistics.*

[3] “Container Statistics | Port of Los Angeles.” [Online]. Available: https://www.portoflosangeles.org/business/statistics/container-statistics.

[4] The Tioga Group, (2002). Empty ocean logistics study. Technical Report, Submitted to the Gateway Cities Council of Governments, CA, USA.

[5] Dessouky, M., & Carvajal, S. (2017). Congestion reduction through efficient empty container movement. National Center for Sustainable Transportation.

[6] Dejax, P. J., & Crainic, T. G. (1987). Survey paper-a review of empty flows and fleet management models in freight transportation. Transportation Science, 21(4), 227-248.

[7] Crainic, T.G., Gendreau, M., & Dejax, P. (1993). Dynamic and stochastic models for the allocation of empty containers. Operations Research, 41(1), 102-126.

[8] Shen, W.S., & Khoong, C.M. (1995). A DSS for the empty container distribution planning. Decision Support Systems, 15(1), 75-82.

[9] Li, L., Wang, B., & Cook, D.P. (2014). Enhancing green supply chain initiatives via empty container reuse. Transportation Research Part E: Logistics and Transportation Review, 70, 190-204.

[10] Song, D.P., & Dong, J.X. (2015) Empty Container Repositioning. Handbook of Ocean Container Transport Logistics. International Series in Operations Research & Management Science, vol 220.

[11] Bourbeau, B., Crainic, T. G., & Gendron, B. (2000). Branch-and-bound parallelization strategies applied to a depot location and container fleet management problem. Parallel Computing, 26(1), 27-46.

[12] Choong, S.T., Cole, M.H., & Kutanoglu, E. (2002). Empty container management for intermodal transportation networks. Transportation Research Part E: Logistics and Transportation Review, 38(6), 423-438.

[13] Choong, S.T. (2000) Container Management for Multimodal Transportation Networks. M.S. Thesis, Department of Industrial Engineering, University of Arkansas, Fayetteville, AR.

[14] Francesco, M.D., Crainic, T.G., & Zuddas, P. (2009). The effect of multi-scenario policies on empty container repositioning. Transportation Research Part E: Logistics and Transportation Review, 45(5), 758-770.

[15] Long, Y., Chew, E.P., & Lee, L.H. (2015). Sample average approximation under non-i.i.d. sampling for stochastic empty container repositioning problem. OR Spectrum, 37, 389-405.

[16] Braekers, K., Caris, A., & Janssens, G. K. (2013). Integrated planning of loaded and empty container movements. OR spectrum, 35(2), 457-478.

[17] Zhang, R., Yun, W.Y., & Moon, I. (2009). A reactive search algorithm for the multi-depot container truck transportation problem. Transportation Research Part E: Logistics and Transportation Review, 45(6), 904-914

[18] Funke, J., & Kopfer, H. (2016). A model for multi-size inland container transportation problem. Transportation Research Part E: Logistics and Transportation Review, 89, 70-85.

[19] Vidovic, M., Popovic, D., Ratkovic, B., & Radivojevic, G. (2017). Generalized mixed integer and VNS heuristic approach to solving the multisize containers drayage problem. International Transactions in Operational Research, 24, 583-614.

[20] Popovic, D., Vidovic, M., & Nikolic, M. (2014). The variable neighborhood search heuristic for the containers drayage problem with time windows. Advances in Intelligent Systems and Computing, 223, 351-364.

[21] Mladenovic, N., & Hansen, P. (1997). Variable neighborhood search. Computers & Operations Research, 24(11), 1097-1100.

[22] Tan, K.C., Chew, Y.H., & Lee, L.H. (2006). A hybrid multi-objective evolutionary algorithm for solving truck and trailer vehicle routing problems. European Journal of Operational Research, 173(3), 855-885.

[23] Sterzik, S., & Kopfer, H. (2013). A tabu search heuristic for the inland container transportation problem. Computers & Operations Research, 40(4), 953-962.

[24] Zhang, R., Yun, W.Y., & Moon, I. (2009). A reactive search algorithm for the multi-depot container truck transportation problem. Transportation Research Part E: Logistics and Transportation Review, 45(6), 904-914.

[25] Sterzik, S., & Kopfer, H. (2015). Reducing hinterland transportation costs through container sharing. Flexible Services and Manufacturing Journal, 27(2-3), 382-402.

[26] Clarke, G., & Wright, J. (1964). Scheduling of vehicles from a central depot to a number of delivery points. Operations Research, 12(4), 568-581.

[27] Jula, H., Chassiakos A., & Ioannou, P. (2006). Port dynamic empty container reuse. Transportation Research Part E: Logistics and Transportation Review, 42(1), 43-60.

