UNIVERSITY OF MARYLAND

National Transportation Center

Project ID: NTC2016-MU-R-05

METHODS FOR IMPROVING THE RELIABILITY OF TRANSPORTATION SYSTEMS

Final Report

by

George F. List
gflist@ncsu.edu
919-515-8038
North Carolina State University
909 Capability Drive, Suite 3600
Research Building IV
Raleigh, North Carolina 27606

and

Atefeh Morsali
Jeremy Addison
Elizabeth Byrom
North Carolina State University

for

National Transportation Center at Maryland (NTC@Maryland)
1124 Glenn Martin Hall
University of Maryland
College Park, MD 20742

July 2018

ACKNOWLEDGEMENTS

The research team is deeply appreciative of the support received from the collaboration with Dr. Paul Schonfeld at UMD who helped shape and focus the effort. A deep sense of gratitude is also owed the late Dr. Mark Turnquist, with whom the lead author worked for more than a decade. Many of the ideas presented here, especially in Section 4, are derived from that collaborative work. The project was funded by the National Transportation Center @ Maryland (NTC@Maryland), one of the five National Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT).

DISCLAIMER

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This research project focused on studying ways in which carriers can plan freight services in a stochastic environment, with an eye toward maximizing reliability and efficiency. It used truck system operations as a context although the tools developed can be applied to other modes and multi-modal systems. Studying the operation of transport systems under stochastic conditions is important because the negative service quality impacts of stochasticity can be significant and there are losses in economic value.

As is well recognized, carriers aim to provide services where package pickups and deliveries occur on-time, within allowable windows (OTWs) a very high percentage of the time. This maximizes customer satisfaction and bolsters brand loyalty. To do this, carriers not only have to manage the first and last transport events, but also the intermediate activities that occur at terminals and distribution hubs. That is, to provide the high-quality service, all the intermediate activities, such as the handling of packages at intermediate hubs, must also be punctual. OTWs must pertain to these events as well; and if these intermediate events occur within these on-time windows, then the first and last events are likely to occur on-time as well.

Research efforts have addressed this problem from both a deterministic and stochastic perspective. In the deterministic studies, everything is assumed to be fixed. The service times and travel times are constant, often assumed to be at their average values. The packages scheduled to be picked up and delivered are all known. The times when they should be picked up and dropped off are also known. In the stochastic studies, some or all these values can vary. In the work presented here, the travel times and load and unload times are stochastic as well as the probabilities that package transport requests will materialize in any realization of the system.

Carrier options at the strategic, tactical, and operational levels are examined. At all three levels, there are actions the carrier can take to improve reliability and enhance service quality.

For example, at the strategic level, carriers can select hub locations and determine their truck fleet sizes. In this study, the hub locations are assumed fixed, but the fleet sizes can vary. In an earlier project, see List *et al*. (2017), the effects of hub location choice were examined. In the case of fleet size, larger fleets provide more flexibility, increase the likelihood that on-time windows will be met, and allow door-to-door times to be shorter. More packages can be handled in less time. Delays are reduced.

At the tactical level, carriers can determine how long the planning horizon is and how many hours to work each day. They can decide the distribution of trucks among hubs, and they can choose the service requests to accept. In this study, the planning horizon is assumed to be a week, which is consistent with the load cycle faced by many carriers. The distribution of trucks is treated as an exogenous input whose impact is explored through a sensitivity analysis. The impacts of variations in demand are addressed by randomly choosing the service requests that are extant in each of the problem realizations.

At the operational level, carriers can select the rules by which the system operates and determine how packages are assigned to trucks. In this study, the over-the-road trucks are treated as a free running fleet that is constantly in operation. The trucks start their tours each week from a home depot. Then they migrate from one load to the next, departing either when they are full or when a maximum headway between departures has been reached. The local trucks do both pick-ups and deliveries during a given tour. They make one tour per day. The assignment of packages to trucks in both cases is done using block-building procedures that are like the ones described in prior project reports, see List *et al.* (2018). All these details are discussed more completely in later sections.

Three case study settings are addressed. They reflect the variety of operating conditions that trucking companies can face. The first is a back-and-forth operation between two locations. It represents situations where a dedicated fleet of trucks is involved in a single operation. One example is trucks moving parts from a component manufacturing plant to a main assembly plant. The second setting is a truckload (TL) operation where full-truck shipments are moved from one location to another. One example is the movement of chemicals from manufacturing plants to customers. Another is the distribution of merchandise from warehouses to retail stores. The third setting is a less-than-truckload (LTL) operation where packages are picked up, carried from hub to hub and then delivered. UPS is an example of this type of service. The postal service is another.

In all these settings, one study objective is to see how the truck fleet size effects on-time performance. For example, in the first setting, where the trucks carry truckload shipments from “A” to “B” and vice versa, the travel times vary and so do the load-unload times. For the extreme cases where the travel times and the load-unload times are both large, the fleet size limits the carrier’s ability to have all the shipments be on time. One question is: what fleet size is required to ensure that the quality of service (on-time performance) is at or above a target level (e.g., 95% of the deliveries are on-time). Another is, more generally, how does the fleet size effect the on-time performance? Put differently, what is the fleet size beyond which further quality of service improvements are difficult to achieve?

In the second setting, the focus is again on seeing how the fleet size affects the service quality. But here, the service times, travel times, and fleet size are all user inputs. On-time performance suffers if the travel times are too long, the service times are too lengthy, or the fleet size is too small. Through parametric variation, it is possible to determine the sensitivity of the on-time performance to the values of these inputs. If the fleet size is too small, shipments are delayed. Increasing the fleet size reduces those delays. At some point, since the problem is deterministic, the delays go to zero because the fleet size is adequate. At that point, there is slack in the schedule. Trucks arrive early for loads. Measuring that slack gives a sense of reserve capacity that exists in the system to cope with unexpected delays and abnormally long load and unload times.

In the third setting, packages are picked up and delivered by local trucks and transported between hubs by other, larger trucks. The load and unload times vary, as do the travel times. The packages have a probability of being present in any given problem realization. A major question again is how the fleet size affects service quality.

This study shows that tools can be developed to schedule truck operations in uncertain environments. These tools show how resource levels affect performance, and they illustrate the fact that in stochastic settings no single operating plan is always optimal. An ability to vary the truck dispatching plan is critical to achieving the best performance possible. Put another way, a nimble tool is valuable in achieving the best possible system performance when conditions are stochastic and resources are limited. Moreover, increasing the resource levels improves performance and reduces the sensitivity of system performance to the idiosyncrasies of specific scenarios.

The report is organized as follows. Section 1 provides an overview of the study. Section 2 reviews pertinent literature. Section 3 examines the two-node, single link network and illustrates tradeoffs between resource levels and service quality. Section 4 analyzes the truckload network problem and shows how resource levels affect system performance in a more complex environment. Section 5 examines the less-than-truckload setting where packages are transported between regions. The relationship between resource levels and system performance is again illustrated, this time in a more complex setting. Finally, Section 6 summarizes the effort and identifies opportunities for future work.

1.0 INTRODUCTION

This research focuses on studying ways in which carriers can maximize the quality of their freight services in a stochastic environment, with an eye toward reliability and efficiency. The attention is on developing and testing tools that can help carriers make informed choices about truck system operations although the tools developed can be applied to other modes and multi-modal systems. Studying the operation of transport systems under stochastic conditions is important because stochasticity can have adverse effects on service quality and there are economic consequences from those impacts.

As is probably well recognized, carriers aim to provide services where package pickups and deliveries occur within on-time windows (OTWs) a very high percentage of the time. This helps maximize customer satisfaction and ensure service loyalty. To make this happen, carriers not only have to focus on the first and last transport events, but also the intermediate activities that occur in-between at terminals and distribution hubs. To provide the high-quality service, all intermediate activities must be punctual, like the handling of packages at system hubs. Effectively, "on-time" windows must pertain to these intervening events as well; if these on-time windows are achieved, then the first and last events are more likely to occur on-time as well.

Previous studies have addressed the service planning problem from both a deterministic and stochastic perspective. In the deterministic case, everything is assumed to be fixed. The service times and travel times are constant, often assumed to be at their average values. The demands for service are known. In the stochastic case, values vary. In the work presented here, the travel times and load/unload times are stochastic and package transport requests have a probability of emerging.

Options for enhancing reliability exist at the strategic, tactical, and operational levels. The strategic level choices affect system configuration and size. The operational level choices deal with day-to-day operations. The tactical choices are in-between.

More specifically, at the strategic level, carriers can alter hub locations and change truck fleet sizes. Better hub locations provide access to more reliable routes; more hubs reduce travel distances and times. Larger fleet sizes provide more flexibility, increase the likelihood that on-time windows will be met, and allow the door-to-door times to be shorter. More packages can be handled in less time because system capacity is larger. On-time performance is improved. Unexpected events can be accommodated more easily. In this study, the hub locations are assumed fixed, but the fleet size can vary.

At the tactical level, carriers can set the planning horizon, the workday schedules, the distribution of trucks among hubs, and the service requests to accept. Longer horizons have the potential to produce better plans. Better distributions of the trucks, that align with the hub workloads, produce better service. Careful choices of loads to accept result in service request patterns that are easier to accommodate. In this study, the planning horizon is treated as being either a week or a day. In many truck service settings, the demand patterns repeat on a weekly basis. Most trucking firms operate around-the-clock, but the bulk of the activities, especially pickups and deliveries, occur during the day. In this study, both the planning horizon and the weekday hours are varied through sensitivity analysis. The choices of requests to accept are addressed through a random selection process in which service requests have a probability of arising in each scenario examined.

At the operational level, carriers can determine how the trucks operate and how the packages are assigned to the trucks. Locally, trucks can either be used for pickups, or deliveries, or both. They can make one tour each day. Packages can be assigned to trucks on a geographic basis, a temporal basis, or some other set of criteria. In this study, the local trucks are assigned to hubs. They make one tour per day and can do both pickups and deliveries. They are assigned to loads on a "soonest to arrive" basis. The over-the-road trucks are a free running fleet. They start from a specific hub at the beginning of the week, but after that they travel from hub to hub in response to the loading demands. The earliest available truck that is closest to the demand is selected. Trucks are released from a given location when they are either full or the maximum headway between departures has been reached. No timetable is followed.

1.1 STUDY DESCRIPTION

The study team elected to explore the implications of stochasticity by examining three different operational settings. Studying one would not have been enough. The industry is too complex. The operating modes are too diverse. Even three is marginally adequate. The settings span the gamut of operating conditions typically confronted by service providers. They range from simple to complex, from small to large scale.

The first is a simple two-node network where trucks travel back-and-forth, say A and B. Such services arise when items manufactured or stored at A are inputs to the production process or activity at B. In addition, the shipment volumes are large enough to justify a fleet of dedicated trucks. This situation arises in the automotive industry when component parts such as engines are produced in one plant and then used to assemble vehicles at another. One more example is a situation where grain is stored at a local elevator and then transported to a mill for processing.

The second setting is a multi-hub less-than-truckload (LTL) network where packages are transported from origins to destinations. These operations are typified by carriers like the US Mail, FedEx Freight, XPO Logistics, Old Dominion Freight, and UPS. Regional hubs manage local pickup and delivery services and interface with the over-the-road operations. Shipments between hubs are carried by full truckloads.

The third setting is a large-scale, over-the-road truckload operation where whole truckloads need to be carried from one location to another. This is illustrated by big box retailers where truckloads of merchandise are moved from distribution centers to stores. It also arises in the case of mail trucks moving between processing centers, or the inter-hub shipments of LTL carriers like FedEx Ground and UPS, or truckload carriers like Swift, Schneider, or Chemical Leaman.

Stochasticity affects all three settings. One notable example is variations in network travel times caused by incidents, weather, and traffic congestion. This impact is outside the control of the trucking firm. A second is variability in pickup and delivery times caused by unforeseen challenges in loading and unloading the shipments. Another, more at the tactical level, is variability in the demands which causes changes in the assignment of packages to trucks and, as a result, truck utilization.

Because the conditions of these systems are assumed to vary, there are no unique, optimal ways to operate the system. There is no mathematical problem statement that can be formulated to produce a unique best possible answer. The best solution depends on the conditions that pertain in any given realization of the problem setting. The real insight lies in seeing trends in the solutions, like the probability that a specific operating plan is employed the relationship between fleet size and on-time performance; or the value of providing schedule slack.

Hence, the line of inquiry should be predicated on a sensitivity analysis. The questions should be: what parameters have the most impact? How much impact do they have? Are there combinations of parameter values that produce the best results? How good are the results?

An obvious example is the fleet size. It seems obvious that larger fleets provide more flexibility. The workload for each truck is reduced. Trucks that are not fully utilized can put into play more intensely when the conditions stress the system. Clearly, truck utilization is diminished, but service quality is improved.

Another parameter is schedule slack. This is dead time allowed between the arrival of one truck and the departure of another. It can also be extra time allowed in the hub-to-hub travel time or the local tour to ensure that on-time arrivals can occur. If the slack is small, there is a risk that connections will be missed. Service quality will suffer. But when conditions are favorable, door-to-door travel times will be short. Service quality will be perceived as being quick and highly responsive. If the slack is large, on-time deliveries and pickups will be highly likely, but the door-to-door transit times will be large. The questions are like those pertaining to fleet size. What amount of slack is required to ensure that the quality of service (on-time performance) is at or above a target level (e.g., 95% of the deliveries are on-time). Another is, more generally, how does the amount of slack provided affect the on-time performance? Put differently, is there an amount of slack beyond which further quality of service improvements are difficult to achieve?

1.2 SETTING THE CONTEXT

As a prelude to presenting the study results, it seems useful to describe how reliability is often perceived in the context of freight transportation. This will help the reader understand the reasoning behind the work presented. Moreover, since freight is the focus, the report focuses on the reliability of package shipments and truck movements rather than personal trips and auto movements.

Leemis (2009) offered a recent definition of reliability: “The reliability of an item is the probability that it will adequately perform its specified purpose for a specified period under specified environmental conditions.” This is suitable for assessing the mean time to failure for a physical device, but it does not pertain particularly well to the reliability of transport services.

A slight shift in terminology makes Leemis’s definition fit well with freight. A shipment delivery can be regarded as the “device” and the absolute difference between the arrival time window (ATW) and the actual time of arrival (ATA) can be the metric monitored. A trip termination or customer visit is considered reliable if the ATA is within the DTA window. That is, if t_a is the arrival time, a_b is the beginning of the DTA window, and a_e is the end, then $d_a = \max(a_b - t_a , t_a - a_e, 0)$ is the deviation from being on-time. If $d_a = 0$, then the delivery was on time, and if $d_a > 0$ then it was not.

Evidence that this is the perspective of the freight industry is illustrated by the trade press, which uses on-time performance to highlight the performance of the shipment industry. Success occurs if an arrival or a customer visit begins within the applicable DTA window. Actions that increase the likelihood of that occurring increase the reliability of the service.

The same comments pertain to the departure time window (DTW). If t_d is the departure time, d_b is the beginning of the ATW, and d_e is the end, then $d_d = \max(d_b - t_d, t_d - d_e, 0)$ is the deviation from being on-time. If $d_d = 0$, then the departure is on-time, and if $d_d > 0$ then it is not. Hence, freight travel time reliability is about consistency in on-time arrivals and departures, not travel times per se. The focus needs to be on probability density functions (PDFs) and cumulative distribution functions (CDFs) for d_a and d_d.

A graphical way to think about these ideas is shown in Figure 1.1. If the ATA is within the DTA window, then an OTA has occurred. The freight transportation system's reliability can be measured by the percentage of trips that have ATAs within their DTA windows.

Using this information, carriers can assess the percentage of all ATAs that fall within their DTA windows. Customers can do this as well. They can adjust their DTA windows so that the on-time performance is improved. Carriers can adjust their departure times, or add slack to the trip time, so that the probability of arriving within the DTA window is maximized.

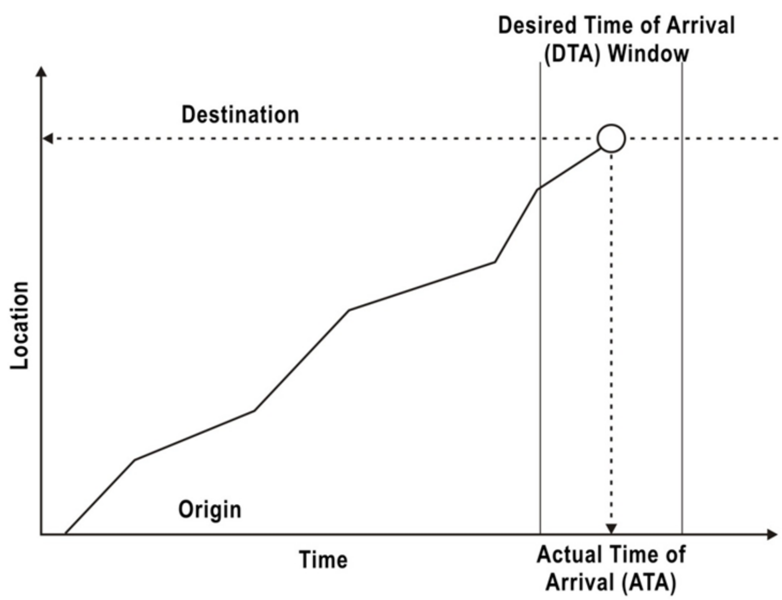

Figure 1.1: Concepts of desired and actual times of arrival

These thoughts have an interpretation that is based in utility theory, as described by Hansson (1994). Each trip has a disutility that reflects the "cost" of making the trip. That cost includes the travel time, tolls, and other expenses; in this case, the "cost" of being either late or early is of the highest importance. That delivery-related cost is zero if the ATA is inside the DTA window. It becomes non-zero if the ATA is outside the DTA window, either before or after. Moreover, the cost of being late may be different from that associated with being early. The slope of the cost curve in figure 1.2, indicates the per-unit-time penalties involved. The steeper the slope, the costlier it is to be late or early. In the aggregate, the on-time costs of the trips can be summed to assess the "societal cost" of the unreliability of the system.

In the situations where the focus is only on the shipments' arrival time, not their departure time, the main questions are: 1) when should the shipment leave to maximize the probability of arriving within the DTA window, or 2) how can one strike a balance between the travel time and on-time performance, or minimize the total generalized cost.

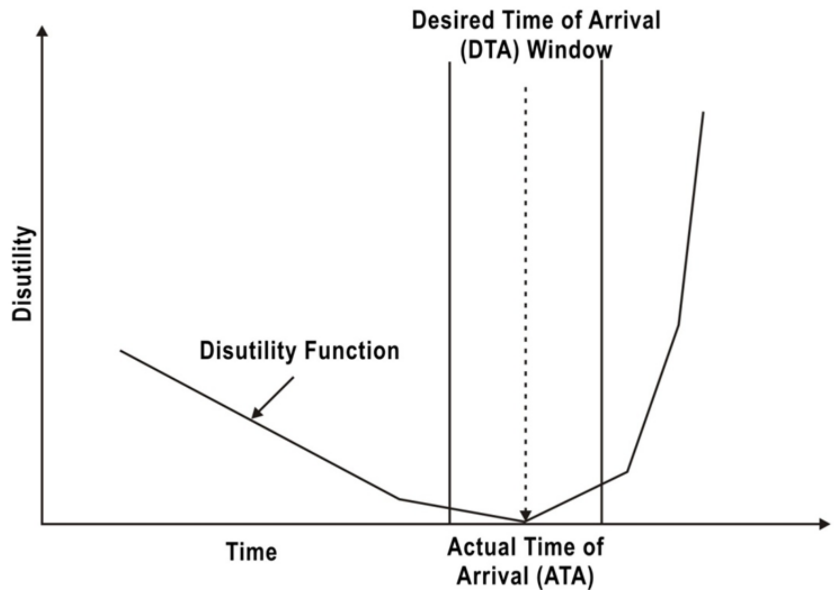

Figure 1.2: Disutility function to characterize desired and actual times of arrival

These notions implicitly suggest that shorter travel times are not necessarily the objective, although reduced travel times are still of great importance. Instead, the focus is on maximizing the likelihood of on-time arrivals, and making intelligent tradeoffs between this and other metrics such as cost. There are benefits to smaller travel times: later departure times or earlier arrivals can be employed, more deliveries of pickups can be made within a given time span, more such activities can occur for a given fleet size, and more customers can be serviced for a given fleet size. In-transit inventory costs and fleet size requirements are also reduced. However, the travel times should not be reduced at the expense of poorer on-time performance.

Thus, a carrier's objective is to arrive and depart within the on-time windows (OTWs) a very high percentage of the time. If a carrier has high percentages for both their on-time arrivals (OTAs) and on-time departures (OTDs), their service will be perceived as high quality. For the OTAs, if the truck arrives early and must wait, less carrier resources could have been used or extant resources could have been better deployed. If it arrives late, customer expectations have not been met. For the OTDs, the same two thoughts still pertain. If the truck departs early, then the same resource issues exist. And if it departs late, customer resources (dock time and space) were tied-up unnecessarily.

As is probably obvious, achieving on-time departures and arrivals 100% of the time is impossible. Systems do not operate deterministically. Travel times vary. Network conditions change. Loading and unloading times vary. Hence, carrier performance must be measured in terms of the probability that on-time events occur.

