Design of Multistar Many-to-Many Distribution Networks

Anton J. Kleywegt, Jinpyo Lee, and Amy R. Ward

School of Industrial and Systems Engineering

Georgia Institute of Technology

December 31, 2006

Abstract

We consider the problem of designing a distribution network to facilitate the repeated movement of shipments from many origins to many destinations. A sufficient number of the origin-destination shipments require less than the capacity of a vehicle, so that consolidation of shipments is economical. We consider the case in which consolidation takes place at terminals, and we assume each shipment moves through exactly one terminal on its way from its origin to its destination. Then, a major design decision is to determine the best number of terminals. We develop a continuous approximation method to estimate transportation costs as a function of the number of terminals. We use the continuous approximation method to choose the number of terminals that minimizes the sum of terminal cost and transportation cost. Numerical results indicate that the design resulting from the continuous approximation method facilitates operations with lower cost than those resulting from a widely used integer programming based design.

1 Introduction

In most distribution systems goods are transported from various origins to various destinations. For example, many retail chains manage distribution systems in which goods are transported from a number of suppliers to a number of retail stores. Much of this transportation takes place by truck, and it often happens that the flow rates for a substantial fraction of the origin-destination pairs are so small that it is not economical to send goods directly from each origin to each destination in a dedicated truck. That is, it is often more economical to consolidate the shipments of various origin-destination pairs, and transport such consolidated shipments in the same truck at the same time. There are many ways in which such consolidation can be accomplished. Next we give a number of examples.

Figure 1 shows the locations of origins (circles) and destinations (squares) for a small example. Suppose that freight has to be moved from each origin to each destination in the example every week. To keep the example simple, we do not specify the volume of freight that has to be moved from each origin to each destination every week, but rather suppose that each week a truck can accommodate the total freight flowing from up to four origins to all destinations, and the total freight flowing from all origins to up to three destinations. The question at hand is how to design the transportation operations to move all freight each week at minimum cost.

There are many alternative designs, and we give only a few examples that are qualitatively different from each other. One alternative is to use out-of-vehicle consolidation, that is, to consolidate freight at terminal facilities, also called consolidation terminals, transshipment terminals, crossdocks, or transfer facilities. There are many alternative ways to design transportation operations involving terminal facilities. Here we present two simple, but important, ways, distinguished based on the number of terminals that each shipment passes through on its way from its origin to its destination. First, even if there are more than one terminal in the system, each shipment may move through only one terminal on its way from its origin to its destination. This type of system is quite widely used in the USA to distribute goods from multiple suppliers to multiple distribution centers or retail stores. In such a system it is typical for each vehicle to be based at a terminal, and to execute routes that take freight from the terminal to particular destinations, and after the freight has been delivered, to visit particular origins and collect freight at these origins and take it back to the terminal, for organization according to delivery routes and later delivery. This type of distribution network is called a *multistar many-to-many distribution network*, and is the type of system for which we propose a design approach in this paper. We
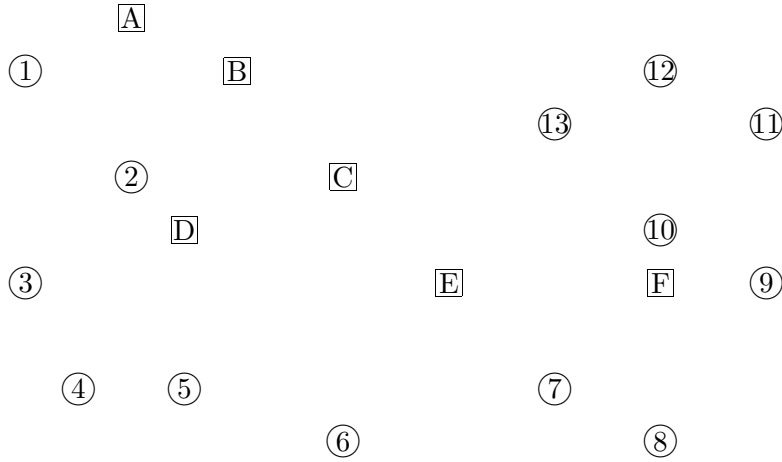
Figure 1: Origins (circles) and destinations (squares) for freight flows.

will point out some of the advantages and disadvantages of such a system soon, but first we give an example. Figure 2 shows a solution using out-of-vehicle consolidation with two terminals denoted by X and Y, in which each shipment moves through only one terminal on its way from its origin to its destination. The solution in Figure 2 consists of the following routes: $X, D, B, A, 1, 2, 3, X$; $X, C, E, F, 6, 5, 4, X$; $Y, C, B, A, 13, 12, 11, 10, Y$; and $Y, D, E, F, 9, 8, 7, Y$. For example, consider a shipment that has to go from origin 1 to destination F. One week the shipment is picked up at origin 1 by the vehicle that executes route $X, C, B, A, 1, 2, 3, X$, then at terminal X the shipment is offloaded from the vehicle and loaded onto the vehicle that will execute the route $X, D, F, E, 6, 5, 4, X$ next, and later the shipment is delivered at destination F while the vehicle executes this route.
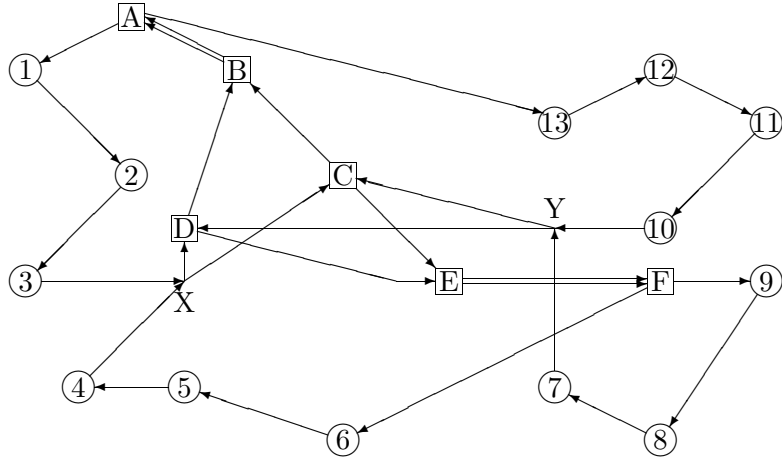

Figure 2: Solution in which each shipment moves through only one terminal on its way from its origin to its destination.

Note that with such a system sufficient routes have to be chosen to ensure that all freight move from its origins to its destinations, without exceeding constraints such as vehicle capacity constraints. In Figure 2, freight can move from every origin to every destination, because each destination receives truck visits from both terminals (but each origin receives truck visits from only one terminal). In practice, it seems quite common to ensure that freight can move from every origin to every destination either by serving each origin from each terminal, or by serving each destination from each terminal. However, such a solution may lead to unnecessarily many visits

during which very little freight is picked up or delivered, and unnecessarily long distances traveled by vehicles. In this paper we search for more economical ways to facilitate all freight flows.

An alternative system that uses out-of-vehicle consolidation is the following. Each origin and each destination is served from one terminal only, typically the terminal closest to the point (origin or destination). To enable freight to move from every origin to every destination, vehicles also move freight between each pair of terminals. The vehicles that move freight between the terminals are often of a different type (usually larger) than the vehicles that pick up and deliver freight at origins and destinations. Such a distribution network is called a *complete topology many-to-many distribution network*. In such a system, each shipment moves through either one or two terminals on its way from its origin to its destination. Figure 3 shows a solution using out-of-vehicle consolidation with two terminals denoted by X and Y, in which each origin and each destination is served from one terminal only. For example, consider a shipment that has to go from origin 1 to destination F. One week the shipment is picked up at origin 1 by the vehicle that executes route $X, D, B, A, 1, 2, 3, X$, then at terminal X the shipment is offloaded from the vehicle and loaded onto the vehicle that goes from terminal X to terminal Y, then at terminal Y the shipment is offloaded from the vehicle and loaded onto the vehicle that will execute the route $Y, F, 9, 8, 7, Y$ next, and later the shipment is delivered at destination F while the vehicle executes this route.
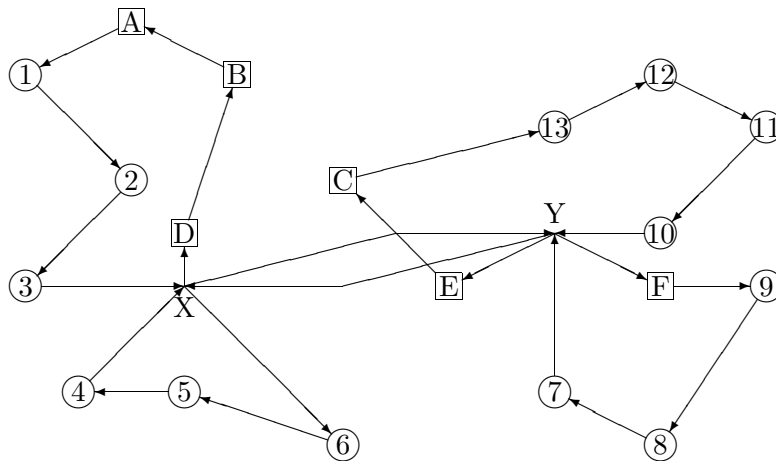

Figure 3: Solution in which each shipment moves through only one or two terminals on its way from its origin to its destination.

There are various other systems in which each shipment moves through one or more terminals on its way from its origin to its destination. Some such systems, such as complete topology systems in which an origin or a destination may be served by more than one nearby terminals, are reviewed in Section 2. Another related design is a *star topology many-to-many distribution network* with a central terminal through which all freight flows. Specifically, with a star topology, each shipment travels from its origin through the pickup and delivery terminal serving the origin to the central terminal, from there to the pickup and delivery terminal serving the destination, and from there to the destination. Thus, in such a system, each shipment travels through three terminals on its way from its origin to its destination. Some practical systems, such as distribution systems for small packages, can be much more complicated than the basic system types described above, involving a hierarchy of terminals and many vehicle types traveling between terminals, with different shipments moving through different numbers of terminals on their way from their origins to their destinations.

Complete topology, star topology, and hierarchical systems may facilitate solutions with less total travel distance than a multistar system. The transportation cost with such systems may also be less if the cost per unit freight and distance is less for the vehicles that travel between the terminals than for the pickup and delivery vehicles. However, when a shipment moves through more than one terminal on its way from its origin to its destination it requires more loading, offloading, and additional load sorting operations, and such additional

3

handling incurs not only more cost, but also increases the risk that shipments are lost or damaged. Which of the types of systems is best depends on the importance of handling related costs relative to transportation costs.

Finally, an alternative that does not require any terminals is to use in-vehicle consolidation. For example, one vehicle may pick up the loads at origins 1, 2, and 3, in that order, that have to go to destinations A, B, and D, and then drive to destinations D, B, and A, in that order, and deliver these loads. From destination A, the vehicle would next return to origin 1, ready to repeat the cycle $1, 2, 3, D, B, A, 1$. Other vehicles would perform similar cycles. Sufficient cycles have to be chosen to ensure that all freight move from its origins to its destinations, without exceeding constraints such as vehicle capacity constraints. For example, if in addition to cycle $1, 2, 3, D, B, A, 1$, cycle $1, 2, 3, E, F, C, 1$ is performed, then all the freight originating from origins 1, 2, and 3 could be moved to its various destinations. In-vehicle consolidation is a reasonable alternative if for many of the origin-destination pairs, the amount of freight is not much less than the vehicle capacity. Figure 4 shows a solution using in-vehicle consolidation with the following cycles: $1, 2, 3, D, B, A, 1$; $1, 2, 3, E, F, C, 1$; $4, 5, 6, D, B, A, 4$; $4, 5, 6, E, F, C, 4$; $7, 8, 9, 10, A, B, D, 7$; $7, 8, 9, 10, F, C, E, 7$; $13, 11, 12, A, B, D, 13$; and $11, 12, 13, C, E, F, 11$.
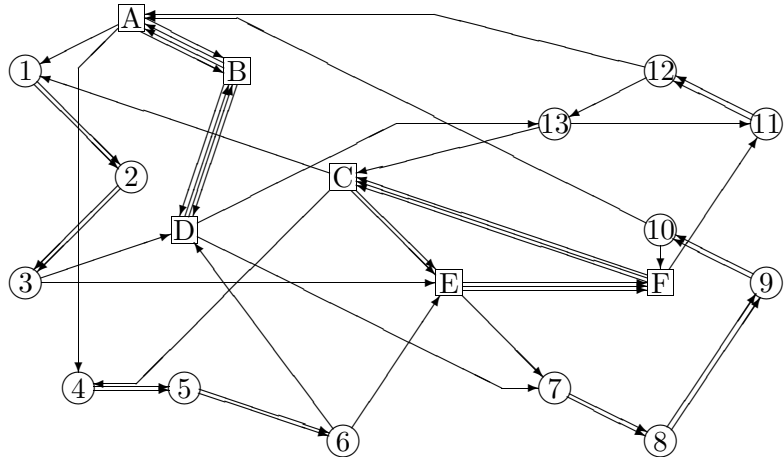

Figure 4: Solution using in-vehicle consolidation.

In this paper we focus on multistar many-to-many distribution systems in which each shipment moves through one terminal on its way from its origin to its destination. These systems are sufficiently widely used to justify an in-depth study focused on such systems. In addition, sometimes there is interest in determining which basic type of system is the best for a particular case, and clearly it is desirable to be able to estimate the cost of the optimal design for each type of system in order to find the best overall system. Our major motivation is to develop a method to design such distribution systems. To facilitate such design, we develop a *continuous approximation* (CA) method to estimate the cost of the operations that will be conducted with a given design, and then we use these estimates to search for a good design. Chosen designs are then evaluated with more detailed calculations of the resulting operations.

As mentioned, often a vehicle makes all the deliveries at the destinations on its route before it picks up any new loads at the origins on its route. Such a practice may increase the travel distance over what would have been possible if pickups and deliveries could have been done in any sequence. However, often this is infeasible or undesirable, sometimes because of scheduling constraints (for example, loads are delivered at the destinations in the morning and picked up from origins in the afternoon), or because loads that are picked up before all deliveries have been done would obstruct access to the loads in the back of the truck that still have to be delivered. We focus on the case in which all deliveries on a route take place before any pickups take place. Some aspects of our approach can also be used if pickups and deliveries can be done in any sequence on a route.

This work was motivated by our collaboration with two companies that manage their own inbound distribution systems. One is a large retailer that obtains merchandize from many suppliers in the USA as well as overseas. For the purpose of the distribution system in the USA, the ports of import (more precisely, warehouses close to the

ports) are regarded as the origins of the imported freight. Thus the origins are suppliers and import warehouses in the USA. The destinations are individual retail stores. The retailer owns and operates a number of terminals. Each shipment moves through one terminal on its way from its origin to its destination with transportation provided by independent contract carriers. In the current system, each origin is served by all the terminals, and each destination is served by the terminal closest to the destination. The greater the travel distances of the carriers' vehicles, the more the retailer has to pay the carriers, and thus the retailer has an incentive to design a distribution system that minimizes the total distance traveled, also taking into account the cost of owning and operating a terminal, and handling related costs. The retailer was considering increasing the number of terminals, and asked for design guidelines. The other company is a distributor of a large variety of products, most of which are sold to retailers, and some to industrial consumers. The origins are the distributor's suppliers, most of whom are located in the USA and Canada, and the destinations are the distributor's distribution centers located close to various cities in the USA. The distributor owns and operates a number of terminals, as well as a fleet of trucks. As for the retailer, each shipment moves through one terminal on its way from its origin to its destination. Unlike the retailer, in the current system, most origins are served by one terminal, and each destination is served by all the terminals. Products are further distributed from the distributor's distribution centers to the distributor's customers with a separate fleet of smaller vehicles.

The first challenge in designing such a system is to determine how many terminals there should be. Transportation cost is decreasing in the number of terminals, but the cost of owning and operating terminals is increasing in the number of terminals. Since we focus on systems in which each shipment moves through one terminal on its way from its origin to its destination, handling costs do not vary with the number of terminals. The second challenge in designing such a system is to determine the best locations for the chosen number of terminals. We argue that one can accurately determine the best number of terminals without exactly determining the best locations of the terminals. We also argue that to determine the best number of terminals it is important to accurately estimate how operational costs, in this case transportation costs, vary as a function of the number of terminals. As the examples above already indicate, important operational variables are the numbers of terminals that serve each origin and each destination. It will be shown that these variables have an important impact on the transportation costs, and thus should be taken into account when choosing the number of terminals. The major part of the work is the development of a new CA approach for the estimation of transportation cost, and especially linehaul cost (the average distance from a terminal to the center of the points on a route), as a function of the number of terminals as well as the numbers of terminals that serve each origin and each destination, before knowing the exact terminal locations. Then we use these estimates to search for the best number of terminals as well as the numbers of terminals that serve each origin and each destination. Thereafter the terminals are located, and the resulting design is evaluated through detailed calculations of the vehicle routes to move all the freight from its origins to its destinations.

The remainder of this paper is organized as follows. We review related literature in Section 2, and we formulate the problem in Section 3. Section 4 provides a qualitative discussion of the factors influencing total cost, and identifies the factors that do and do not seem important in the selection of the best number of terminals. Section 5 describes our CA method for the design of multistar many-to-many distribution networks. We also provide a procedure for making operational decisions (which terminals to use for each origin-destination flow and how to route the vehicles from each terminal) in Section 6 that we use in Section 7 to evaluate our solutions compared with those produced by a widely used approach.

2   Related Literature

Optimization problems that incorporate both facility location and vehicle routing decisions, such as the problem in this paper, are called *location-routing problems.* Laporte (1988) proposed a classification of location-routing problems, and gave an overview of both exact branch-and-bound algorithms and heuristics for a number of specific location-routing problems. However, the paper considered only the case of a single commodity — all shipments were considered as the same product, and thus a shipment picked up from an origin did not have a

specified destination. In the problem we consider, each shipment has a specified origin-destination pair. Also, the paper did not consider the possibility of deliveries and pickups on the same route. The method that we propose is applicable both for the case with deliveries and pickups on separate routes, as well as for the case with deliveries and pickups on the same routes. The type of approach that we consider, namely a two-phase approach in which location decisions are made in the first phase and routing decisions are made repeatedly, possibly with varying data, in the second phase, and future routing costs are approximated when location decisions are made, was mentioned on p. 192 as a promising research area. Laporte et al. (1989) formulated and solved integer programming models of stochastic location-routing problems. The problems involve decisions regarding the location of a single depot among a set of candidate sites, the vehicle fleet size, and the pickup routes. All these decisions are made before the pickup quantities become known. The models constrain either the probability of a route failure or the expected penalty of a route failure due to insufficient capacity to handle the pickups on a route. Vidal and Goetschalckx (1997) surveyed strategic production-distribution problems, with special emphasis on features that are important for international supply chains. Some of the problems in the survey involve facility location and distribution, and an overview is given of some mixed integer programming formulations and tools for solving these problems.

The approach that we consider can be regarded as a *continuous approximation* (CA) approach. In the transportation literature, the term "continuous approximation" refers to an approach in which some problem data, typically discrete data such as the locations of origins, destinations, or facilities, are approximated with distributions, typically continuous distributions such as the uniform distribution. Kantorovitch (1942), Beckmann (1952), and Beardwood et al. (1959) did some path-breaking work on CA in transportation. Newell (1973) described the use of CA methods to provide insight into the qualitative behavior of various discrete optimization problems and sometimes to find good solutions for such problems. Various researchers developed these ideas further. Here we only give a brief overview of this line of research, and we compare our work with some of the more closely related work. Langevin et al. (1996) surveyed and classified the literature on CA methods in freight distribution.

Daganzo (1984b) proposed a method to construct good tours in rectangles and presented formulas, similar to that of Beardwood et al. (1959), to approximate the lengths of the tours. Daganzo (1984a) proposed a cluster-first route-second method to construct vehicle routes and presented formulas to approximate the lengths of the routes. The effect of the proportions of a rectangular district for each cluster on the sum of the linehaul and detour travel distances for the cluster was studied. The results in Daganzo (1984b) were used to approximate the detour travel distances. We also follow the approach of separately approximating the linehaul and detour travel distances.