[28] Chang, H., Jula, H., Chassiakos, A., & Ioannou, P. (2008). A heuristic solution for the empty container substitution problem. Transportation Research Part E: Logistics and Transportation Review, 44(2), 203-216.

[29] Lam, S.W., Lee, L.H., & Tang, L.C. (2007). An approximate dynamic programming approach for the empty container allocation problem. Transportation Research Part C: Emerging Technologies, 15(4), 265-277.

[30] Dam Hanh, Le. 2003. The logistics of empty cargo container in the southern California region. Metrans Research Project.

[31] Islam, S., Arthanari, T., & Olsen, T. (2010). Empty container-truck movement problem: at Ports of Auckland. 45th Annual Conference of the ORSNZ.

[32] Bandeira, D. L., Becker, J. L., & Borenstein, D. (2009). A DSS for integrated distribution of empty and full containers. Decision Support Systems, 47(4), 383-397.

[33] Erera, A.L., Morales J. C., & Savelsbergh, M. (2009). Robust optimization for empty repositioning problems. Operations Research, 57(2), 468-483.

[34] Ropke, S., & Pisinger, D. (2006). An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. Transportation Science, 40(4), 455-472.

9. Data Management

Products of Research

The primary data source was publicly available from the report "SCAG Regional Travel Demand Model and 2008 Model Validation" (LSA Associates, 2012), which was conducted by the Southern California Association of Governments (SCAG) in 2012. Within this model the Southern California region is divided into 4, 109 blocks and each block is associated with a pair of longitude and latitude, which represents its location. Also, randomly generated data sets were used.

Data Format and Content

All data sets were stored as excel files.

Data Access and Sharing

The data is stored in the Dryad open source repository.

Reuse and Redistribution

The citation for this data set is:

> Dessouky, Maged (2020), Data from: Congestion reduction through efficient container movement under stochastic demand, v3, Dryad, Dataset, https://doi.org/10.5061/dryad.mcvdncjxf

10. Appendix

Table A1. Transitional probability distribution 1

Tomorrow's State	Today's state	Yesterday's state	Probability
1	1	1	0.1
2	1	1	0.45
3	1	1	0.45
1	1	2	0.05
2	1	2	0
3	1	2	0.95
1	1	3	0.05
2	1	3	0.95
3	1	3	0
1	2	1	0.7
2	2	1	0.05
3	2	1	0.25
1	2	2	0
2	2	2	0
3	2	2	1
1	2	3	0.9
2	2	3	0.01
3	2	3	0.09
1	3	1	0.7
2	3	1	0.25
3	3	1	0.05
1	3	2	0.9
2	3	2	0.09
3	3	2	0.01
1	3	3	0
2	3	3	1
3	3	3	0

Table A2. Transitional probability distribution 2

Tomorrow's State	Today's state	Yesterday's state	Probability
1	1	1	0.2
2	1	1	0.4
3	1	1	0.4
1	1	2	0.1
2	1	2	0
3	1	2	0.9
1	1	3	0.1
2	1	3	0.9
3	1	3	0
1	2	1	0.6
2	2	1	0.1
3	2	1	0.3
1	2	2	0
2	2	2	0
3	2	2	1
1	2	3	0.8
2	2	3	0.05
3	2	3	0.15
1	3	1	0.6
2	3	1	0.3
3	3	1	0.1
1	3	2	0.8
2	3	2	0.15
3	3	2	0.05
1	3	3	0
2	3	3	1
3	3	3	0

Table A3. Transitional probability distribution 3

Tomorrow's State	Today's state	Yesterday's state	Probability
1	1	1	0.3
2	1	1	0.35
3	1	1	0.35
1	1	2	0.2
2	1	2	0
3	1	2	0.8
1	1	3	0.2
2	1	3	0.8
3	1	3	0
1	2	1	0.5
2	2	1	0.15
3	2	1	0.35
1	2	2	0
2	2	2	0
3	2	2	1
1	2	3	0.7
2	2	3	0.1
3	2	3	0.2
1	3	1	0.5
2	3	1	0.35
3	3	1	0.15
1	3	2	0.7
2	3	2	0.2
3	3	2	0.1
1	3	3	0
2	3	3	1
3	3	3	0

Table A4. Demand Distributions 1

Location	State 1	State 2	State 3
Importer 1	(59,63)	(63,67)	(28,32)
Importer 2	(69,73)	(73,77)	(43,47)
Importer 3	(46,50)	(78,82)	(38,42)
Importer 4	(63,67)	(68,72)	(41,45)
Importer 5	(52,56)	(83,87)	(43,47)
Importer 6	(49,53)	(62,66)	(38,42)
Importer 7	(62,66)	(81,85)	(33,37)
Exporter 1	(48,52)	(53,57)	(83,87)
Exporter 2	(58,62)	(48,52)	(78,82)
Exporter 3	(38,42)	(43,47)	(93,97)
Exporter 4	(53,57)	(41,45)	(73,75)
Exporter 5	(43,57)	(58,62)	(86,90)
Port	(400,425)	(365,385)	(468,493)