When only an ATW exists, the prior DTW is implicitly identified by percentile bounds on when departures must occur for the ATAs to be within the ATW. Assessing this window is best done backwards because the ATW is the constraint, but in many studies, it is done in the forward direction instead. The "best" path maximizes the likelihood that an OTA will occur.

Carriers, such as trucking companies face situations where both DTWs and ATWs exist. When this is the case, the carrier measures its on-time performance in terms of both departure and arrival events, separately and in combination. Shipments (and vehicle moves) are deemed to be "on-time" if they both depart during the DTWs and arrive within the ATWs. A joint density function can be used to track this performance.

Maximizing the on-time performance is sensitive to the operating conditions under which the trips are made. Different departure times may be considered best for different operating conditions. For example, earlier departures may be preferable when the weather is inclement, network maintenance is underway, or the network is heavily congested.

Maximizing ATW performance requires decision-maker actions. The packages cannot move by themselves; they cannot make decisions about when to depart, what route to follow, or what truck to use. The shipments must be managed, handled, and transported by people. People must take actions. Hence, the objective in this research has been to develop tools that help those decision makers. The tools should facilitate their decision making.

Basic ideas about measuring travel times for reliability are presented in List *et al*. (2017). Trip times are described as comprising 1) transport times across links and 2) processing times at nodes. Sometimes the processing times are significant, as in the case of pickup and delivery times.

Figure 1.3 shows a carrier's hub-to-hub network. The regional hubs are dots. They are surrounded by boxes. Each regional hub has a letter designation (A through H). Connections between the regional hubs are shown as links. The word "link" means a two-way connection. Arc means one-way and implies directionality as in the arc DB, which originates at D and terminates at B.

When a truck leaves a hub, it passes through the box on the departing arc. It arrives at a hub as it passes through the box on the arriving arc. Arc travel times arise between the boxes (on the arc). Processing times occur between the arriving and departing boxes at the node.

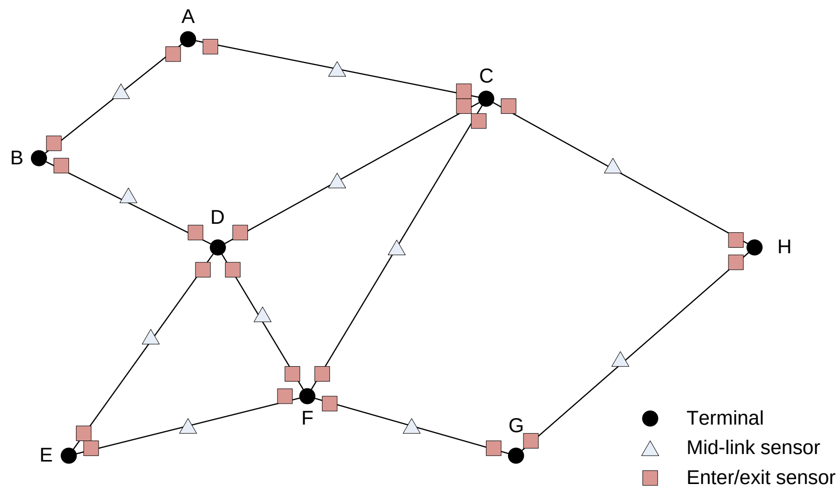

Figure 1.3: A hypothetical hub-to-hub network and event monitoring locations

For truck trips, there are arc transit times and nodal processing times. If the truck makes a trip from B to H, there is the loading time at B, a travel time on arc BD, loading/unloading time at D, a travel time on arc DC, loading/unloading time at C, a travel time on arc CH, and unloading time at H.

Insofar as the arcs are concerned, their travel rates (inverses of the space-based speeds) may vary temporally and spatially. The travel time from B to D on arc BD may be produced by travel rates that vary by time and location. These rates vary because of changes in capacity, congestion, weather, incidents, maintenance work, etc. These "operating environment" variables determine the travel rates that are achieved. These variables describe the "operating condition."

1.3 REPORT OVERVIEW

The remainder or the report is organized as follows. Section 2 reviews pertinent literature. Section 3 examines the two-node, single link network and illustrates tradeoffs between resource levels and service quality. Section 4 analyzes the truckload network problem and shows how resource levels affect system performance in a more complex environment. Section 5 examines the less-than-truckload setting where packages are transported between regions. The relationship between resource levels and system performance is again illustrated; this time in a more complex setting. Finally, Section 6 summarizes the effort and identifies opportunities for future work.

2.0 PERTINENT RESEARCH

This section reviews prior research efforts that have focused on topics that are the same as or like the one addressed here. A prior project report, see List *et al*. (2017), provided a more comprehensive review of the literature focused on truck system reliability.

2.1 TRANSPORT SYSTEM RELIABILITY

Efforts focused on transport system reliability are the ones most closely aligned with the study presented here. They focus on the integrative manifestations of the reliability that arises on the links and at the nodes. They reflect the effects of operational and strategic decisions made to mitigate the impacts of travel time variability.

For example, just-in-time delivery systems typically buffer travel time variability by planning on early arrivals and time spent waiting. That is a strategy examined in this study as well. Systems that have one or more transshipment points can buffer the variability in link travel times by carefully setting the arrival and departure times of vehicles. The challenge is to time them so that the connection times for transferring the freight will be adequate. For rail systems, it is setting the schedules for inbound and outbound trains, as with the trucks, plus the sequencing of train classifications for inbound trains and the assignment of yard tracks to blocks. Of course, short connection times produce missed connections and delays; long connection times, while they improve the connection probability, add to the time shipments spend in the system. In that regard, for systems that are well instrumented, dynamic decisions can adapt the operating plan to the evolving conditions.

One of the earliest papers on system reliability is Detmold (1972). He discussed the issue of specifying timetables for train operations. His assertion was that looking at the timetable alone was insufficient. It was important to employ an integrated approach that combined shipper and carrier costs when determining the operational strategy. He considered how to set the standards of rail service at levels that would offer the best compromise between shipper needs and rail costs. From the railway's perspective, he perceived that this meant creating service schedules and timetables that ensure there is sufficient slack to make up time after delays. From the shippers' perspective, he believed it involved holding inventories that guard against stock-outs. The monograph describes some routines Detmold developed for assessing the optimal combinations of service to offer.

Contemporaneous with Detmold (1972), other authors were providing perspectives on the connection between carrier operations and reliability. Lang and Reid (1970), Martland and Sussman (1972), Williams (1972), and Shamberger (1975) all suggested ideas about how the reliability of railroad operations could be improved. All of them related to the reliability of train-to-train connections in yards and to equipment reliability for over-the-road trains. Martland and Sussman (1972) examined the impacts on the reliability of railroad services that could be created by improved operations practices. Their paper can be viewed as 'scoping' railroad performance with respect to service time reliability. They analyzed railroad data to understand the level of service then being provided by the railroads as well as the reasons for the observed differences in the service provided to individual O-D pairs.

Sines (1972) described, the freight car scheduling (FCS) system being developed by the Missouri Pacific Railroad (MoPac). The intent of this system was to allow the railroad to understand when cars would arrive at their destination, estimate travel times, and reorganize train schedules to improve reliability. Ten years later, Sierleja *et al.* (1981) reviewed this system under the sponsorship of the Federal Railroad Administration; subsequently, List and Bongaardt (1981) explored the potential benefits of its implementation on ConRail and Buchan *et al.* (1981) examined its value for the railroads in New England. The findings from these studies were that: 1) FCS improved freight train management but not necessarily freight car transit times, 2) it would have significant value on ConRail for managing freight car movements, and 3) its value to the New England railroads (specifically the Boston & Maine and the Maine Central) would be limited because the number of freight cars being handled was too small. Findings from this analysis form part of the backdrop for this present study.

A recognition that rail trip time reliability and profitability was intensely coupled to operations led to a new body of literature focused on blocking plans. A blocking plan indicates how railcars are to be handled at the classification yards (nodes, sorting locations) as the cars transit the network. Turnquist and Daskin (1982) examined the use of queuing models to represent and study freight car delays. Daganzo *et al.* (1983) devised a mixed integer linear programming model that could identify "optimal" blocking plans for a given objective and pattern of flows. Yagar *et al.* (1983) devised a similar model as did Crainic and Rousseau (1986) and Daganzo (1986). A practical solution to the problem was described by Van Dyke (1986) and that model has been employed by many railroads. Kraay *et al.* (1991) explored the optimal spacing of trains to avoid meet/pass delays caused by poor dispatching. Martland (1992) portrayed the role that control systems play in managing operations and reliability. Kraft (2002a, 2002b) presented models for scheduling the classification activities at rail yards and for scheduling railway operations.

Improving shipment connections at terminals has been the focus of Chen (2010) and Chen and Schonfeld (2011). Chen (2010) examines transfer coordination in intermodal and intra-modal logistic networks. One model is developed for coordinating vehicle schedules and cargo transfers at freight terminals. A mixed integer nonlinear programming problem of a multi-mode, multi-hub, and multi-commodity network is formulated and solved using sequential quadratic programming (SQP), genetic algorithms (GA) and a hybrid GA-SQP heuristic algorithm. This is done primarily by optimizing service frequencies and slack times for system coordination while also considering loading and unloading, storage, and cargo processing operations at the transfer terminals. A second model focuses on counteraction strategies for schedule disruptions. The dispatching control method proposed determines whether each ready outbound vehicle should be dispatched immediately or held waiting for some late incoming vehicles with connecting freight. An additional sub-model deals with the freight left over due to missed transfers.

The ability for a system to overcome adverse conditions in the system was also investigated. In 2002, Armacost, Barnhart and Ware discussed using composite variable formulations in the logistical planning for companies such as UPS. Using composite variables, the network design becomes robust and exercises greater flexibility. It was also found that the optimal solution of the network allows for the number of packages to be less than the capacity of the aircraft, allowing an opportunity for recourse, if needed. Armacost, Barnhart, Ware, and Wilson further investigated UPS in 2004. Utilizing a planning system called Volume, Location, and Aircraft Network Optimizer (VOLCANO), UPS can plan for anticipated changes in the schedule and allow flexibility in the schedule. VOLCANO makes use of composite variables that will leaves room in the system if an adverse system arises.

The spectrum of these studies has since increased to include other aspects of the service development process and other modes and multi-modal environments as well. Beginning in 1978, Bevilacqua discussed the relationships between using alternative modes of transport service, economic efficiency, and energy consumption. Crainic and Rousseau (1986) presented a multi-commodity, multi-modal service design model that emphasized reliability as well as cost and operational efficiency. Barnhart, Belobaba, and Odoni (2003) studied issues of reliable service design in the context of the air transportation industry. Chen and Schonfeld (2011) examined ways to use real-time dispatching control to alleviate schedule disruptions at intermodal terminals. Markovic and Schonfeld (2011) examined scheduling issues for single-hub intermodal freight systems.

In a highway context, Taniguchi *et al*. (2001) examined the issue from a city logistics perspective, including the potential impacts from capitalizing on the information sharing from intelligent transportation systems. Jones and Sedor (2006) described a study of reliability from an interstate trucking perspective. Lomax *et al*. (2001) began the process of doing national-level assessments of congestion and reliability in urban areas. Washburn and Ko (2007) focused on understanding the impacts of travel time reliability on the perceptions of highway level-of-service held by trucking companies. List *et al*. (2006) examined the repercussions of considering reliability in truck fleet sizing decisions for national-level services. Yang and Regan (2012) described a multi-criteria decision support system for trucking operations. Hsu and Wang (2013) examined reliability in the context of hub-and-spoke air cargo networks.

2.2 SUPPLY CHAIN RELIABILITY

When supply chains are the focus of a reliability analysis, the end-to-end process is brought into the picture. The travel times are one source of stochasticity, but there are others: manufacturing disruptions, component delivery delays, and demand variability. In cases like shipping coal where one can assume an infinite supply of material at the source, managing the variability in the travel times becomes the major focus. In other instances, some of the other sources of stochasticity may be more important. Arvis, Raballand, and Marteau (2007) provide a thoughtful commentary on the impacts of supply chain reliability on logistics costs, especially for enterprises located in landlocked, third-world countries. Vernimmen, Dullaert, and Engelen (2007) provide a similar review of the impacts of schedule reliability for shippers worldwide that rely on containerships in their logistics networks.

Whybark (1974) appears to have been the first to focus on reliability in the context of a supply chain. He asserts that while most organizations set inventory policies and choose transportation alternatives separately, there is an interaction between these decisions when the transportation alternatives have different speed, reliability and cost characteristics. Whybark presents a heuristic procedure for jointly determining the reorder points, order quantities and transportation alternatives for minimum total transportation and inventory costs for a receiving facility.

Allen, Mahmoud, and McNeil (1985) present a model that shows how a cost-minimizing shipper should adjust its economic order quantity (EOQ) as reliability and/or time in transit changes. A matrix shows the minimum cost attainable with each combination of mean and variance values for the transit time distribution. In addition, by comparing one cell with another, the matrix shows how costs are affected by changes in the mean transit time and the variance. It shows how reductions in cost can in some cases be achieved by improving reliability while increasing average transit time. The paper also shows how points can be reordered and adjusted in response to changes in the travel time mean and variance.

Muthuraman, (1991) presents an analysis like that of Allen, Mahmoud, and McNiel. They combinate a continuous time back-ordered inventory system with stochastic demand and stochastic delivery lags for placed orders. By modeling demand as a diffusion process, they reformulate the inventory control problem as an impulse control problem which they then simplify into a Quasi-Variation Inequality (QVI). This allows them to obtain the optimal ordering policy, the limiting distribution of the inventory level, and the long run average cost. Computational experiments show that significant losses can be incurred in approximating a stochastic lead time system with a fixed lead time system, highlighting the need to consider the stochasticity in the lead time.

Later articles consider variations on these ideas. Ouyang and Chang (2001) create a formulation that uses fuzzy set theory to determine the backorder rate. Zhao and Simchi-Levi (2006) consider an assemble-to-order problem where the lead times are stochastic. Hnaien, Delorme, and Dolgui (2008) examine supply planning for two-level assembly systems with stochastic component delivery times. They study the trade-off between holding cost and service level. A basic multi-objective meta-heuristic is used to explore the trade-off between holding cost and stockout probability. Louly, Dolgui, and Hnaien (2008) explore ways to address problems involving supply planning for single-level assembly systems with stochastic component delivery times and a high service level constraint.

A recent study of the impacts of reliability on logistics managers and freight operators is reported by McKinnon *et al*. (2008). The project examined the impacts of congestion for nine sectors of the economy and inquired about company reactions to the significant deterioration in road traffic conditions in the United Kingdom (UK). Thirty-seven managers were interviewed in twenty-eight companies or divisions of companies. Five visits were made to distribution centers. Detailed inquiries were made about the impact of congestion on the logistics operations, the relationship between congestion and other sources of unreliability and any measures companies were taking to mitigate the effects of congestion. Very few of the companies could provide statistical information about the operational and financial impacts. The interview data suggested that there were wide variations in the impact of congestion. There was little evidence of congestion causing companies to restructure their logistics systems. Nor was it causing companies to run more vehicles, increase tractor-trailer ratios, carry more inventory, or modify internal warehouse design and capacity. Companies learn to compensate by altering schedules, building in extra slack, making internal processes more flexible and, in some cases, upgrading their dispatching systems. There were companies, however, for which congestion did have a significant impact. And for them, congestion was clearly having serious impacts on cost and quality of service.

Jane (2011) presents a technique for estimating the reliability of a distribution network. The measure is identified as $R_{\$b,d}$. It is the probability (R) of successfully delivering a demand d to a specific destination given a budget constraint of $\$b$. Jane presents a hybrid approach for computing this metric and demonstrates that it is effective and efficient. It is interesting for several reasons: 1) it defines reliability in the classical sense of on-time deliveries, 2) it combines the reliability of the source location with that of the transportation system, and 3) it applies a budget constraint to achieve the assessment. It represents a holistic perspective on the reliability of the supply chain.

Lai (2011) presents a similar analysis based on the semiconductor industry. Multiple manufacturing sites are involved, stochastic travel times are assumed, the vehicles have capacity constraints, and time windows are imposed for both pick-up and delivery. A chance constrained programming model is employed. It uses two categories of chance constraints, one for the time windows and another for the duration of driver service. The objective is to find a set of routes and schedules for the vehicles that minimizes travel distance without violating time windows, product/vehicle compatibility, pickup and delivery, driver duration, and vehicle capacity constraints.

Lee and Song (2011) provide a commentary on the importance of reliability in the context of the maritime industry. The authors say that within the context of the maritime industry, maritime logistics value represents the quality with which shipping and port operators meet reliability. An exploratory case study within the Korean maritime transport industry is used to conclude that the most valuable knowledge is acquired through having maritime transport operators embedded in the supply chain decision-making, which then improves the value of the maritime logistics offered.

Kim and Simchi-Levi (2012) consider what could be regarded as a "realistic, real-world problem" in the logistics world of today. A company is viewed as operating multiple delivery modes such as standard freight shipping and air. The lead times (travel times) are viewed as being stochastic (subject to delays). A model is presented that shows how to make the best use of these multiple delivery modes to minimize the impacts of lead time variability. As might be expected, the model depends on an order tracking system so that expedited handling becomes possible. The goods move stochastically among the installations and the system faces a stochastic demand. Kim and Simchi-Levi identify systems that result in simple and tractable optimal policies, in which both regular ordering and expedited handling follow a variant of the base stock policy. They show that optimal expediting results in a significant reduction in the total logistics cost and the reduction increases as variability in delivery lead time increases. They also show that expediting allows the system to be operated in a leaner way due to the reduced regular order amount and provide various managerial insights linking expediting, base stock levels, and expediting costs. This study exactly illustrates the holistic perspective on supply chain management that is needed to mitigate the impacts of stochasticity in the transportation system.

Hayya *et al*. (2013) present an interesting study focused on stochastic lead times for JIT delivery systems. They point out that FIFO (first-in-first-out) may not be preserved when stochasticity is present. They find that with stochastic lead times there is a possibility of order crossover, and what order crossover does is to transform the original lead times into effective lead times. The mean value of arrivals at the destination is the same as that at the origin, but the variance can be smaller. The implication is that when order crossovers are considered, the cost can be less than it would be without the crossovers. They demonstrate some important properties of the effective delivery time.

Li and Savachkin (2013) study an integrated logistics network design problem and endeavors to optimize the assignment of supplier locations to terminal facilities. It allows for expedited shipments. They first formulate elementary models for certain special case problems and discusses their properties and solution methods. They then propose a mathematical programming model that minimizes the sum of supplier set-up costs, expected shipment costs for both regular and expedited services, and expected inventory holding cost under stochastic demand rates and transportation lead times. This represents a general network logistics system working against a planning horizon. After studying the properties of the problem, Li and Savachkin (2013) develop a solution approach that makes use of Lagrangian relaxation. This approach is tested on three logistics networks of different scales. From these analyses, they note that under the optimal design, utilizing expedited shipment services actually does not produce much extra cost while it guarantees service reliability. They also finds that with the integrated design, all planning and operational components complement each other in an optimal, holistic manner.

System-wide reliability has been studied for other modes as well. The annual urban mobility assessments (see for example Schrank *et al*., 2012), while not explicitly focused on freight, portray the impacts of congestion on travel times in major metropolitan areas. More from a freight standpoint, Washburn and Ko (2007) and Dowling *et al*. (2014) examined reliability in developing trucking level-of-service models. Xu *et al*. (2008) explored trip and network reliability by incorporating truck trip assignment into a dynamic traffic assignment model. Czuch and McDaniel (2011) propose a methodology for measuring reliability on the highway network. One of the preferred measures is the buffer index. This is the extra time that a driver must add to the average travel time to ensure an on-time arrival. Ogawa *et al.* (2012) study a set of managed motorways around Birmingham, United Kingdom, with an aim of finding ways to address congestion and improve journey time reliability. The tool is advanced intelligent transport system applications. They present the high-level results obtained from the traffic data analysis undertaken for the three schemes. The results demonstrate that traffic conditions within the scheme extents can be improved by the introduction of managed motorways and further identified findings which can be applied to assist the development of future schemes.

McCormack *et al*. (2010) and subsequently Ma *et al*. (2011) collected and analyzed global positioning system (GPS) data for trucks operating in the central Puget Sound region. The researchers examined truck freight performance measures that could be extracted from travel times and speeds. The utility of spot speeds and the GPS data in general was evaluated in in the context of a three-week construction project on Interstate-90.

Figliozzi *et al*. (2011) developed a programming logic that uses GPS data to a) identify natural segments or regions in a corridor between urban centers, Interstate junctions, or rural areas and b) estimate corridor-wide reliability. While the study focused on an I-5 corridor in Oregon, the methodology is applicable on a network-wide basis. The researchers applied statistical techniques to compute vehicle travel time and reliability for freight movements within each segment. The methodology successfully identified distinct segments and characteristics of travel time reliability in freight corridors. Travel time information was used to compute cost effects of delays within rural and urban areas along the I-5 corridor.

3.0 DEDICATED TRUCKLOAD SERVICE

This first case study focuses on a simple, dedicated truck service, such as the one depicted in Figure 3.1. Trucks travel back-and-forth between nodes A and B, carrying loads in one or both directions. The travel times are stochastic as are the load and unload times. The objective is to see how the size of the truck fleet affects the on-time performance.