The approximations in Daganzo (1984a) were used to study various *one-to-many* distribution problems, that is, problems with one origin and many destinations. (Note that a result for a one-to-many problem gives a result for a many-to-one problem that is symmetric to it and vice versa). For example, Burns et al. (1985) developed formulas to approximate the transportation and inventory costs for one-to-many direct shipping and peddling (routes with multiple stops) distribution strategies. Daganzo (1985) derived expressions for transportation and inventory costs, argued that vehicles should be fully loaded (assuming that inventory costs are not very high), and then derived an expression for the optimal frequency with which to pick up loads for different classes of items. Daganzo (1988a) showed that if products going to the different destinations are homogeneous, the same vehicles operate everywhere, and there are no route length restrictions, then in-vehicle consolidation is cheaper than out-of-vehicle consolidation, that is, transshipments are not economical. Daganzo (1988b) considered a setting with more general vehicle capacity constraints than Daganzo (1988a). Also, different items have different characteristics, so that optimal vehicle utilization (minimizing total vehicle miles traveled) may require careful selection of the mix of items to place on different vehicles. In such a setting, in contrast with Daganzo (1988a), it may be optimal to consolidate some items at a terminal into more efficient load combinations. Campbell (1990a) considered a similar setting as Daganzo (1988a), except that Campbell (1990a) considered the case in which there are two vehicle sizes, and the larger vehicles cannot be used for delivering at the destinations. Distribution networks with transshipment terminals, in which larger vehicles transport freight from the origin directly to

terminals, and smaller vehicles transport freight from terminals on routes to destinations, were compared with distribution networks without transshipment terminals, in which smaller vehicles transport freight from the origin on routes to destinations. Campbell (1993b) considered the same setting as Campbell (1990a), except that in Campbell (1990a), each larger vehicle can visit only one terminal on a trip (that is, larger vehicles do not make multiple stop routes), whereas in Campbell (1993b) larger vehicles may visit multiple terminals on a route. Unlike many of the other papers, and similar to ours, Daganzo and Newell (1986) considered a problem that involves both design and operational decisions. Specifically, the design of a hierarchical distribution network with one origin and many destinations was studied. Also unlike many of the other papers, the case in which vehicles with different transportation costs per distance operate at different levels in the hierarchy was considered. As a result, it may be optimal for a shipment to undergo multiple transshipments from the origin to its destination.

Continuous approximations have also been used to study *many-to-many* distribution problems, that is, problems with many origins and many destinations such as the problem that we consider. Daganzo (1978) considered many-to-many demand responsive transportation systems, such as taxicab systems with at most one request being transported at a time and dial-a-bus systems that allow multiple requests to be transported at a time. Three routing algorithms were modeled, and approximate expressions were derived for the average waiting times and average riding times of customers. Blumenfeld et al. (1985) derived cost expressions for the following cases: (1) Direct shipping from one origin to one destination. (2) Direct shipping from many origins to one destination. (3) Direct shipping from one origin to many destinations. (4) Direct shipping from many origins to many destinations. (5) Shipping from many origins to many destinations with all loads moving through a single consolidation terminal. Vehicle routing is not considered — shipments move directly from each origin to the terminal, and directly from the terminal to each destination. (6) Shipping from many origins to many destinations with some loads moving directly from origin to destination and other loads moving through a single consolidation terminal. Vehicle routing is not considered. Our work differs from Blumenfeld et al. (1985) as follows:

1. We consider the problem of shipping from many origins to many destinations only, and not the other (easier) cases.

2. In our problem, the number and locations of terminals are to be determined, but in cases (5) and (6) above, it is given that there is a single terminal with a given location.

3. We consider transportation costs and terminal costs, but not production setup costs and inventory costs. Instead, as is the case with all the applications that we have worked on, it is given that the transportation schedule repeats periodically (daily or weekly), and thus the inventory costs are not affected by the distribution decisions.

4. We take vehicle capacity constraints into account.

5. We consider vehicle routing — we allow a vehicle to visit multiple origins and/or multiple destinations on a route.

Blumenfeld et al. (1987) described a number of simple models that were used to evaluate strategies for the distribution of parts from suppliers to General Motors assembly plants. A variety of strategies involving direct shipping, shipping through a single terminal, as well as peddling, were evaluated for the shipments from a single supplier (Delco) with 3 plants to 30 GM assembly plants. The models considered the trade-off between transportation costs that favor fewer larger shipments and inventory costs that favor more frequent smaller shipments. A decomposition approach that exploited the small number of origins was proposed to find the best combination of direct shipping and shipping through the terminal.

Hall (1984), Daganzo (1987), Hall (1987), Campbell (1990b), and Campbell (1993a) are closely related to each other. These papers all considered a setting with an equal number of origins and destinations independently and uniformly distributed in a square (sometimes rectangular) region. The flow rate is the same for all origin-destination pairs. Terminals are arranged in a square (sometimes rectangular) grid. The following four distribution strategies were considered by several of these papers (as described in more detail below, some papers considered only some of these strategies, and some also considered other strategies):

7

1. One-terminal-nearest-terminal: Each shipment moves through exactly one terminal. All shipments from an origin move through the terminal closest to the origin irrespective of the location of the destination of the shipment.

2. One-terminal-minimum-distance: Each shipment moves through exactly one terminal. A shipment from an origin moves through one of the (typically four) terminals that surround the origin in the grid, that minimizes the travel distance from the origin through the terminal to the destination.

3. Two-terminal-nearest-terminal: Each shipment moves through one or two terminals. All shipments from an origin move through the terminal closest to the origin irrespective of the location of the destination of the shipment, and all shipments to a destination move through the terminal closest to the destination irrespective of the location of the origin of the shipment.

4. Two-terminal-minimum-distance: Each shipment moves through one or two terminals. A shipment from an origin moves through one of the (typically four) terminals that surround the origin in the grid, and a shipment to a destination moves through one of the (typically four) terminals that surround the destination in the grid, to minimize the travel distance from the origin through the terminals to the destination.

Hall (1984) derived expressions for the average distance from origin to destination for the one-terminal-nearest-terminal and the two-terminal-nearest-terminal strategies. The expressions were compared with average travel distances computed between the 37 largest standard metropolitan statistical areas in the United States. Daganzo (1987) considered the case of one-to-many distribution routes, without transhipment terminals, and the case of many-to-many distribution routes, without transhipment terminals, in addition to the four strategies mentioned above. Hall (1987) compared the four strategies in terms of (a) average travel distance, (b) number of terminals, and (c) number of links. For a given average travel distance, the number of terminals and number of links were regarded as measures of consolidation. Campbell (1990b) considered the case in which the transportation cost per unit distance between terminals may be less than the transportation cost per unit distance between origins and terminals and between terminals and destinations. The paper compared the two-terminal-nearest-terminal strategy, the two-terminal-minimum-distance strategy, and the two-terminal-minimum-cost strategy, that takes into account the difference between the local transportation cost and the transportation cost between terminals. The expressions for the minimum cost were used to derive the optimal spacing between the terminals in the grid for a given number of terminals. Unlike the other papers, the resulting service areas of the terminals were not equal. Campbell (1993a) evaluated the accuracy of the approximation formula for minimum average transportation cost derived in Campbell (1990b). The approximation formula was quite accurate, taking into account the extent with which the idealized assumptions of the approximation were violated. With two terminals, the errors were around 5%. However, as the number of terminals increased, the errors tended to increase as well. Our paper differs from these papers as follows:

1. These papers assume that origins and destinations are independently and uniformly distributed in a square or rectangular region, whereas we allow arbitrary distributions of origins and destinations in a rectangle.

2. These papers assumed that the number of origins equals the number of destinations, or equivalently that the density of origins equals the density of destinations, whereas we allow the numbers of origins and destinations to be different.

3. These papers assumed that the flow rate from origins to destinations is a deterministic constant for all origin-destination pairs, whereas we allow flow rates that are random with different marginal distributions for different origin-destination pairs.

4. These papers required terminals to be located on a square or rectangular grid in the region (and thus that the number of terminals be a square number or the product of the numbers of rows and columns in the grid), whereas we allow any number of terminals, but we assume that, except for a single centrally located terminal, the terminals are independently and uniformly distributed in the rectangular region.

5. We focus on the case in which each shipment moves through exactly one terminal on its way from its origin to its destination, whereas these papers consider various strategies in which each shipment moves through

two terminals on its way from its origin to its destination.

6. For the case in which each shipment moves through exactly one terminal, the papers that considered this case required that each shipment moves through the terminal closest to the origin (one-terminal-nearest-terminal routing), or each shipment moves through one of the (typically four) terminals closest to the origin (one-terminal-minimum-distance routing). That is, for one-terminal-nearest-terminal routing, each origin is served by vehicles from the closest terminal and each destination is served by vehicles from all terminals, and for one-terminal-minimum-distance routing, each origin is served by vehicles from the closest four terminals, and each destination is served by vehicles from all terminals. It is easily seen that it can be very inefficient to serve each destination by vehicles from all terminals. We allow each origin and each destination to be served by vehicles from a chosen subset of terminals in addition to the central terminal, where the number of terminals that serves an origin/destination depends on the total flow rate from/to the origin/destination.

As mentioned before, our approach allows pickups on the same routes as deliveries. Some CA work has addressed similar ideas. Hall (1991) considered a distribution problem with two terminals that serve as origins, and multiple destinations. The items originating at the two terminals are different, so that items may be sent from a terminal to destinations close to the other terminal. As a result, after delivering the items originating at a terminal to some destinations, it may be better for a vehicle to next proceed to the other terminal to pick up loads there. The paper develops approximate expressions for the linehaul and detour distances if a vehicle travels from one terminal to a delivery district, and thereafter travels to the other terminal. Daganzo and Hall (1993) derived expressions to approximate the cost of vehicle routes to do both pickups and deliveries, where the deliveries on a route are completed before any pickups are done. It is assumed that the pickup and delivery points are independent and uniformly distributed in a region that can be partitioned into approximate rectangles, that each route can make at most C delivery stops, and that there is no bound on the number of pickups that a route can make.

As mentioned above, very few papers address CA methods for network design, that is, to determine the number and locations of terminals. It was mentioned that Daganzo and Newell (1986) considered hierarchical network design for one-to-many distribution. In addition, Hall (1993) considered the design of a many-to-many freight distribution network in a local area, such as a metropolitan area. The area has one gateway terminal located in the center through which all shipments to and from locations outside the area pass. Unlike our problem, origins and destinations in the area are independently and uniformly distributed. Similar to our problem, pickup and delivery terminals are uniformly distributed, and the number of pickup and delivery terminals is a design variable. Each origin and each destination is served on vehicle routes from the pickup and delivery terminal in which service district it is located. Pickup and delivery routes are separate, and the vehicles used for pickup and delivery routes are different from the vehicles used to transport freight between terminals. As is the case in our problem, transportation and terminal costs were considered, and the headway between successive routes was given. The optimal number of pickup and delivery terminals and the optimal number of stops on a vehicle route were determined for two distribution systems, namely a star topology and a complete topology. The costs for the two systems were compared, and a number of conclusions were made, such as that the star topology is better if the interterminal vehicles are large relative to the pickup and delivery vehicles, and if handling cost is small. Both Dasci and Verter (2001) and Ouyang and Daganzo (2006) considered problems in which a number of facilities and their locations are to be selected, and the region is to be partitioned into service areas such that each facility supplies the destinations in one of the service areas. Dasci and Verter (2001) considered only outbound transportation costs from the facilities to the destinations, in addition to fixed facility costs and facility capacity costs. Ouyang and Daganzo (2006) considered both inbound transportation costs from a single origin to the facilities as well as outbound transportation costs from the facilities to the destinations; thus the problem considered by Ouyang and Daganzo (2006) corresponded to a special case of the one-to-many distribution network design problem considered by Daganzo and Newell (1986) in which each shipment moves through one terminal from the origin to its destination. Dasci and Verter (2001) developed a CA that required a service area size, or equivalently a terminal density, to be selected at each point in the region. Ouyang and

Daganzo (2006) proposed an algorithm to convert the service area size as a function of the point in the region to a solution with discrete terminal locations, and they also evaluated the differences between the CA costs and the costs resulting from their algorithm. In all the design problems described above, the input data include the spatial density of destination demand, and not many-to-many origin-destination flows as in our problem. Recently, Dasci and Laporte (2005) studied a Stackelberg game in which location decisions are made by two competitors. Each competitor can decide to locate many facilities, and the location decisions are represented with location density functions, instead of with discrete variables that specify the exact location of each facility.

3 Model Formulation

In this section we give a formulation of the problem we want to solve. The purpose is threefold. First, we want to give a precise statement of one version of the problem that we want to solve. Second, we want to determine what size instances can be solved with available software. Third, we want to introduce a common approximation to our problem that we will later compare with our CA method.

Distribution operations take place repeatedly over multiple time periods. We consider distribution systems in which (1) each shipment moves through one terminal on its way from its origin to its destination, and (2) in each time period, the vehicles based at each open terminal transport goods to be delivered from the terminal to the destinations of the shipments, and thereafter pick up goods at their origins and return again to the terminals where the vehicles are based. Specifically, in each time period, goods have to be moved from origins in a set $\mathcal{O}$ to destinations in a set $\mathcal{D}$. There is a finite set Ω of flow scenarios. Each scenario $\omega \in \Omega$ has a probability or weight $p(\omega)$. The set Ω may represent the support of an input probability distribution, or may be obtained from an input probability distribution by sampling, or may represent predictable differences in flow rates in different time periods such as seasonal differences, or a combination of the above. For each scenario $\omega \in \Omega$, let flow rate $q_{ij}(\omega) \geq 0$ denote the quantity of goods per time period that must be moved from origin $i \in \mathcal{O}$ to destination $j \in \mathcal{D}$ in scenario ω. Each vehicle has the same capacity Q_v. There is a set $\mathcal{X}_E$ of existing terminals, and a set $\mathcal{X}_P$ of potential terminals. For each terminal $m \in \mathcal{X} := \mathcal{X}_E \cup \mathcal{X}_P$, let c_m denote the difference in cost per time period between having terminal m open and operating, and not having terminal m open and operating. For each $i, j \in \mathcal{O} \cup \mathcal{D} \cup \mathcal{X}$, let d_{ij} denote the cost to move a vehicle from point i to point j. We assume that the vehicle movement cost does not depend on the load carried by the vehicle. In addition to vehicle movement costs, there is a cost of C_v per time period for each vehicle based at each terminal, whether the vehicle is used or not, and a cost of c_v for each vehicle that is used during a time period, independent of the distance traveled by the vehicle.

We have to decide which of the existing and potential terminals should be open for all scenarios. Let binary decision variable u_m denote whether terminal m is open, that is,

$$u_m := \begin{cases} 1 & \text{if terminal } m \in \mathcal{X} \text{ is open} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Let integer decision variable n_v^m denote the number of vehicles assigned to terminal m.

For each scenario $\omega \in \Omega$, we have to decide how to move each shipment from its origin to its destination through the open terminals. That is, for each origin-destination pair $(i, j) \in \mathcal{O} \times \mathcal{D}$ with $q_{ij}(\omega) > 0$, we have to determine through which open terminal the shipment will move, and we have to decide how all the movements will be handled with vehicle routes. Let the binary decision variable $z_{ij}^m(\omega)$ denote whether the shipment from i to j moves through terminal m, that is,

$$z_{ij}^m(\omega) := \begin{cases} 1 & \text{if terminal } m \in \mathcal{X} \text{ is used in moving the shipment in scenario } \omega \in \Omega \\ & \text{from origin } i \in \mathcal{O} \text{ to destination } j \in \mathcal{D} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The definition of $z_{ij}^m(\omega)$ implies that the assignment of an origin-destination flow to a terminal may vary according to the scenario. If the assignment of origin-destination flows to terminals must be the same for all scenarios, then one only needs z_{ij}^m variables that do not depend on ω.

Vehicle routing decisions are formulated in more detail later in the section. At this stage, it is sufficient to specify that $\tau(\mathcal{O}', \mathcal{D}', Q', d', n_v)$ denotes the optimum cost of the vehicle routing problem with a particular terminal, set $\mathcal{O}' \subset \mathcal{O}$ of origins, set $\mathcal{D}' \subset \mathcal{D}$ of destinations, quantities to be picked up and delivered given by $Q' \in \mathbb{R}_+^{|\mathcal{O}'|+|\mathcal{D}'|}$ (where for each origin $i \in \mathcal{O}'$, Q_i' denotes the quantity to be picked up at i and brought to the terminal, and for each destination $j \in \mathcal{D}'$, Q_j' denotes the quantity to be taken from the terminal and delivered at j), vehicle movement costs between the terminal and the considered origins and destinations given by $d' \in \mathbb{R}^{(1+|\mathcal{O}'|+|\mathcal{D}'|)^2}$, and n_v vehicles with capacity Q_v each. The arguments of interest of the function τ depend on the decision variables in (2), as follows: For each $m \in \mathcal{X}$ and $z^m(\omega) \in \{0,1\}^{|\mathcal{O}| \times |\mathcal{D}|}$, let

$$
\mathcal{O}^m(z^m(\omega)) \; := \; \left\{ i \in \mathcal{O} \; : \; \sum_{j \in \mathcal{D}} z_{ij}^m(\omega) > 0 \right\}
$$

$$
\mathcal{D}^m(z^m(\omega)) \; := \; \left\{ j \in \mathcal{D} \; : \; \sum_{i \in \mathcal{O}} z_{ij}^m(\omega) > 0 \right\}
$$

$$
Q_i^m(z^m(\omega), \omega) \; := \; \sum_{j \in \mathcal{D}^m(z^m(\omega))} q_{ij}(\omega) z_{ij}^m(\omega) \quad \text{for } i \in \mathcal{O}^m(z^m(\omega))
$$

$$
Q_j^m(z^m(\omega), \omega) \; := \; \sum_{i \in \mathcal{O}^m(z^m(\omega))} q_{ij}(\omega) z_{ij}^m(\omega) \quad \text{for } j \in \mathcal{D}^m(z^m(\omega))
$$

$$
Q^m(z^m(\omega), \omega) \; := \; [Q_l^m(z^m(\omega), \omega) \; : \; l \in \mathcal{O}^m(z^m(\omega)) \cup \mathcal{D}^m(z^m(\omega))]
$$

$$
d^m(z^m(\omega)) \; := \; [d_{ij} \; : \; i, j \in \mathcal{O}^m(z^m(\omega)) \cup \mathcal{D}^m(z^m(\omega)) \cup \{m\}].
$$

Then the optimization problem of interest is

$$
\min_{u \in \{0,1\}^{|\mathcal{X}|}, n_v \in \mathbb{N}^{|\mathcal{X}|}} \left\{ \sum_{m \in \mathcal{X}} (c_m u_m + C_v n_v^m) + \sum_{\omega \in \Omega} p(\omega) V(u, n_v, \omega) \right\} \tag{3}
$$

where

$$
V(u, n_v, \omega) \; := \; \min_{z(\omega)} \; \sum_{m \in \mathcal{X}} \tau(\mathcal{O}^m(z^m(\omega)), \mathcal{D}^m(z^m(\omega)), Q^m(z^m(\omega), \omega), d^m(z^m(\omega)), n_v^m) \tag{4}
$$

$$
\text{subject to} \quad \sum_{m \in \mathcal{X}} z_{ij}^m(\omega) \; = \; \mathbb{I}_{\{q_{ij}(\omega) > 0\}} \qquad \text{for all } i \in \mathcal{O}, \; j \in \mathcal{D} \tag{5}
$$

$$
z_{ij}^m(\omega) \; \leq \; u_m \qquad \text{for all } i \in \mathcal{O}, \; j \in \mathcal{D}, \; m \in \mathcal{X} \tag{6}
$$

$$
z_{ij}^m(\omega) \; \in \; \{0,1\} \qquad \text{for all } i \in \mathcal{O}, \; j \in \mathcal{D}, \; m \in \mathcal{X} \tag{7}
$$

gives the minimum cost distribution plan for scenario ω given the open terminals specified by u, vehicle fleet sizes specified by n_v, and the origin-destination flows specified by $q(\omega)$.

Note that in the definition of $Q^m(z^m(\omega), \omega)$ above, the quantities that have to be picked up and delivered in a time period in scenario ω are specified by the origin-destination flows $q_{ij}(\omega)$. In practice, in a time period it is typical to deliver goods that were picked up in the previous time period and then to pick up goods and bring them to a terminal to be delivered in the next time period. If the origin-destination flows vary from week-to-week, then the total amount picked up in a particular week may not equal the total amount delivered in the same week. For example, if q is positive every second time period and zero every other period, then in alternating time periods goods will be picked up (but not delivered), and delivered (but not picked up). To capture such behavior accurately, a multistage formulation is needed instead of the two-stage formulation (3)–(7).

The formulation above can accommodate various definitions of the function τ, and thus various types of vehicle routing constraints. Below we give a definition of τ that allows multiple vehicles to visit an origin or a destination during a time period, and that requires deliveries to be completed before any pickups are done on a route. This definition was motivated by the applications described in the introduction. Such a vehicle routing

problem is a combination of two problems that have been studied in the literature. The one problem is called the vehicle routing problem with backhauls; see, for example, Goetschalckx and Jacobs-Blecha (1989), Anily (1996), Potvin et al. (1996), Thangiah et al. (1996), Toth and Vigo (1997, 1999), Mingozzi et al. (1999), and Osman and Wassan (2002). The other problem is called the vehicle routing problem with split deliveries (sometimes split pickups); see, for example, Dror and Trudeau (1989), Dror et al. (1994), Mullaseril et al. (1997), Archetti et al. (2006), and Lee et al. (2006). As far as we know, the vehicle routing problem with backhauls and split pickups and deliveries has not been studied in the literature.