Table A5. Demand Distributions 2

Location	State 1	State 2	State 3
Importer 1	(56,66)	(60,70)	(25,35)
Importer 2	(66,76)	(70,80)	(40,50)
Importer 3	(46,50)	(75,85)	(35,45)
Importer 4	(60,70)	(65,75)	(38,48)
Importer 5	(50,58)	(80,90)	(42,48)
Importer 6	(48,55)	(58,70)	(38,43)
Importer 7	(58,70)	(78,88)	(30,40)
Exporter 1	(45,55)	(50,60)	(80,90)
Exporter 2	(55,65)	(45,55)	(75,85)
Exporter 3	(35,45)	(40,50)	(85,95)
Exporter 4	(50,60)	(38,48)	(70,80)
Exporter 5	(40,50)	(55,65)	(83,93)
Port	(375,450)	(350,400)	(443,518)

Table A6. Demand Distributions 3

Location	State 1	State 2	State 3
Importer 1	(54,68)	(58,72)	(23,37)
Importer 2	(64,78)	(68,82)	(38,52)
Importer 3	(44,52)	(73,87)	(33,47)
Importer 4	(58,72)	(63,77)	(36,50)
Importer 5	(48,60)	(78,92)	(40,50)
Importer 6	(46,57)	(56,72)	(36,45)
Importer 7	(56,72)	(76,90)	(28,42)
Exporter 1	(43,57)	(48,62)	(78,92)
Exporter 2	(53,67)	(43,57)	(73,87)
Exporter 3	(33,47)	(38,52)	(83,97)
Exporter 4	(48,62)	(36,50)	(68,82)
Exporter 5	(38,52)	(53,63)	(81,95)
Port	(370,455)	(345,405)	(438,523)

Table A7. Demand Distributions 4

Location	State 1	State 2	State 3
Importer 1	(51,71)	(55,75)	(20,40)
Importer 2	(61,81)	(65,85)	(35,55)
Importer 3	(41,55)	(70,90)	(30,50)
Importer 4	(55,75)	(60,80)	(33,53)
Importer 5	(45,63)	(75,95)	(37,53)
Importer 6	(43,60)	(53,75)	(33,48)
Importer 7	(53,75)	(73,93)	(25,45)
Exporter 1	(40,60)	(45,65)	(75,95)
Exporter 2	(50,70)	(40,60)	(70,90)
Exporter 3	(30,50)	(35,55)	(80,100)
Exporter 4	(45,65)	(33,53)	(65,85)
Exporter 5	(35,55)	(50,70)	(78,98)
Port	(350,475)	(325,425)	(418,543)

Table A8. Demand Distributions 5

Location	State 1	State 2	State 3
Importer 1	(49,73)	(53,77)	(18,42)
Importer 2	(59,83)	(63,87)	(33,57)
Importer 3	(39,57)	(68,92)	(28,52)
Importer 4	(53,77)	(58,82)	(31,55)
Importer 5	(43,65)	(73,97)	(35,55)
Importer 6	(41,62)	(51,77)	(31,50)
Importer 7	(51,77)	(71,95)	(23,47)
Exporter 1	(38,62)	(43,67)	(73,97)
Exporter 2	(48,72)	(38,62)	(68,92)
Exporter 3	(28,52)	(33,57)	(78,102)
Exporter 4	(43,67)	(31,55)	(63,87)
Exporter 5	(33,57)	(48,72)	(76,100)
Port	(345,480)	(320,430)	(413,548)

Table A9. Demand Distributions 1 for the San Pedro Ports Scenarios

Location	State 1	State 2	State 3
Importer locations	(38,42)	(43,47)	(33,37)
Exporter locations	(28,32)	(23,27)	(33,37)
Port	(195,205)	(185,195)	(205,215)

Table A10. Demand Distributions 2 for the San Pedro Ports Scenarios

Location	State 1	State 2	State 3
Importer locations	(36,44)	(41,49)	(31,39)
Exporter locations	(26,34)	(21,29)	(31,39)
Port	(190,210)	(180,200)	(200,220)

Table A11. Demand Distributions 3 for the San Pedro Ports Scenarios

Location	State 1	State 2	State 3
Importer locations	(34,46)	(39,51)	(29,41)
Exporter locations	(24,36)	(19,31)	(29,41)
Port	(185,215)	(175,205)	(195,225)