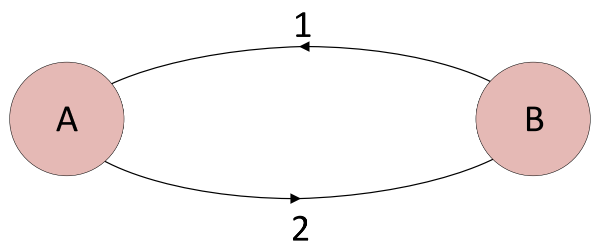

Figure 3.1: A hypothetical dedicated truckload service network

This type of trucking service exists in many settings. Quarry operations are a good example, using trucks like the one shown in Figure 3.2. The trucks cycle back-and-forth between the excavating location and the processing plant or storage area. The travel times can vary as can the loading and unloading times. Another example (not shown) is trucks carrying product components like engines from a component manufacturing plant to a final assembly plant.

Figure 3.2: Quarry trucks in a dedicated fleet

While no fleet size can ensure that maximum throughput is achieved (in the case of the mine trucks) or all pick-ups and deliveries are accomplished on-time (in the case of the trucks carrying engines), larger fleet sizes, up to a point, improve system performance. So, an important question is: what fleet size is required to ensure that the quality of service (on-time performance) is at or above a target level (e.g., 95% of the deliveries are on-time). Another is how does the resource level affect the quality of service provided? A third is what is the fleet size beyond which further quality of service improvements are very small?

3.1 DETERMINISTIC ANALYSIS

For a deterministic analysis, the assessment is simple. The load and unload times, t_l and t_u are known, as are the travel times, t_{AB} and t_{BA}. The flow rate objective, *V*, (say, tons per day) is divided by the truck capacity, *C*, (say, in tons) and the duration of the work day, *H*, to determine how many truck loadings, *q*, are needed per hour:

$$q = \frac{V}{C * H} \tag{3.1}$$

This then allows computation of the headway, *h*, between departures:

$$h = \frac{1}{d} \tag{3.2}$$

The cycle time for an individual truck, *T*, is the sum of the load and unload times, t_l and t_u, the travel times, t_{AB} and t_{BA}, and any slack, *s*, allowed for driver rest time and other considerations:

$$T = t_l + t_u + t_{AB} + t_{BA} + s \tag{3.3}$$

The slack time does not have to occur every cycle. Rather, it represents an average amount of time per cycle needed to ensure that adequate rest time is provided as well as time for other contingencies. (In this sense, the values for the other times represent typical times, which may be averages or a higher percentile in the distributions of observed times.)

The fleet size, *F*, is then the number of trucks needed to cover the departures at headway *h* that occur on a per-cycle basis, *T*, where $\lceil \ \rceil$ is the ceiling operator:

$$F = \left\lceil \frac{T}{h} \right\rceil \tag{3.4}$$

To give a numerical example, if the objective was to mine one unit-train load of material each day (V = 10,000 tons) and the capacity of each truck, *C*, was 100 tons, then 100 loads would have to be carried; if the workday was 8 hours, the flow rate would need to be 12.5 loads per hour:

$$q = \frac{10,000}{100 * 8} = 12.5 \text{ loads/hr} \tag{3.5}$$

This would mean a headway of 4.8 minutes (60 minutes/12.5 loads).

If the load time is 4 minutes, the unload time is 6 minutes, the travel time from shovel to tipple is 15 minutes, from tipple to shovel is 10 minutes, and the slack time is 5 minutes, then the cycle time is:

$$T = 4 + 6 + 15 + 10 + 5 = 40 \text{ minutes} \tag{3.3}$$

Finally, this means the fleet size needs to be:

$$F = \left\lceil \frac{40}{12.5} \right\rceil = \lceil 3.2 \rceil = 4 \text{ trucks} \tag{3.4}$$

Since all the values are fixed, a single answer exists for the problem. Maybe the slack time allows for contingencies, but since no Monte Carlo-type analysis has been performed, the ability of the fleet to handle unforeseen delays is unknown.

3.2 STOCHASTIC ANALYSIS

If the times t_l, t_u, t_{AB} and t_{BA}, are treated as being stochastic, then the cycle time, T, is also stochastic. If the distributions of these times are independent (in a statistical sense), then the distribution of T can be obtained through convolution. In a Monte Carlo simulation sense, independently drawn values of the travel times can be added to obtain observations of T and the result can be examined statistically to determine its distribution.

To provide an illustration, the network shown below is examined. Trucks travel back-and-forth between A and B. In-between, they traverse 3 arcs in each direction (arcs 1, 2, and 3 eastbound; and 4, 5, and 6 westbound). The load and unload times are stochastic as are the travel times for the arcs.

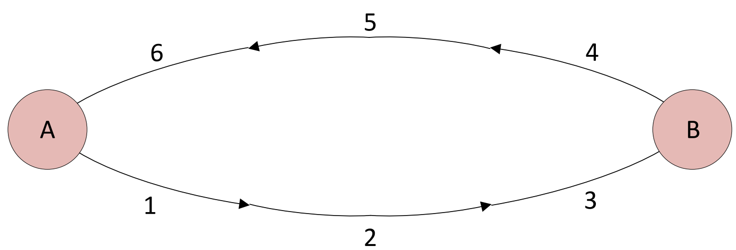

Figure 3.3: A dedicated truck network

The equations used to compute load, unload, and travel times, are shown in Table 3.1. The distributions are uniform with the mean values and ranges shown. Clearly, some of the time distributions have very little variation, like the loading and unloading times, while others have a wide range, such as the travel time on arc 4.

Element	Time Equation	Mean	Range
Load at A	= 7.65 + 2*RAND()	8.65	± 1
Arc 1	= 18 + 4*RAND()	20	± 2
Arc 2	= 15 + 5*RAND()	17.5	± 2.5
Arc 3	= 45 + 10*RAND()	50	± 5
Unload at B	= 7.6 5 + 2*RAND()	8.65	± 1
Slack	= 4 + 1*RAND()	5	± 1
Arc 4	= 30 + 16*RAND()	38	± 8
Arc 5	= 16 + 4*RAND()	18	± 2
Arc 6	= 18 + 8*RAND()	22	± 4

Table 3.1: Equations for computing load, unload, and travel times

Since it is assumed that the load, unload, and travel time variables are independent, convolution can be used to obtain the distribution of the cycle time. However, Monte Carlo simulation can also be used. That is what has been done. An Excel worksheet was developed to conduct a Monte Carlo simulation of the system. Values for 100 cycle times were computed based on the formulas shown in Table 3.1. These values were then summarized statistically to identify the number of trucks required to cover ranges of the cycle times identified.

The distribution of the cycle length, *T*, is shown in Figure 3.4. It ranges from a minimum of 157.3 (the sum of the minimum values) to a maximum of 208.3 (the sum of the maximums).

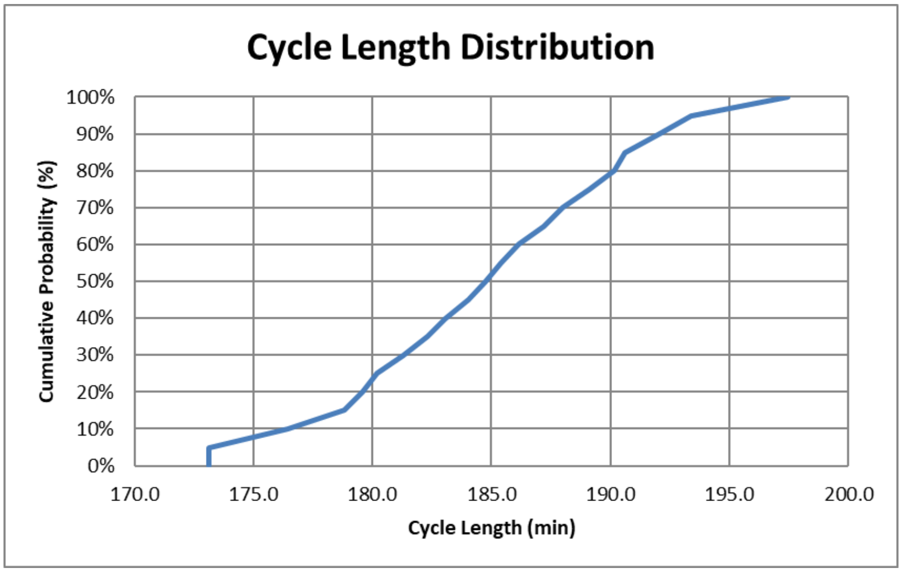

Figure 3.4: Cycle length distribution for the dedicated truck network

In this case, since the distributions are bounded, it is possible to determine how large the fleet needs to be to cover all contingencies. The headway between departures is assumed to be 3 minutes. As Figure 3.5 shows, the maximum fleet size is 66 trucks.

It is also possible to determine number of trucks required in order to cover a specific percentiles of the cycle times. For example, to cover up to 50% of the cycle times, 62 trucks are required. This number of trucks is needed to cover percentiles ranging from 45% to 55%. (Of course, this number of trucks can also cover all percentiles below 45%, but fewer trucks are required for lower percentiles.) To cover up to 95% of the cycle times, 65 trucks are required, and that number of trucks is needed to cover percentiles from 90% to 99%. A 66th truck is needed to cover the 100th percentile.

From a system management perspective, a conclusion that can be drawn from this analysis is: to cover the average condition for the cycle time (at about the 50th percentile), 62 trucks are needed. However, to be able to cover all contingencies, an additional 4 would be needed, absent any maintenance spares. The trucking company could use this analysis method, and the results obtained, to determine how many trucks to acquire to provide a specific quality of service.

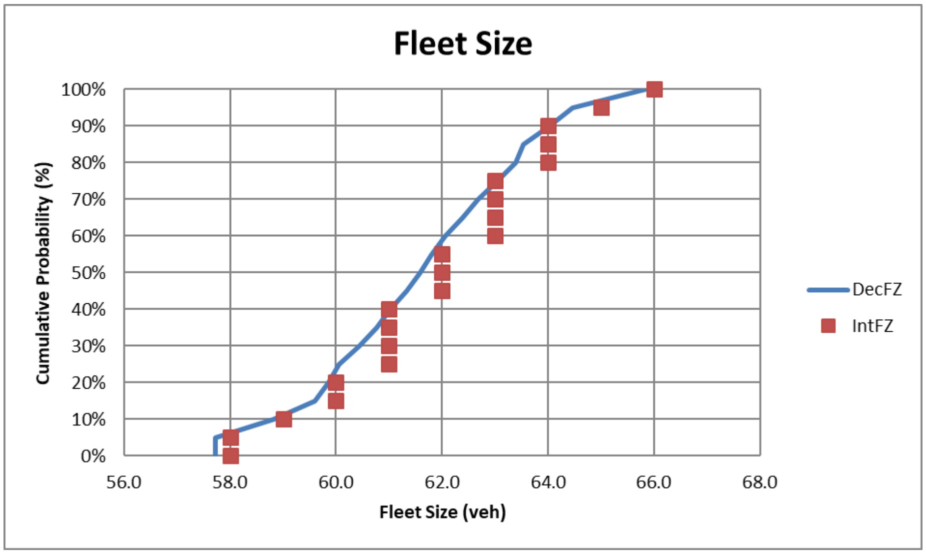

Figure 3.5: Fleet size requirements by percentile for the dedicated truck network

If, for example, the trucking company decided to purchase 64 trucks, then about once in every 10 days (10% of the time), the fleet would not be large enough to ensure that the daily volume throughput targets will be met. At 63 trucks, the failure rate will be about 25%, or slightly more than one day a week. This helps to illustrate the disadvantage of performing an "average value" analysis. If just the average times were used, it would have been determined that 62 trucks were needed. With that many trucks, the fleet would be inadequate to meet the target throughput every other day. To work around that shortcoming, using a deterministic analysis, the trucking company should assume conservative values for the load, unload, and travel times (such as the maximum values). The result would be 66 trucks. But the 66th truck is needed only for the last percentile. The firm might not realize that buying 65 trucks would cover 99% of the conditions. It would also have no assessment of the value provided by the 63rd, 64th or 65th trucks. It would only know that 66 trucks were needed to cover all contingencies (absent spares).

3.3 SUMMARY

This section has studied a simple, dedicated service where trucks travel back-and-forth between two locations. They carry loads in both directions. The travel times are stochastic as are the load and unload times. The objective was to see how the size of the truck fleet affects the on-time performance.

Clearly, the analysis shows that the stochastic analysis has value. For example, to cover the average conditions, 62 trucks are sufficient. But, that fleet is only capable of covering only up to and including the 50th percentile of the conditions that arise. Beyond that, the cycle times that arise cannot be covered. Load departures are missed. Total throughput suffers. However, by increasing the fleet size to 66 trucks, by adding only 4 more, slightly more than 6%, all the cycle times that arise can be covered. In addition, with 65 trucks, or just 3 more, the fleet can cover up to 95% of the cycle time. This information should be very valuable to trucking firms that place a value on contingency planning.

4.0 TRUCKLOAD SERVICE

The second setting examined in this study is that of a truckload carrier providing service between a set of locations. The network studied, without loss of generality, is the one shown in Figure 4.1. Loads are carried between the various locations shown. The travel times, service times, and service demands vary by problem realization. The planning horizon is several weeks. Performance is assessed based on the ability to make on-time deliveries. The number of trucks is a user input. Fleet sizes that are too small cause shipment delays, and increasing the fleet size mitigates that impact.

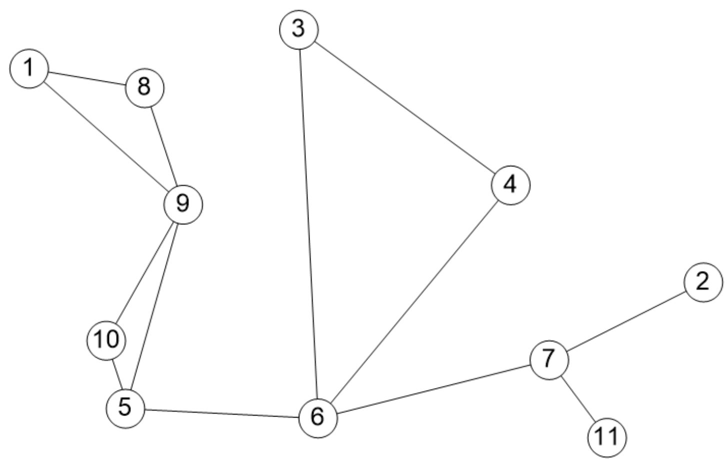

Figure 4.1: A hypothetical truckload network

Most of the shipments originate at nodes 1-10 and they are destined for node 11. They follow sequences of segments (i.e., the travel through specified sequences of node-to-node links). For example, a shipment from node 9 to node 11 can travel via segments 9-5, 5-6, 6-7 and 7-11. It can also travel via segments 9-10, 10-5, 5-6, 6-7, and 7-11.

Trucks are assigned to the shipments; drivers are then assigned to the trucks. Both these assignments are done on a next-available/closest to the load basis. Trucks deadhead (move empty) if necessary from the end of one shipment to the beginning of the next. Drivers are assigned to both the loaded and empty truck moves. Drivers also deadhead (on either loaded or empty trucks) as needed to reposition themselves for the next truck assignment. Nodes 5, 6, and 7 serve as home nodes (bases) for the trucks and the drivers. The drivers return home at the end of each week. To ensure they can do that, driver deadheading movements are created at the end of each week.

4.1 METHODOLOGY

The methodology employed is much like one that would be employed by a truckload trucking company. It is derived from work done earlier by List *et al.* (2006). Customers want to move shipments from one location to another, at different times, and expect that the trucking company (as the service provider) assigns appropriate resources to make those shipments happen.

Meeting the shipment demands requires planning. Over the long-run (i.e., several years), the company must make decisions about resource levels – determining how many trucks and drivers (crews) to have to meet the anticipated demand. In the shorter run (i.e., several weeks), it must plan how to use its existing resources to accommodate the shipment requests.

The focus of this study is on the short-term planning: determining how to assign the trucks and drivers to shipments over the planning horizon of a few weeks while maintaining a daily resolution in resource allocations. In most cases, the planning horizon is 4-6 weeks.

To do this, a hierarchical model is employed that has two sequential steps: 1) assign trucks to the shipments by creating truck tours that form a sequence of loaded and empty moves and 2) assign drivers to those truck trips. This hierarchical structure implicitly assumes that the trucks are the limiting resource, not the drivers. That is, truck tours must be formed first, based on the number of trucks available, to ensure that a feasible solution is developed; shipments must be delayed on a prioritized basis if the fleet size is insufficient. Once the truck trips are developed, drivers are assigned; the number of drivers available is assumed to be "infinite". That is, while the user is asked to indicate how many drivers are available at each home node, these values are not treated as a hard constraint; a reserve pool of drivers is assumed to exist.

In the first stage, the truck trips are developed by minimizing the delay (i.e., failure to initiate the shipment within its allowed time window) for all shipments. Ideally, sufficient trucks are available to allow each shipment to begin and end within its specified time window. But that may not be the case. If not, shipment pick up times are delayed until a truck is available and low priority shipments are deferred.

In the second stage, drivers are assigned to the truck tour segments in a way that "covers" all the segments while minimizing the total driver requirements and deadheading miles. The assignments also reflect the regional home bases of the drivers and ensure that the drivers all return home at the end of each week.

The model produces the following outputs:

- A report that shows the actual (versus desired) departure and arrival times for all shipments;
- Itineraries for each truck over the planning horizon (4-6 weeks), including flags that indicate whether individual segments are loaded or empty moves;
- Driver work assignments for each home node over the planning horizon;
- A list of the problems encountered (e.g., shipments delayed, drivers used more than the indicated pool sizes).

The following sections describe the sub-models more completely.

4.1.1 Truck Tour Formation

Truck tours are formed by assigning shipments to trucks interspersed with empty movements. A "vehicle blocking model" based on successful methods used in other transportation applications is used to create the tours. This model originated as the "concurrent scheduler" created by Bodin and Dial (1980).

Shipment i has priority p_i. It has an intended start time t_i and can originate as late as $t_i + \tau_i$ where τ_i is a buffer, ithout being considered late. Both t_i and τ_i are assumed to be an integer numbers of days. q_i days are required for loading, r_i days for travel, and s_i days for unloading. Decimal values for these three parameters are allowed. The duration of the shipment, that is, the total time d_i required to accomplish these three activities is given by:

$$d_i = \lceil q_i + r_i \rceil + s_i \quad (1)$$

where $\lceil\ \rceil$ is the ceiling operator.

A generalized cost function is employed to construct the truck trips. It is sensitive to shipment-days of delay and resource reassignments. Shipment days are weighted by shipment priority. The time between shipments must allow the trucks to deadhead from the end of their previous shipment to the beginning of the next.

Truck trips are created for each week based on the following procedure. The length of a week is defined as T:

1) Sort the shipments in ascending order by start time, and then in descending order based on a combination of duration d_i and priority p_i. Call this the sorted list.
2) Assign truck j to the first shipment (i.e., initiate a partial block).
3) Assuming the first n shipments have been assigned to partial blocks for m trucks ($m \leq M$, where M is the fleet size), add shipment $n+1$ to the partial block for truck j (j = 1,2, ..., m) if:

$$z(n,j) \leq z(n,k) \qquad \forall \ k \neq j \quad (2)$$

and

$$t(n+1) \geq e(j) + dh(j,n+1) + t_{min} \quad (3)$$

where:

$z(n,j)$ = "cost" of assigning truck j to shipment n
$t(n+1)$ = start time of shipment $n+1$
$e(j)$ = end of the current partial block for truck j
$dh(j,n+1)$ = time for truck j to deadhead from the endpoint of its current partial block to the beginning point of trip $n+1$
t_{min} = minimum allowable schedule slack before the next load event

4) Update $e(j)$ for the truck selected:

$$e(j)_{new} = \lceil e(j)_{old} + dh(j,n+1) + t_{min} \rceil + \lceil q_{n+1} + r_{n+1} \rceil + s_{n+1} \quad (4)$$

5) If equations (2) and (3) yield more than one qualified truck, chose the one which minimizes the difference between $d(n+1)$ and $e(j)$.
6) If no truck is available to assign to shipment $n+1$ without exceeding the maximum allowable delay, τ, and $m < M$, then increment m and begin a partial block for a new truck. This partial block begins with segment $n+1$. If $m = M$ (i.e., the fleet size has been reached), then delay segment $n+1$ to the earliest time when it can be scheduled. (This is the next time that a truck is free.) Then repeat step 3.
7) Repeat steps 2-5 for all shipments in the sorted list.

The result is a set of truck trips, each of which contains one or more shipments. Every shipment has a designated sequence of segments, so the truck trips can be broken down into flows on segments over the course of the planning period. This truck segment information is the input needed to assign the drivers.

4.1.2 Driver Assignment

Driver assignment to the truck trip segments is done so that they meet the requirements for each truck trip segment. Sometimes, two drivers are required. As an additional objective, the assignments are created so that driver workloads are balanced. This problem can be described using an integer programming formulation like the so-called *set covering* models used by airlines to schedule flight crews (see, for example, Barnhart, *et al*., 1998; Barnhart *et al*. 1999).

The mathematical formulation can be created as follows:

$$\min \quad \sum_j \phi_j y_j + \theta v \tag{8}$$

such that:

$$v - \phi_j y_j \quad \geq \quad 0 \qquad \forall\, j \tag{9}$$

$$\sum_j d_{ij} y_j \quad \geq \quad f_i \qquad \forall\, i \tag{10}$$

$$y_j \quad = \quad 0 \; or \; 1 \qquad \forall\, j \tag{11}$$

where: f_i = number of drivers required for truck trip segment *i*
d_{ij} = 1 if driver trip *j* includes truck trip segment *i*; 0 if not
φ_j = time associated with driver trip *j*
θ = weighting parameter
v = maximum time assigned to any driver
y_j = 1 if driver trip *j* is selected; 0 if not.