Consider a set $\mathcal{O}'$ of origins, set $\mathcal{D}'$ of destinations, quantities to be picked up and delivered given by Q', vehicle movement costs between the considered origins, destinations, and a particular terminal given by d', and n_v' vehicles with capacity Q_v each. Let 0 denote the terminal. Let $\mathcal{V}' := \{0\} \cup \mathcal{O}' \cup \mathcal{D}'$ denote the set of nodes. Because deliveries must be completed before any pickups are done on a route, the feasible arc set on a route is

$$\mathcal{A}' := \left\{ (i,j) \in (\mathcal{V}')^2 \setminus \mathcal{O}' \times \mathcal{D}' \ : \ i \neq j \right\}.$$

The decision variables are as follows:

$$v_k := \begin{cases} 1 & \text{if vehicle } k \text{ is in use} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ijk} := \begin{cases} 1 & \text{if vehicle } k \text{ travels on arc } (i,j) \\ 0 & \text{otherwise} \end{cases}$$

$$a_{ik} \geq 0, \text{ the amount of goods picked up at } i \text{ if } i \in \mathcal{O}' \text{ or}$$
$$\text{the amount of goods delivered to } i \text{ if } i \in \mathcal{D}', \text{ by vehicle } k.$$

Then

$$\tau\left(\mathcal{O}', \mathcal{D}', Q', d', n_v'\right) := \min_{x,v,a} \sum_{(i,j)\in\mathcal{A}'} \sum_{k=1}^{n_v'} d_{ij}' x_{ijk} + c_v \sum_{k=1}^{n_v'} v_k \tag{8}$$

$$\text{subject to} \quad \sum_{\{j \,:\, (j,i)\in\mathcal{A}\}} x_{jik} = \sum_{\{j \,:\, (i,j)\in\mathcal{A}\}} x_{ijk} \quad \text{for all } i \in \mathcal{O}' \cup \mathcal{D}', \ k \in \{1,\dots,n_v'\} \tag{9}$$

$$a_{ik} \leq \left(\sum_{\{j \,:\, (i,j)\in\mathcal{A}\}} x_{ijk} \right) Q_v \quad \text{for all } i \in \mathcal{O}' \cup \mathcal{D}', \ k \in \{1,\dots,n_v'\} \tag{10}$$

$$\sum_{i\in\mathcal{D}'} a_{ik} \leq Q_v v_k \quad \text{for all } k \in \{1,\dots,n_v'\} \tag{11}$$

$$\sum_{i\in\mathcal{O}'} a_{ik} \leq Q_v v_k \quad \text{for all } k \in \{1,\dots,n_v'\} \tag{12}$$

$$\sum_{k=1}^{n_v'} a_{ik} = Q_i' \quad \text{for all } i \in \mathcal{O}' \tag{13}$$

$$\sum_{k=1}^{n_v'} a_{jk} = Q_j' \quad \text{for all } j \in \mathcal{D}' \tag{14}$$

$$\sum_{\{(i,j)\in\mathcal{A} \,:\, i,j\in S\}} x_{ijk} \leq |S| - 1 \quad \text{for all } k \in \{1,\dots,n_v'\},$$
$$S \subset \mathcal{O}' \text{ or } S \subset \mathcal{D}', |S| \geq 2 \tag{15}$$

$$v_k \in \{0,1\} \quad \text{for all } k \in \{1,\dots,n_v'\} \tag{16}$$

$$x_{ijk} \in \{0,1\} \quad \text{for all } (i,j) \in \mathcal{A}, \ k \in \{1,\dots,n_v'\} \tag{17}$$

$$a_{ik} \geq 0 \quad \text{for all } i \in \mathcal{O}' \cup \mathcal{D}', \ k \in \{1,\dots,n_v'\} \tag{18}$$

With τ given by (8)–(18), problem (3)–(7) can be formulated as a two-stage mixed integer linear program, or simply as a large mixed integer linear program. Very small instances of such a problem can be solved with available software. In computational tests, instances with up to 5 origins, 5 destinations, 3 candidate terminals, and a single scenario, could be solved. Most instances with 6 origins, 6 destinations, 3 candidate terminals, and a single scenario, could not be solved — the computer's memory was insufficient to complete the branch-and-bound procedure. Problems in applications have hundreds of origins and destinations, and thus it seems that in the foreseeable future it will be impractical to solve the mixed integer linear program.

Next we describe an approach to problem (3) that seems to be widely used. It is natural to attempt to improve the tractability of problem (3) by simplifying the vehicle routing problem (8)–(18). A popular way to do this is to model freight movements as flows on arcs instead of vehicle routes, and to ignore fixed vehicle costs. The resulting problem is the following two echelon multicommodity (TEMC) location problem (for a deterministic version, see for example Ghiani et al. 2004). Decision variable u_m is the same as in (1). Decision variable $y_{ij}^m(\omega)$ denotes the quantity of goods flowing from origin i to destination j through terminal m in scenario ω. Then the TEMC location problem is as follows:

$$\min_{u \in \{0,1\}^{|\mathcal{X}|}} \left\{ \sum_{m \in \mathcal{X}} c_m u_m + \sum_{\omega \in \Omega} p(\omega) W(u, \omega) \right\} \tag{19}$$

where

$$W(u, \omega) \quad := \quad \min_{y(\omega)} \quad \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \sum_{m \in \mathcal{X}} (d_{im} + d_{mj}) \, y_{ij}^m(\omega) \tag{20}$$

$$\text{subject to} \quad \sum_{m \in \mathcal{X}} y_{ij}^m(\omega) \quad = \quad q_{ij}(\omega) \qquad \text{for all } i \in \mathcal{O}, \ j \in \mathcal{D} \tag{21}$$

$$y_{ij}^m(\omega) \quad \leq \quad q_{ij}(\omega) u_m \qquad \text{for all } i \in \mathcal{O}, \ j \in \mathcal{D}, \ m \in \mathcal{X} \tag{22}$$

$$y_{ij}^m(\omega) \quad \geq \quad 0 \qquad \text{for all } i \in \mathcal{O}, \ j \in \mathcal{D}, \ m \in \mathcal{X} \tag{23}$$

Note that, for any $u \in \{0,1\}^{|\mathcal{X}|}$, problem (20)–(23) has an easy solution. For each origin-destination pair $(i,j) \in \mathcal{O} \times \mathcal{D}$, let $m(i,j) := \arg\min_{m \in \mathcal{X}} \{d_{im} + d_{mj} : u_m = 1\}$ denote the cheapest open (given u) terminal to use from origin i to destination j. Then

$$W(u, \omega) \quad = \quad \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \left(d_{i,m(i,j)} + d_{m(i,j),j} \right) q_{ij}(\omega)$$

Thus

$$\sum_{\omega \in \Omega} p(\omega) W(u, \omega) \quad = \quad \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \left(d_{i,m(i,j)} + d_{m(i,j),j} \right) \sum_{\omega \in \Omega} p(\omega) q_{ij}(\omega)$$

$$= \quad \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \left(d_{i,m(i,j)} + d_{m(i,j),j} \right) \bar{q}_{ij}$$

$$= \quad \min_{\bar{y}} \quad \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \sum_{m \in \mathcal{X}} (d_{im} + d_{mj}) \, \bar{y}_{ij}^m$$

$$\text{subject to} \quad \sum_{m \in \mathcal{X}} \bar{y}_{ij}^m \quad = \quad \bar{q}_{ij} \qquad \text{for all } i \in \mathcal{O}, \ j \in \mathcal{D}$$

$$\bar{y}_{ij}^m \quad \leq \quad \bar{q}_{ij} u_m \qquad \text{for all } i \in \mathcal{O}, \ j \in \mathcal{D}, \ m \in \mathcal{X}$$

$$\bar{y}_{ij}^m \quad \geq \quad 0 \qquad \text{for all } i \in \mathcal{O}, \ j \in \mathcal{D}, \ m \in \mathcal{X}$$

where $\bar{q}_{ij} := \sum_{\omega \in \Omega} p(\omega) q_{ij}(\omega)$. Therefore problem (19)–(23) reduces to the following problem:

$$\min_{u, \bar{y}} \quad \left\{ \sum_{m \in \mathcal{X}} c_m u_m + \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \sum_{m \in \mathcal{X}} (d_{im} + d_{mj}) \bar{y}_{ij}^m \right\} \tag{24}$$

$$\text{subject to} \quad \sum_{m \in \mathcal{X}} \bar{y}_{ij}^m \;=\; \bar{q}_{ij} \qquad \text{for all } i \in \mathcal{O}, \; j \in \mathcal{D} \tag{25}$$

$$\bar{y}_{ij}^m \;\leq\; \bar{q}_{ij} u_m \qquad \text{for all } i \in \mathcal{O}, \; j \in \mathcal{D}, \; m \in \mathcal{X} \tag{26}$$

$$\bar{y}_{ij}^m \;\geq\; 0 \qquad \text{for all } i \in \mathcal{O}, \; j \in \mathcal{D}, \; m \in \mathcal{X} \tag{27}$$

$$u_m \;\in\; \{0, 1\} \qquad \text{for all } m \in \mathcal{X} \tag{28}$$

Formulation (24)–(28) is used in several commercial software packages for distribution network design.

4 Qualitative Discussion of Important Factors in Distribution Network Design

Recall the first stage problem (3), which can be rewritten as follows:

$$\min_{N \in \{1,2,\dots\}} \; \min_{\{u \in \{0,1\}^{|\mathcal{X}|} \,:\, \sum_{m \in \mathcal{X}} u_m = N\}} \; \min_{n_v \in \mathbb{N}^{|\mathcal{X}|}} \; \left\{ \sum_{m \in \mathcal{X}} (c_m u_m + C_v n_v^m) + \sum_{\omega \in \Omega} p(\omega) V(u, n_v, \omega) \right\} \tag{29}$$

$$= \min_{N \in \{1,2,\dots\}} \; f(N)$$

where

$$f(N) \;:=\; \min_{\{u \in \{0,1\}^{|\mathcal{X}|} \,:\, \sum_{m \in \mathcal{X}} u_m = N\}} \; \min_{n_v \in \mathbb{N}^{|\mathcal{X}|}} \; \left\{ \sum_{m \in \mathcal{X}} (c_m u_m + C_v n_v^m) + \sum_{\omega \in \Omega} p(\omega) V(u, n_v, \omega) \right\} \tag{30}$$

In words, one can first choose the number of terminals, then the locations of the terminals, and then the number of vehicles at each terminal. Of course, when one chooses the number N of terminals, one should take into account how the following optimization problems depend on N. Next we describe an approach for choosing the number N of terminals, approximating how the following optimization problems depend on N.

Suppose that the locations most likely to be chosen for the terminals (typically the locations with smaller values of c_m) have approximately the same fixed costs $c_m \approx c$. Then the first term $\sum_{m \in \mathcal{X}} c_m u_m$ in the objective function f can be replaced with cN. Next, note that the total number $\sum_{m \in \mathcal{X}} n_v^m$ of vehicles needed can be estimated quite accurately with the flow data and the vehicle capacity only, for example, by $\max_{\omega \in \Omega} \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij}(\omega) / Q_v$, so that the total fixed vehicle cost $\sum_{m \in \mathcal{X}} C_v n_v^m$ does not depend much on the chosen number N of terminals. Selection of the optimal number n_v^m of vehicles at each terminal is addressed later. Next, suppose that we approximate the remaining part of the objective function, $\min_{\{u \in \{0,1\}^{|\mathcal{X}|} \,:\, \sum_{m \in \mathcal{X}} u_m = N\}} \min_{n_v \in \mathbb{N}^{|\mathcal{X}|}} \sum_{\omega \in \Omega} p(\omega) V(u, n_v, \omega)$, with $\bar{V}(N) := \sum_{\omega \in \Omega} p(\omega) \hat{V}(N, \omega)$. Then one obtains an approximating problem

$$\min_{N \in \{1,2,\dots\}} \; \left\{ \hat{f}(N) \;:=\; cN + \bar{V}(N) \right\} \tag{31}$$

It remains to show how an approximation $\bar{V}(N)$ can be constructed that is both accurate and easy to compute. In the remainder of this section, we make a few observations that guide the development of our CA method regarding (1) the locations of the origins and destinations and the flow rates between them, (2) the importance of terminal location, and (3) the selection of how many terminals serve each origin and each destination.

In most applications, the distribution of origins and destinations is quite different from the uniform distribution. A relevant question is whether approximation of the locations of origins and destinations with a uniform

distribution may lead, or is likely to lead, to choosing an incorrect number N of terminals. It is not our purpose to formulate precise questions and give precise answers in this regard; let us just point out that it is easy to construct examples in which approximation of the locations of origins and destinations with a uniform distribution leads to an incorrect number of terminals being selected. Also, in most applications, the origin-destination flow rates vary greatly for different origin-destination pairs. It can also easily be seen that approximating the origin-destination flow rates with a constant rate for all origin-destination pairs may lead to an incorrect number of terminals being selected. We also note that in the applications that motivated this work, data on locations of origins and destinations, and historical flow rates, were easy to obtain. Thus, to select the number of terminals, it is important to take into account the locations of origins and destinations, and the origin-destination flow rates, with more accurate detail than a uniform distribution. For example, suppose that most origins and destinations are located along the east coast and west coast. If most flows are between origin-destination pairs on opposite coasts, then it may be good to have a central terminal, with some vehicles delivering and picking up loads on the east coast, and other vehicles delivering and picking up loads on the west coast, with the loads being exchanged at the central terminal. On the other hand, if most flows are between origin-destination pairs on the same coast, then it may be good to have two terminals, one on the east coast and one on the west coast. A uniform distribution cannot capture the distinction between the two cases above.

Next we address the question of whether it is important to accurately take into account how the terminals will be located when the number N of terminals is chosen. First, note that a crucial difference between this question and the questions of the previous paragraph is that the locations of the terminals are obviously not known before the number of terminals is chosen, but as already pointed out, often one has data about the locations of origins and destinations, and the flow rates. Also, it is reasonable to expect that, given N, the function $V(u, n_v, \omega)$ will not be very sensitive with respect to the locations u, the intuition being that the optimization in the second stage problem (4)–(7) that defines $V(u, n_v, \omega)$ allows the second stage decisions to adjust to the locations u. Next we describe a crude but simple experiment to illustrate the intuition regarding the effect of terminal location. We took the locations of the origins and destinations as well as the flow rates from the data set for one of the applications that we worked on. The locations of the origins and destinations are shown in Figure 5. The location and flow data are given in the online appendix. Given any set of terminal locations, for each origin-destination pair the least great-circle distance from the origin to one of the terminals and from the terminal to the destination can easily be calculated. For each origin-destination pair we calculate the following weighted distance between the origin and destination:

$$\left(\frac{\text{origin-destination flow rate}}{\text{vehicle capacity}}\right) \times (\text{least great-circle distance from origin through a terminal to destination}).$$

Thus the total weighted distance over all origin-destination pairs can be calculated for any given set of terminal locations. The number N of terminals was varied from 1 to 10. For each number N of terminals, the following procedure was repeated 10,000 times. One terminal was located centrally in a rectangular area covering all the origin and destination locations in our dataset. $N - 1$ terminals were located independently and uniformly distributed in the rectangle that contains all origins and destinations. Then the total weighted distance over all origin-destination pairs was calculated as described above. The total cost per time period was equal to the total weighted distance in miles plus $25,000 \times N$ for the terminals. We used 2000 units for the vehicle capacity. Table 4 shows the average total cost over the 10,000 replications, and the minimum total cost over the 10,000 replications (approximating the total cost of the best terminal locations) for each number N of terminals. The number of terminals minimizing the average cost is equal to the number of terminals minimizing the minimum cost, namely 4 terminals, which suggests that the optimum number of terminals can be determined quite accurately without optimizing the locations of the terminals.

One further observation regarding terminal location is that in practice there are many factors that play a role in the selection of exact locations of facilities that do not lead to simple models, such as the local transportation infrastructure, availability of individual sites, property prices, taxes, cost of local labor, climate, aesthetics of the location, and other subjective preferences. We agree with Dasci and Laporte (2005) that models that "entail a

Figure 5: Origins (circles) and destinations (pluses) for one of the applications.

Number of	Cost		
Terminal Facilities	Mean	Minimum	Standard Deviation
1	415,201	N/A	N/A
2	407,240	317,864	48,997
3	391,429	305,709	55,679
4	385,652	300,914	53,914
5	389,874	320,991	50,023
6	398,968	342,544	44,989
7	420,964	359,528	47,112
8	438,693	379,866	44,242
9	458,251	401,393	39,517
10	477,503	418,487	36,558

Table 1: The total cost of locating N terminal facilities.

very precise representation of the locations" are "a fixation in pursuit of theoretical accuracy that has relatively little practical value". We do not attempt to model factors such as those mentioned above, and therefore we do not address the exact location of terminals. Fortunately, the optimal number of terminals does not seem to be very sensitive with respect to the exact locations to be chosen for the terminals.

We also want to address the effect of the number of terminals that serve each origin and each destination. Recall the one-terminal-nearest-terminal distribution strategy described in Section 2. With that strategy, shipments from each origin are taken to the terminal closest to the origin, and from there the shipment is transported to its destination. Thus, with such a strategy, each origin is served by exactly one terminal, and each destination is served by all the terminals. The larger the number of terminals that serves an origin or destination, the larger the number of vehicles that have to stop at the origin or destination per time period. It was shown in Beardwood et al. (1959) that the optimal tour length increases proportionally to the square root of the number of points to be visited on the tour. Therefore, the detour distance for pickups or deliveries increases approximately proportionally to the square root of the number of terminals that serves origins or destinations. At the same time, the larger the number of terminals that serves origins or destinations, the smaller the linehaul portion of the transportation distance between origins and terminals, or terminals and destinations, can be made by careful selection of the terminal to be used to flow goods for each origin-destination pair. Therefore, there is a trade-off between linehaul distance and detour distance as a function of the number of terminals that serves origins and destinations, and we want to capture this trade-off with our approach, instead of fixing a strategy such as the one-terminal-nearest-terminal strategy. Also, it seems reasonable that the larger the total quantity of goods that should be picked up at an origin or delivered to a destination in a time period, the larger the number of terminals that serves that origin or destination should be in the time period. As an extreme example, if the total quantity of goods that should be picked up at an origin in a time period is very small, then only one terminal should serve that origin in that time period, so that only one vehicle has to make a stop at that origin in that time period. It is typical for the total quantity of goods that should be picked up at an origin or delivered to a destination to vary significantly among origins and destinations in the same time period, and also among time periods for the same origin or destination. Figure 6 shows the distribution over origins and destinations of the average total quantity of goods picked up or delivered per week for one of the applications that we worked on. Therefore, we want to allow the number of terminals that serves an origin or destination to be different for different origins and destinations in the same time period, and also different for the same origin or destination in different time periods.
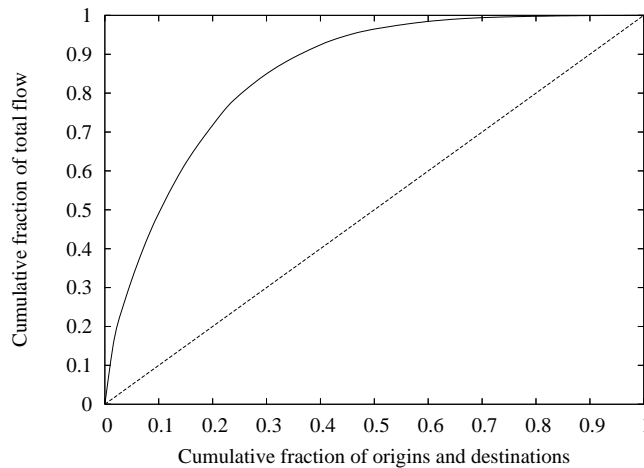

Figure 6: The distribution over origins and destinations of the average total quantity of goods picked up or delivered per week.

Motivated by the observations above, we construct an approximation $\hat{V}(N, \omega)$ that takes into account detailed

data regarding locations of origins and destinations, and the origin-destination flows, that approximate terminal locations with a uniform distribution, and that chooses the number of terminals that serve an origin or destination in a time period based on the total quantity of goods that should be picked up at the origin or delivered to the destination in the time period. Note that this approach is in some sense the opposite of the approach followed in Hall (1984), Daganzo (1987), Hall (1987), Campbell (1990b), and Campbell (1993a), where it was assumed that origins and destinations are uniformly distributed, and origin-destination flows are the same for all origin-destination pairs, and the terminals were carefully located on a rectangular grid. Also note that the purpose of the papers above was to obtain qualitative insight without data, and not to develop a method that can produce good solutions, whereas we want to develop a method that can produce good solutions.

5   Continuous Approximation Approach

In this section we describe a CA approach for designing a distribution network in which each shipment moves through one terminal on its way from its origin to its destination. First we provide an overview of the approach in Section 5.1. Thereafter we describe the steps of the method.

5.1   Overview

It was shown in (29), (4), and (8) that we can make decisions in the following sequence, and that it should be taken into account how each decision will affect the subsequent decisions and objective values. The following decisions are made before the scenario ω (and thus the flow realization $q(\omega)$) is known:

 (i) Select the number of terminals.

 (ii) Select the location of each terminal.

 (iii) Select the number of vehicles at each terminal.

The following decisions are made after the flow realization $q(\omega)$ is known:

 (iv) For each origin-destination pair with flow $q_{ij}(\omega) > 0$, determine which terminal will be used to move the shipment from the origin to the destination.

 (v) For each terminal, decide how vehicles are routed from the terminal to do the pickups and deliveries of the flows assigned to the terminal in the previous step.

Embarking on this work, we were primarily interested in the first design decision, namely how to select the best number of terminals for a distribution network. However, inspection of (3), (4), and (8), reveals that the design decision (i) should not be made without taking into account design decisions (ii) and (iii), and operational decisions (iv) and (v). It was also argued that taking design decisions (ii) and (iii) and operational decisions (iv) and (v) into account by solving mixed integer programs was impractical, and hopefully unnecessary. Thus a major part of the work is aimed at the development of tractable approximations that take design decisions (ii) and (iii) and operational decisions (iv) and (v) into account, so that design decision (i) can be made relatively easily and well.

We also want to test the quality of the decisions resulting from the developed approximations. To do so, we would like to take the value of design decision (i) resulting from the approximations and that resulting from formulation (24)–(28), for each solve the mixed integer programs that produce decisions (ii)–(v), and then compare the total cost resulting from the two values of design decision (i). The challenge that the mixed integer programs that produce decisions (ii)–(v) are intractable remain, and therefore we developed heuristics to produce decisions (ii)–(v). These heuristics require significantly more computational effort than the approximations developed to make decision (i), but are tractable enough to do the comparison for given values of decision (i). Some of these heuristics may be of interest by themselves.

Section 5.2 describes a CA method for selecting the number of terminals. Section 5.3 describes how the terminals are located in computational experiments to test the quality of the solutions produced with the CA method. Section 5.4 provides a heuristic for determining the vehicle fleet sizes.

5.2 Selection of Number of Terminals

The two major cost components considered in the selection of the number of terminals are terminal fixed cost and transportation cost. Terminal fixed cost increases proportionally with the number of terminals. Transportation cost should decrease as the number of terminals increases, because there are more terminals to choose from when routing shipments from their origins to their destinations through terminals. Transportation cost is taken to be proportional to the transportation distance. The transportation distance results from the routes that vehicles travel from each terminal to perform the pickups and deliveries, and is partitioned into two components, namely the linehaul distance and the detour distance. The linehaul distance associated with a route is the distance from the terminal to the center of the points that are visited, and back again. The detour distance is the remaining distance on the route, and is approximately the length of a tour through the points that are visited, excluding the terminal.

As explained in Section 4, as the number of terminals that serves each origin or destination increases, the average linehaul distance, or the average distance from the origin of a shipment through the terminal used for the shipment to the destination, decreases, but the total number of stops on vehicle routes increases, and thus the detour distance increases. As also explained in Section 4, the larger the total flow from an origin or to a destination, the larger the number of terminals that should serve that origin or destination. We control the number of terminals that serves an origin or a destination as follows. Suppose there are N open terminals. Then we select $N-1$ thresholds, $0 \leq Q_1 \leq \cdots \leq Q_{N-1}$. Consider an origin i and a scenario ω. If $\sum_{j\in\mathcal{D}} q_{ij}(\omega) \in (0, Q_1]$, then 1 terminal serves origin i in scenario ω. If $\sum_{j\in\mathcal{D}} q_{ij}(\omega) \in (Q_{N-1}, \infty)$, then all N terminals serve origin i in scenario ω. Otherwise, if $\sum_{j\in\mathcal{D}} q_{ij}(\omega) \in (Q_{k-1}, Q_k]$, then k terminals serve origin i in scenario ω. The number of terminals that serve each destination in each scenario is determined in the same way. Although the number of terminals that serves an origin or a destination is allowed to depend on the scenario, the thresholds do not depend on the scenario.

Since not all terminals serve each origin and destination, care has to be taken to ensure that for each origin-destination pair (i, j) with $q_{ij}(\omega) > 0$, there is at least one terminal that serves both i and j. We do that by designating one terminal, called the center terminal, to serve all origins and destinations.

An outline of the method for selecting the number of terminals and the thresholds is as follows:

1. A method is developed to approximate total linehaul distance as a function of the number of terminals and the thresholds. This method is described in Section 5.2.1.

2. Section 5.2.2 describes a method to approximate total detour distance as a function of the number of terminals and the thresholds.

3. The approximations of total linehaul distance and total detour distance as a function of the number of terminals and the thresholds are used to search for the optimal (as measured by the approximations) number of terminals and thresholds. The search method is described in Section 5.2.3.

5.2.1 Linehaul Distance Estimation

Suppose there are N open terminals, $n = 0, 1, \ldots, N - 1$, located in some region that does not have to include the locations of all origins and destinations. The center terminal, used by all origins and destinations, is indexed by 0. This section describes the estimation of the total linehaul distance for given values $Q := (Q_1, \ldots, Q_{N-1})$ of the thresholds.

Consider a scenario $\omega \in \Omega$, and an origin $i \in \mathcal{O}$. Suppose that origin i is served by a set $\mathcal{N}_i$ of N_i terminals, including the center terminal. The selection of the $N_i - 1$ terminals that serve origin i in addition to the center terminal from the set $\{1, 2, \ldots, N-1\}$ of open terminals besides the center terminal is described later. Similarly,

consider a destination $j \in \mathcal{D}$, and suppose that destination j is served by a set $\mathcal{N}_j$ of N_j terminals, also including the center terminal. The numbers N_i and N_j depend on the thresholds Q and on the scenario ω, but the dependence is not shown in the notation. Let $\mathcal{N}_{ij} := \mathcal{N}_i \cap \mathcal{N}_j$ denote the set of terminals that serve both origin i and destination j. The sets $\mathcal{N}_i$, $\mathcal{N}_j$, and $\mathcal{N}_{ij}$ depend on the thresholds Q, on the scenario ω, and on the selection of the terminals, but as before the dependence is not shown in the notation.

Let $\lambda_{i,n,j}$ denote the distance from origin i through terminal n to destination j. Distance $\lambda_{i,n,j}$ depends on the locations of the terminals, but the dependence is not shown in the notation. Then the minimum distance from origin i to destination j through a terminal that serves both i and j is given by

$$\Lambda_{i,j} \quad := \quad \min_{n \in \mathcal{N}_{ij}} \lambda_{i,n,j}. \tag{32}$$

Distance $\Lambda_{i,j}$ depends on $\mathcal{N}_{ij}$ and thus on the thresholds Q, on the scenario ω, and on the selection of the terminals; and on distances $\lambda_{i,n,j}$, and thus on the locations of the terminals.

For the given number N of terminals, and the given thresholds Q, the expected total linehaul distance $L(N,Q)$ is then calculated as follows:

$$L(N,Q) \quad := \quad \sum_{\omega \in \Omega} p(\omega) \sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} \frac{q_{ij}(\omega)}{Q_v} \mathbb{E}[\Lambda_{i,j}] \tag{33}$$

where $\mathbb{E}[\Lambda_{i,j}]$ denotes the expected value of $\Lambda_{i,j}$ with respect to random parameters involved in the selection of the terminals.

Next we consider the calculation of $\mathbb{E}[\Lambda_{i,j}]$ in greater detail. Suppose that the locations of the $N-1$ terminals $\{1, 2, \ldots, N-1\}$ besides the center terminal are independently and identically distributed in some region. Suppose that the set $\mathcal{N}_i \setminus \{0\}$ of terminals that serve origin i in addition to the center terminal are selected from terminals $1, 2, \ldots, N-1$ by making $N_i - 1$ random selections without replacement from $\{1, 2, \ldots, N-1\}$, each time selecting each remaining element with equal probability. The set $\mathcal{N}_j \setminus \{0\}$ of $N_j - 1$ terminals that serve destination j in addition to the center terminal are selected in the same way. This selection of terminals is independent for all origins and destinations.

Note that the linehaul distance $\Lambda_{i,j}$ is decreasing in both N_i and N_j, as it should be. Specifically, let $\varpi_i := (\varpi_{i,1}, \ldots, \varpi_{i,N-1})$ and $\varpi_j := (\varpi_{j,1}, \ldots, \varpi_{j,N-1})$ be two independent and identically distributed random permutations of $\{1, 2, \ldots, N-1\}$, with distribution such that each of the $(N-1)!$ permutations has probability $1/(N-1)!$. Consider any $n_i < N$ and $n_j \leq N$. Let $\tilde{\mathcal{N}}_i := \{0, \varpi_{i,1}, \ldots, \varpi_{i,n_i-1}\}$, $\tilde{\mathcal{N}}_i^+ := \{0, \varpi_{i,1}, \ldots, \varpi_{i,n_i}\}$, $\tilde{\mathcal{N}}_j := \{0, \varpi_{j,1}, \ldots, \varpi_{i,n_j-1}\}$, $\tilde{\mathcal{N}}_{ij} := \tilde{\mathcal{N}}_i \cap \tilde{\mathcal{N}}_j$, $\tilde{\mathcal{N}}_{ij}^+ := \tilde{\mathcal{N}}_i^+ \cap \tilde{\mathcal{N}}_j$, $\tilde{\Lambda}_{i,j} := \min_{n \in \tilde{\mathcal{N}}_{ij}} \lambda_{i,n,j}$, and $\tilde{\Lambda}_{i,j}^+ := \min_{n \in \tilde{\mathcal{N}}_{ij}^+} \lambda_{i,n,j}$. Then, $\tilde{\mathcal{N}}_i$, $\tilde{\mathcal{N}}_j$, $\tilde{\mathcal{N}}_{ij}$, and $\tilde{\Lambda}_{i,j}$ have the same distributions as $\mathcal{N}_i$, $\mathcal{N}_j$, $\mathcal{N}_{ij}$, and $\Lambda_{i,j}$ respectively, given that $N_i = n_i$ and $N_j = n_j$. Also, $\tilde{\mathcal{N}}_i^+$, $\tilde{\mathcal{N}}_{ij}^+$, and $\tilde{\Lambda}_{i,j}^+$ have the same distributions as $\mathcal{N}_i$, $\mathcal{N}_{ij}$, and $\Lambda_{i,j}$ respectively, given that $N_i = n_i + 1$ and $N_j = n_j$. Note that, w.p.1, $\tilde{\mathcal{N}}_i \subset \tilde{\mathcal{N}}_i^+$ and $\tilde{\mathcal{N}}_{ij} \subset \tilde{\mathcal{N}}_{ij}^+$. Thus $\tilde{\Lambda}_{i,j} \geq \tilde{\Lambda}_{i,j}^+$ w.p.1. Hence, the conditional distribution of $\Lambda_{i,j}$ given $N_i = n_i$ is stochastically decreasing in n_i.

Let $\mathcal{N}_{ij}^< := \{n \in \mathcal{N}_{ij} : \lambda_{i,n,j} < \lambda_{i,0,j}\}$ denote the set of terminals in $\mathcal{N}_{ij}$ that give a distance from i to j strictly less than the distance from i to j through the center terminal. We calculate $\mathbb{E}[\Lambda_{i,j}]$ by conditioning on $|\mathcal{N}_{ij}|$ and $|\mathcal{N}_{ij}^<|$, where for a set S, $|S|$ denotes its cardinality:

$$\mathbb{E}[\Lambda_{i,j}] \quad = \quad \sum_{k=1 \vee (N_i+N_j-N)}^{N_i \wedge N_j} \mathbb{P}[|\mathcal{N}_{ij}| = k] \sum_{\ell=0}^{k-1} \mathbb{P}\left[|\mathcal{N}_{ij}^<| = \ell \;\Big|\; |\mathcal{N}_{ij}| = k\right] \mathbb{E}\left[\Lambda_{i,j} \;\Big|\; |\mathcal{N}_{ij}| = k, |\mathcal{N}_{ij}^<| = \ell\right] \tag{34}$$

(Note that it always holds that $|\mathcal{N}_{ij}| \geq N_i + N_j - N$, irrespective of how $\mathcal{N}_i$ and $\mathcal{N}_j$ are selected.) First, note that since the terminals in $\mathcal{N}_j \setminus \{0\}$ are selected without replacement and with equal probabilities from the terminals $1, 2, \ldots, N-1$, independently of $\mathcal{N}_i$, it follows that the random variable $|\mathcal{N}_{ij}|$ follows a hypergeometric distribution. Specifically, given any set $\mathcal{N}_i$, the conditional probability $\mathbb{P}[|\mathcal{N}_{ij}| = k \,|\, \mathcal{N}_i]$ for any $k \in \{1 \vee (N_i +$

$N_j - N), \ldots, N_i \wedge N_j\}$ satisfies

$$\mathbb{P}[|\mathcal{N}_{ij}| = k \mid \mathcal{N}_i] \;\; = \;\; \frac{\binom{N_i - 1}{k - 1}\binom{N - N_i}{N_j - k}}{\binom{N - 1}{N_j - 1}} \;\; = \;\; \frac{(N_i - 1)!(N_j - 1)!(N - N_i)!(N - N_j)!}{(k - 1)!(N_i - k)!(N_j - k)!(N - N_i - N_j + k)!(N - 1)!}$$

(It is the same as the probability of drawing $k - 1$ red balls from a jar containing $N - 1$ balls, $N_i - 1$ of which are red, in a sample of size $N_j - 1$ drawn without replacement.) Since the right side above is the same for all realizations of $\mathcal{N}_i$, it follows that the probability that exactly $k \in \{1 \vee (N_i + N_j - N), \ldots, N_i \wedge N_j\}$ terminals serve both i and j is

$$\mathbb{P}[|\mathcal{N}_{ij}| = k] \;\; = \;\; \frac{(N_i - 1)!(N_j - 1)!(N - N_i)!(N - N_j)!}{(k - 1)!(N_i - k)!(N_j - k)!(N - N_i - N_j + k)!(N - 1)!}$$

As before, $\mathbb{P}[|\mathcal{N}_{ij}| = k]$ depends on the thresholds Q and the scenario ω. Next, since the locations of terminals $\{1, 2, \ldots, N - 1\}$ are independent and identically distributed, and the terminals in $\mathcal{N}_i \setminus \{0\}$ and $\mathcal{N}_j \setminus \{0\}$ are selected independently of the locations of the terminals, it follows that given that $|\mathcal{N}_{ij}| = k$, the number of the $k - 1$ terminals n other than the center terminal that have distance $\lambda_{i,n,j}$ strictly less than $\lambda_{i,0,j}$ has a binomial distribution. Specifically, for $\ell \in \{0, 1, \ldots, k - 1\}$,

$$\mathbb{P}\left[|\mathcal{N}_{ij}^{<}| = \ell \,\middle|\, |\mathcal{N}_{ij}| = k\right] \;\; = \;\; \binom{k - 1}{\ell} \mathbb{P}[\lambda_{i,1,j} < \lambda_{i,0,j}]^{\ell} \mathbb{P}[\lambda_{i,1,j} \geq \lambda_{i,0,j}]^{k - 1 - \ell}$$

Since $\Lambda_{i,j}$ is a nonnegative random variable, it holds for $\ell \in \{0, 1, \ldots, k - 1\}$ that

$$\mathbb{E}\left[\Lambda_{i,j} \,\middle|\, |\mathcal{N}_{ij}| = k, |\mathcal{N}_{ij}^{<}| = \ell\right]$$

$$= \int_0^{\infty} \mathbb{P}\left[\Lambda_{i,j} > \alpha \,\middle|\, |\mathcal{N}_{ij}| = k, |\mathcal{N}_{ij}^{<}| = \ell\right] d\alpha$$

$$= \int_0^{\infty} \mathbb{P}\left[\min_{n \in \mathcal{N}_{ij}} \lambda_{i,n,j} > \alpha \,\middle|\, |\mathcal{N}_{ij}| = k, |\mathcal{N}_{ij}^{<}| = \ell\right] d\alpha$$

$$= \int_0^{\lambda_{i,0,j}} \mathbb{P}\left[\min\{\lambda_{i,1,j}, \ldots, \lambda_{i,\ell,j}\} > \alpha \,\middle|\, \lambda_{i,1,j} < \lambda_{i,0,j}, \ldots, \lambda_{i,\ell,j} < \lambda_{i,0,j}\right] d\alpha$$

$$= \int_0^{\lambda_{i,0,j}} \mathbb{P}\left[\lambda_{i,1,j} > \alpha, \ldots, \lambda_{i,\ell,j} > \alpha \,\middle|\, \lambda_{i,1,j} < \lambda_{i,0,j}, \ldots, \lambda_{i,\ell,j} < \lambda_{i,0,j}\right] d\alpha$$

$$= \int_0^{\lambda_{i,0,j}} \mathbb{P}\left[\lambda_{i,1,j} > \alpha \,\middle|\, \lambda_{i,1,j} < \lambda_{i,0,j}\right]^{\ell} d\alpha$$

The third and fifth equalities follow from terminals $\{1, 2, \ldots, N - 1\}$ being independent and identically distributed, and the terminals in $\mathcal{N}_i \setminus \{0\}$ and $\mathcal{N}_j \setminus \{0\}$ being selected independently of the locations of the terminals. As a special case, note that it holds for $\ell = 0$ that

$$\mathbb{E}\left[\Lambda_{i,j} \,\middle|\, |\mathcal{N}_{ij}| = k, |\mathcal{N}_{ij}^{<}| = 0\right] \;\; = \;\; \lambda_{i,0,j}$$

In summary, it follows from (34) that

$$\mathbb{E}[\Lambda_{i,j}] \;\; = \;\; \sum_{k = 1 \vee (N_i + N_j - N)}^{N_i \wedge N_j} \frac{(N_i - 1)!(N_j - 1)!(N - N_i)!(N - N_j)!}{(N_i - k)!(N_j - k)!(N - N_i - N_j + k)!(N - 1)!}$$

$$\times \sum_{\ell = 0}^{k - 1} \frac{1}{\ell!(k - 1 - \ell)!} \mathbb{P}[\lambda_{i,1,j} < \lambda_{i,0,j}]^{\ell} \mathbb{P}[\lambda_{i,1,j} \geq \lambda_{i,0,j}]^{k - 1 - \ell}$$

$$\times \int_0^{\lambda_{i,0,j}} \mathbb{P}\left[\lambda_{i,1,j} > \alpha \,\middle|\, \lambda_{i,1,j} < \lambda_{i,0,j}\right]^{\ell} d\alpha \tag{35}$$

21

It remains to explain how $\mathbb{P}[\lambda_{i,1,j} < \lambda_{i,0,j}]$ and $\mathbb{P}[\lambda_{i,1,j} > \alpha \,|\, \lambda_{i,1,j} < \lambda_{i,0,j}]$ for $\alpha \in (0, \lambda_{i,0,j})$ can be computed. We provide the details of these calculations in the appendix, where it is shown that if the terminals $\{1, 2, \ldots, N-1\}$ are uniformly distributed in a rectangular region, and distance $\lambda_{i,1,j}$ is given by the L_1-metric, then $\mathbb{P}[\lambda_{i,1,j} > \alpha \,|\, \lambda_{i,1,j} < \lambda_{i,0,j}]$ is a piecewise polynomial in α of degree at most two, and thus

$$\int_0^{\lambda_{i,0,j}} \mathbb{P}\left[\lambda_{i,1,j} > \alpha \,\Big|\, \lambda_{i,1,j} < \lambda_{i,0,j}\right]^\ell d\alpha$$

is conceptually simple. Thus, for a given number N of terminals, and given thresholds Q, the expected total linehaul distance $L(N, Q)$ can be calculated quite easily using (33).

5.2.2 Detour Distance Estimation

Beardwood et al. (1959) established the following result that is widely used in the CA literature. Consider an independent sequence of uniformly distributed points in a set $A \subset \mathbb{R}^k$ with Lebesque measure $\mu(A) > 0$. Let T^n denote the shortest tour length, as measured by the L_2 Euclidean distance, through the first n points in the sequence. Then there exists a constant β_k, independent of the sequence and of A, such that with probability 1, $\lim_{n \to \infty} n^{-(k-1)/k} T^n = \beta_k k^{1/2} [\mu(A)]^{1/k}$. Specifically, for $\mathbb{R}^2$, there exists a constant $\beta_2 \in [0.44, 0.65]$ (the exact value is not yet known), such that with probability 1, $\lim_{n \to \infty} n^{-1/2} T^n = \beta_2 2^{1/2} [\mu(A)]^{1/2}$. The approximation $T^n \approx \beta \sqrt{n\mu(A)}$, for some β depending on the distance metric, has been used in many vehicle routing applications, and in this section we do the same for the purpose of obtaining a tractable approximation of the detour distance as a function of the number N of terminals and the thresholds Q.

Consider a given number N of terminals, given values $Q := (Q_1, \ldots, Q_{N-1})$ of the thresholds, and a given scenario $\omega \in \Omega$. Assume that the vehicles are fully loaded with goods to be delivered when departing from a terminal and fully loaded with goods that were picked up when arriving back at the terminal (one can multiply vehicle capacity Q_v with a factor between 0 and 1 to compensate for vehicles not being fully loaded on average). Then the total number of vehicle routes is equal to $\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij}(\omega)/Q_v$. Let the total region be denoted by $\bar{A}$ with area $\mu(\bar{A})$. Suppose that each terminal serves origins and/or destinations in most of $\bar{A}$ (this is the case even for distribution strategies such as one-terminal-nearest-terminal described in Section 2). The average number of vehicle routes per terminal is equal to $\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij}(\omega)/(NQ_v)$. Thus, if different vehicle routes from the same terminal do not overlap, then the average area served per vehicle route is equal to

$$\mu(A) = \frac{\mu(\bar{A})NQ_v}{\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij}(\omega)}$$

Note that $\mu(A)$ depends on N and ω, but the notation does not indicate the dependence.