This formulation selects *driver trips* to employ (in the airline applications, they are called “bid lines”) based on a set of possible choices. The column vectors in the *D* matrix are those options. Each potential driver trip (indexed by *j*) represents a feasible set of work (assignments to truck segments) for driver *i* for a specified time interval (e.g., a week). The optimization model selects from these potential driver trips the subset that minimizes the weighted combination of total time and maximum time represented in the objective function, subject to the constraint that sufficient drivers must be provided for each truck trip segment (indexed by *i*).

The two differences from typical *set covering* models are 1) typically $f_i = 1 \;\forall\, i$, not a variable number, and 2) the first constraint (which defines the maximum time assigned) is not included.

The key to an effective implementation of this driver assignment model is a driver trip generator that creates “good” pieces of work for individual drivers. A driver trip (*j*) for an individual is defined by the column of the *D* matrix associated with it and the value of φ_j that results. The column of the *D* matrix is a series of zeroes and ones, with ones in the rows corresponding to the truck trip segments included in that work assignment. The driver trip generator must create columns that are feasible, both from a scheduling standpoint (i.e., the driver cannot be in two places at once) and from a work rule standpoint.

Implementation of this formulation requires two principal algorithmic parts – a driver trip generator and an evaluator that takes a proposed driver trip and determines if it should be used in the optimal solution. For small test problems, the evaluator part can be provided by a standard mathematical programming package, but this is unlikely to be computationally effective in full-scale implementation. In practice, proving exact optimality of the solutions is very difficult, but there is extensive experience using set covering formulations to produce good assignments of people to tasks, particularly in the airline industry.

For this study, an approximate solution is used instead of a math programming formulation. It involves a heuristic much like the "vehicle blocking model" that is used to develop the truck trips.

4.1.3 Driver Trip Creation

To illustrate how the driver trips are created, consider the following example. It is based on the network shown previously, in Figure 4.1.

Assume the travel times for the various segments in the network are shown in Table 4.1. Time is measured in decimal days. For a typical five-day week, time starts at t = 0 and ends at t = 5.0. Times for loading, unloading and travel are also measured in decimal days.

Network				
Link	From	To	Miles	Days
1	1	9	1400	0.9
2	8	9	700	0.5
3	9	10	300	0.3
4	9	5	450	0.4
5	10	5	100	0.1
6	5	6	300	0.2
7	3	4	1000	0.9
8	3	6	1500	1.1
9	4	6	1000	0.8
10	6	7	1100	0.8
11	7	11	300	0.3
12	7	2	450	0.5
13	1	8	850	0.6

Table 4.1: Example problem network

Perhaps obviously, the driver trips arise because of truck trips that must be made; those truck trips in turn arise because of the shipments. It is thus useful to begin by seeing how the truck trips are developed from the shipments.

Shipments have an availability time that is an integer value (i.e., at t = 0.0, 1.0, etc.) corresponding to the beginning of a day. When a load or unload event ends, travel can commence, regardless of the hour. Unloading commences when the shipment reaches its destination. Times for loading, unloading, and travel are measured in decimal days.

To illustrate creation of the truck trips, consider the two shipments described in Table 4.2. Both are available for pickup at the beginning of the week and have an allowable "window" that requires only that they be moved within the week. Table 4.2 also shows the load and unload times for the shipments – these values might also be determined by the cargo type (and potentially the origin and destination points).

ID #	From	To	Trucks	Drivers	At (time)	Until	Load Time	Unload Time
1	9	5	1	1	0.0	5	0.3	0.1
2	5	10	1	2	0.0	5	0.2	0.1

Table 4.2: Summary characteristics of the shipments in the example

Two truck trips can be created by combining these two shipments. The one where shipment 2 is connected to shipment 1 is shown graphically in Figure 4.2. The shipment numbers are listed in parentheses after each activity. The darkest blocks represent loaded travel segments, the medium gray blocks are empty travel segments, and the light gray blocks are load/unload activities. In this case, the truck trip originates and terminates at node 5 (a home node). In constructing this trip, it is necessary to provide a rest period at node 9, because the total trip exceeds 1.25 days. No driver can be on the road for more than 32 hours (1.25 days) without a break. Moreover, when a break does occur, it lasts 8 hours (0.33 days). It is assumed that the rest period can occur during the loading activity at node 9, so that when the trucks are loaded, the drivers are ready to depart for node 5. It is also assumed that the drivers and must stay with the trucks until unloading is completed.

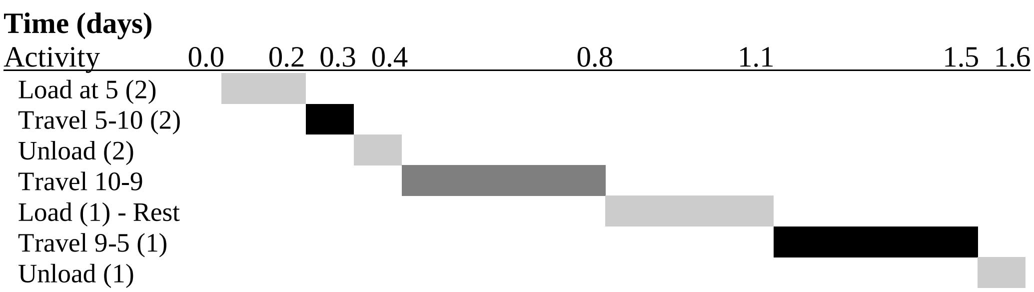

Figure 4.2: Illustration of a truck trip that can be formed from two shipments

For the portion of the trip from node 5 to node 10, two drivers are required; so, a separate, additional driver trip needs to be created so that this trip option can be selected. The itinerary for the second driver is shown in Figure 4.3.

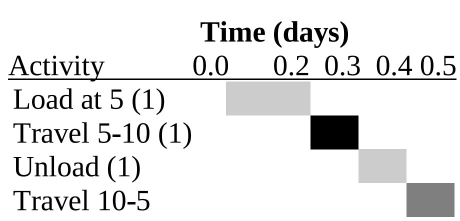

Figure 4.3: Itinerary for the second driver

In summary, insofar as the driver trips are concerned, to make sure the original truck trip can be selected, two driver trips need to be included in the list of driver trip choices. One is the same as the truck trip shown in Figure 4.1. The other is shown in Figure 4.2.

To illustrate additional characteristics of this trip creation methodology, suppose two more truck trips are added, as shown in Table 4.3. The first is from 6 to 1. The second is from 1 to 6. The first can be picked up at any time during the week. The second has a pickup window from day 1 until day 4.

ID #	From	To	Trucks	Drivers	At	Until	Load Time	Unload Time
3	6	1	1	2	0.0	5	0.1	0.1
4	1	6	1	2	1.0	4	0.1	0.1

Table 4.3: Two additional example shipments

Based these two shipments, two additional truck trips can be constructed, as shown in Figure 4.4. Because the distance from 6 to 1 is quite long, rest times must be inserted between the travel segments (at node 9). A rest time is also required at node 1. Some of this rest time occurs in parallel with the unload/load time, but the remaining 0.1 day is shown as a rest block to complete the required 0.3 days of rest time.

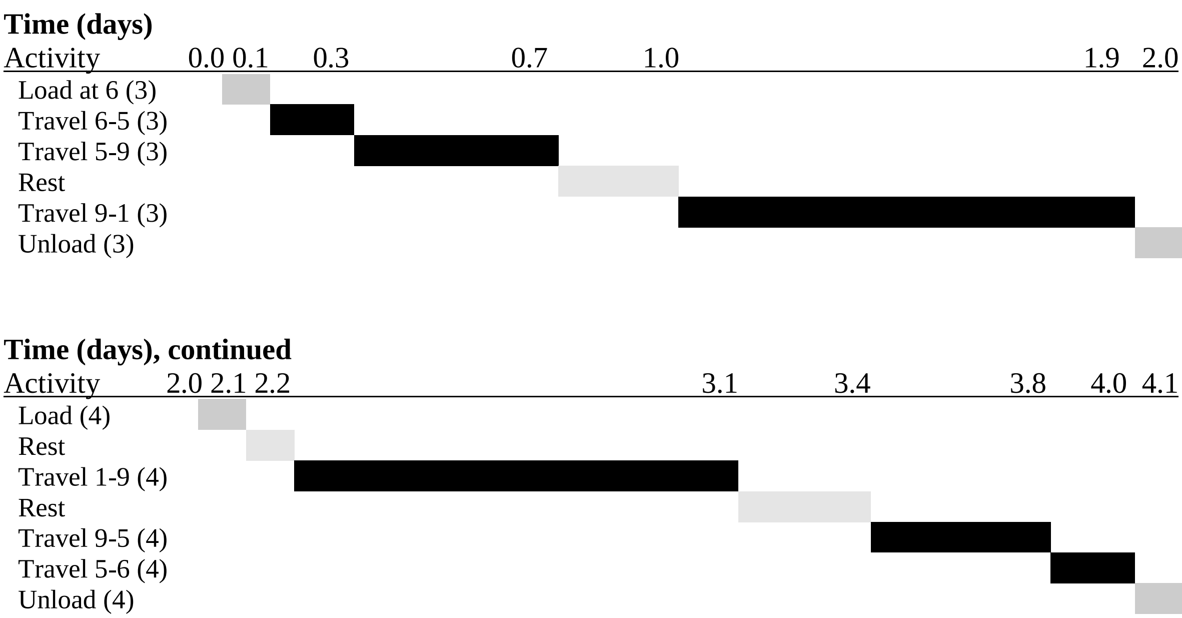

Figure 4.4: Illustration of truck trips formed to handle shipments 3 and 4

Because these two truck trips constitute a round-trip loaded movement with identical driver requirements, the driver trips for these truck trips look just like the ones shown in Figure 4.4. There is no need for an additional driver trip as there was with the first two trips.

4.2 CASE STUDY

The case study involves 41 shipments as shown in Table 4.4. The from (Fm) and to (To) locations are widely varied. The priorities (Prio) range from 1 up to 5. Some require just one driver; others require two (nDrv). The loading (tLoad) and unloading times (tUnld) range from 0.2 to 0.5 days. The beginning windows (bWin), in days, range from day 0 up to day 18. The tolerances for pickup range from 1 to 3 days.

The analysis timeframe encompasses 4 weeks, with 6 days in each week. The seventh day each week is ignored (the system rests). Therefore, the analysis involves a total of 24 days.

The distances and times for the network links are the same as those shown in Table 4.1. Each node in the network has a home node. The home nodes are as follows:

- For node 5: 1, 5, 8, 9, and 10
- For node 6: 3, 4, and 6; and
- For node 7: 2, 7, and 11.

The number of trucks available is an input. It can range up to 10 trucks, but the values that produce reasonable results for this problem range from 5 to 9 trucks.

The number of drivers is not an input. It is computed by the model. The number required is tied to the truck trips. In experiments with this set of demands, the value ranges up to 22 drivers. It is typically between 10-20 drivers.

The main performance criterion is lateness among the individual shipments. Ideally, the fleet size is large enough that no shipments are delayed, but often the repositioning of the trucks from one shipment to the next means that delays are introduced.

Another metric is the workload balance between the drivers. Ideally, the drivers all work the same amount of time each week. That is difficult to achieve since the worktime accumulates from the load, unload, travel, and mandatory rest times.

Parametric analyses where the truck fleet size is varied provides insight into the sensitivity of system performance to the resources available.

		Shipments						
Load	Prio	Fm	To	nDrv	tLoad	tUnld	bWin	Tol
1	2	8	11	1	0.3	0.3	0	1
2	1	9	6	2	0.3	0.4	0	1
3	1	7	6	1	0.4	0.2	0	1
4	5	6	1	1	0.3	0.2	17	1
5	5	2	6	1	0.4	0.2	9	3
6	1	8	6	2	0.2	0.2	6	1
7	5	10	1	1	0.3	0.3	17	1
8	4	6	11	1	0.4	0.3	15	2
9	5	4	6	2	0.4	0.2	12	1
10	5	4	6	2	0.2	0.2	4	2
11	3	9	6	1	0.2	0.3	16	3
12	5	10	1	2	0.2	0.5	3	1
13	5	6	1	2	0.5	0.4	4	3
14	5	6	2	1	0.2	0.4	13	1
15	5	1	2	1	0.2	0.4	6	3
16	5	1	6	2	0.4	0.2	8	2
17	3	8	11	1	0.2	0.2	2	2
18	2	7	6	1	0.3	0.3	16	1
19	5	6	2	1	0.2	0.2	7	2
20	2	9	10	2	0.3	0.5	3	2
21	3	8	6	2	0.3	0.4	17	1
22	5	3	6	2	0.3	0.3	15	3
23	5	7	6	2	0.4	0.2	2	1
24	3	9	10	2	0.3	0.3	1	1
25	4	6	7	1	0.4	0.2	5	2
26	3	6	11	1	0.2	0.4	10	1
27	1	6	7	2	0.3	0.3	8	2
28	5	6	3	2	0.2	0.2	9	1
29	5	6	3	1	0.2	0.2	5	1
30	5	2	6	2	0.3	0.3	10	1
31	1	9	11	2	0.3	0.3	0	1
32	5	1	2	1	0.5	0.4	12	2
33	2	10	6	1	0.2	0.5	13	1
34	5	4	3	2	0.4	0.3	18	1
35	1	10	6	2	0.2	0.2	14	1
36	5	3	6	1	0.3	0.3	3	2
37	3	10	11	2	0.4	0.2	12	3
38	5	6	4	2	0.3	0.4	16	3
39	1	9	11	1	0.2	0.3	0	2
40	2	8	6	1	0.3	0.2	15	1
41	1	6	10	1	0.4	0.2	0	1

Table 4.4: Case study shipments

Examining the impacts of variations in the travel times and load/unload times is useful as well. As the case study in Section 3 showed, predicating fleet size decisions on average values can lead to under-investment decisions whose consequences are quite undesirable.

For this case study, and a fleet size of 9 trucks, the assignment of trucks to shipments is shown in Table 4.5. Looking at the columns, some of the trucks are busy with 4-6 shipments carried. Two have much lighter assignments with just two shipments.

Load	From	To	nDrvr	bWin	Dep	Arr	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9
1	8	11	1	0.0	0	2.8					X				
2	9	6	2	0.0	3.7	4.9						X			
3	7	6	1	0.0	2.4	3.9							X		
4	6	1	1	17.0	18.2	20.3									X
5	2	6	1	9.0	12.3	14.1		X							
6	8	6	2	6.0	8.9	10.4				X					
7	10	1	1	17.0	18.1	20.1					X				
8	6	11	1	15.0	18.2	20							X		
9	4	6	2	12.0	13	14.4							X		
10	4	6	2	4.0	4	5.2			X						
11	9	6	1	16.0	16	17			X						
12	10	1	2	3.0	6.1	8.2				X					
13	6	1	2	4.0	6	8.5			X						
14	6	2	1	13.0	13	14.8					X				
15	1	2	1	6.0	6	9.4								X	
16	1	6	2	8.0	8	10.2	X								
17	8	11	1	2.0	8	10.6		X							
18	7	6	1	16.0	18	19.5		X							
19	6	2	1	7.0	7	8.6					X				
20	9	10	2	3.0	3	4.1				X					
21	8	6	2	17.0	18.9	20.7						X			
22	3	6	2	15.0	15	16.8	X								
23	7	6	2	2.0	2	3.5		X							
24	9	10	2	1.0	1	1.9				X					
25	6	7	1	5.0	6.2	7.7						X			
26	6	11	1	10.0	12.2	13.9						X			
27	6	7	2	8.0	8.6	10.1						X			
28	6	3	2	9.0	12	13.6	X								
29	6	3	1	5.0	6.2	7.8							X		
30	2	6	2	10.0	13.2	15			X						
31	9	11	2	0.0	0	2.2							X		
32	1	2	1	12.0	12	15.7									X
33	10	6	1	13.0	14.4	15.4		X							
34	4	3	2	18.0	18.8	20.3			X						
35	10	6	2	14.0	14.7	15.4							X		
36	3	6	1	3.0	3	4.8	X								
37	10	11	2	12.0	12.1	14.1				X					
38	6	4	2	16.0	18	19.5	X								
39	9	11	1	0.0	0	2.1						X			
40	8	6	1	15.0	15	16.6								X	
41	6	10	1	0.0	3.5	4.4		X							

Table 4.5: Truck assignments to shipments for the case study

The assignment of drivers to shipments is shown in Table 4.6. As with the trucks, looking at the columns gives a sense of the workload among the drivers. Most of the drivers are involved in 3-6 shipments. A few of them have lighter assignments. Looking at the columns, it is possible to see what drivers are involved for what loads. For example, driver #12 is involved in transporting shipment #5. One driver is needed and one is assigned. But looking at shipment #1, two drivers are used even though only one is needed. The reason is that the shipment involves multiple segments and different drivers are used on the segments. The same is true for shipment #2. Three drivers are used while 2 are needed. There happen to be two segments; and one driver is used for both segments while the other driver is different.

Load	From	To	nDrvr	D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8	D-9	D-10	D-11	D-12	D-13	D-14	D-15	D-16	D-17	D-18
1	8	11	1	X							X										
2	9	6	2		X					X					X						
3	7	6	1				X														
4	6	1	1								X										
5	2	6	1												X						
6	8	6	2					X					X								
7	10	1	1										X								
8	6	11	1									X									
9	4	6	2	X	X																
10	4	6	2	X								X									
11	9	6	1												X						
12	10	1	2					X					X								
13	6	1	2													X	X				
14	6	2	1			X															
15	1	2	1								X										
16	1	6	2	X	X																
17	8	11	1											X							
18	7	6	1												X						
19	6	2	1			X															
20	9	10	2						X	X											
21	8	6	2	X	X																
22	3	6	2													X	X				
23	7	6	2										X	X							
24	9	10	2						X	X											
25	6	7	1									X									
26	6	11	1									X									
27	6	7	2									X						X			
28	6	3	2													X	X				
29	6	3	1				X														
30	2	6	2											X							X
31	9	11	2			X	X														
32	1	2	1																X		
33	10	6	1												X						
34	4	3	2													X	X				
35	10	6	2	X	X																
36	3	6	1					X													
37	10	11	2				X	X													
38	6	4	2																	X	X
39	9	11	1		X																
40	8	6	1							X											
41	6	10	1										X								

Table 4.6: Driver assignments to shipments for the case study

It is a bit difficult to present the truck and driver assignments by segment, but Table 4.7 endeavors to do that. Only drivers up to #14 are shown because, for these shipments, drivers beyond #14 are not used. By viewing the second column, the shipment number, it is possible to see that the number is the same for multiple rows. For example, the shipment number is the same for the first 5 rows. The truck assignment is the same, T-5, as it should be, but the driver assignments change. Driver #8 is used for the first 4 segments, and driver #1 is used for the last.

Seg	Load	From	To	nDrvr	Dep	Arr	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9		D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8	D-9	D-10	D-11	D-12	D-13	D-14
1	1	7	11	1	2.3	2.8					X													X						
2	1	6	7	1	1.4	2.3					X													X						
3	1	5	6	1	1.2	1.4					X													X						
4	1	9	5	1	0.9	1.2					X													X						
5	1	8	9	1	0	0.9					X						X													
6	2	5	6	2	4.3	4.9						X						X					X							
7	2	9	5	2	3.7	4.3						X						X										X		
8	3	7	6	1	2.4	3.9							X							X										
9	4	9	1	1	19	20.3									X									X						
10	4	5	9	1	18.7	19									X									X						
11	4	6	5	1	18.2	18.7									X									X						
12	5	7	6	1	13	14.1		X																				X		
13	5	2	7	1	12.3	13		X																				X		
14	6	5	6	2	10	10.4				X											X					X				
15	6	9	5	2	9.7	10				X											X					X				
16	6	8	9	2	8.9	9.7				X											X					X				
17	7	9	1	1	18.7	20.1					X															X				
18	7	10	9	1	18.1	18.7					X															X				
19	8	7	11	1	19.5	20							X												X					
20	8	6	7	1	18.2	19.5							X												X					
21	9	4	6	2	13	14.4							X				X	X												
22	10	4	6	2	4	5.2			X								X								X					
23	11	5	6	1	16.5	17			X																			X		
24	11	9	5	1	16	16.5			X																			X		
25	12	9	1	2	6.6	8.2				X											X					X				
26	12	10	9	2	6.1	6.6				X											X					X				
27	13	9	1	2	7	8.5			X																				X	X
28	13	5	9	2	6.7	7			X																				X	X
29	13	6	5	2	6	6.7			X																				X	X
30	14	7	2	1	14.1	14.8					X								X											
31	14	6	7	1	13	14.1					X								X											
32	15	7	2	1	8.7	9.4								X										X						
33	15	6	7	1	7.8	8.7								X										X						
34	15	5	6	1	7.6	7.8								X										X						
35	15	9	5	1	7.3	7.6								X										X						
36	15	1	9	1	6	7.3								X										X						
37	16	5	6	2	9.8	10.2	X										X	X												
38	16	9	5	2	9.5	9.8	X										X	X												
39	16	1	9	2	8	9.5	X										X	X												
40	17	7	11	1	10.2	10.6		X																			X			
41	17	6	7	1	9.3	10.2		X																			X			
42	17	5	6	1	9.1	9.3		X																			X			
43	17	9	5	1	8.8	9.1		X																			X			
44	17	8	9	1	8	8.8		X																			X			
45	18	7	6	1	18	19.5		X																				X		
46	19	7	2	1	8.1	8.6					X								X											

Table 4.7: Truck and driver assignments to load segments for the case study

The segments for deadheading the trucks and drivers back to their home bases are added after the segments for the shipments. And the deadheading segments for the trucks are added first. Table 4.8 presents a picture of these assignments. The segments up to #139 are for deadheading trucks back to their home nodes. And as can be seen, both a truck and the requisite number of drivers are assigned to each of those segments. The segments beyond #139 are for drivers, and no truck is assigned. Presumably, some other conveyance is used to transport the drivers back to their home nodes. In some instances, multiple drivers are assigned to the same deadheading segment.