Next we calculate the average number of delivery stops and the average number of pickup stops on a vehicle route, as a function of N, Q, and ω. The number of vehicle stops at an origin $i \in \mathcal{O}$ is at least $\max\left\{N_i, \lceil \sum_{j \in \mathcal{D}} q_{ij}(\omega)/Q_v \rceil\right\}$, and the number of vehicle stops at a destination $j \in \mathcal{D}$ is at least $\max\left\{N_j, \lceil \sum_{i \in \mathcal{O}} q_{ij}(\omega)/Q_v \rceil\right\}$ (recall that N_i and N_j depend on N, Q, and ω). Thus, the average number of pickup stops per vehicle route is approximately

$$n_p = \frac{\sum_{i \in \mathcal{O}} \max\left\{N_i, \lceil \sum_{j \in \mathcal{D}} q_{ij}(\omega)/Q_v \rceil\right\}}{\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij}(\omega)/Q_v}$$

and the average number of delivery stops per vehicle route is approximately

$$n_d = \frac{\sum_{j \in \mathcal{D}} \max\left\{N_j, \lceil \sum_{i \in \mathcal{O}} q_{ij}(\omega)/Q_v \rceil\right\}}{\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij}(\omega)/Q_v}$$

Then, the approximate expected total detour distance $D(N, Q)$ is calculated as follows:

$$D(N, Q) := \sum_{\omega \in \Omega} p(\omega) \frac{\sum_{i \in \mathcal{O}} \sum_{j \in \mathcal{D}} q_{ij}(\omega)}{Q_v} \beta\left[\sqrt{(n_p - 1)\mu(A)} + \sqrt{(n_d - 1)\mu(A)} + \sqrt{\mu(A)}\right]$$

22

The approximation is obtained by substituting expressions for the average number of stops and the average area of the region served into the tour length approximation $T^n \approx \beta\sqrt{n\mu(A)}$. The reason $n_p - 1$ and $n_d - 1$ are used in the tour length calculations is because the vehicle does not have to return to the first pickup point or the first delivery point after completing pickups or deliveries. The term $\beta\sqrt{\mu(A)}$ approximates the average distance from the last delivery point to the first pickup point on a vehicle route. Since $\sqrt{x}$ is a concave function, it follows from Jensen's inequality that this overestimates the average tour length. In numerical experiments, this overestimation did not seem to have much of an effect on the selection of the optimal number of terminals. In practice, such overestimation is likely to be dominated by the amount by which actual route lengths on road networks differ from the lengths according to simple metrics.

Note that n_p is increasing in the numbers N_i of terminals serving origins i, n_d is increasing in the numbers N_j of terminals serving destinations j, and $D(N, Q)$ is increasing in n_p and n_d. Also, the average area served per vehicle route $\mu(A)$ is increasing in the number N of terminals, and $D(N, Q)$ is increasing in $\mu(A)$. Thus the expected total detour distance $D(N, Q)$ is increasing in the number N of terminals and in the numbers of terminals serving origins and destinations, as it should be.

5.2.3 Search for Number of Terminals and Thresholds

Recall that we want to construct and solve an approximating problem (31):

$$\min_{N \in \{1, 2, \dots\}} \left\{ \hat{f}(N) := cN + \bar{V}(N) \right\}$$

taking into account the terminal fixed cost cN and the approximate expected transportation cost $\bar{V}(N)$. Without loss of generality, suppose that the unit of cost or the unit of distance has been scaled to make transportation cost per distance equal to 1. The approximate expected transportation cost $\bar{V}(N)$ is given by minimizing the sum of the linehaul cost and the detour cost over the thresholds:

$$\bar{V}(N) := \min_{0 \leq Q_1 \leq \cdots \leq Q_{N-1}} \{L(N, Q) + D(N, Q)\} \tag{36}$$

Note that the larger the value of N, the larger the set of thresholds that can be selected, and thus the smaller the value of $\bar{V}(N)$.

Note that if the total flows $\sum_{j \in \mathcal{D}} q_{ij}(\omega)$ for all $i \in \mathcal{O}$ and $\sum_{i \in \mathcal{O}} q_{ij}(\omega)$ for all $j \in \mathcal{D}$ are sorted, then all values of a threshold Q_k between two successive sorted values of the total flow give the same values of $L(N, Q)$ and $D(N, Q)$. Thus, if N is small, say $N \leq 4$, then problem (36) can easily be solved by enumerating all relevant values of the $N - 1$ thresholds. If N is large, then problem (36) can be solved approximately by a neighborhood search on the set of relevant values of the $N - 1$ thresholds. In addition, if a threshold Q_k is changed from one interval in the sorted list of total flows to a neighboring interval, the resulting change in the values of $L(N, Q)$ and $D(N, Q)$ can be computed very quickly, because the value of N_i or N_j for only one origin i or destination j is affected by the change.

Finally, problem $\min_{N \in \{1, 2, \dots\}} \left\{ cN + \bar{V}(N) \right\}$ can be solved by enumerating a range of reasonable values of N. The optimal value N^* is the number of terminals obtained with the CA method described above.

5.3 Terminal Location

Recall from Section 4 that it is not our purpose to model all the factors that should enter or do enter into the location of terminals. It is our purpose to test the CA method for selection of the number of terminals described in Section 5.2 by more detailed calculation of transportation costs. To facilitate such detailed calculation of vehicle routing costs, we need to calculate locations for the chosen number N of terminals. For that purpose,

we choose a set $\mathcal{X}$ of candidate locations, and solve the following problem:

$$\min_{u,\bar{y}} \quad \left\{ \sum_{m\in\mathcal{X}} c_m u_m + \sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}}\sum_{m\in\mathcal{X}} (d_{im} + d_{mj})\, \bar{y}_{ij}^m \right\}$$

$$\begin{aligned}
\text{subject to} \quad & \sum_{m\in\mathcal{X}} \bar{y}_{ij}^m = \bar{q}_{ij} && \text{for all } i\in\mathcal{O},\ j\in\mathcal{D} \\
& \bar{y}_{ij}^m \leq \bar{q}_{ij} u_m && \text{for all } i\in\mathcal{O},\ j\in\mathcal{D},\ m\in\mathcal{X} \\
& \sum_{m\in\mathcal{X}} u_m = N && \\
& \bar{y}_{ij}^m \geq 0 && \text{for all } i\in\mathcal{O},\ j\in\mathcal{D},\ m\in\mathcal{X} \\
& u_m \in \{0,1\} && \text{for all } m\in\mathcal{X}
\end{aligned}$$

5.4 Choosing the Vehicle Fleet Sizes

The final design decision required to test the CA method for selection of the number of terminals described in Section 5.2 is to choose the number of vehicles stationed at each terminal. Our approach is simple enumeration, using the operational decision procedures described in Section 6. First, for every scenario ω, we assign values to the decision variables $z_{ij}(\omega)$ that satisfy constraints (5)–(7), and such that the number of terminals that serve each origin i or destination j does not exceed the numbers N_i or N_j obtained from the thresholds Q and the flows $q(\omega)$ as described in Section 5.2. To do this, we use a simple heuristic described in Section 6.1. A lower bound on the number of vehicles required at each open terminal m is then given by

$$L_m \quad := \quad \max_{\omega\in\Omega} \left\lceil \frac{\sum_{i\in\mathcal{O}}\sum_{j\in\mathcal{D}} z_{ij}^m(\omega)q_{ij}(\omega)}{Q_v} \right\rceil .$$

We use the detailed routing cost calculations described in Section 6.2 to calculate the cost of routing $L_m, L_m + 1, L_m + 2, \ldots, L_m + k$ vehicles from terminal m for some small number k. We then choose the number of vehicles for terminal m that minimizes the sum of vehicle cost and transportation cost from terminal m over the $k+1$ fleet sizes. For each fleet size, the detailed vehicle routing cost calculations for all scenarios described in Section 6.2 take a large amount of time. Therefore, as described in Section 5.2, the CA is based on an assumption of full vehicles, unlike the more detailed choice of vehicle fleet sizes described in this section.

6 Operational Decisions

Section 5 described how to make the design decisions, namely selection of the number of terminals, location of the terminals, and choice of the vehicle fleet sizes. In the process thresholds were also selected that control how many terminals serve an origin or destination, depending on the total flow from/to the origin/destination. In this section we describe methods for making operational decisions. Specifically, Section 6.1 describes a method for selecting which terminal to use for each origin-destination flow, and Section 6.2 describes a method for routing the vehicles from each terminal to do the pickups and deliveries.

6.1 Selection of Terminal for Each Origin-Destination Flow

This section describes how to decide through which terminal to route each origin-destination flow for a given set of flows $q(\omega)$. Ideally, for given open terminals u, vehicle fleet sizes n_v, and origin-destination flows $q(\omega)$, one would like to solve the integer linear program (4)–(7). However, solving (4)–(7) exactly does not seem to be practical. In this section, we describe a heuristic for choosing the decision variables $z(\omega)$ that specify through which terminal each origin-destination flow is routed. The heuristic does not require knowledge of the vehicle fleet sizes n_v, but does require knowledge of the thresholds Q.

Let $\mathcal{N}_i(k)$ and $\mathcal{N}_j(k)$ represent the sets of terminals serving origin i and destination j respectively at the kth iteration of the algorithm. We require that $|\mathcal{N}_i(k)| \leq N_i$, $|\mathcal{N}_j(k)| \leq N_j$, and $\mathcal{N}_i(k) \subset \mathcal{U}$, $\mathcal{N}_j(k) \subset \mathcal{U}$, where $\mathcal{U} := \{m \in \mathcal{X} : u_m = 1\}$ denotes the set of open terminals. (Recall from Section 5.2 how the thresholds Q and the origin-destination flows $q(\omega)$ are used to determine the maximum numbers N_i and N_j of terminals serving origin i and destination j respectively.) For each origin $i \in \mathcal{O}$, destination $j \in \mathcal{D}$, and set $\mathcal{N} \subset \mathcal{U}$, let

$$m(i, j, \mathcal{N}) \quad \in \quad \arg\min_{m \in \mathcal{N}} \{d_{im} + d_{mj}\}$$

denote a terminal from the set $\mathcal{N}$ that minimizes the distance from the origin i through a terminal in the set $\mathcal{N}$ to the destination j.

Algorithm to Select Terminal for Each Origin-Destination Flow:

(0) Initially, every origin and destination is served by only the center terminal, so that

$$\mathcal{N}_i(0) \quad = \quad \mathcal{N}_j(0) \quad := \quad \{0\} \text{ for all } i \in \mathcal{O} \text{ and } j \in \mathcal{D}.$$

(1) For each $i \in \mathcal{O}$ and $j \in \mathcal{D}$,

    (a) If $|\mathcal{N}_i(k)| < N_i$ and $|\mathcal{N}_j(k)| < N_j$, then set

$$\mathcal{N}_i^+(k) \quad := \quad \mathcal{N}_i(k) \cup m(i, j, \mathcal{U}) \quad \text{and} \quad \mathcal{N}_j^+(k) \quad := \quad \mathcal{N}_j(k) \cup m(i, j, \mathcal{U});$$

    (b) Else if $|\mathcal{N}_i(k)| < N_i$ and $|\mathcal{N}_j(k)| = N_j$, then set

$$\mathcal{N}_i^+(k) \quad := \quad \mathcal{N}_i(k) \cup m(i, j, \mathcal{N}_j(k)) \quad \text{and} \quad \mathcal{N}_j^+(k) \quad := \quad \mathcal{N}_j(k);$$

    (c) Else if $|\mathcal{N}_i(k)| = N_i$ and $|\mathcal{N}_j(k)| < N_j$, set

$$\mathcal{N}_i^+(k) \quad := \quad \mathcal{N}_i(k) \quad \text{and} \quad \mathcal{N}_j^+(k) \quad := \quad \mathcal{N}_j(k) \cup m(i, j, \mathcal{N}_i(k));$$

    (d) Else $|\mathcal{N}_i(k)| = N_i$ and $|\mathcal{N}_j(k)| = N_j$. Set

$$\mathcal{N}_i^+(k) \quad := \quad \mathcal{N}_i(k) \quad \text{and} \quad \mathcal{N}_j^+(k) \quad := \quad \mathcal{N}_j(k).$$

(Note that $|\mathcal{N}_i^+(k)| \leq N_i$ and $|\mathcal{N}_j^+(k)| \leq N_j$ for all $i \in \mathcal{O}$ and $j \in \mathcal{D}$.)
If

$$\mathcal{N}_i^+(k) \quad = \quad \mathcal{N}_i(k) \quad \text{and} \quad \mathcal{N}_j^+(k) \quad = \quad \mathcal{N}_j(k)$$

for all $i \in \mathcal{O}$ and $j \in \mathcal{D}$, then terminate the algorithm.

(2) Choose

$$(i', j')(k) \quad \in \quad \arg\max_{i \in \mathcal{O}, j \in \mathcal{D}} q_{ij}(\omega) \left[\min_{m \in \mathcal{N}_i(k) \cap \mathcal{N}_j(k)} (d_{im} + d_{mj}) - \min_{m \in \mathcal{N}_i^+(k) \cap \mathcal{N}_j^+(k)} (d_{im} + d_{mj}) \right].$$

Set

$$\mathcal{N}_{i'}(k+1) \quad := \quad \mathcal{N}_{i'}^+(k) \quad \text{and} \quad \mathcal{N}_{j'}(k+1) \quad := \quad \mathcal{N}_{j'}^+(k).$$

For all $i \in \mathcal{O} \setminus \{i'\}$ and $j \in \mathcal{D} \setminus \{j'\}$, set

$$\mathcal{N}_i(k+1) \quad := \quad \mathcal{N}_i(k) \quad \text{and} \quad \mathcal{N}_j(k+1) \quad := \quad \mathcal{N}_j(k).$$

Return to step 1.

At the completion of the algorithm at finite iteration $k^\star$, let

$$z_{ij}^m(\omega) \quad = \quad \mathbb{I}_{\{m = m(i, j, \mathcal{N}_i(k^\star) \cap \mathcal{N}_j(k^\star))\}}.$$

6.2 Vehicle Routing with Backhauls, Split Pickups and Deliveries

In this section we describe a method for routing the vehicles from each terminal to do given pickups and deliveries. The vehicle routes also provide a more accurate estimate of the transportation cost resulting from a given network design. The method described in this section is also, as far as we know, the first heuristic proposed for the vehicle routing problem with backhauls and split pickups and deliveries, and thus may be of interest in itself.

The vehicle routing problem with backhauls and split pickups and deliveries (VRPBS) was given in (8)–(18). Recall that the problem input is a terminal indexed by 0, a set $\mathcal{O}'$ of origins, a set $\mathcal{D}'$ of destinations, quantities to be picked up and delivered given by Q', vehicle movement costs between the origins, destinations, and the considered terminal given by d', and n'_v vehicles with capacity Q_v each. Note that the origin-destination flows $q(\omega)$ and the output of the methods described in the previous sections provide the input for the vehicle routing problem, except for the fleet size n'_v, which can be determined by enumeration, as described in Section 5.4.

We describe a cluster-first, route-second heuristic for the VRPBS that uses various ideas of the heuristic for the vehicle routing problem with backhauls (VRPB) proposed by Toth and Vigo (1999). In both the VRPB and the VRPBS, deliveries have to be performed before pickups on the same route. The heuristic of Toth and Vigo (1999) has to be modified for the following reasons. First, as already pointed out, in the VRPBS multiple vehicles are allowed to visit each origin and destination, whereas in the VRPB exactly one vehicle must visit each origin and destination. One reason this modification is needed is because it often holds that $Q'_i > Q_v$ for some origins or destinations $i \in \mathcal{O}' \cup \mathcal{D}'$, and thus some origins and destinations must be visited by more than one vehicle (recall the typical nonuniform distribution of pickup and delivery quantities shown in Figure 6). Second, in the VRPB considered by Toth and Vigo (1999), each of the n'_v vehicles must visit at least one destination, and thus no vehicle may travel directly from the terminal to an origin. In the VRPBS, fewer than n'_v vehicles may visit origins or destinations, and vehicles may travel directly to origins (and thus not visit any destinations).

6.2.1 Initial Splitting Step

As pointed out, for some points $i \in \mathcal{O}' \cup \mathcal{D}'$, it may hold that $Q'_i > Q_v$. If point i has quantity $Q'_i > Q_v$, then it should be served by multiple vehicles, and all these vehicles except possibly one should carry a full load associated with this point, and the remaining quantity should be carried by another vehicle which can combine the remaining quantity with additional loads associated with other points. For example, if a destination j has $Q'_j = 3.5Q_v$, then each of 3 vehicles should deliver full vehicle loads to j and afterward move to pick-up points, and one vehicle should deliver a load of size $0.5Q_v$ to j and the rest of the vehicle's space could be used for other deliveries on its route. Thus, in the first step each point i and its quantity Q'_i is split, after which each new point i' has quantity $Q''_{i'} \leq Q_v$. Specifically, for each point $i \in \mathcal{O}' \cup \mathcal{D}'$, create $\lceil Q'_i/Q_v \rceil$ copies of point i, of which $\lfloor Q'_i/Q_v \rfloor$ new points i' have quantity $Q''_{i'} = Q_v$, and if $\lfloor Q'_i/Q_v \rfloor < \lceil Q'_i/Q_v \rceil$, then the remaining point i' has quantity $Q''_{i'} = Q'_i - \lfloor Q'_i/Q_v \rfloor Q_v$. Let $\mathcal{O}''$ denote the set of new origins, $\mathcal{D}''$ denote the set of new destinations, $\mathcal{V}'' := \{0\} \cup \mathcal{O}'' \cup \mathcal{D}''$ denote the new set of nodes, $\mathcal{A}'' := \left\{ (i,j) \in (\mathcal{V}'')^2 \setminus \mathcal{O}'' \times \mathcal{D}'' : i \neq j \right\}$ denote the new set of arcs, and Q''_i and Q''_j denote the new quantities to be picked up and delivered. For $(i,j) \in \mathcal{A}''$, the vehicle movement costs $d''_{i,j}$ are obtained from the given vehicle movement costs d in the obvious way, except if i and j correspond to the same original point, in which case $d''_{i,j} := \varepsilon$ for some chosen $\varepsilon > 0$.