Seg	Load	From	To	nDrvr	Dep	Arr	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9		D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8	D-9	D-10	D-11	D-12	D-13	D-14	D-15	D-16	D-17	D-18
92	-36	6	3	2	3	3	X														X				X									
93	-10	6	4	2	4	4			X											X							X							
94	-24	5	9	2	1	1				X											X				X									
95	-1	5	8	2	0	0					X														X	X								
96	-20	10	9	2	2.7	3				X												X	X											
97	-39	5	9	2	0	0						X											X	X										
98	-2	11	9	2	2.1	3.7						X						X										X						
99	-31	5	9	2	0	0							X								X	X												
100	-3	11	7	2	2.2	2.4							X						X	X														
101	0	6	5	1	3.9	4.1											X					X											
102	0	10	5	1	4.1	4.2																X											
103	0	10	7	1	4.4	5.6																					X						
104	0	11	5	1	2.8	4.1																	X										
105	-16	6	1	2	6.4	8	X										X	X																
106	-12	5	10	2	6	6.1				X											X					X								
107	-29	5	6	2	6	6.2							X				X	X																
108	-15	5	1	2	6	6								X									X	X										
109	-19	5	6	2	6.8	7					X								X			X												
110	-25	5	6	2	6	6.2						X								X					X									
111	-17	7	8	2	6	8		X																			X	X						
112	-6	1	8	2	8.2	8.9				X											X					X								
113	-27	7	6	2	7.7	8.6						X													X						X			
114	0	1	6	1	8.5	10.1																							X				
115	0	2	5	1	9.4	10.8												X					X										
116	0	3	5	1	7.8	9.2													X														
117	0	6	5	1	10.4	10.6														X													
118	0	7	5	1	10.1	11.2																		X									
119	0	11	7	1	10.6	10.8																				X							
120	-5	7	2	2	12	12.3		X																			X	X						
121	-14	5	6	2	12.8	13					X								X				X											
122	-32	5	1	2	12	12										X							X				X								
123	-30	6	2	2	12	13.2			X																								X	X
124	-9	5	4	2	12	13							X				X	X																
125	-26	5	6	2	12	12.2						X												X	X									
126	-11	6	9	2	15.5	16			X									X										X						
127	-37	5	10	2	12	12.1				X										X	X													
128	-40	5	8	2	14.1	15								X									X	X										
129	-33	6	10	2	14.1	14.4		X																				X			X			
130	-35	6	10	2	14.4	14.7							X				X	X																
131	0	2	5	1	14.8	16.2												X													X		
132	0	6	5	1	16.6	16.8																X				X							
133	0	6	7	1	15.4	16.3										X																	
134	0	11	5	1	13.9	15.2													X	X													
135	-7	5	10	2	18	18.1					X											X				X								
136	-34	6	4	2	18	18.8			X																				X	X				
137	-4	5	6	2	18	18.2										X					X	X													
138	-8	5	6	2	18	18.2							X											X	X									
139	-21	5	8	2	18	18.9						X					X	X																
140	0	1	5	1	20.1	21.5																	X		X								
141	0	3	6	1	20.3	21.5																							X				
142	0	4	6	1	19.5	20.3																											X
143	0	6	5	1	20.7	20.9										X																	
144	0	6	7	1	19.5	20.4														X													
145	0	11	5	1	20	21.3																		X									
146	0	5	8	1	0	0										X																	
147	0	5	9	1	0	0											X																
148	0	5	9	1	0	0												X	X														
149	0	8	5	1	0	0.9																		X									
150	0	9	5	1	0	0.3														X													
151	0	8	7	1	0	2																			X								
152	0	7	11	1	2.1	2.1																					X						
153	0	9	6	1	1	1.5														X				X									
154	0	3	4	1	3	3.8																		X									
155	0	9	4	1	0.9	2.2										X																	
156	0	5	10	1	4.3	4.4																					X						
157	0	4	5	1	4	5													X														
158	0	4	7	1	4	5.7																				X							
159	0	6	5	1	5.2	5.4										X				X				X									
160	0	7	5	1	2.4	3.5												X															
161	0	10	5	1	4.4	4.5															X				X								

Table 4.8: Truck and driver assignments to deadheading segments for the case study

The performance of this solution is portrayed in Figure 4.5. The horizontal axis shows when the shipment was made available for pickup. The vertical axis shows when it was picked up. If the solution were ideal, all the points would lie on the diagonal line where the two times are equal. Clearly, while some of the shipments were picked up on-time, many of them were not, being delayed by as much as 6 days in one case.

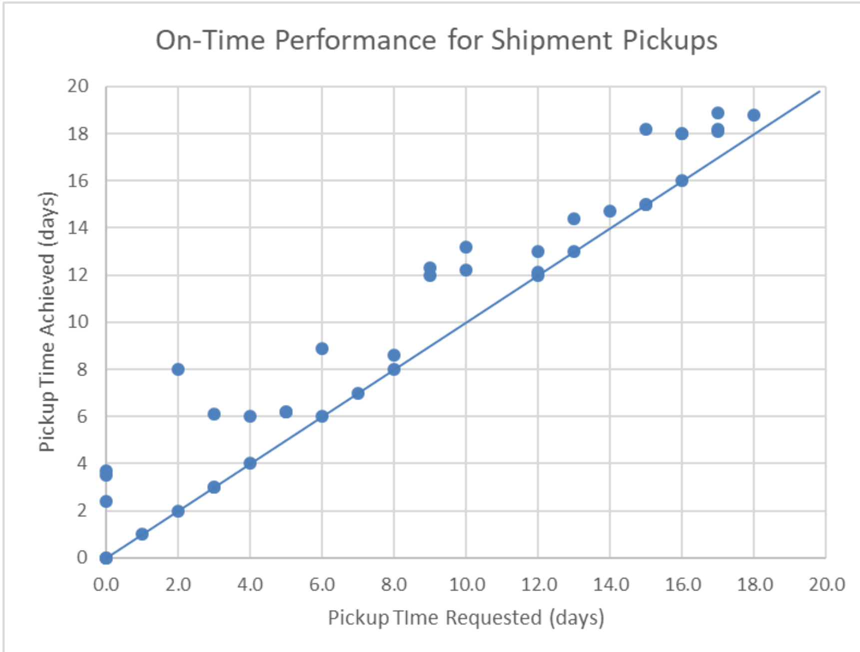

Figure 4.5: Performance of the truck and driver assignments

Figure 4.6 shows the cumulative distribution function (CDF) for the pickup delays. Clearly, many of them are zero, but 60% of them are not, ranging from less than one day up to 6.

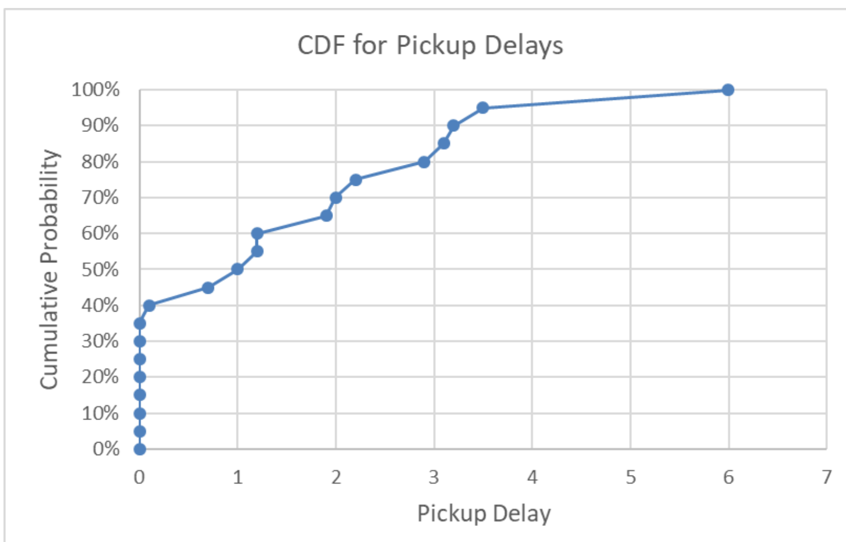

Figure 4.6: Performance of the truck and driver assignments

4.3 SUMMARY

This section has studied a setting where a truckload carrier is providing service between a set of locations. Loads are carried between specific locations. The travel times, service times, and service demands vary by problem realization. The planning horizon is several weeks. Performance is assessed based on the ability to make on-time deliveries. The number of trucks is a user input. Fleet sizes that are too small cause shipment delays, and increasing the fleet size mitigates that impact.

The methodology is derived from work done earlier by List *et al.* (2006). Customers ask to move shipments from one location to another, at different times, and expect that the trucking company (as the service provider) assigns appropriate resources to make those shipments happen.

A hierarchical model is employed that assigns trucks first and drivers second. This ordering implicitly assumes that the trucks are the limiting resource. The truck tours must be formed first, based on the number of trucks available; shipments are delayed, on a prioritized basis, if the fleet size is insufficient. Once the truck trips have been developed, drivers are assigned. The number of drivers is assumed to be "infinite". Or, less dramatically, that a reserve pool of drivers is assumed to exist.

Trucks are assigned so that shipment delays are minimized (i.e., failure to initiate the shipment within its allowed time window). The first available truck that is closest to the origin of the load is selected. Pick up times are delayed if no truck can cover the load at its desired departure time. Low priority shipments are deferred.

Drivers are assigned in a way that "covers" all the segments while minimizing the total driver person-hours and deadheading time. The assignments also respect the regional home bases of the drivers and ensure that the drivers all return home at the end of each week.

The case study example shows that the methodology works and produces defensible, useful results. The truck and driver assignments are easy to follow and understand; they make sense; and they are consistent with the tenets by which the solutions are to be developed. Carriers need to use such procedures to generate truck and driver tours that work and are sensitive to the nuances of both the shipment demands and the operating rules and regulations.

5.0 LESS-THAN-TRUCKLOAD SERVICE

This third setting focuses on the operations of a less-than-truckload (LTL) carrier, which is a very common operating condition for many trucking firms. Packages are picked up at commercial or residential locations, carried from one region to another, and then delivered to a recipient. In-between, those packages are consolidated into truckloads and then disaggregated for delivery.

A hypothetical LTL carrier's network is depicted in Figure 5.1. Triangles 1-18 are places where shipments (packages) can originate or terminate. Circles 19-21 are distribution center (regional terminal) hubs in each of the three regions. Local "box" trucks are used to pick up and deliver packages in the three regional areas. Tractor trailers (larger trucks) are used to carry aggregations of packages from one region to another.

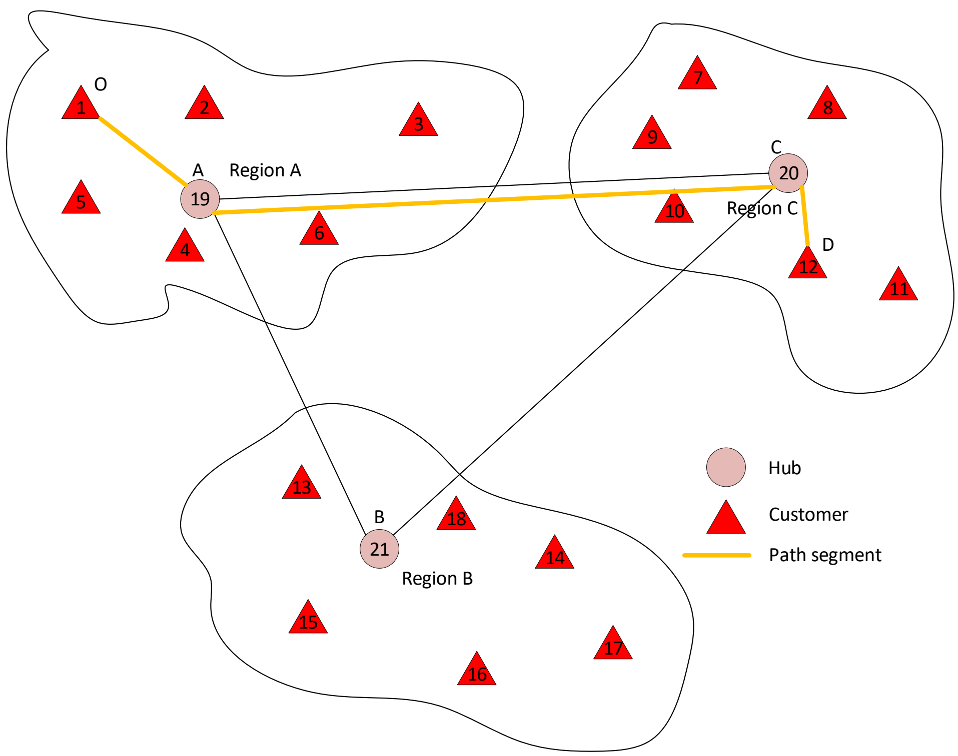

Figure 5.1: A hypothetical LTL network

In this study, the question is: how does stochasticity affect the performance of the network and the investment and operational decisions the carrier must make. As was seen in Section 3, accounting for the stochasticity explicitly suggests making decisions that are more conservative so that resources are available to deal with contingency (stressful) situations.

As has been the situation previously in the study, the impacts of fleet size are of major interest. To deal with contingencies, the carrier is better off with more rather than less trucks; providing slack in the schedule has merit as well. The questions are: how large should the fleet size be? And, how much slack should be included to reach a high percentile of on-time pick-ups and deliveries?

As indicated by the literature review in Section 2, this type of problem has been studied extensively by prior research teams. Tools exist to develop deterministically-based operating plans in response to daily, weekly, and longer timeframes. Some work has been done to address these issues in stochastic situations.

This study is focused on developing and testing a simple, pragmatic tool that can be used to assess system performance in response to fleet size decisions; and in the process of doing that, develop practical, implementable operating plans. It also treats the problem in such a manner that the impacts of resource levels and operating decisions can be assessed against "ideal" service plans.

5.1 METHODOLOGY

The method for developing service plans involves three sub-steps: a) a purely package-based solution, b) the incremental impacts of region-to-region shipping patterns, and c) local truck routing and scheduling plans that result in pick-up and delivery times for the packages. These are part of step 3) in the following analysis procedure:

1) Set fleet sizes for the inter-hub truck fleet (ITF) and the regional pick-up and delivery fleets (RTFs).
2) Establish the baseline conditions for the network: a) shipment demands, b) load and unload times, c) node-to-node travel times, d) deviations from the nominal values of the load, unload, and travel times and likelihoods that they will occur in a problem realization, and e) probabilities that each of the shipment demands will materialize in those realizations.
3) Repeatedly, for many problem realizations
    a. Compute the purely package-based solution
    b. Adjust that solution based on the size of the ITF
    c. Further adjust that solution based on the sizes of the RTFs
4) Summarize the results from the problem realizations

The methods used in each of the sub-steps a), b), and c) in step 3 are described in the sections that follow. The methods for steps 1), 2), and 4) are described in the context of a case study example.

5.1.1 The Purely Package-Based Solution

In this sub-step, the packages are individually examined to determine when they can ideally be delivered based on the stochastic variations and the operating plan employed. This solution serves as a benchmark against which to compare the plan developed subsequently. It makes use of a) an earliest time when the package can be picked up, b) the travel time from the origin location to the regional hub, c) the travel time between the regional hubs, d) the travel time from the destination hub to the delivery location, and e) the operating rules by which the shipments are handled.

It will be seen that the latter aspect of the assessment has a significant effect on the results obtained. For example, consistent with most LTL carrier operations, it is assumed that intra-regional shipments received on day k are scheduled for delivery on day k +1 even though, in principle, they could be delivered the same day. This builds considerable slack into the delivery schedule for intra-regional shipments. Second, and in a similar fashion, packages that arrive at a destination hub after a pre-specified cut-off time t_c on day k (in time to be loaded onto the RTFs for delivery) are scheduled for delivery on day $k + 1$. They may sit at the regional hub for as much as a little more than a day until t_d, the earliest point in time when trucks can leave the hub in the morning to make deliveries. (This means $t_c < t_d$ on any given day, and the hours between t_c on day k until t_d on day $k + 1$ is more than 24. There are more operating plan details that affect the operation of the ITF and the RTFs, but they will be addressed later.

The process for this first sub-step, then, is as follows:

1) Select a shipment that is part of this realization.
2) Identify the beginning of its pick-up window. Treat this as the pick-up time for the shipment.
3) Move the package to the origin hub based on the load time and the travel time to the origin hub.
4) Move the package to the destination hub based on the inter-hub travel time.
5) Determine when the package will be loaded for delivery. If t_c is the cut-off time for packages to be scheduled for loading, then a package arriving after t_c on day k is scheduled to be depart the hub for delivery starting at t_d on day k +1.
6) Move the package to its destination based on t_d for day $k + 1$.
7) Record the deliver time
8) Repeat steps 1) through 7) for all shipments in this realization
9) Record the results for the realization

This process is repeated for each of the problem realizations considered. An examination of the $n + 1^{st}$ realization commences once the two subsequent sub-steps (the inter-hub transfers and the local pick-ups and deliveries) have been accomplished for the n^{th} realization.

Effectively, this method assumes that one truck is dedicated to the movement of each shipment. Not an implementable solution but an interesting perspective; each truck picks up a package, takes it to the origin hub, moves it to the destination hub, and delivers it to the destination. What it provides is a "very best case" benchmark against which to assess more realistic solutions. It also provides a mechanism to assess the direct, incremental impacts of operational decisions that affect how quickly packages can be moved from origin to destination.

5.1.2 The Inter-Hub Transfer Plan

The second sub-step is to account for the impact of the inter-hub transport activities. Basically, the size of the ITF affects the ability of the system to have packages arrive at the destination hub as quickly as the solution from step #1 suggests that should be possible. If the ITF size is sufficiently large, there is no impact. That is, despite the non-zero headways between ITF truck departures from the origin hub r to the destination hub s, the day-to-day operating slack from t_c on day k until t_d on day $k + 1$ provides enough buffer for there to be no effective difference between the solution from sub-step #1 and the one that can be achieved (assuming the RTF sizes are also large). But, if the ITF size is smaller, the package arrival time at s may be later than t_c on day k. And, if so, this cascades into delays to the departure time for delivery until t_d on day $k + 2$, $k + 3$, etc.

Clearly, focusing on the ITF first is a choice. The RTF problem could have been addressed first, or the two could have been addressed simultaneously. There are interactions between impacts of the sizes of the ITF and RTF fleet sizes, and these would be manifest in an iterative or simultaneous problem solution methodology, but in the case, like here, where the problem is solved sequentially and incrementally, assessing the impacts of the ITF size first seems more logical. First, the regional pick-up and delivery problem cannot be solved until the package arrival times at the destination hub are known. And, second, while solutions involving less-than-adequate RTF sizes are interesting, intellectually, they are not practical for implementation. So, to solve the regional pick-up and delivery problem, and to have a sense of the impact of fleet size on the solution, realistic package arrival times at the destination hub are needed, and those times are affected by the size (and utilization) of the ITF trucks.

The structure of the inter-hub package transport problem is much like the one addressed in Section 4. ITF trucks carry "truckloads" of packages from one hub to another. They are then released from that truckload shipment and dispatched to the next truckload (trailer-load) transport task. The difference is, that in this case, there is consideration of loading the packages into the trucks, and unloading them. There is also the consideration that, unlike the truckload problem in Section 4, the carrier is in control of the departure times for the trucks going from r to s. It can have them depart when either a) they are full, or b) a maximum headway has been reached since the last truck to travel from r to s.

Hence, the solution strategy is to treat this problem as a variant of the problem described in Section 4 in which a) packages are assembled and loaded into waiting ITF trucks as the packages arrive at the origin hub, and b) trucks depart the origin hub for their destination hub when either they are full or a maximum departure headway is reached (24, 12, 8, 6, etc. hours).