Next one can define the following VRPB with input data $\mathcal{O}''$, $\mathcal{D}''$, $\mathcal{V}''$, $\mathcal{A}''$, Q'', d'', n'_v, and Q_v. The decision variables are

$$
\begin{aligned}
x_{ij} &:= \begin{cases} 1 & \text{if a vehicle travels on arc } (i,j) \\ 0 & \text{otherwise} \end{cases} \\
n_v^O &:= \text{number of vehicles performing pickups} \\
n_v^D &:= \text{number of vehicles performing deliveries}
\end{aligned}
$$

Note that the number of vehicles used is given by $\max\{n_v^O, n_v^D\}$. For each set $S \subset \mathcal{O}''$ or $S \subset \mathcal{D}''$, let $\sigma(S)$ denote the minimum number of vehicles needed to serve each point in S if exactly one vehicle must visit each

point, that is, $\sigma(S)$ is the optimal value of the bin packing problem with item sizes given by Q_i'' for $i \in S$ and bin size Q_v. For each combination of $n_v^O \in \{\lceil \sum_{i \in \mathcal{O}''} Q_i''/Q_v \rceil, \ldots, n_v'\}$ and $n_v^D \in \{\lceil \sum_{j \in \mathcal{D}''} Q_j''/Q_v \rceil, \ldots, n_v'\}$, let

$$\tau'\left(\mathcal{O}'', \mathcal{D}'', Q'', d'', n_v^O, n_v^D\right) \quad := \quad \min_x \sum_{(i,j) \in \mathcal{A}''} d_{ij}'' x_{ij} \tag{37}$$

$$\text{subject to} \quad \sum_{\{j \,:\, (j,i) \in \mathcal{A}''\}} x_{ji} \quad = \quad 1 \qquad \text{for all } i \in \mathcal{O}'' \cup \mathcal{D}'' \tag{38}$$

$$\sum_{\{j \,:\, (i,j) \in \mathcal{A}''\}} x_{ij} \quad = \quad 1 \qquad \text{for all } i \in \mathcal{O}'' \cup \mathcal{D}'' \tag{39}$$

$$\sum_{i \in \mathcal{O}''} x_{i0} \quad = \quad n_v^O \tag{40}$$

$$\sum_{i \in \mathcal{O}''} x_{0i} \quad = \quad \max\{0, n_v^O - n_v^D\} \tag{41}$$

$$\sum_{j \in \mathcal{D}''} x_{0j} \quad = \quad n_v^D \tag{42}$$

$$\sum_{j \in \mathcal{D}''} x_{j0} \quad = \quad \max\{0, n_v^D - n_v^O\} \tag{43}$$

$$\sum_{\{(i,j) \in \mathcal{A}'' \,:\, i \in S, j \notin S\}} x_{ij} \quad \geq \quad \sigma(S) \quad \text{for all } S \subset \mathcal{O}'', \ |S| \geq 2 \tag{44}$$

$$\sum_{\{(i,j) \in \mathcal{A}'' \,:\, i \notin S, j \in S\}} x_{ij} \quad \geq \quad \sigma(S) \quad \text{for all } S \subset \mathcal{D}'', \ |S| \geq 2 \tag{45}$$

$$x_{ij} \quad \in \quad \{0,1\} \qquad \text{for all } (i,j) \in \mathcal{A}'' \tag{46}$$

denote the optimal value of the VRPB given that exactly n_v^O vehicles serve the origins $\mathcal{O}''$ and exactly n_v^D vehicles serve the destinations $\mathcal{D}''$. If n_v^O or n_v^D is too small so that (38)–(46) is infeasible, then $\tau''\left(\mathcal{O}'', \mathcal{D}'', Q'', d'', n_v^O, n_v^D\right) := \infty$. Then the VRPB is given by

$$\tau''\left(\mathcal{O}'', \mathcal{D}'', Q'', d'', n_v'\right) \quad := \quad \min_{n_v^O, n_v^D} \quad \left\{c_v \max\{n_v^O, n_v^D\} + \tau'\left(\mathcal{O}'', \mathcal{D}'', Q'', d'', n_v^O, n_v^D\right)\right\}$$

$$\text{subject to} \quad n_v^O \quad \in \quad \left\{\left\lceil \sum_{i \in \mathcal{O}''} Q_i''/Q_v \right\rceil, \ldots, n_v'\right\}$$

$$n_v^D \quad \in \quad \left\{\left\lceil \sum_{j \in \mathcal{D}''} Q_j''/Q_v \right\rceil, \ldots, n_v'\right\}$$

6.2.2   Initial Clustering Step

Note that, for any $i \in \mathcal{O}''$, if $Q_i'' = Q_v$, then (38)–(46) imply that i is visited by exactly one vehicle that does not visit any other origin. Similarly, for any $j \in \mathcal{D}''$, if $Q_j'' = Q_v$, then j is visited by exactly one vehicle that does not visit any other destination. Let $\tilde{\mathcal{O}}'' := \{i \in \mathcal{O}'' \,:\, Q_i'' = Q_v\}$ and $\tilde{\mathcal{D}}'' := \{j \in \mathcal{D}'' \,:\, Q_j'' = Q_v\}$. Then for all $i \in \tilde{\mathcal{O}}''$, (38) can be replaced by

$$x_{0i} + \sum_{j \in \mathcal{D}''} x_{ji} \quad = \quad 1 \tag{47}$$

$$x_{ji} \quad = \quad 0 \qquad \text{for all } j \in \mathcal{O}'' \tag{48}$$

and for all $j \in \tilde{\mathcal{D}}''$, (38) can be replaced by

$$x_{0j} \quad = \quad 1 \tag{49}$$

$$x_{ij} \quad = \quad 0 \qquad \text{for all } i \in \mathcal{D}'' \tag{50}$$

Also, for all $i \in \tilde{\mathcal{O}}''$, (39) can be replaced by

$$x_{i0} = 1 \tag{51}$$
$$x_{ij} = 0 \qquad \text{for all } j \in \mathcal{O}'' \tag{52}$$

and for all $j \in \tilde{\mathcal{D}}''$, (39) can be replaced by

$$x_{j0} + \sum_{i \in \mathcal{O}''} x_{ji} = 1 \tag{53}$$
$$x_{ji} = 0 \qquad \text{for all } i \in \mathcal{D}'' \tag{54}$$

Note that, given (51), constraint (40) holds if and only if $\sum_{i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}''} x_{i0} = n_v^O - |\tilde{\mathcal{O}}''|$. If $\mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \neq \varnothing$, then it follows from $n_v^O \geq \left\lceil \sum_{i \in \mathcal{O}''} Q_i''/Q_v \right\rceil$ that $n_v^O - |\tilde{\mathcal{O}}''| \geq \left\lceil \sum_{i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}''} Q_i''/Q_v \right\rceil \geq 1$. Similarly, given (49), constraint (42) holds if and only if $\sum_{j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''} x_{0j} = n_v^D - |\tilde{\mathcal{D}}''|$, and if $\mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \neq \varnothing$, then it follows that $n_v^D - |\tilde{\mathcal{D}}''| \geq \left\lceil \sum_{j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''} Q_j''/Q_v \right\rceil \geq 1$. Also note that constraints (47) and (51) imply that constraint (44) holds for all $S \subset \tilde{\mathcal{O}}''$. Similarly, constraints (49) and (53) imply that constraint (45) holds for all $S \subset \tilde{\mathcal{D}}''$. Furthermore, constraints (48) and (52) imply that constraint (44) is required only for $S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$, and not for $S \subset \mathcal{O}''$ such that $S \cap \tilde{\mathcal{O}}'' \neq \varnothing$ and $S \cap (\mathcal{O}'' \setminus \tilde{\mathcal{O}}'') \neq \varnothing$. Similarly, constraints (50) and (54) imply that constraint (45) is required only for $S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$, and not for $S \subset \mathcal{D}''$ such that $S \cap \tilde{\mathcal{D}}'' \neq \varnothing$ and $S \cap (\mathcal{D}'' \setminus \tilde{\mathcal{D}}'') \neq \varnothing$.

Note that it follows from constraints (38)–(43) that $\sum_{j \in \mathcal{D}''} \sum_{i \in \mathcal{O}''} x_{ji} = \sum_{j \in \mathcal{D}''} \sum_{i \notin \mathcal{D}''} x_{ji} - \sum_{j \in \mathcal{D}''} x_{j0} = \sum_{i \notin \mathcal{D}''} \sum_{j \in \mathcal{D}''} x_{ij} - \sum_{j \in \mathcal{D}''} x_{j0} = \sum_{j \in \mathcal{D}''} x_{0j} - \sum_{j \in \mathcal{D}''} x_{j0} = n_v^D - \max\{0, n_v^D - n_v^O\} = \min\{n_v^O, n_v^D\}$. Thus we can add the following redundant constraint:

$$\sum_{j \in \mathcal{D}''} \sum_{i \in \mathcal{O}''} x_{ji} = \min\{n_v^O, n_v^D\} \tag{55}$$

Next, we add the following redundant constraints that follow from constraints (38) and (39) respectively:

$$x_{0i} + \sum_{j \in \mathcal{D}''} x_{ji} \leq 1 \qquad \text{for all } i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \tag{56}$$
$$x_{j0} + \sum_{i \in \mathcal{O}''} x_{ji} \leq 1 \qquad \text{for all } j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \tag{57}$$

Also, we add the following redundant constraints that follow from constraints (44) and (45) respectively:

$$\sum_{\{(i,j) \in \mathcal{A}'' \,:\, i \in S, j \notin S\}} x_{ij} \geq 1 \qquad \text{for all } S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}'', |S| \geq 2 \tag{58}$$
$$\sum_{\{(i,j) \in \mathcal{A}'' \,:\, i \notin S, j \in S\}} x_{ij} \geq 1 \qquad \text{for all } S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}'', |S| \geq 2 \tag{59}$$

Next we formulate a Lagrangian relaxation for problem (37)–(59). Let the multipliers associated with constraint (38) for $i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$ and associated with constraint (39) for $i \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$ be denoted by λ_i, and let the multipliers associated with constraint (44) for $S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$ and associated with constraint (45) for $S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$

be denoted by $\mu_S \geq 0$. Then the corresponding Lagrangian relaxation is as follows:

$$L\left(\mathcal{O}'', \mathcal{D}'', Q'', d'', n_v^O, n_v^D, \lambda, \mu\right) \quad := \quad \min_x \quad \sum_{(i,j)\in\mathcal{A}''} d_{ij}'' x_{ij} + \sum_{i\in\mathcal{O}''\setminus\tilde{\mathcal{O}}''} \lambda_i \left(\sum_{\{j\,:\,(j,i)\in\mathcal{A}''\}} x_{ji} - 1 \right)$$

$$+ \sum_{i\in\mathcal{D}''\setminus\tilde{\mathcal{D}}''} \lambda_i \left(\sum_{\{j\,:\,(i,j)\in\mathcal{A}''\}} x_{ij} - 1 \right)$$

$$+ \sum_{S\subset\mathcal{O}''\setminus\tilde{\mathcal{O}}''} \mu_S \left(\sigma(S) - \sum_{\{(i,j)\in\mathcal{A}''\,:\,i\in S, j\notin S\}} x_{ij} \right)$$

$$+ \sum_{S\subset\mathcal{D}''\setminus\tilde{\mathcal{D}}''} \mu_S \left(\sigma(S) - \sum_{\{(i,j)\in\mathcal{A}''\,:\,i\notin S, j\in S\}} x_{ij} \right) \tag{60}$$

subject to

$$x_{0i} + \sum_{j\in\mathcal{D}''} x_{ji} \;=\; 1 \qquad \text{for all } i \in \tilde{\mathcal{O}}'' \tag{61}$$

$$x_{0i} + \sum_{j\in\mathcal{D}''} x_{ji} \;\leq\; 1 \qquad \text{for all } i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \tag{62}$$

$$x_{ij} \;=\; 0 \qquad \text{for all } i \in \mathcal{O}'', \; j \in \tilde{\mathcal{O}}'' \tag{63}$$

$$x_{0j} \;=\; 1 \qquad \text{for all } j \in \tilde{\mathcal{D}}'' \tag{64}$$

$$x_{ij} \;=\; 0 \qquad \text{for all } i \in \mathcal{D}'', \; j \in \tilde{\mathcal{D}}'' \tag{65}$$

$$\sum_{\{i\,:\,(i,j)\in\mathcal{A}''\}} x_{ij} \;=\; 1 \qquad \text{for all } j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \tag{66}$$

$$x_{i0} \;=\; 1 \qquad \text{for all } i \in \tilde{\mathcal{O}}'' \tag{67}$$

$$x_{ij} \;=\; 0 \qquad \text{for all } i \in \tilde{\mathcal{O}}'', \; j \in \mathcal{O}'' \tag{68}$$

$$\sum_{\{j\,:\,(i,j)\in\mathcal{A}''\}} x_{ij} \;=\; 1 \qquad \text{for all } i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \tag{69}$$

$$x_{j0} + \sum_{i\in\mathcal{O}''} x_{ji} \;=\; 1 \qquad \text{for all } j \in \tilde{\mathcal{D}}'' \tag{70}$$

$$x_{j0} + \sum_{i\in\mathcal{O}''} x_{ji} \;\leq\; 1 \qquad \text{for all } j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \tag{71}$$

$$x_{ij} \;=\; 0 \qquad \text{for all } i \in \tilde{\mathcal{D}}'', \; j \in \mathcal{D}'' \tag{72}$$

$$\sum_{i\in\mathcal{O}''\setminus\tilde{\mathcal{O}}''} x_{i0} \;=\; n_v^O - \left|\tilde{\mathcal{O}}''\right| \tag{73}$$

$$\sum_{i\in\mathcal{O}''} x_{0i} \;=\; \max\{0, n_v^O - n_v^D\} \tag{74}$$

$$\sum_{j\in\mathcal{D}''\setminus\tilde{\mathcal{D}}''} x_{0j} \;=\; n_v^D - \left|\tilde{\mathcal{D}}''\right| \tag{75}$$

$$\sum_{j\in\mathcal{D}''} x_{j0} \;=\; \max\{0, n_v^D - n_v^O\} \tag{76}$$

$$\sum_{j\in\mathcal{D}''}\sum_{i\in\mathcal{O}''} x_{ji} \;=\; \min\{n_v^O, n_v^D\} \tag{77}$$

$$\sum_{\{(i,j)\in\mathcal{A}''\,:\,i\in S, j\notin S\}} x_{ij} \;\geq\; 1 \qquad \text{for all } S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}'', |S| \geq 2 \tag{78}$$

$$\sum_{\{(i,j)\in\mathcal{A}''\,:\,i\notin S, j\in S\}} x_{ij} \;\geq\; 1 \qquad \text{for all } S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}'', |S| \geq 2 \tag{79}$$

$$x_{ij} \;\in\; \{0,1\} \qquad \text{for all } (i,j) \in \mathcal{A}'' \tag{80}$$

29

For each $(i,j) \in \mathcal{A}''$, let

$$
\bar{d}''_{ij} \;\; := \;\;
\begin{cases}
d''_{ij} & \text{if} \quad i,j \in \{0\} \cup \tilde{\mathcal{O}}'' \cup \tilde{\mathcal{D}}'' \\
d''_{ij} + \lambda_j & \text{if} \quad i \in \{0\} \cup \tilde{\mathcal{O}}'' \cup \tilde{\mathcal{D}}'', j \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \\
d''_{ij} - \sum_{\{S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \, : \, j \in S\}} \mu_S & \text{if} \quad i \in \{0\} \cup \tilde{\mathcal{O}}'' \cup \tilde{\mathcal{D}}'', j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \\
d''_{ij} - \sum_{\{S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \, : \, i \in S\}} \mu_S & \text{if} \quad i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}'', j \in \{0\} \cup \tilde{\mathcal{O}}'' \\
d''_{ij} + \lambda_i & \text{if} \quad i \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'', j \in \{0\} \cup \tilde{\mathcal{O}}'' \cup \tilde{\mathcal{D}}'' \\
d''_{ij} + \lambda_j - \sum_{\{S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \, : \, i \in S, j \notin S\}} \mu_S & \text{if} \quad i,j \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \\
d''_{ij} + \lambda_i - \sum_{\{S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \, : \, i \notin S, j \in S\}} \mu_S & \text{if} \quad i,j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \\
d''_{ij} + \lambda_i + \lambda_j & \text{if} \quad i \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'', j \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}''
\end{cases}
\tag{81}
$$

Then the objective function (60) is equal to

$$
\sum_{(i,j) \in \mathcal{A}''} \bar{d}''_{ij} x_{ij} - \sum_{i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}''} \lambda_i - \sum_{i \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''} \lambda_i + \sum_{S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}''} \mu_S \sigma(S) + \sum_{S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}''} \mu_S \sigma(S)
$$

To interpret feasible solutions of the Lagrangian relaxation (60)–(80), partition the arcs $\mathcal{A}''$ into the following subsets:

$$
\mathcal{A}''_1 \;\; := \;\; \left\{ (i,0) \, : \, i \in \tilde{\mathcal{O}}'' \right\} \cup \left\{ (0,j) \, : \, j \in \tilde{\mathcal{D}}'' \right\}
$$

$$
\mathcal{A}''_2 \;\; := \;\; \left\{ (i,j) \, : \, i \in \mathcal{O}'', j \in \tilde{\mathcal{O}}'' \right\} \cup \left\{ (i,j) \, : \, i \in \tilde{\mathcal{O}}'', j \in \mathcal{O}'' \right\} \cup \left\{ (i,j) \, : \, i \in \mathcal{D}'', j \in \tilde{\mathcal{D}}'' \right\} \cup \left\{ (i,j) \, : \, i \in \tilde{\mathcal{D}}'', j \in \mathcal{D}'' \right\}
$$

$$
\mathcal{A}''_3 \;\; := \;\; \left\{ (i,j) \, : \, i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}'', j \in \{0\} \cup \mathcal{O}'' \setminus \tilde{\mathcal{O}}'' \right\}
$$

$$
\mathcal{A}''_4 \;\; := \;\; \left\{ (i,j) \, : \, i \in \{0\} \cup \mathcal{D}'' \setminus \tilde{\mathcal{D}}'', j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \right\}
$$

$$
\mathcal{A}''_5 \;\; := \;\; \left\{ (i,j) \, : \, i \in \{0\} \cup \mathcal{D}'', j \in \{0\} \cup \mathcal{O}'' \right\}
$$

First, observe that each of the constraints (61)–(79) involves decision variables x_{ij} for arcs (i,j) in only one of the subsets above. Specifically, constraints (64) and (67) involve arcs in $\mathcal{A}''_1$ only; constraints (63), (65), (68), and (72) involve arcs in $\mathcal{A}''_2$ only; constraints (69), (73), and (78) involve arcs in $\mathcal{A}''_3$ only; constraints (66), (75), and (79) involve arcs in $\mathcal{A}''_4$ only; and constraints (61), (62), (70), (71), (74), (76), and (77) involve arcs in $\mathcal{A}''_5$ only. Also, clearly each individual constraint in (80) involves an arc in only one of the subsets above. Thus, the Lagrangian relaxation (60)–(80) decomposes into 5 subproblems, corresponding to the sets of arcs and constraints identified above. Next we consider each of these 5 subproblems in turn.

For the subproblem involving $\mathcal{A}''_1$, it follows from constraints (64) and (67) that $x_{ij} = 1$ for all $(i,j) \in \mathcal{A}''_1$.

For the subproblem involving $\mathcal{A}''_2$, it follows from constraints (63), (65), (68), and (72) that $x_{ij} = 0$ for all $(i,j) \in \mathcal{A}''_2$.

For the subproblem involving $\mathcal{A}''_3$, it follows from constraint (69) that exactly one arc out of each node $i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$ must be chosen, it follows from constraint (73) that exactly $n^O_v - \left| \tilde{\mathcal{O}}'' \right|$ arcs from nodes $i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$ to the terminal node 0 must be chosen, and it follows from constraint (78) that for each subset $S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$, at least one arc out of the subset must be chosen, and thus it follows from constraints (69), (73), and (78) that the chosen arcs may not form any cycles in $\mathcal{O}'' \setminus \tilde{\mathcal{O}}''$. In other words, the chosen arcs in $\mathcal{A}''_3$ must form a spanning anti-arborescence on the nodes $\{0\} \cup \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$ with the terminal node 0 being the root node and with fixed indegree $n^O_v - |\tilde{\mathcal{O}}''|$ at node 0. Thus the subproblem involving $\mathcal{A}''_3$ is a shortest spanning anti-arborescence problem with fixed indegree $K^O := n^O_v - |\tilde{\mathcal{O}}''|$ at the terminal node 0 (K^O-SSAA), and with arc costs given by $\bar{d}''_{ij}$.

Similarly, it follows from constraints (66), (75), and (79) that the subproblem involving $\mathcal{A}''_4$ is a shortest spanning arborescence problem with fixed outdegree $K^D := n^D_v - |\tilde{\mathcal{D}}''|$ at the terminal node 0 (K^D-SSA), and with arc costs given by $\bar{d}''_{ij}$. Problems K^O-SSAA and K^D-SSA can be solved in $O(|\mathcal{O}'' \setminus \tilde{\mathcal{O}}''|^2)$ and $O(|\mathcal{D}'' \setminus \tilde{\mathcal{D}}''|^2)$ time respectively, for example with the algorithm of Gabow and Tarjan (1984).

30

Next we show that the subproblem involving $\mathcal{A}_5''$ can be represented as a network flow problem. The network flow problem has a node for each node in $\mathcal{V}''$, as well as an additional source node s and sink node t. The supply at the source node and the demand at the sink node are both equal to $\max\{n_v^O, n_v^D\}$. There is an arc (s, j) from the source node s to each node $j \in \mathcal{D}''$ with cost 0. The lower bound of the flow on each arc (s, j) is 1 if $j \in \tilde{\mathcal{D}}''$ and 0 if $j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$. The upper bound of the flow on each arc (s, j) is 1. There is an arc (j, i) from each node $j \in \mathcal{D}''$ to each node $i \in \mathcal{O}''$ with cost $\bar{d}_{ji}''$. The lower bound of the flow on each such arc (j, i) is 0, and the upper bound of the flow on each such arc (j, i) is 1. There is an arc (i, t) from each node $i \in \mathcal{O}''$ to the sink node t with cost 0. The lower bound of the flow on each arc (i, t) is 1 if $i \in \tilde{\mathcal{O}}''$ and 0 if $i \in \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$, and the upper bound of the flow on each arc (i, t) is 1. Suppose that $n_v^O > n_v^D$. Then there is an arc $(s, 0)$ with cost 0. The lower bound and the upper bound of the flow on arc $(s, 0)$ are both equal to $n_v^O - n_v^D$. There is also an arc $(0, i)$ from node 0 to each node $i \in \mathcal{O}''$ with cost $\bar{d}_{0i}''$. The lower bound of the flow on each such arc $(0, i)$ is 0, and the upper bound of the flow on each such arc $(0, i)$ is 1. If $n_v^O < n_v^D$, then there is an arc $(j, 0)$ from each node $j \in \mathcal{D}''$ to node 0 with cost $\bar{d}_{j0}''$, lower bound 0, and upper bound 1, and an arc $(0, t)$ with cost 0, and lower bound and upper bound both equal to $n_v^D - n_v^O$. It is easy to see that for every solution that satisfies constraints (61), (62), (70), (71), (74), (76), (77), and (80) (for $(i, j) \in \mathcal{A}_5''$), there is a feasible integer flow for the network flow problem described above with the same cost, and vice versa. The network flow problem can be solved in $O(\max\{n_v^O, n_v^D\}(|\mathcal{O}''| + |\mathcal{D}''|)^2)$ time, for example with a shortest augmenting path algorithm; see, for example, Ahuja et al. (1993).