The solution methodology can be described as follows:

1) Import the package shipment plan from sub-step #1 for the current problem realization. Treat this as the sequence of package arrivals to be accommodated (regardless of origin hub). Treat the hub arrival time from the package sub-step analysis as the target time for loading the packages.
2) For each package, j, in the sequence of arrivals at any one of the regional hubs:
    a. Assign it to truck i for transport to s if a truck i, has already been designated for transporting packages from r to s.
    b. If no truck i is designated for transporting packages from r to s, identify, then, from among all the trucks in the ITF, select the next one that can be designated for that purpose (sent from wherever it is to r at the earliest possible time). Make that truck be truck i in the context of a) above and assign package j to this new truck i.
    c. Check if truck i is now full; and if it is, allow it to depart.
    d. Also check to see if the current time t is at or beyond t_{rs}, the next time when a truck can depart from r to s based on the policy headway h_{rs}. If $t \geq t_{rs}$, then allow the truck i to depart and update the allowable departure time to be one headway later: $t_{rs} \leftarrow t_{rs} + h_{rs}$.
    e. If truck i is released from r, have it travel to s, and then unload its packages. Identify the hub arrival times.
    f. Assess whether the loading (pick-up) event was early, late, or delayed. The truck was early if it was available to load the package before the target time from 1) above. It was late if it was available later than the target time plus a user-specified on-time window. It was delayed if the departure of the truck (as opposed to the loading time) was beyond the target time plus the on-time window.
    g. Determine the earliest time when each package could be delivered to its destination. Consistent with the rules followed in sub-step #1, the earliest point in time when it can be loaded onto an RTF is t_d on day k or $k + 1$ depending upon whether the hub arrival time is before or after t_c on the arrival day k. Then, the earliest possible deliver time is t_d plus the travel time from the

destination hub to the destination location. Record this best possible delivery time and use it as the target delivery time in sub-step #3 instead of the delivery time produced by sub-step #1.

3) Repeat step #2 above until all packages have been handled.

Once this processing is completed, the main outcome is that new, more defensible times have been established for the arrival of packages at their destination hubs. These times are then used in sub-step #3 that focuses on the pick-up and delivery activities at the regional hubs.

5.1.3 The Regional Pick-Ups and Deliveries

The third sub-step is to solve the daily pick-up and delivery time problem for each regional hub. From sub-steps #1 and #2 above, the desired times for pick-up are known (they were given in the specification of the problem realization) and the possible times for delivery departure are also known. (They were produced as an outcome from sub-step #2.)

Operationally, the algorithm assumes that the pick-up and delivery activities on day k commence at t_d, the earliest time when any RTF truck leaves the regional hub. Also, it assumes that all trucks are to return to the hub by t_b. The latter time is a target, though, not a hard constraint, because the time at which the last truck returns to the hub is affected by the size of the regional fleet. (A smaller fleet size means a later last RTF truck return time.)

Also, the algorithm assumes that the RTF trucks are in regional fleets based on user inputs. These fleet sizes are fixed. The trucks are not shareable between regions. RTF trucks can make both pickups and deliveries during their tours. Moreover, each RTF truck makes one tour per day, i.e., it leaves the hub loaded with packages to deliver and returns with packages it has picked up.

In the context of target times, the customers set the target times for pick-ups. They are inputs from the problem data. The delivery times are computed based on the results from sub-step #2.

For a given problem realization is as follows:

1) Assemble the time-sequenced list of packages J to be picked-up or delivered during the day.
2) For each package, j, in the time-sequenced set J:
   a. Identify the package's pick up or deliver time, t_j, and delivery location s.
   b. Find the best truck for picking up or delivering package j. This is the truck i which has a current partial tour (including null) that ends at a location r, at time t_{ir} such that its arrival time at s, $t_{is} = t_{ir} + \Delta t_{rs}$ is the smallest among all trucks, where Δt_{rs} is the travel time from r to s. That is, select truck i if its t_{is} is the earliest possible arrival time at s.This truck might be a new truck dispatched from the regional hub.
   c. Move truck i to s and add package j to its list of packages to be picked up or delivered. Also keep track of the total load on the truck (space and weight) so this can be compared with the truck's capacity(ies).
3) Repeat step #2 above until all packages have been handled.

Since the RTF sizes are user inputs, there is no guarantee that the capacities of the trucks will not be exceeded. The routine does not defer pickups or deliveries (place them later in the sequence) because a capacity limit has been reached. Hence, a hard capacity constraint is not imposed (unless for a price, new trucks can be added to the RTFs, which is an option not presently available). Rather, when the results are analyzed, it is important to assess the extent to which the RTF truck capacities have been exceeded.

5.2 CASE STUDY

The case study is hypothetical and predicated on the network shown in Figure 5.1. There are 18 nodes in three regions between which shipments can take place, and each region has a hub.

5.2.1 Basic Input Data

The timespan examined is 6 days (from $t = 0$ to $t = 6$). A rectilinear coordinate system sits behind the network and it is used to obtain Euclidean distances between the nodes. Those distances, in conjunction with a random circuity ranging uniformly between 1.0 and 1.2 is used to create nominal travel distances and times. The solutions have the packages travel from their origin node to the origin region hub node via a portion of an origin region RTF truck tour, then from the origin region hub to the destination region hub on a single ITF truck, in a non-stop move, and then from the destination hub to the destination again via a portion of a destination region RTF truck tour.

There are 52 loads as shown in Table 5.1. For each one, the table shows a shipment number, an origin, a destination, an origin hub, a destination hub (which can be checked using Figure 5.1), a desired pickup time (in decimal days), an ending time for desired pickup, and a service time. A time value of 0.5 implies noon on the first day (12 hours); 1.5 is noon on the second day. The increments of time are one-hundredth of a day, which is 14.4 minutes.

Ship	Orig	Dest	oHub	dHub	Size	bWin	eWin	tSvc
1	12	6	20	19	17	1.70	1.71	0.02
2	14	5	21	19	19	1.68	1.69	0.01
3	11	2	20	19	26	0.67	0.68	0.04
4	11	8	20	20	12	1.61	1.62	0.02
5	2	10	19	20	22	2.56	2.56	0.01
6	4	3	19	19	27	3.63	3.64	0.04
7	10	13	20	21	20	1.69	1.70	0.04
8	10	7	20	20	24	1.67	1.68	0.02
9	13	2	21	19	28	3.40	3.40	0.04
10	14	15	21	21	28	4.46	4.47	0.02
11	3	14	19	21	28	2.63	2.63	0.01
12	2	6	19	19	21	0.44	0.46	0.03
13	15	5	21	19	14	2.44	2.45	0.02
14	12	4	20	19	18	4.62	4.62	0.04
15	5	13	19	21	30	4.47	4.48	0.01
16	17	5	21	19	19	1.67	1.69	0.04
17	6	8	19	20	27	0.60	0.61	0.01
18	9	6	20	19	21	0.60	0.61	0.02
19	17	15	21	21	14	0.68	0.69	0.01
20	7	8	20	20	25	2.47	2.48	0.03
21	7	11	20	20	22	4.38	4.40	0.01
22	16	10	21	20	17	0.66	0.67	0.03
23	7	7	20	20	26	0.56	0.57	0.01
24	8	4	20	19	14	1.39	1.40	0.03
25	5	9	19	20	12	4.71	4.71	0.01
26	12	7	20	20	14	3.46	3.48	0.03
27	8	10	20	20	21	4.54	4.56	0.02
28	16	5	21	19	14	1.57	1.58	0.02
29	1	4	19	19	22	0.60	0.60	0.03
30	13	4	21	19	23	1.49	1.51	0.02
31	9	12	20	20	18	4.47	4.48	0.01
32	17	9	21	20	13	1.55	1.57	0.03
33	10	7	20	20	18	1.50	1.51	0.02
34	17	3	21	19	27	2.44	2.46	0.01
35	1	1	19	19	11	3.41	3.41	0.02
36	6	2	19	19	22	1.39	1.40	0.04
37	3	5	19	19	17	3.39	3.41	0.02
38	4	13	19	21	23	2.45	2.45	0.03
39	4	12	19	20	28	4.44	4.45	0.03
40	12	8	20	20	12	1.53	1.53	0.04
41	19	1	19	19	22	0.69	0.7	0.03
42	19	3	19	19	23	0.74	0.74	0.02
43	19	5	19	19	18	0.61	0.62	0.01
44	19	4	19	19	13	0.54	0.55	0.03
45	20	7	20	20	18	0.52	0.53	0.02
46	20	12	20	20	27	0.48	0.49	0.01
47	20	9	20	20	11	0.53	0.54	0.02
48	20	8	20	20	22	0.47	0.48	0.04
49	21	17	21	21	17	0.68	0.69	0.02
50	21	13	21	21	23	0.48	0.49	0.03
51	21	15	21	21	28	0.48	0.49	0.03
52	21	16	21	21	12	0.55	0.56	0.04

Table 5.1: Packages to be transported

The first 42 shipments have both an origin and a destination among the 18 pickup and delivery locations. Those shipments occur during the week. The last 10 are deliveries only. Their "origin" location is the hub in the region to which the package is going. These last 10 packages are hold-overs from the prior week that are waiting to be delivered.

The hours of operation are 8am – 6pm. That is, the box trucks start their pickup and delivery tours at 8am and are expected to have returned by 6pm. This is a 10-hour day for the drivers. The cutoff time for incoming packages is 6am. That is, if a package is at the destination terminal by 6am on a given day, then there is enough time to get it set for delivery that day. (This pertains primarily to packages arriving on ITF trucks from other hubs.) If the package arrives after 6am on day k, it is deferred for delivery until day $k + 1$.

The last 10 packages are instructive to review in terms of timing. Since they are deliveries only, the bWin value is the earliest time the package could reach the destination location if a truck left the hub at the beginning of the day and carried it directly to the destination location. As explained before, in the first and second sub-steps (for packages and ITF trucks) these times are calculated both by the "pure package" analysis and the ITF truck analysis and then passed on to the RTF routine. They are the earliest possible delivery times when packages could be delivered.

5.2.2 Deterministic Results

Outputs from the model accrue in two spreadsheets within a single Excel workbook. The "Plans" spreadsheet shows the operating plans developed by the model for each of the problem realizations considered. There are three plan tables: 1) consideration of the packages alone (created first), 2) consideration of the ITF trucks moving inter-regional packages from one location to another (created second), and 3) consideration of the RTF trucks doing pickups and deliveries in their respective regions (created third).

The model can be run in one of two modes. In the first, a single realization is examined predicated on the load, unload, and travel times that appear in the input data worksheet. That one is illustrated here. In the second, multiple realizations are analyzed; how many is chosen by the analyst, and the plans for each realization are placed in the "Plans" spreadsheet with a marker indicating which realization produced entries within the plans. That one is illustrated in the next sub-section. In the stochastic analyses, averages of the results (e.g., actual pickup and delivery times) are placed in the input worksheet that contains the shipment information as well as performance information about the RTF and ITF trucks employed. In this analysis, those values are deterministic, because they are the result of a single realization analysis.

An example of the output from the package sub-step analysis are shown in Table 5.2. Two columns of output are shown. The first column presents results for packages 1-4. The second shows packages 5-8. For each record in both columns, the entries are the scenario number, the event number (for tracking event sequences), the location, the associated regional hub, the package number, and the event number (1 is a pickup at the origin, 2 is arrival at the origin hub, 3 is arrival at the destination hub, 4 is scheduled for delivery from the destination hub, and 5 is arrival at the destination hub). Five records can be seen for each package corresponding to the event numbers just described: 1) pickup at the origin, 2) arrival at the origin hub, 3) arrival at the destination hub, 4) available for delivery from the destination hub, and 5) delivery to the destination location. All times are integers in hundredths of a day. So, 2.64 is 64% of the way through day 3 (after day 2). This is about about 3:30 on day 3. In the table, package #1 is picked up at t = 170 (tEv), which is the same as 1.70 days (or just shy of 5pm on day 2). It arrives at the origin region hub at t = 1.82; since there is no constraint on the subsequent departure, it immediately leaves that location and travels to the destination hub. It arrives there at t = 2.45 (about midday on day 3). It then gets scheduled to depart for delivery at 3.33 (8am on day 4), and it arrives at the destination location at t = 3.56 (just after noon on day 4). As explained before, these are times that assume, implicitly, that the package can move itself, or that a truck is dedicated to moving the package from origin to destination. This is constrained only by the travel times on the network and the operating rules for when packages can leave the destination hub for delivery. Even though these arrival times are not likely to be achievable, they are benchmarks against which to compare the results from arrival times predicated on the ITF and RTF trucks available and their utilization.

Scen	nEv	tEv	loc	hub	pNum	eNbr
1	113	170	12	20	1	1
1	117	182	20	20	1	2
1	150	245	19	19	1	3
1	177	333	19	19	1	4
1	194	356	6	19	1	5
1	109	168	14	21	2	1
1	123	188	21	21	2	2
1	126	213	19	19	2	3
1	132	233	19	19	2	4
1	159	256	5	19	2	5
1	34	67	11	20	3	1
1	52	88	20	20	3	2
1	91	151	19	19	3	3
1	131	233	19	19	3	4
1	165	264	2	19	3	5
1	100	161	11	20	4	1
1	115	180	20	20	4	2
1	116	181	20	20	4	3
1	136	233	20	20	4	4
1	146	242	8	20	4	5
1	158	256	2	19	5	1
1	172	288	19	19	5	2
1	192	351	20	20	5	3
1	208	433	20	20	5	4
1	216	446	10	20	5	5
1	198	363	4	19	6	1
1	205	383	19	19	6	2
1	206	384	19	19	6	3
1	213	433	19	19	6	4
1	236	469	3	19	6	5
1	111	169	10	20	7	1
1	121	186	20	20	7	2
1	144	240	21	21	7	3
1	179	333	21	21	7	4
1	188	343	13	21	7	5
1	106	167	10	20	8	1
1	118	182	20	20	8	2
1	119	183	20	20	8	3
1	139	233	20	20	8	4
1	155	247	7	20	8	5

Table 5.2: Illustrative Event Times Based on the Packages Alone

After the package-based plan has been created, its results are passed to the sub-step that develops transport plans for the inter-hub package transfers. Table 5.3 presents a partial picture of the results from that sub-step. It shows information for packages 1, 2, 3, 5. 7, 9. 11, 13, 14, and 15. The columns are: the scenario number, the number of the event (to keep track of event sequence), the time at which the event occurred, the location, the associated hub, the ITF truck employed (from the free-running fleet), the package number, the event number (7 is a loading, 8 is a delivery), and assessments as to whether the loading was early, late, or delayed (descriptions of these assessments are provided in the next paragraph). In the case of package 3, for example, the load time for inter-hub transfer at the origin hub is t = 0.88. This matches the arrival time at the origin hub for package 3 in Table 5.2 (the second record), as it should. Then, it arrives at the destination hub at t = 1.58 given the Δt_{rs} travel time of 0.63 days between the hubs. Since this is after t = 1.25, the cutoff time for loading packages for delivery on day 2, it is scheduled for loading for delivery at 2.33 (8am on day 3). The sub-step also extends this result to create a target for when the package should be delivered to its destination. This value is computed by adding the travel time from the destination hub to the destination location to the “available for loading” time, which was just illustrated. So, in the case of package 3, which is destined for node 2, the travel time from the region A hub to node 2, which is 0.31 hours is added to the “available for loading” time of 2.33 so that the target time for delivery is 2.64 (just after 3pm on day 3). This result is not shown in the Table 5.3, but it is recorded by the software for use by the RTF scheduling sub-step.

Scen	nEv	tEv	loc	hub	ttNum	pNum	eNbr	early	late	delay
1	14	261	20	20	2	1	7	0	69	61
1	15	328	19	19	2	1	8	0	0	0
1	20	206	21	21	4	2	7	0	8	0
1	23	335	19	19	4	2	8	0	0	0
1	7	88	20	20	1	3	7	25	0	0
1	8	158	19	19	1	3	8	0	0	0
1	30	383	19	19	2	5	7	0	85	76
1	31	448	20	20	2	5	8	0	0	0
1	18	296	20	20	1	7	7	0	100	94
1	19	358	21	21	1	7	8	0	0	0
1	34	361	21	21	1	9	7	0	0	0
1	36	436	19	19	1	9	8	0	0	0
1	32	424	19	19	3	11	7	0	114	105
1	33	451	21	21	3	11	8	0	0	0
1	25	353	21	21	2	13	7	0	86	78
1	26	382	19	19	2	13	8	0	0	0
1	41	604	20	20	4	14	7	0	118	112
1	42	675	19	19	4	14	8	0	0	0
1	39	534	19	19	1	15	7	0	53	44
1	40	561	21	21	1	15	8	0	0	0

Table 5.3: Partial Results for the Inter-Hub Transfers

An assessment is also performed as to whether the ITF truck was early, late, or delayed insofar as the inter-hub loading event is concerned. The truck was late if it was available later than the target time plus a user-specified on-time window. It was delayed if the departure of the truck (as opposed to the loading time) was beyond the target time plus the on-time window.

Once the inter-hub transfer analysis has been performed, the RTF scheduling sub-step is performed. The concurrent scheduler described in Section 5.1.3 is employed to develop tours for each RTF truck, every day in each region. The target pick-up time targets are obtained from the package sub-step analysis (which are the original requested pick-up times). The delivery time targets are obtained from one of two sources. For the inter-hub packages, they are from the transfer analysis step. For the intra-hub packages, they are from the package analysis sub-step.

Partial results from this analysis are shown in Table 5.4. Each row shows the scenario number, the event number (for tracking sequence), the time of the event (in hundredths of a day), the location, the associated regional hub number, the box truck (RTF truck) number, the package number, the event number (7 is a pickup, 8 is a delivery), and an assessment as to whether the event was early, late, or delayed.

Focusing again on package #3, two events are listed, the pickup at the origin and the delivery at the destination. The pickup time is t = 0.67, which means it occurred on-time. And it is consistent with the package-based analysis shown in Table 5.2. In addition, Table 5.4 shows that the truck arrived about an hour early (the early column value is 4 (or 0.04 days). The delivery was at t = 2.75, which is again consistent with the desired time of at t = 2.64 which was shown in Table 5.2. More particularly, the truck was 15 minutes late (the late value is 1, or 0.01 days). And, since the service time is 0.04 days (about an hour), it was 1.25 hours late leaving, (the delay value is 5, or 0.05 days). Most of the packages are not delivered late. Two are – packages 5 and 9. But, even though this is true, there are 9 packages for which the truck tour logistics caused the pickup event to involve delay.

And, there were 4 delivery events where delays also occurred. Two of those delays were sizeable for packages 5 and 7. The extent to which these events occur early or late, or create delay depends on the size of the ITF and RTF fleets.

Scen	nEv	tEv	loc	hub	btNum	pNum	eNbr	early	late	delay
1	58	170	12	20	6	1	7	10	0	1
1	28	356	6	19	3	1	8	14	0	0
1	95	168	14	21	9	2	7	5	0	0
1	22	260	5	19	2	2	8	0	0	0
1	46	67	11	20	5	3	7	4	0	3
1	25	275	2	19	2	3	8	0	1	5
1	55	161	11	20	5	4	7	0	0	1
1	61	242	8	20	5	4	8	8	0	0
1	18	256	2	19	4	5	7	22	0	1
1	74	534	10	20	7	5	8	0	78	79
1	29	363	4	19	4	6	7	29	0	3
1	37	475	3	19	1	6	8	0	0	0
1	57	170	10	20	5	7	7	0	0	4
1	100	434	13	21	9	7	8	0	81	85
1	56	167	10	20	5	8	7	3	0	1
1	65	264	7	20	5	8	8	0	7	9
1	98	340	13	21	10	9	7	6	0	4
1	35	464	2	19	1	9	8	16	0	0
1	101	446	14	21	8	10	7	12	0	1
1	103	544	15	21	8	10	8	8	0	0

Table 5.4: Partial Results for the Local RTF analysis

Even though this case study is deterministic, since it is only one realization, CDFs are still very useful in showing the results. Figure 5.2 shows CDFs for packages being delivered early, late or delayed.

It is important not to become confused here. There is no stochasticity in this analysis. The figure simply shows the cumulative percentage of packages whose delivery events were early, late, or delayed.

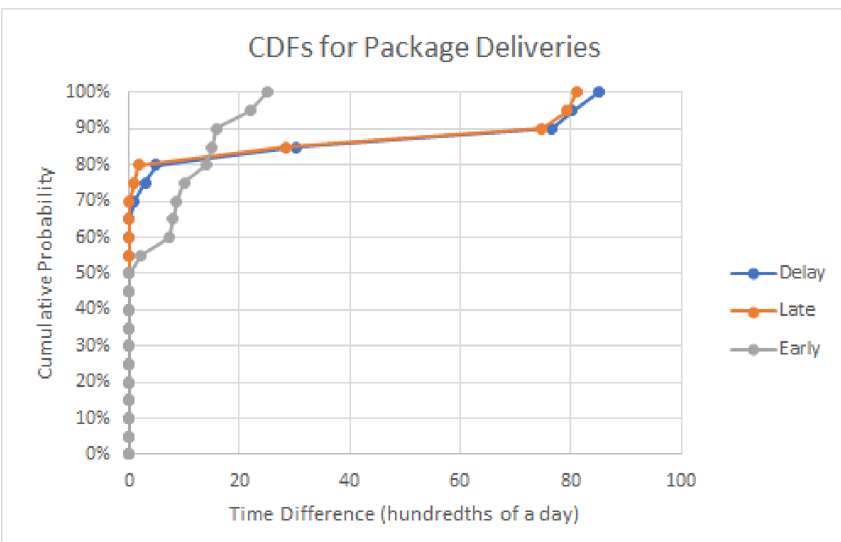

Figure 5.2: CDFs for Package Deliveries

Numerically, Figure 5.2 shows that about 70% of the packages were delivered on-time. The CDF for being late does not diverge from 0 until the 70th percentile. A slightly higher percentage of the deliveries resulted in delay. The "delay" CDF departs from 0 at about the 65th percentile. The next 10% of the delays are small, but then they increase substantially. The largest delays, for about 10%, of the packages involve delays of about 0.75 days or more. The "early" distribution shows that for about 50% of the packages, the delivery event was simply "on-time", the truck did not arrive early. For the remaining 50%, the amount of time by which the truck was early grows to a maximum of 0.25 days (6 hours).