Next we briefly address the following two issues. First, the number of subsets $S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$ and $S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$ in the Lagrangian objective (60) may be very large. Second, we want to find multipliers λ, μ that solve the Lagrangian dual problem

$$\max_{\lambda, \mu} \left\{ L\left(\mathcal{O}'', \mathcal{D}'', Q'', d'', n_v^O, n_v^D, \lambda, \mu\right) \ : \ \mu \geq 0 \right\}$$

We use the same subgradient optimization procedure described in Toth and Vigo (1997) to simultaneously address both issues. Briefly, instead of enumerating all subsets $S \subset \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$ and $S \subset \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$, at each major iteration the procedure identifies subtrees in the K^O-SSAA and K^D-SSA that violate constraints (44) and (45) respectively, and adds the terms corresponding to the subsets S of nodes in the violating subtrees to the Lagrangian objective (60). The identification of the violating subtrees can be done in $O(|\mathcal{O}'' \setminus \tilde{\mathcal{O}}''|)$ and $O(|\mathcal{D}'' \setminus \tilde{\mathcal{D}}''|)$ time respectively. During each major iteration, the multipliers λ and μ are updated by performing several minor iterations of a subgradient search. For additional details, we refer to Toth and Vigo (1997).

6.2.3 Second Splitting Step

When progress made by the subgradient optimization procedure in solving the Lagrangian dual problem slows down, the procedure is stopped. The last K^O-SSAA and K^D-SSA constructed are considered. If there are no subarborescences in the K^O-SSAA and K^D-SSA that violate constraints (44) and (45) respectively, then set $\tilde{\mathcal{O}}''' := \tilde{\mathcal{O}}''$, $\mathcal{O}''' \setminus \tilde{\mathcal{O}}''' := \mathcal{O}'' \setminus \tilde{\mathcal{O}}''$, $\tilde{\mathcal{D}}''' := \tilde{\mathcal{D}}''$, $\mathcal{D}''' \setminus \tilde{\mathcal{D}}''' := \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$, and continue with the assignment-routing step described in Section 6.2.5. Otherwise, a second round of node splitting is performed, as described in this section.

Suppose there are subarborescences in the K^D-SSA constructed on nodes $\{0\} \cup \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$ that violate constraint (45). The set $\mathcal{D}'' \setminus \tilde{\mathcal{D}}''$ is partitioned into $K^D := n_v^D - |\tilde{\mathcal{D}}''|$ subsets $\mathcal{D}_k''$, $k = 1, \ldots, K^D$, corresponding to the K^D subarborescences in the K^D-SSA rooted at node 0. For each $j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$ and $k \in \{1, \ldots, K^D\}$, let $\underline{d}_{jk}'' := \min\{d_{ji}'' : i \in \mathcal{D}_k''\}$ denote the distance between node j and subarborescence k. Note that if $j \in \mathcal{D}_k''$, then $\underline{d}_{jk}'' = 0$. Next the following transportation problem is solved to assign loads to subarborescences in such a way that the total load assigned to each subarborescence is less than the vehicle capacity, and such that the total

assignment cost is minimized.

$$\min_y \quad \sum_{j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''} \sum_{k=1}^{K^D} \underline{d}''_{jk} y_{jk} \tag{82}$$

$$\text{subject to} \quad \sum_{k=1}^{K^D} y_{jk} \;=\; Q''_j \qquad \text{for all } j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'' \tag{83}$$

$$\sum_{j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''} y_{jk} \;\leq\; Q_v \qquad \text{for all } k \in \{1, \dots, K^D\} \tag{84}$$

$$y_{jk} \;\geq\; 0 \qquad \text{for all } j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}'', \; k \in \{1, \dots, K^D\} \tag{85}$$

Let y^* denote an optimal solution of problem (82)–(85). The new set $\mathcal{D}'''$ of nodes and their loads $Q'''_i, i \in \mathcal{D}'''$ are determined as follows: $\tilde{\mathcal{D}}''' := \tilde{\mathcal{D}}''$ and $Q'''_i := Q_v$ for all $i \in \tilde{\mathcal{D}}'''$. Also, for each $j \in \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$ and each $k \in \{1, \dots, K^D\}$ such that $y^*_{jk} > 0$, there is a node $i \in \mathcal{D}''' \setminus \tilde{\mathcal{D}}'''$ with load size $Q'''_i := y^*_{jk}$. If there are no subarborescences in the K^D-SSA that violate constraint (45), then $\tilde{\mathcal{D}}''' := \tilde{\mathcal{D}}''$, $\mathcal{D}''' \setminus \tilde{\mathcal{D}}''' := \mathcal{D}'' \setminus \tilde{\mathcal{D}}''$, and $Q'''_i := Q''_i$ for all $i \in \mathcal{D}'''$. The sets $\tilde{\mathcal{O}}'''$, $\mathcal{O}''' \setminus \tilde{\mathcal{O}}'''$, and the loads Q'''_i for $i \in \mathcal{O}'''$, are determined similarly based on the K^O-SSAA. Note that after this second splitting, we are guaranteed to find a feasible solution of the VRPBS, as long as $n_v^O \geq \left\lceil \sum_{i \in \mathcal{O}'} Q'_i / Q_v \right\rceil$ and $n_v^D \geq \left\lceil \sum_{j \in \mathcal{D}'} Q'_j / Q_v \right\rceil$.

Let $\mathcal{V}''' := \{0\} \cup \mathcal{O}''' \cup \mathcal{D}'''$ denote the new set of nodes, and $\mathcal{A}''' := \left\{ (i,j) \in (\mathcal{V}''')^2 \setminus \mathcal{O}''' \times \mathcal{D}''' \; : \; i \neq j \right\}$ denote the new set of arcs. For $(i,j) \in \mathcal{A}'''$, the vehicle movement costs $d'''_{i,j}$ are obtained from the given vehicle movement costs d, and as before if i and j correspond to the same original point, then $d'''_{i,j} := \varepsilon$ for some chosen $\varepsilon > 0$.

6.2.4   Second Clustering Step

After the second splitting step, a second clustering step is performed. The second clustering step is the same as the initial clustering step described in Section 6.2.2, but with $\mathcal{O}'''$, $\mathcal{D}'''$, $\mathcal{V}'''$, $\mathcal{A}'''$, Q''', and d''' instead of $\mathcal{O}''$, $\mathcal{D}''$, $\mathcal{V}''$, $\mathcal{A}''$, Q'', and d'' respectively.

6.2.5   Assignment-Routing Step

Consider the K^O-SSAA and K^D-SSA constructed in the last clustering step. The set $\mathcal{D}'''$ is partitioned into n_v^D subsets as follows. First, $\tilde{\mathcal{D}}'''$ is partitioned into its singletons: for each $k = 1, \dots, |\tilde{\mathcal{D}}'''|$, choose without replacement $i \in \tilde{\mathcal{D}}'''$ and set $\mathcal{D}'''_k := \{i\}$. Second, the set $\mathcal{D}''' \setminus \tilde{\mathcal{D}}'''$ is partitioned into $K^D := n_v^D - |\tilde{\mathcal{D}}'''|$ subsets $\mathcal{D}'''_k, k = |\tilde{\mathcal{D}}'''| + 1, \dots, n_v^D$, corresponding to the K^D subarborescences in the K^D-SSA rooted at node 0. Similarly, set $\mathcal{O}'''$ is partitioned into n_v^O subsets $\mathcal{O}'''_l, l = 1, \dots, n_v^O$. For each $k \in \{1, \dots, n_v^D\}$ and $l \in \{1, \dots, n_v^O\}$, consider the traveling salesman problem (TSP) with node set $\mathcal{V}'''_{kl} := \{0\} \cup \mathcal{D}'''_k \cup \mathcal{O}'''_l$, arc set $\mathcal{A}'''_{kl} := \left\{ (i,j) \in (\mathcal{V}'''_{kl})^2 \setminus \mathcal{O}'''_l \times \mathcal{D}'''_k \; : \; i \neq j \right\}$, and arc costs $d'''_{ij}, (i,j) \in \mathcal{A}'''_{kl}$. Note that by construction the TSP has the precedence constraint that after 0, all nodes in $\mathcal{D}'''_k$ must be visited before any nodes in $\mathcal{O}'''_l$ are visited, and is thus called the traveling salesman problem with backhauls (TSPB). Let $\underline{d}'''_{kl}$ denote the optimal objective value, or an estimate of the optimal objective value, of the TSPB. Similar to Toth and Vigo (1999), we use the farthest insertion heuristic to obtain a solution of the TSPB, and set $\underline{d}'''_{kl}$ equal to the objective value of the solution produced by the heuristic. In addition, for each $k \in \{1, \dots, n_v^D\}$, consider the TSP with node set $\mathcal{V}'''_{k0} := \{0\} \cup \mathcal{D}'''_k$, arc set $\mathcal{A}'''_{k0} := \left\{ (i,j) \in (\mathcal{V}'''_{k0})^2 \; : \; i \neq j \right\}$, and arc costs $d'''_{ij}, (i,j) \in \mathcal{A}'''_{k0}$. Let $\underline{d}'''_{k0}$ denote the optimal objective value, or an estimate of the optimal objective value, of the TSP. Similarly, for each $l \in \{1, \dots, n_v^O\}$, consider the TSP with node set $\mathcal{V}'''_{0l} := \{0\} \cup \mathcal{O}'''_l$, arc set $\mathcal{A}'''_{0l} := \left\{ (i,j) \in (\mathcal{V}'''_{0l})^2 \; : \; i \neq j \right\}$, and arc costs $d'''_{ij}, (i,j) \in \mathcal{A}'''_{0l}$. Let $\underline{d}'''_{0l}$ denote the optimal objective value, or an estimate of the optimal objective value, of the TSP.

Next the following assignment problem is solved to combine subsets of $\mathcal{D}'''$ and $\mathcal{O}'''$ into vehicle routes.

$$\min_z \quad \sum_{k=1}^{n_v^D} \underline{d}_{k0}''' z_{k0} + \sum_{l=1}^{n_v^O} \underline{d}_{0l}''' z_{0l} + \sum_{k=1}^{n_v^D} \sum_{l=1}^{n_v^O} \underline{d}_{kl}''' z_{kl} \tag{86}$$

$$\text{subject to} \quad z_{k0} + \sum_{l=1}^{n_v^O} z_{kl} \;=\; 1 \qquad\qquad \text{for all } k \in \{1, \ldots, n_v^D\} \tag{87}$$

$$z_{0l} + \sum_{k=1}^{n_v^D} z_{kl} \;=\; 1 \qquad\qquad \text{for all } l \in \{1, \ldots, n_v^O\} \tag{88}$$

$$z_{k0}, z_{0l}, z_{kl} \;\in\; \{0,1\} \qquad\qquad \text{for all } k \in \{1, \ldots, n_v^D\},\; l \in \{1, \ldots, n_v^O\} \tag{89}$$

Let z^* denote an optimal solution of problem (86)–(89). If $z_{kl}^* = 1$, then the nodes in $\{0\}$, $\mathcal{D}_k'''$ and $\mathcal{O}_l'''$ are included in a vehicle route given by the TSPB discussed above. Similarly, if $z_{k0}^* = 1$, then the nodes in $\{0\}$ and $\mathcal{D}_k'''$ are included in a vehicle route, and if $z_{0l}^* = 1$, then the nodes in $\{0\}$ and $\mathcal{O}_l'''$ are included in a vehicle route, given by the TSP. The resulting routes may violate vehicle capacity constraints. The improvement step discussed in the next section attempts to modify routes to eliminate violations of capacity constraints.

6.2.6 Improvement Heuristic

After construction of the vehicle routes as described in the previous section, an improvement heuristic is applied to modify routes to eliminate violations of capacity constraints, and to reduce the costs. (Note that another reasonable option is to first obtain a feasible solution by applying the assignment-routing step in Section 6.2.5 to the feasible solution obtained in the second splitting step in Section 6.2.3, and to then use an improvement heuristic to find a solution with a better objective value.) The improvement heuristic is the same as the post-optimization procedure described in Toth and Vigo (1999). Briefly, intra-route 2-exchanges and 3-exchanges are performed in each route to reduce the cost of the route. Also, inter-route 1-exchanges and 2-exchanges are performed to reduce the amount by which capacity constraints are violated, and to reduce the costs of the routes. We refer to Toth and Vigo (1999) for additional details.

Let $\tilde{\tau}\left(\mathcal{O}', \mathcal{D}', Q', d', n_v^O, n_v^D\right)$ denote the total cost of the resulting solution of the VRPBS, given that exactly n_v^O vehicles serve the origins $\mathcal{O}'$ and exactly n_v^D vehicles serve the destinations $\mathcal{D}'$. (Note that $\mathcal{O}'''$, $\mathcal{D}'''$, $\mathcal{V}'''$, $\mathcal{A}'''$, Q''', and d''' are determined by $\mathcal{O}'$, $\mathcal{D}'$, Q', and d', and thus we may denote $\tilde{\tau}$ as a function of $\mathcal{O}'$, $\mathcal{D}'$, Q', and d'.) Then let

$$\hat{\tau}\left(\mathcal{O}', \mathcal{D}', Q', d', n_v'\right) \;:=\; \min_{n_v^O, n_v^D} \;\left\{c_v \max\{n_v^O, n_v^D\} + \tilde{\tau}\left(\mathcal{O}', \mathcal{D}', Q', d', n_v^O, n_v^D\right)\right\}$$

$$\text{subject to} \quad n_v^O \;\in\; \left\{\left\lceil \sum_{i \in \mathcal{O}'} Q_i'/Q_v \right\rceil, \ldots, n_v'\right\}$$

$$n_v^D \;\in\; \left\{\left\lceil \sum_{j \in \mathcal{D}'} Q_j'/Q_v \right\rceil, \ldots, n_v'\right\}$$

denote the total cost of the resulting solution of the VRPBS.

Recall that, as referred to in Section 5.4, the number n_v^m of vehicles for each terminal m can be chosen by solving

$$\min_{n_v^m} \quad \left\{C_v n_v^m + \sum_{\omega \in \Omega} p(\omega)\hat{\tau}\left(\mathcal{O}^m(z^m(\omega)), \mathcal{D}^m(z^m(\omega)), Q^m(z^m(\omega), \omega), d^m(z^m(\omega)), n_v^m\right)\right\}$$

$$\text{subject to} \quad n_v^m \;\in\; \{L_m, L_m + 1, L_m + 2, \ldots, L_m + k\}$$

where $z^m(\omega)$ can be chosen as described in Section 6.1, and $\mathcal{O}^m(z^m(\omega))$, $\mathcal{D}^m(z^m(\omega))$, $Q^m(z^m(\omega), \omega)$, and $d^m(z^m(\omega))$ are defined in Section 3.

33

7 Evaluation of the Proposed CA Solutions

To evaluate our solutions, we considered eight combinations of small and large datasets, with origins and destinations uniformly and nonuniformly distributed, and with origin-destination flows correlated and uncorrelated, as described in more detail below. For each set of origin and destination locations, we generated ten flow scenarios according to a specified distribution as described below, and used these ten design scenarios to make the design decisions, namely selection of the number of terminals and the thresholds, and the location of the terminals, using the method described in Section 5. We also solved the TEMC problem (24)–(28) using the same ten design scenarios, which provided a design for comparison. We selected the same vehicle fleet sizes, namely $L_m + 2$, to evaluate both the CA solution and the TEMC solution. Using the same flow distribution, we then generated flow scenarios for evaluation, independent of the design scenarios, and calculated the cost for each evaluation scenario using the approach described in Section 6 to make the operational decisions. Specifically, for each evaluation scenario we selected which terminal served each origin-destination flow and how to route the vehicles from each terminal to do the pickups and deliveries. Because the time involved in performing the detailed routing calculations described in Section 6.2 was large, it was not practical to generate a huge number of evaluation scenarios and calculate the cost associated with each one. Therefore, we only generated enough scenarios (around 35) to ensure that the sample standard deviation of the sample average cost difference was less than half the sample average cost difference. Specifically, let

$$
\begin{aligned}
\{\omega_1, \ldots, \omega_M\} &:= \quad \text{set of evaluation scenarios} \\
c_{CA}(\omega_m) &:= \quad \text{cost of the CA design for scenario } \omega_m \\
c_{TEMC}(\omega_m) &:= \quad \text{cost of the TEMC design for scenario } \omega_m.
\end{aligned}
$$

Then, the evaluation sample size M is chosen large enough so that

$$
\sqrt{\frac{1}{M}\frac{1}{(M-1)}\left[\sum_{m=1}^{M}\left(c_{CA}(\omega_m)-c_{TEMC}(\omega_m)\right)^2 - \frac{\left(\sum_{m=1}^{M}\left(c_{CA}(\omega_m)-c_{TEMC}(\omega_m)\right)\right)^2}{M}\right]}
$$
$$
< \quad \frac{1}{2}\left|\frac{\sum_{m=1}^{M}\left(c_{CA}(\omega_m)-c_{TEMC}(\omega_m)\right)}{M}\right|. \tag{90}
$$

The origin and destination locations. The industry dataset with 148 origins and 36 destinations, shown in Figure 5, provided a set of origins and destinations for evaluation, namely for the two cases (correlated and uncorrelated flows) with small datasets and nonuniform distributions of origins and destinations. To generate origin and destination locations "uniformly", we selected 1097 cities roughly uniformly distributed across the USA. For the two cases with small datasets and uniform distributions of origins and destinations, we sampled 148 of these 1097 cities without replacement to be origins, and 41 of these 1097 cities without replacement to be destinations. For the two cases with large datasets and uniform distributions of origins and destinations, the numbers of origins and destinations chosen were 243 and 105 respectively. Finally, for the two cases with large datasets and nonuniform distributions of origins and destinations, we used only the 788 of the 1097 cities that were approximately west of Colorado and east of Illinois, and then chose 248 origins and 134 destinations from the 788 cities without replacement. The reason for having more origins than destinations was to mimic the observed situations in the motivating industrial datasets.

Flow generation. Next we describe how we generated flows between origins and destinations for each of the eight cases. We first decided which origin-destination pairs would have positive flow. Because long distance positive flows tend to be less frequent than positive flows between origin-destination pairs located closer together, we selected each origin-destination pair with the origin located approximately west of Colorado (east of Illinois) and with the destination located approximately east of Illinois (west of Colorado) with probability 0.15 to have

Uncorrelated Demands

| Origins and Destinations | | Number of Terminals | | Cost | | | |
Total number	Location Distribution	TEMC	CA	TEMC	CA	STD	% Decrease
Small	Uniform	2	1	644,489	631,008	5150	2.09%
	Nonuniform (Industry dataset)	3	1	382,792	353,982	5688	7.53%
Large	Uniform	2	3	873,348	858,519	5288	1.70%
	Nonuniform	3	4	1,018,598	990,303	8890	2.78%

Correlated Demands

| Origins and Destinations | | Number of Terminals | | Cost | | | |
Total number	Location Distribution	TEMC	CA	TEMC	CA	STD	% Decrease
Small	Uniform	2	1	688,519	673,474	5124	2.19%
	Nonuniform (Industry dataset)	3	1	616,531	589,997	5333	4.30%
Large	Uniform	3	1	675,098	599,387	5383	11.21%
	Nonuniform	3	4	989,678	965,099	9691	2.48%

Table 2: Numbers of terminals chosen and resulting costs.

zero flow, independently for all origin-destination pairs. In addition, for the four large cases, each remaining origin-destination pair, including the above mentioned "cross-country" pairs as well as pairs closer to each other, was assigned zero flow with probability 0.5, again independently for all origin-destination pairs. The remaining origin-destination pairs were assigned positive flows as described below. The expected number of positive flows was thus

$$\frac{1}{2}\mathbb{I}(\text{``large'' case}) \times \left( \begin{array}{l} \text{total number of origin-destination pairs} \\ -0.15 \times \text{ the number of ``cross-country'' origin-destination pairs} \end{array} \right),$$

which yielded approximately 5000 positive flows for the small cases and 10,000 positive flows for the large cases.

For the origin-destination pairs having positive flow, we generated the size of the flow according to a uniform distribution between 0 and an origin-specific upper bound that follows approximately an exponential distribution among origins. Recall from Figure 6 that the industrial dataset shows that a small fraction of the origin-destination pairs accounts for a large fraction of the total flow. Similarly, a small fraction of the origins, mostly the origins of bulky goods, accounts for a large fraction of the total flow. To mimic this skewed flow distribution, we generated the flow upper bounds of the origins as follows. We first generated a random permutation of the origins to label the origins from 1 to $|\mathcal{O}|$. Then the flow from each origin i to any destination with positive origin-destination flow was distributed uniformly on $[0, 1500 \exp(-\mu(i-1))]$, for $i \in \{1, \ldots, |\mathcal{O}|\}$. Here, μ was chosen as 0.06 for the small uniform cases, 0.055 for the small nonuniform cases, 0.045 for the large uniform cases, and 0.043 for the large nonuniform cases.

Finally, for the four cases with uncorrelated flows, we generated independent random variables with uniform distributions as described above for every origin-destination pair with positive flow. For the four cases with correlated flows, we first generated independent uniform random variables as described above, and then multiplied all these independently generated flows by a single realization of a uniform $(0.7, 1.3)$ random variable. It was of interest to also test our approach with correlated flow distributions because in practice flows are often correlated. For example, for home improvement retailers, shipments between origins and destinations tend to be relatively large during spring and summer.