As has been observed several times, these results are highly dependent upon the sizes of the ITF and RTFs. (They are also dependent, in a relative basis, on the number of packages.)

There are many more analysis results that could be presented, but this report will focus only on two. One is the early, late, and delay results for the inter-hub transfer activities. Figure 5.3 presents those results. Most of the inter-hub pickup events were on-time, with no delay, no lateness, and no early arrivals. But, about 15% of them did involve some degree of early arrival, up to nearly a day. And, about 30-35% of them were late, delayed, or both. The greatest late arrival was slightly more than a day. The same is true for the delays. The reason is that there was a juncture during the week when the ITF trucks were all busy and packages had to wait until they could be picked up and/or loaded onto waiting trucks for transport to their destination hubs.

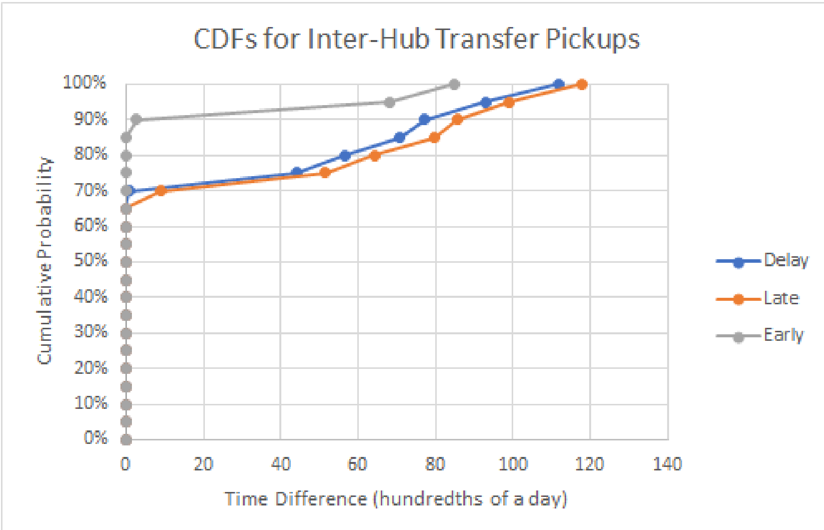

Figure 5.3: CDFs for Inter-Hub Transfer Pickups

The last analysis results to present shows the planned delivery days/times versus the actual values. This is shown in Figure 5.4. The *x* axis shows the intended delivery day/time in hundredths of a day. For example, 150 is noon on the second day. If all the deliveries occurred at the times they were planned, all of these points would lie on the 1:1 diagonal. But, they do not. Reflecting the "delay" results shown in Figure 5.2, some of the packages have small delays. The actual delivery time is slightly greater than the planned value. But, for several of the packages, the delivery day/time is about a day later than was planned. This is an artifact of the inter-hub transfers, i.e., the availability of ITF trucks, and the fact that the arrival times at the destination hub occur after the cutoff time for being loaded onto RTF trucks for delivery. That is, if the package

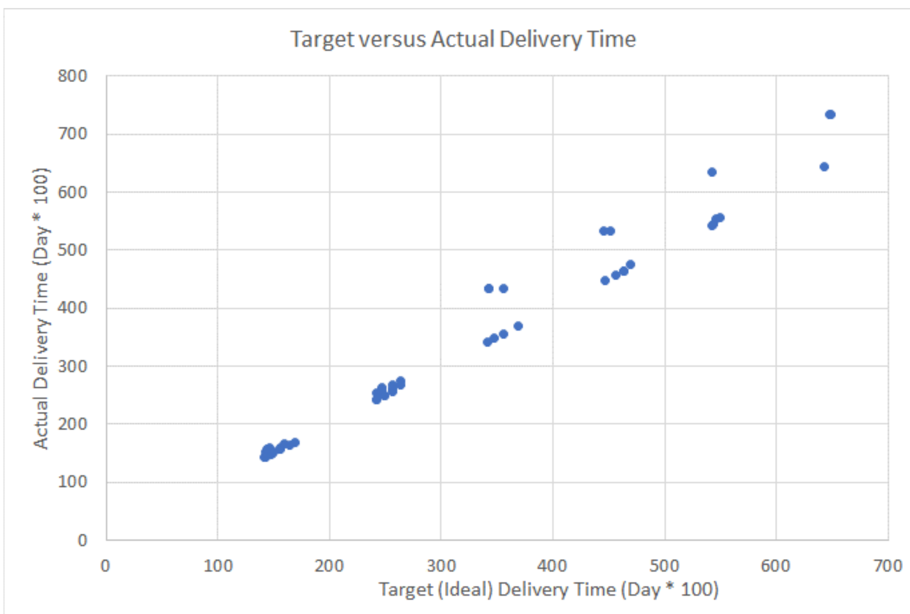

Figure 5.4: Planned versus Actual Delivery Days/Times

analysis, where every package has a truck, has the package arrive at the destination hub at a time before t_c on day k, but the inter-hub transfer ITF truck delivers it after t_c on day k, consistent with Figure 5.3, then the delivery is pushed off to day $k + 1$.

These results are interesting and instructive, but they do not capture the effects of stochastic variations in the problem parameters. They don't provide much information about the reliability of the system. The way in which they can do that is by varying the ITF and RTF fleet sizes and seeing the extent to which smaller (or larger) fleet sizes produce worse (or better) results. In the sense of contingency plans for a deterministic (average) setting, the benefits (costs) of increasing (decreasing) the ITF and RTF fleet sizes can be assessed. But, still, the stochastics are not captured.

5.2.3 Stochastic Results

This section presents the results of a stochastic analysis of the problem setting described and analyzed in the preceding section. Instead of the loading, unloading, and travel times being fixed, they can vary. As with the approach described in List *et al.* (2018), the values can be different from the nominal values by 20% less, 20% more, and 50% more, 10%, 30% and 10% of the time (or other combinations selected by the analyst). For the remaining 50% of the time, the values match the nominal values. These selections are done independently for the load/unload and travel times, treating them as being statistically independent. Each realization of the problem has a different combination of these values. The analyst also determines how many realizations will be examined to develop the stochastic results. For the results presented here, 20 realizations were examined. That number could be made much larger.

Results from the analysis are captured three ways. First, the early, late, and delay results are captured for three events in every realization: pickup at the origin, pickup at the origin node for transfer to the destination hub, and delivery to the destination location. These results can be analyzed statistically to see trends in the distribution of results. Also, the average values for these results are recorded in the original input worksheet. Moreover, the average workloads experienced by ITF and RTF trucks are recorded in the input worksheet. (Clearly, more and additional output results could be obtained.)

Figure 5.5 shows the distribution of averages for the deliveries being early, late, or delayed among the packages. These are the average values among the 20 realizations analyzed. Many of the delay averages are still zero (about 55% of them), but the maximum delay reaches up to more than a day. The same is true for the lateness results. The distribution of early arrivals immediately becomes non-zero and ranges up to 0.17 days (slightly more than 4 hours). These results can be contrasted with those shown in Figure 5.2 for a single analysis at the nominal values. The distributions in Figure 5.5 involve larger values and more values that are non-zero. This is to be expected since the load/unload and travel times are varying.

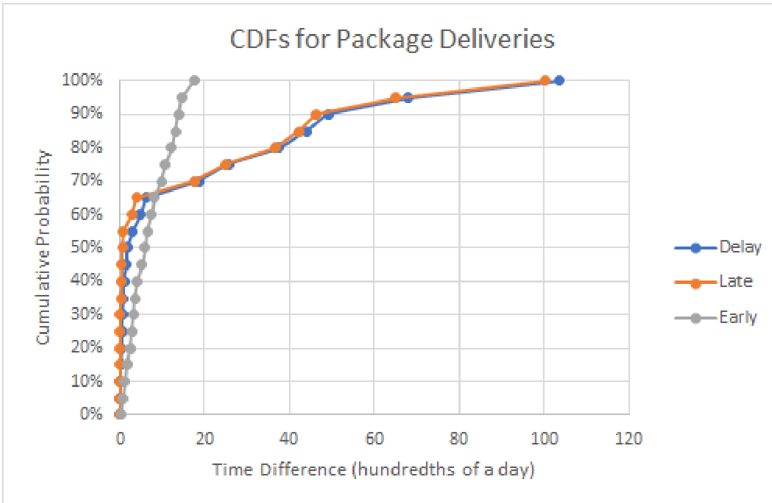

Figure 5.5: Distribution of the Averages for Deliveries

The more revealing results are found in the distributions of the individual values for the early, late, and delay results for each event in every realization. Those results for the package delays are shown in Figure 5.6. There are 1037 deliveries upon which the distributions are based (3 shy of the 1040 that should have been accomplished). While many of the deliveries are accomplished on-time, with no delay and no lateness (as well as no early arrivals), about 40% of the delivery events did involve some amount of lateness or delay or both. Moreover, the maximum delay value is 276 (2.76 days late). This is clearly an impact of the size of the ITF and RTF fleets.

It is important to recognize that this result cannot be seen in the distributions of the average delivery results. That is, while the information presented in both Figures 5.5 and 5.6 are based on the same set of problem realizations, Figure 5.5 only shows the distribution of the average values, while Figure 5.6 presents the distribution of the individual values. This illustrates the risk of predicating fleet size decisions on average values, as was highlighted in Section 3 for the dedicated fleet example. Here, one might conclude from Figure 5.5 that the system performance, while not good, is "adequate" if funding for additional ITF or RTF trucks is not available. But, the distributions in Figure 5.6 show that there are going to be some customers, for whom the quality of the deliveries is not good at all, and twice as bad as indicated by the average values.

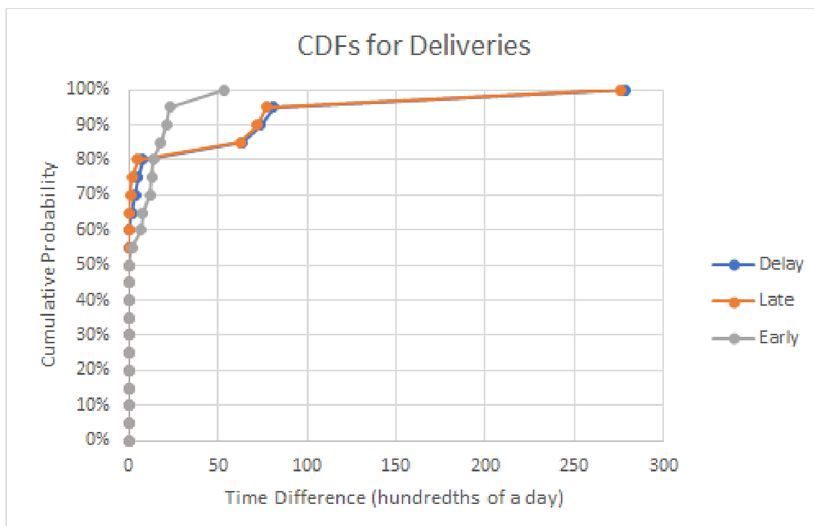

Figure 5.6: CDFs for the Delivery Events

As a final look at this stochastic analysis, Figure 5.7 presents a comparison of the planned delivery times against the actual ones. This is the same result that was presented in Figure 5.4 for the deterministic analysis. Except, in this case, the *average* planned times (because the load/unload and travel times vary) are plotted against the *average* actual values achieved. As can be seen, the same pattern of up to a day of delay is evident, but the scatter in the data points is much richer.

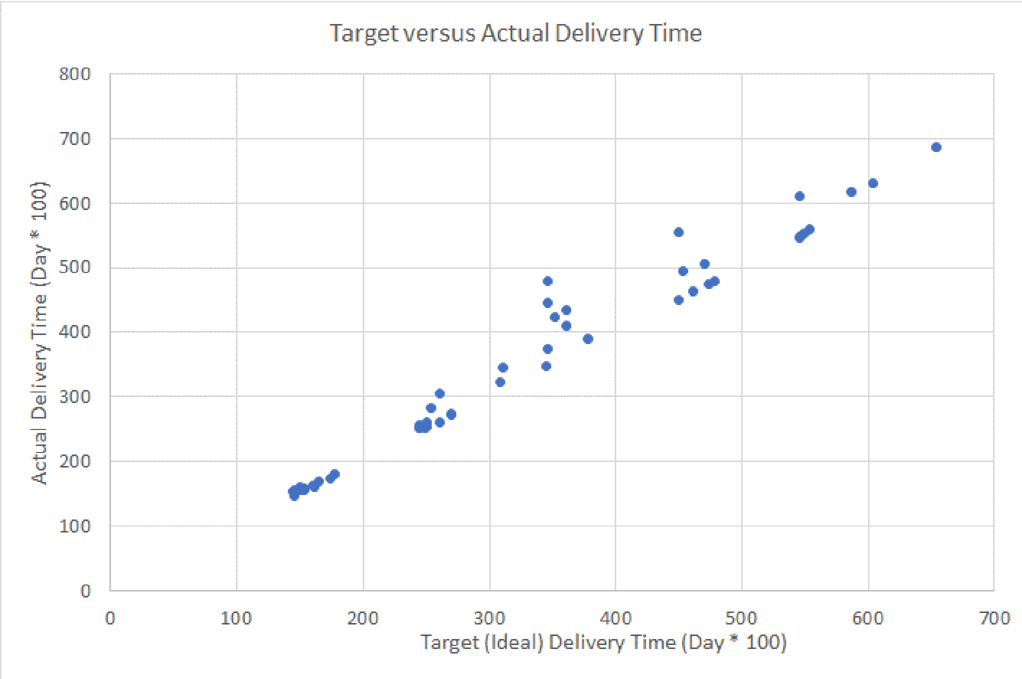

Figure 5.7: Average Planned versus Achieved Delivery Times for the Stochastic Analysis

That the poor performance is due to delays in the inter-hub transfers is evident from Figure 5.8. It shows the early, late, and delay distributions for the inter-hub pickup events at the origin hubs. Like Figure 5.6, although many of the inter-hub pickup events involved no lateness and/or no delay, some of them do, and the maximum lateness and/or delay is up to 289 (nearly three days). Clearly, these delays are the source of the delivery delays.

In addition, as Figure 5.8 shows, increasing the ITF number of trucks form 4 (the original value) to 8 (twice as many) all but eliminates the delays in delivery. (Increasing the ITF fleet to 6 trucks also helps, but it does not all-but eliminate the delays.)

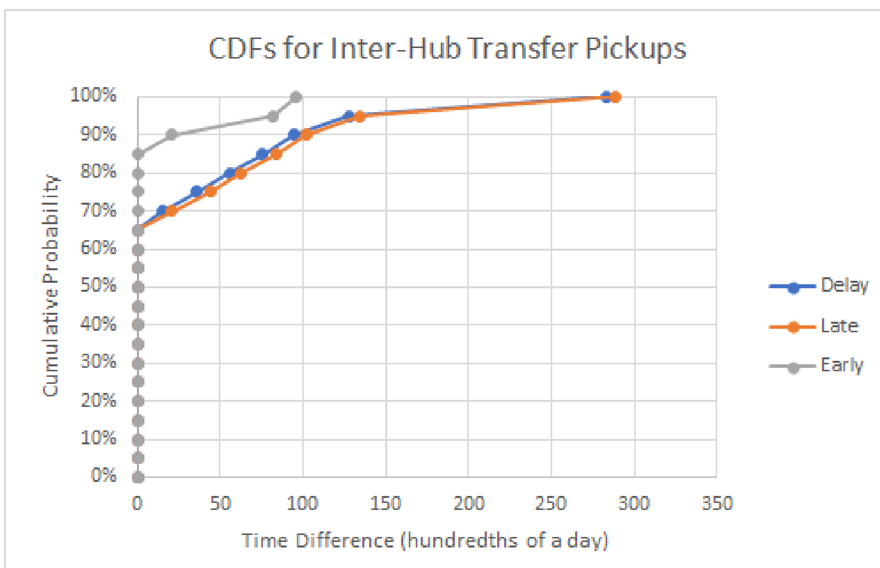

Figure 5.8: CDFs for the Inter-Hub Pickups

Now, with 8 ITF trucks, the distribution of delivery delays has far more zero values and the maximum value shrinks to 75 (18 hours), as is shown in Figure 5.9.

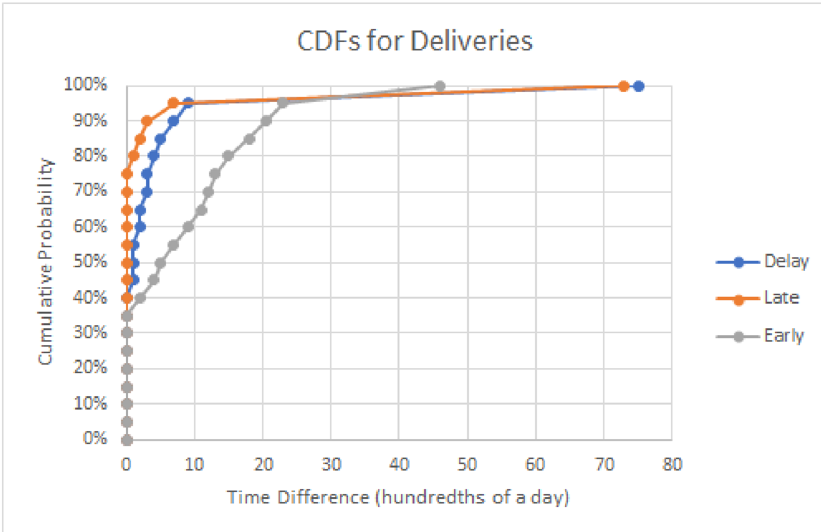

Figure 5.9: CDFs for Deliveries where 8 ITF trucks are used instead of 4

This change illustrates the value in conducting the stochastic analyses. The trucking company can see the value of increasing the ITF fleet size because it has a sense of what percentage of the time packages will be delivered late if more ITF trucks are employed. The same sensitivity analysis can be conducted for changes in the RTF fleet sizes, overall and by region. Similar results are obtained. Increasing the fleet sizes reduces the delays and lateness, and increases the extent to which trucks can arrive at the pickup and/or delivery events early.

To reinforce the point about the changes caused by increasing the fleet size, Figure 5.10 shows the correspondence between the average planned and actual delivery times for the packages if, in addition to the 8 ITF trucks, the RTF fleet sizes are 5, 4 and 3 trucks; this amounts to one more truck each for regions A and B. Now, although some of the packages are still delivered a little late, the points effectively all lie along the 1:1 diagonal.

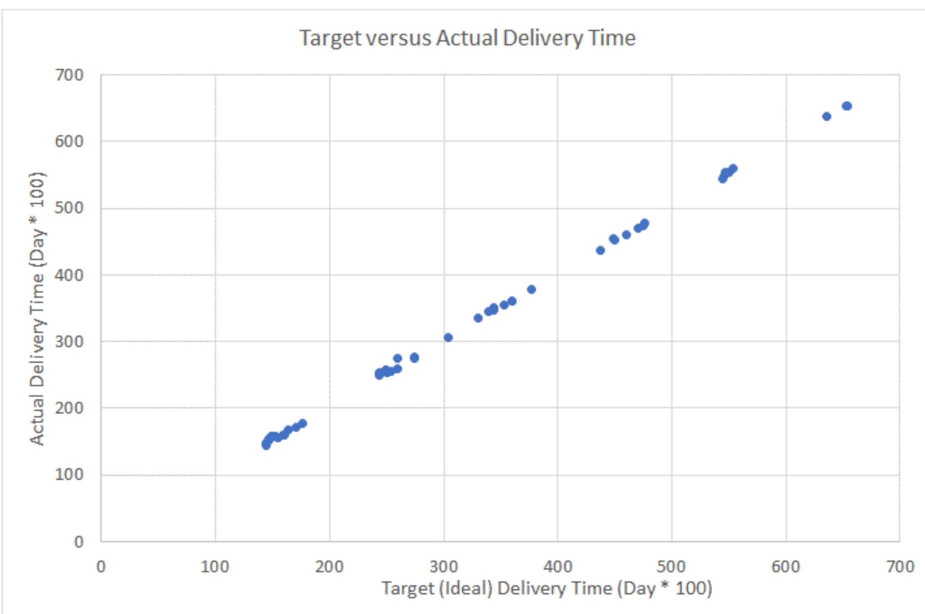

Figure 5.10: Average Planned versus Achieved Delivery Times for the Stochastic Analysis when 8 ITF trucks are used instead of 4. And the RTF fleet sies are 5,4, and 3 trucks respectively

5.3 SUMMARY

This section has presented the study of a less-than-truckload (LTL) operation, which is the mode in which many trucking firms operate. Packages are picked up by local trucks, carried from one regional hub to another by inter-hub trucks, and then delivered to their destination by local trucks in the destination region. For the inter-hub transfers, the packages are consolidated into truckloads and then handled individually at the regional hubs.

The method that creates the solutions presented here uses three sub-steps to develop the package handling plan: a) a purely package-based solution, b) the incremental impacts of region-to-region shipping patterns, and c) local truck routing and scheduling plans that result in pick-up and delivery times for the packages. The methods used in each of the sub-steps were described in Section 5.1.

The method can be applied in both a deterministic and stochastic mode. In the deterministic mode, a single realization is examined predicated on the load, unload, and travel times that appear in the input data worksheet. In the stochastic mode, multiple realizations are analyzed, based on how many are requested by the analyst, and the plans for each realization are placed in a "Plans" spreadsheet with a marker indicating which realization produced entries within the plans. The averages of from these realizations (e.g., actual pickup and delivery times) are placed in the input worksheet that contains the shipment information as well as performance information about the RTF and ITF trucks employed.

A hypothetical operation was studied to illustrate use of the method. Both deterministic and stochastic applications are presented. The case study network breaks down into three regions. Each region has a hub and six origin/destination locations. Local "box" trucks pick up and deliver packages in the three regional areas. Tractor trailers (larger trucks) carry aggregations of packages from one region to another. The timespan examined was 6 days (from $t = 0$ to $t = 6$). The solutions show how the packages travel from their origin node to the origin region hub node via a portion of an origin region RTF truck tour, then from the origin region hub to the destination region hub on a single ITF truck, in a non-stop move, and then from the destination hub to the destination again via a portion of a destination region RTF truck tour.

The main question of interest was: how does stochasticity affect the performance of the network and the investment and operational decisions the carrier must make. As was seen in Section 3, accounting for the stochasticity explicitly suggests making decisions that are more conservative so that resources are available to deal with contingency (stressful) situations.

One of the main insights from the case studies is that the relationship between the reliability of the service provided and the size of the fleet employed. As would be expected, it shows that higher reliability is provided by larger fleet sizes. But more importantly, it shows that the stochastic analysis allows quantification of the extent of that improvement. It also allows the analyst to see if there are consistent patterns in the assignment of customer visits to trucks across the problem realizations examined.

The contrast between the deterministic and stochastic analyses shows the value in conducting the latter. It is easy to see and quantify the value of increasing the ITF fleet size because the model gives a sense of how much the system performance will improve if the fleet size is increased.

6.0 SUMMARY AND FUTURE WORK

6.1 SUMMARY

This research project focused on studying ways in which carriers can plan freight services in a stochastic environment, with an eye toward maximizing reliability and efficiency. It used truck system operations as a context although the tools developed can be applied to other modes and multi-modal systems. Studying the operation of transport systems under stochastic conditions is important because the negative service quality impacts of stochasticity can be significant and there are losses in economic value.

As is well recognized, carriers aim to provide services where package pickups and deliveries occur on-time, within allowable windows (OTWs) a very high percentage of the time. This maximizes customer satisfaction and bolsters brand loyalty. To do this, carriers not only have to manage the first and last transport events, but also the intermediate activities that occur at terminals and distribution hubs. That is, to provide the high-quality service, all the intermediate activities must also be punctual, like the handling of packages at intermediate hubs. OTWs must pertain to these events as well; and if these intermediate events occur within these on-time windows, then the first and last events are likely to occur on-time as well.

Carrier options at the strategic, tactical, and operational levels were examined. At all three levels, there are actions the carrier can take to improve reliability and enhance service quality.

For example, at the strategic level, carriers can select hub locations and determine their truck fleet sizes. In this study, the hub locations are assumed fixed, but the fleet sizes can vary. In an earlier project, see List *et al.* (2017), the effects of hub location choice were examined. In the case of fleet size, larger fleets provide more flexibility, increase the likelihood that on-time windows will be met, and allow door-to-door times to be shorter. More packages can be handled in less time. Delays are reduced.

At the tactical level, carriers can determine how long the planning horizon is and how many hours to work each day. They can decide the distribution of trucks among hubs, and they can choose the service requests to accept. In this study, the planning horizon is assumed to be a week, which is consistent with the load cycle faced by many carriers. The distribution of trucks is treated as an exogenous input whose impact is explored through sensitivity analysis. The impacts of variations in demand is addressed by randomly choosing the service requests that are extant in each of the problem realizations.

At the operational level, carriers can select the rules by which the system operates and determine how packages are assigned to trucks. In this study, the over-the-road trucks are treated as a free running fleet that is constantly in operation. The trucks start their tours each week from a home depot. Then they migrate from one load to the next, departing either when they are full or when a maximum headway between departures has been reached. The local trucks do both pick-ups and deliveries during a given tour. They make one tour per day. The assignment of packages to trucks in both cases is done using block-building procedures that are like the ones described in prior project reports, see List *et al.* (2018).

Three case study settings were addressed. They reflect the variety of operating conditions that trucking companies can face. The first is a back-and-forth operation between two locations. It represents situations where a dedicated fleet of trucks is involved in a single operation. One example is trucks moving parts from a component manufacturing plant to a main assembly plant. The second setting is a truckload (TL) operation where full-truck shipments are moved from one location to another. One example is the movement of chemicals from manufacturing plants to customers. Another is the distribution of merchandise from warehouses to retail stores. The third setting is a less-than-truckload (LTL) operation where packages are picked up, carried from hub to hub and then delivered. UPS and the postal service are examples of this type of service.

In all these settings, one study objective is to see how the truck fleet size effects on-time performance. For example, in the first setting, where the trucks carry truckload shipments from “A” to “B” and vice versa, the travel times vary and so do the load-unload times. For the extreme cases where the travel times and the load-unload times are both large, the fleet size limits the carrier’s ability to have all the shipments be on time. One question is: what fleet size is required to ensure that the quality of service (on-time performance) is at or above a target level (e.g., 95% of the deliveries are on-time). Another is, more generally, how does the fleet size affect the on-time performance? Put differently, what is the fleet size beyond which further quality of service improvements are difficult to achieve? In the second setting, the focus is again on seeing how the fleet size affects the service quality. But here, the service times, travel times, and fleet size are all user inputs. On-time performance suffers if the travel times are too long, the service times are too lengthy, or the fleet size is too small. Through parametric variation, it is possible to determine the sensitivity of the on-time performance to the values of these inputs. If the fleet size is too small, shipments are delayed. Increasing the fleet size reduces those delays. At some point, since the problem is deterministic, the delays go to zero because the fleet size is adequate. At that point, there is slack in the schedule. Trucks arrive early for loads. Measuring that slack gives a sense of reserve capacity that exists in the system to cope with unexpected delays and abnormally long load and unload times. In the third setting, packages are picked up and delivered by local trucks and transported between hubs by other, larger trucks. The load and unload times vary as do the travel times. The packages have a probability of being present in any given problem realization. A major question, again, is how the fleet size affects service quality.

The study shows that tools can be developed to schedule truck operations in uncertain environments. Those tools show how resource levels affect performance; they illustrate the fact that in stochastic settings, no single operating plan is always optimal. An ability to vary the truck dispatching plan is critical to achieving the best performance possible. Put another way, a nimble tool is valuable in achieving the best possible system performance when conditions are stochastic and resources are limited. Moreover, increasing the resource levels improves performance and reduces the sensitivity of system performance to the idiosyncrasies of specific scenarios.

6.2 FUTURE WORK

Much future work can be carried out based on the analyses conducted so far. Some important examples of these efforts are as follows:

Real-World Tests. As is often the case, the methodological advances presented here have been tested using a blend of empirical data and hypothetical situations. One natural extension for future work is to test these methods based on datasets that are more representative and reflective of real-world conditions. This pertains to all methods presented, from the assessment of reliability for segments and routes to the selection of plans for truck routing and locations for distribution centers.

Additional Sensitivity Analyses. As is often the situation, the case study analyses focus on some aspects of the problem and not others. In this instance, there are several aspects of the problem that were not explored. In the instance of Section 3, with the dedicated service, the impact of variations in the slack time was only addressed briefly. This could be studied more intensely. Also, the probability distribution functions used to create the randomness could be changed. In the results presented here, those density functions are all uniform with an upper and lower bound. Exploring density functions that are skewed toward lower values and have a significant tail (more typical of real world conditions) would be useful. In Section 4 it would be useful to introduce stochasticity in a more significant manner, perhaps following the analysis paradigm in Sections 3 and 5 where model parameter values can vary and be sampled for multiple problem realizations. This would diminish the emphasis on the actual operating plan employed, because it would vary from one realization to another. But, it would make it possible to see the impacts of increasing and decreasing the truck fleet size. It would also be interesting to introduce buffer times into that model, and vary those values, in combination with either the deterministic or stochastic analyses, to see how changes in those values affect system performance. In Section 5 it would be interesting to randomly sample the package requests, instead of always including all of them, to see whether patterns in the demands influence the result and whether that element of uncertainty results in more conservative (larger) fleet size suggestions. Another option, as with the other case studies, is to include explicitly a consideration of buffer times. This is done, presently, because of the cutoff criteria for scheduling packages for delivery, and by the hours of operation that are assumed. While these restrictions capture the reality of typical truck firm operations, they build significant buffers into the travel times, and mitigate the impacts of variability in the travel times.

REFERENCES

Allen, W. B., Mahmoud, M. M., & McNeil, D. (1985). The Importance of Time in Transit and Reliability of Transit Time for Shippers, Receivers, and Carriers. *Transportation Research Part B: Methodological, 19B*(5), pp. 445-447.

Armacost, A. P., Barnhart, C., & Ware, K. A. (2002). Composite variable formulations for express shipment service network design. *INFORMS*, *36*(1), pp. 1-20.

Armacost, A. P., Barnhart, C., Ware, K. A., & Wilson, A. M. (2004). UPS optimizes its air network. *INFORMS*, *34*(1), pp. 15-25.

Arvis, J.F., Raballand, G., & Marteau, J.F. (2007). The Cost of Being Landlocked: Logistics Costs and Supply Chain Reliability, pp. 81.

Barnhart, C., N. Boland, L. Clarke, E. Johnson, G. Nemhauser, and R. Senoi (1998). Flight String Models for Aircrafty Fleeting and Routing. *Transportation Science*, 32:3, pp. 208-220.

Barnhart, C., E. Johnson, G. Nemhauser, and P. Vance, 1999. Crew Scheduling. In Handbook of Transportation Science. R. W. Hall (editor). Kluwer Scientific Publishers, 493–521.

Barnhart, C., Belobaba, P., & Odoni, A. R. (2003). Applications of Operations Research in the Air Transport Industry. *Transportation Science*, *37*(4), pp. 368-391.

Bevilacqua, O. M. (1978). Energy conservation and intercity freight transport. *Transportation Planning and Technology*, *4*(4), pp. 227-240.

Bodin L., and R. Dial (1980). Hierarchical Procedures for Determining Vehicle and Crew Requirements for Mass Transit Systems. TRB, Transportation Research Record 746, pp. 58-64.

Buchan, A. B., List, G. F., Bongaardt, J. H. L., & Pipas, G. (1981). Expansion of the Use of Freight Car Scheduling: Development of a Strategic Plan for New England. pp. 113.

Chen, C.C., & Schonfeld, P. (2010). Modeling and Performance Assessment of Intermodal Transfers at Cargo Terminals. *Transportation Research Record: Journal of the Transportation Research Board*, *2162*(-1), pp. 53-62. doi: 10.3141/2162-07.

Chen, C.C., & Schonfeld, P. (2011). Alleviating Schedule Disruptions at Intermodal Freight Transfer Terminals: Real-Time Dispatching Control. *Transportation Research Record: Journal of the Transportation Research Board, 2238*, pp. 32-43.

Chen, C. F. (2010). Intermodal Transfer Coordination in Logistic Networks. *PhD Thesis, University of Maryland*.

Crainic, T. G., & Rousseau, J-M. (1986). Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem. *Transportation Research Part B: Methodological, 20*(3), pp. 225-242.

Czuch, C., & McDaniel, Z. (2011). *Measuring Goods Movement Travel Time and Reliability: Proposed Methodology*.

Daganzo, C. F. (1986). Static blocking at railyards: Sorting implications and track requirements. *Transportation Science, 20*(3), pp. 189-199.

Daganzo, C. F., Dowling, R. G., & Hall, R. W. (1983). Railroad classification yard throughput: The case of multistage triangular sorting. *Transportation Research Part A: General*, *17*(2), pp. 95-106.

Detmold, P. J. (1972). Some Thoughts on the Reliability of Railroading, *13*.

Dowling, R., List, G., Yang, B., Witzke, E., & Flannery, A. (2014). *NCHRP Report 31: Incorporating Truck Analysis into the Highway Capacity Manual*. Transportation Research Board, the National Academies, Washington, DC.

Figliozzi, M. A., Wheeler, N., Albright, E., Walker, L., Sarkar, S., & Rice, D. (2011). Algorithms for Studying the Impact of Travel Time Reliability Along Multisegment Trucking Freight Corridors. *Transportation Research Record: Journal of the Transportation Research Board, 2224*(1), pp. 26-34. doi: 10.3141/2224-04.

Hansson, S. O. (1994). Decision theory. *A Brief Introduction. Department of Philosophy and the History of Tecnology, Royal Institute of Technology, Stockholm*.

Hayya, J. C., Ramasesh, R. V., Tyworth, J. G., Kim, J. G., & Sun, D. (2013). JIT delivery with stochastic lead time. *Journal of the Operational Research Society, 64*(1), pp. 97-105. doi: 10.1057/jors.2012.27.

Hnaien, F., Delorme, X., & Dolgui, A. (2008). Supply planning for two-level assembly systems with stochastic component delivery times: trade-off between holding cost and service level. pp. 1-7.

Hsu, C.-I., & Wang, C.-C. (2013). Reliability analysis of network design for a hub-and-spoke air cargo network. *International Journal of Logistics Research and Applications, 16*(4), pp. 257-276.

Jane, C. C. (2011). Performance Evaluation of Logistics Systems Under Cost and Reliability Considerations. *Transportation Research Part E: Logistics and Transportation Review, 47*(2), pp. 130-137.

Jones, C., & Sedor, J. (2006). Improving the reliability of freight travel. *Public roads, 70*(1).

Kim, C., & Simchi-Levi, D. (2012). Optimal Expediting Policies for a Serial Inventory System with Stochastic Lead Time and Visibility of Orders In Transit. pp. 1-50.

Kraay, D., Harker, P. T., & Chen, B. (1991). Optimal pacing of trains in freight railroads: model formulation and solution. *Operations Research, 39*(1), pp. 82-99.

Kraft, E. R. (2002). Priority-Based Classification for Improving Connection Reliability in Railroad Yard. Part I of II: Integration with Car Scheduling. *Journal of the Transportation Research Forum, 56*(1), pp. 93-105.

Kraft, E. R. (2002). Scheduling railway freight delivery appointments using a bid price approach. *Transportation Research Part A: Policy and Practice, 36*(2), pp. 145-165.

Lai, C. M. (2011). *Pickup and Delivery Problem with Stochastic Travel Times for Semiconductor Supply Chains*.

Lang, A. S., & Reid, R. M. (1970). Studies in Railroad Operations and Economics. *Volume 1. Railroad Car Movement Reliability: A Preliminary Study of Line-Haul Operations*, pp. 33.

Lee, E. S., & Song, D. W. (2011). Knowledge management systems in maritime logistics. pp. 50-67.

Leemis, L. (2009). Reliability: Probabilistic Models and Statistical Methods, On-Line Books, available at https://prkwrhnfx07.storage.googleapis.com/MDEzNzIwNTE3MQ==07.pdf.

Li, Q., Nie, Y., Vallamsundar, S., Lin, J., & Homem-de-Mello, T. (2013). *Incorporating Environmental Measures into a Reliable Freight Routing Model*.

Li, Q., & Savachkin, A. (2013). A heuristic approach to the design of fortified distribution networks. *Transportation Research Part E: Logistics and Transportation Review, 50*, pp. 138-148.

List F, G., & Bongaardt, H. L. J. (1981). Freight Car Scheduling Benefit Project on CONRAIL: The Pillsburge to St. Louis Corridor, pp. 107.

List, G. F., Wood, B., Turnquist, M. A., Nozick, L. K., Jones, D. A., & Lawton, C. R. (2006). Logistics planning under uncertainty for disposition of radioactive wastes. *Computers & operations research, 33*(3), pp. 701-723.

List, G., E. Williams, J. Addison, and A. Morsali, 2017. Improving the Efficiency and Reliability of Freight Transportation, prepared for the National Transportation Center at Maryland, Project NTC2014-SU-R-06.

List, G., E. Byrom, J. Addison, and A. Morsali, 2018. Improving the Reliability of Freight Transportation: Final Report, prepared for the National Transportation Center at Maryland, Project NTC2015-MU-R-10.

Lomax, T., Turner, S., Hallenbeck, M., Boon, C., & Margiotta, R. (2001). *Traffic Congestion and Travel Reliability: How bad is the situation and what is being done about it?*

Louly, M. A., Dolgui, A., & Hnaien, F. (2008). Supply planning for single-level assembly system with stochastic component delivery times and service-level constraint. *International Journal of Production Economics, 115*(1), pp. 236-247. doi: 10.1016/j.ijpe.2008.06.005.

Ma, X., McCormack, E. D., & Wang, Y. (2011). Processing commercial global positioning system data to develop a web-based truck performance measures program. *Transportation Research Record, 2246*, pp. 92-100. doi: 10.3141/2246-12.

Markovic, N., & Schonfeld, P. (2011). Scheduling Under Uncertainty for Single-Hub Intermodal Freight System. *Transportation Research Record: Journal of the Transportation Research Board, 2238*, pp. 24-31.

Martland, C. D. (1992). Rail Freight Service Productivity from the Manager's Perspective. *Transportation Research Part A: Policy and Practice, 26*(6), pp. 457-469.

Martland, C. D., & Sussman, J. M. (1972). Rail Service Time Reliability, *An Analysis of Operating Data, 13*.

McCormack, E. D., Ma, X., Klocow, C., Currarei, A., & Wright, D. (2010). Developing a GPS-based truck freight performance measures platform, pp. 61.

McKinnon, A., Palmer, A., Edwards, J., & Piecyk, M. (2008). Reliability of Road Transport from the Perspective of Logistics Managers and Freight Operators.

Muthuraman, K. (1991). Inventory Management with Stochastic Lead Times. *Mathematics of Operations Research* 40(2), pp. 302-327.

Ogawa, M. J., Poole, A. J., Meekums, R. J., Cooke, D., & Self, S. (2012). *Managed Motorways: Keeping Business Moving*.

Ouyang, L. Y., & Chang, H. C. (2001). The variable lead time stochastic inventory model with a fuzzy backorder rate. *44*(1).

Schrank, D., Eisele, B., & Lomax, T. (2012). TTI's 2012 urban mobility report. *Texas A&M Transportation Institute. The Texas A&M University System*.

Shamberger, R. C. (1975, 1975). *Freight Car Management Systems: A New Way Out of the Car Utilization Dilemma.*

Sierleja, E. J., Pipas, G., & List, G. F. (1981). Evaluation of MOPAC's Freight Car Scheduling System, pp. 72.

Sines, G. (1972). MOPAC'S Transportation Control System. A Systems Approach to Achieving Service Reliability, *13*.

Taniguchi, E., Thompson, R. G., Yamada, T., & van Duin, R. (2001). City Logistics. *Network Modelling and Intelligent Transport Systems*, pp. 252.

Taniguchi, E., Yamada, T., & Tamagawa, D. (2001). Reliability of Transport Networks. Chapter 6: Probabilistic Routing and Scheduling of Urban Pickup/Delivery Trucks with Variable Travel Times. *Publication of: RESEARCH STUDIES PRESS LTD*, pp. 73-89.

Taniguchi, E., Yamada, T., & Tamaishi, M. (2001). *ITS Based Dynamic Vehicle Routing and Scheduling.*

Turnquist, M. A., & Daskin, M. S. (1982). Queuing models of classification and connection delay in railyards. *Transportation Science*, *16*(2), pp. 207-230.
Van Dyke, C. (1986). The automated blocking model: A practical approach to freight railroad blocking plan development. *Paper presented at the Proceedings of the Transportation Research Forum.*
Vernimmen, B., Dullaert, W., & Engelen, S. (2007). Schedule Unreliability in Liner Shipping: Origins and Consequences for the Hinterland Supply Chain. *Maritime Economics & Logistics*, *9*(3), pp. 193-213.
Washburn, S. S., & Ko, B. (2007). Travel Time Reliability and Truck Level of Service on the Strategic Intermodal System Part B: Truck Level of Service, pp. 303.
Whybark, D. C. (1974). The Combined Transportation and Inventory Policy Decision.
Williams, J. H. (1972). Some Techniques for Controlling Rail Service Reliability, *13*.
Xu, J., Hancock, K. L., & Southworth, F. (2008). Dynamic Freight Traffic Simulation Providing Real-Time Infimation, pp. 1711-1719.
Yagar, S., Saccomanno, F., & Shi, Q. (1983). An efficient sequencing model for humping in a rail yard. *Transportation Research Part A: General*, *17*(4), pp. 251-262.
Yang, C. H., & Regan, A. C. (2012). A multi-criteria decision support methodology for implementing truck operation strategies. *Transportation*, *40*(3), pp. 713-728. doi: 10.1007/s11116-012-9432-7.
Zhao, Y., & Simchi-Levi, D. (2006). Performance Analysis and Evaluation of Assemble-to-Order Systems with Stochastic Sequential Lead Times. *Operations Research*, *54*(4), pp. 706-724. doi: 10.1287/opre.1060.0302
~~for Solving Hazardous Materials Logistical Problems. *Transportation Research Record*, *1783*, pp. 158-166.~~