Results. Table 2 displays the number of terminals chosen in both the CA and TEMC solutions, the average cost of both solutions, the sample standard deviation of the sample average cost difference given in (90), and

the percentage cost improvement of the CA solution over the TEMC solution. All the cases investigated showed cost improvement of at least 1%, with one case showing a cost improvement of more than 10%. Observe that the cases with the largest percentage improvements (4.30%, 7.53%, and 11.21%) are also the cases in which the number of terminals chosen by the CA vs. the TEMC solution differs by more than 1.

Acknowledgements

References

R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

S. Anily. The vehicle-routing problem with delivery and back-haul options. *Naval Research Logistics Quarterly*, 43:415–434, 1996.

C. Archetti, M. G. Speranza, and A. Hertz. A tabu search algorithm for the split delivery vehicle routing problem. *Transportation Science*, 40(1):64–73, 2006.

J. Beardwood, J. H. Halton, and J. M. Hammersley. The shortest path through many points. *Proceedings of the Cambridge Philosophical Society*, 55:299–327, 1959.

M. Beckmann. A continuous model of transportation. *Econometrica*, 20(4):643–660, 1952.

D. E. Blumenfeld, L. D. Burns, C. F. Daganzo, M. C. Frick, and R. W. Hall. Reducing logistics costs at general motors. *Interfaces*, 17(1):26–47, 1987.

D. E. Blumenfeld, L. D. Burns, J. D. Diltz, and C. F. Daganzo. Analyzing trade-offs between transportation, inventory and production costs on freight networks. *Transportation Research*, 19B(5):361–380, 1985.

L. D. Burns, R. W. Hall, D. E. Blumenfeld, and C. F. Daganzo. Distribution strategies that minimize transportation and inventory costs. *Operations Research*, 33(3):469–490, 1985.

J. F. Campbell. Designing logistics systems by analyzing transportation, inventory and terminal cost tradeoffs. *Journal of Business Logistics*, 11(1):159–179, 1990a.

J. F. Campbell. Freight consolidation and routing with transportation economies of scale. *Transportation Research*, 24B:345–361, 1990b.

J. F. Campbell. Continuous and discrete demand hub location problems. *Transportation Research*, 27B(6): 473–482, 1993a.

J. F. Campbell. One-to-many distribution with transshipments: An analytic model. *Transportation Science*, 27: 330–340, 1993b.

C. F. Daganzo. An approximate analytic model of many-to-many demand responsive transportation systems. *Transportation Research*, 12B:325–333, 1978.

C. F. Daganzo. The distance traveled to visit N points with a maximum of C stops per vehicle: An analytic model and an application. *Transportation Science*, 18:331–350, 1984a.

C. F. Daganzo. The length of tours in zones of different shapes. *Transportation Research*, 18B(2):135–145, 1984b.

C. F. Daganzo. Supplying a single location from heterogeneous sources. *Transportation Research*, 19B(5): 409–419, 1985.

C. F. Daganzo. The break-bulk role of terminals in many-to-many logistic networks. *Operations Research*, 35: 543–555, 1987.

C. F. Daganzo. A comparison of in-vehicle and out-of-vehicle freight consolidation strategies. *Transportation Research*, 22B:173–180, 1988a.

C. F. Daganzo. Shipment composition enhancement at a consolidation center. *Transportation Research*, 22B: 103–124, 1988b.

C. F. Daganzo and R. Hall. A routing model for pickups and deliveries: No capacity restrictions on the secondary items. *Transportation Science*, 27:315–329, 1993.

C. F. Daganzo and G. F. Newell. Configuration of physical distribution networks. *Networks*, 16:113–132, 1986.

A. Dasci and G. Laporte. A continuous model for multistore competitive location. *Operations Research*, 53(2): 263–280, 2005.

A. Dasci and V. Verter. A continuous model for production-distribution system design. *European Journal of Operational Research*, 129(2):287–298, 2001.

M. Dror, G. Laporte, and P. Trudeau. Vehicle routing with split deliveries. *Discrete Applied Mathematics*, 50: 239–254, 1994.

M. Dror and P. Trudeau. Savings by split delivery routing. *Transportation Science*, 23(2):141–145, 1989.

H. N. Gabow and R. E. Tarjan. Efficient algorithms for a family of matroid intersection problems. *Journal of Algorithms*, 5:80–131, 1984.

G. Ghiani, G. Laporte, and R. Musmanno. *Introduction to Logistics Systems Planning and Control*. John Wiley & Sons, Ltd., ChiChester, England, 2004.

M. Goetschalckx and C. Jacobs-Blecha. The vehicle routing problem with backhauls. *European Journal of Operational Research*, 42(1):39–51, 1989.

R. W. Hall. Travel distance through consolidation terminals on a rectangular grid. *Journal of the Operational Research Society*, 35:1067–1078, 1984.

R. W. Hall. Comparison of strategies for routing shipments through transportation terminals. *Transportation Research*, 21A:421–429, 1987.

R. W. Hall. Characteristics of multi-stop/multi-terminal delivery routes, with backhauls and unique items. *Transportation Research*, 25B:391–403, 1991.

R. W. Hall. Design for local area freight networks. *Transportation Research*, 27B:79–95, 1993.

L. Kantorovitch. On the translocation of masses. *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, 37(7–8):199–201, 1942. reprinted in *Management Science*, **5(1)**, pp. 1–4, 1958.

A. Langevin, P. Mbaraga, and J. F. Campbell. Continuous approximation models in freight distribution: An overview. *Transportation Research*, 30B(3):163–188, 1996.

G. Laporte. Location-routing problems. In B. L. Golden and A. A. Assad, editors, *Vehicle Routing: Methods and Studies*, chapter 9, pages 163–197. North-Holland, Amsterdam, Netherlands, 1988.

G. Laporte, F. Louveaux, and H. Mercure. Models and exact solutions for a class of stochastic location-routing problems. *European Journal of Operational Research*, 39(1):71–78, 1989.

C.-G. Lee, M. A. Epelman, C. C. White, and Y. A. Bozer. A shortest path approach to the multiple-vehicle routing problem with split pick-ups. *Computers and Operations Research*, 40(4):265–284, 2006.

A. Mingozzi, S. Giorgi, and R. Baldacci. An exact method for the vehicle routing problem with backhauls. *Transportation Science*, 33(3):315–329, 1999.

P. A. Mullaseril, M. Dror, and J. Leung. Split-delivery routing heuristics in livestock feed distribution. *The Journal of the Operational Research Society*, 48(2):107–116, 1997.

G. F. Newell. Scheduling, location, transportation, and continuum mechanics; some simple approximations to optimization problems. *SIAM Journal on Applied Mathematics*, 25(3):346–360, 1973.

I. H. Osman and N. A. Wassan. A reactive tabu search meta-heuristic for the vehicle routing problem with backhauls. *Journal of Scheduling*, 5(4):263–285, 2002.

Y. Ouyang and C. F. Daganzo. Discretization and validation of the continuum approximation scheme for terminal system design. *Transportation Science*, 40(1):89–98, 2006.

J.-Y. Potvin, C. Duhamel, and F. Guertin. A genetic algorithm for vehicle routing problem with backhauling. *Applied Intelligence*, 6:345–355, 1996.

S. R. Thangiah, J.-Y. Potvin, and T. Sun. Heuristic approaches to vehicle routing with backhauls and time windows. *Computers and Operations Research*, 23(11):1043–1057, 1996.

P. Toth and D. Vigo. An exact algorithm for the vehicle routing problem with backhauls. *Transportation Science*, 31:372–385, 1997.

P. Toth and D. Vigo. A heuristic algorithm for the symmetric and asymmetric vehicle routing problems with backhauls. *European Journal of Operational Research*, 113:528–543, 1999.

C. J. Vidal and M. Goetschalckx. Strategic production-distribution models: A critical review with emphasis on global supply chain models. *European Journal of Operational Research*, 98(1):1–18, 1997.

8   Appendix

For a given origin i and destination j, we show how to calculate the probability

$$\mathbb{P}[\lambda_{i,1,j} < \lambda_{i,0,j}] \tag{91}$$

and the conditional probability

$$\mathbb{P}\left[\lambda_{i,1,j} > \alpha \,|\, \lambda_{i,1,j} < \lambda_{i,0,j}\right] \text{ for } \alpha \in (0,\, \lambda_{i,0,j}), \tag{92}$$

which is the key to completing the calculation of the expected linehaul distance $\mathbb{E}[\Lambda_{i,j}]$ in (35).

Recall that the $N-1$ terminals besides the center terminal are independent and identically distributed in a region. To facilitate the calculation of $\mathbb{E}[\Lambda_{i,j}]$, we use the uniform distribution on a rectangular region $[-a,a] \times [-b,b] \subset \mathbb{R}^2$, with $a > 0$ and $b > 0$. Specifically, terminal $n = 1, 2, \ldots, N-1$ has coordinates $U_n \in \mathbb{R}^2$, where U_n is uniformly distributed in $[-a,a] \times [-b,b]$, and $U_1, \ldots, U_{N-1}$ are independent. The center terminal is located at $(0,0)$.

We use the L_1-norm

$$\|(x, y)\|_1 \quad := \quad |x| + |y|$$

and the associated metric, as in Hall (1984), Daganzo (1987), Hall (1987), Campbell (1990b), and Campbell (1993a). The L_1-norm is used for simplicity; for example, calculations using the L_1-norm are much simpler than when using the L_2-norm $\|(x, y)\|_2 := \sqrt{x^2 + y^2}$. Furthermore, the ratios between the L_1 and L_2-metrics between pairs of points do not vary much. To illustrate this, we show the results of the following simple experiment. We generate 10,000 pairs of points independently and uniformly distributed in the unit square, $(x_1(i), y_1(i)), (x_2(i), y_2(i))), i \in \{1, \ldots, 10,000\}$, and calculate the L_1 and L_2-metrics between every pair of points. Figure 7 shows the cumulative distribution function of the ratio $\|(x_2(i), y_2(i)) - (x_1(i), y_1(i))\|_2 / \|(x_2(i), y_2(i)) - (x_1(i), y_1(i))\|_1$. Note that the ratio varies between $1/\sqrt{2}$ and 1, and that 0.8 is a pretty good overall approximation for the ratio.
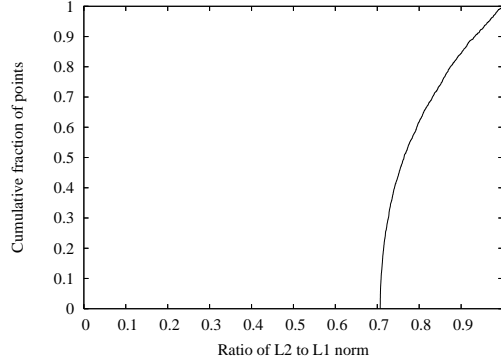

Figure 7: Cumulative distribution function of ratio $(\|x_2, y_2\|_1)^{-1} \|x_1, y_1\|_2$.

Let (x_1, y_1) denote the coordinates of origin i, and let (x_2, y_2) denote the coordinates of destination j. Then, for any terminal n,

$$\lambda_{i,n,j} \quad := \quad \|U_n - (x_1, y_1)\|_1 + \|(x_2, y_2) - U_n\|_1$$

The triangle inequality holds for any terminal n, i.e.,

$$\lambda_{i,n,j} \quad \geq \quad \|(x_2, y_2) - (x_1, y_1)\|_1 \quad = \quad |x_2 - x_1| + |y_2 - y_1|$$

and $\lambda_{i,n,j} = |x_2 - x_1| + |y_2 - y_1|$ when terminal n is located within the rectangle with sides parallel with the coordinate axes and having opposite corners (x_1, y_1) and (x_2, y_2). Note that travel through the center terminal results in a total distance

$$\lambda_{i,0,j} \quad = \quad |x_1| + |x_2| + |y_1| + |y_2|.$$

Thus, the random variable $\Lambda_{i,j}$ defined in (32) satisfies

$$|x_2 - x_1| + |y_2 - y_1| \quad \leq \quad \Lambda_{i,j} \quad \leq \quad |x_1| + |x_2| + |y_1| + |y_2|.$$

To calculate the probabilities in (91) and (92), we consider the following three cases separately:

1. Origin i and destination j are located in opposite quadrants.

2. Origin i and destination j are located in the same quadrant.

3. Origin i and destination j are located in adjacent quadrants.

If origin i and destination j are located in opposite quadrants, then

$$|x_2 - x_1| + |y_2 - y_1| \quad = \quad |x_1| + |x_2| + |y_1| + |y_2|$$

and thus

$$\Lambda_{i,j} \;=\; \lambda_{i,0,j} \;=\; |x_1| + |x_2| + |y_1| + |y_2|$$

for all selections of $\mathcal{N}_i$ and $\mathcal{N}_j$. Thus $\mathbb{E}[\Lambda_{i,j}] = |x_1| + |x_2| + |y_1| + |y_2|$.

It remains to consider the cases in which origin i and destination j are not located in opposite quadrants. Note that since the distribution of $\Lambda_{i,j}$ has support on $[|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|]$, it follows that the integral in the expression for $E[\Lambda_{i,j}]$ in (35) can be written as follows:

$$\int_0^{\lambda_{i,0,j}} \mathbb{P}\left[\lambda_{i,1,j} > \alpha \,\Big|\, \lambda_{i,1,j} < \lambda_{i,0,j}\right]^{\ell} d\alpha$$

$$= \quad |x_2 - x_1| + |y_2 - y_1| + \int_{|x_2-x_1|+|y_2-y_1|}^{|x_1|+|x_2|+|y_1|+|y_2|} \mathbb{P}\left[\lambda_{i,1,j} > \alpha \,\Big|\, \lambda_{i,1,j} < \lambda_{i,0,j}\right]^{\ell} d\alpha \tag{93}$$

Note that, because terminals $1, 2, \ldots, N-1$ are uniformly distributed in the rectangle $[-a, a] \times [-b, b]$, the probabilities in (91) and (92) can be expressed as

$$\mathbb{P}\left[\lambda_{i,1,j} < \lambda_{i,0,j}\right]$$
$$= \quad \frac{\text{Area}\left(\{(x,y) \in [-a,a] \times [-b,b] \;:\; \|(x,y) - (x_1,y_1)\|_1 + \|(x_2,y_2) - (x,y)\|_1 < |x_1| + |x_2| + |y_1| + |y_2|\}\right)}{4ab} \tag{94}$$

and

$$\mathbb{P}\left[\lambda_{i,1,j} > \alpha \,\Big|\, \lambda_{i,1,j} < \lambda_{i,0,j}\right]$$
$$= \quad \frac{\text{Area}\left(\{(x,y) \in [-a,a] \times [-b,b] \;:\; \alpha < \|(x,y) - (x_1,y_1)\|_1 + \|(x_2,y_2) - (x,y)\|_1 < |x_1| + |x_2| + |y_1| + |y_2|\}\right)}{\text{Area}\left(\{(x,y) \in [-a,a] \times [-b,b] \;:\; \|(x,y) - (x_1,y_1)\|_1 + \|(x_2,y_2) - (x,y)\|_1 < |x_1| + |x_2| + |y_1| + |y_2|\}\right)} \tag{95}$$

for $\alpha \in (0, \lambda_{i,0,j})$. The areas in the right side of (94) and (95) are straightforward to calculate. Figures 8 and 9 show the regions, with origin i and destination j located in the same quadrant and in adjacent quadrants respectively, in which location of terminal n provides a linehaul distance $\lambda_{i,n,j}$ that satisfies $\lambda_{i,n,j} \le \alpha$ for $\alpha \in (|x_2-x_1|+|y_2-y_1|, |x_1|+|x_2|+|y_1|+|y_2|)$. In particular, if terminal n is located on the boundary of the eight-sided region, then $\lambda_{i,n,j} = \alpha$, and if terminal n is located in the interior of the eight-sided region, then $\lambda_{i,n,j} < \alpha$. Setting $\alpha = \lambda_{i,0,j} = |x_1| + |x_2| + |y_1| + |y_2|$ gives the region in which location of a terminal provides a smaller linehaul distance than through the center terminal. We show that $\mathbb{P}[\lambda_{i,1,j} > \alpha \,|\, \lambda_{i,1,j} < \lambda_{i,0,j}]$ is a piecewise polynomial in α of degree at most two, from which the calculation of $\int_0^{\lambda_{i,0,j}} \mathbb{P}[\lambda_{i,1,j} > \alpha \,|\, \lambda_{i,1,j} < \lambda_{i,0,j}]^{\ell} d\alpha$ follows in a straightforward manner (although it is numerically dangerous due to catastrophic cancellation).

The remainder of the appendix is organized as follows. We provide the detailed calculations for the case in which the rectangle $[-a, a] \times [-b, b]$ contains (x_1, y_1) and (x_2, y_2). The calculations for the case in which the rectangle does not contain (x_1, y_1) and (x_2, y_2) are similar, but involve more subcases. Section 8.1 shows the calculations for the case in which the origin and destination are located in the same quadrant, and Section 8.2 for the case in which they are located in adjacent quadrants.

8.1 Origin and Destination Located in the Same Quadrant

First, suppose that $0 \leq x_1 \leq x_2$ and $0 \leq y_1 \leq y_2$. Then i and j are both in quadrant I. For $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$, let

$$R_1^S(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ -a \leq x \leq a \text{ and } -b \leq y \leq \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2}\right\}\right)$$

$$R_2^S(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} -a \leq x \leq \frac{-\alpha + x_1 + x_2 - y_1 + y_2}{2} \\ \text{and } \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2} \leq y \leq \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \end{array}\right\}\right)$$

$$R_3^S(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ -a \leq x \leq a \text{ and } \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \leq y \leq b\right\}\right)$$

$$R_4^S(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} \frac{\alpha + x_1 + x_2 + y_1 - y_2}{2} \leq x \leq a \\ \text{and } \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2} \leq y \leq \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \end{array}\right\}\right)$$

$$R_5^S(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} 2(x - y) \geq \alpha + x_1 + x_2 - y_1 - y_2 \text{ and } x_2 \leq x \leq \frac{\alpha + x_1 + x_2 + y_1 - y_2}{2} \\ \text{and } \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2} \leq y \leq y_1 \end{array}\right\}\right)$$

$$R_6^S(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} -2(x + y) \geq \alpha - x_1 - x_2 - y_1 - y_2 \text{ and } \frac{-\alpha + x_1 + x_2 - y_1 + y_2}{2} \leq x \leq x_1 \\ \text{and } \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2} \leq y \leq y_1 \end{array}\right\}\right)$$

$$R_7^S(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} 2(y - x) \geq \alpha - x_1 - x_2 + y_1 + y_2 \text{ and } \frac{-\alpha + x_1 + x_2 - y_1 + y_2}{2} \leq x \leq x_1 \\ \text{and } y_2 \leq y \leq \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \end{array}\right\}\right)$$

$$R_8^S(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} 2(x + y) \geq \alpha + x_1 + x_2 + y_1 + y_2 \text{ and } x_2 \leq x \leq \frac{\alpha + x_1 + x_2 + y_1 - y_2}{2} \\ \text{and } y_2 \leq y \leq \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \end{array}\right\}\right).$$

Figure 8 shows the regions defining $R_1^S(\alpha), \ldots, R_8^S(\alpha)$ for a value of $\alpha \in (|x_2-x_1|+|y_2-y_1|, |x_1|+|x_2|+|y_1|+|y_2|)$. The lines 1–8 in Figure 8 form the boundaries of the regions defining $R_1^S(\alpha), \ldots, R_8^S(\alpha)$, and it will be convenient to refer to the line numbers in the figure.

For the special case with $\alpha = \lambda_{i,0,j} = |x_1| + |x_2| + |y_1| + |y_2|)$, let

$$R_k^S \quad := \quad R_k^S(\lambda_{i,0,j}).$$

Then, the probabilities in (94) and (95) are given by

$$P[\lambda_{i,1,j} < \lambda_{i,0,j}] \quad = \quad \frac{4ab - \sum_{k=1}^{8} R_k^S}{4ab}$$

and

$$P[\lambda_{i,1,j} > \alpha \mid \lambda_{i,1,j} < \lambda_{i,0,j}] \quad = \quad \frac{\sum_{k=1}^{8} R_k^S(\alpha) - \sum_{k=1}^{8} R_k^S}{4ab - \sum_{k=1}^{8} R_k^S} \tag{96}$$

Next, we allow i and j to be in any quadrant (as long as i and j are in the same quadrant), and we allow $x_1 > x_2$ or $y_1 > y_2$. Using Figure 8, it is easy to observe that regardless of the quadrant in which i and j are, the areas $R_1^S(\alpha), \ldots, R_8^S(\alpha)$ depend only on the minimum and maximum of the distances of i and j from the x and y axes. Hence, let

$$\underline{x} := |x_1| \wedge |x_2|, \quad \overline{x} := |x_1| \vee |x_2|, \quad \underline{y} := |y_1| \wedge |y_2|, \quad \text{and} \quad \overline{y} := |y_1| \vee |y_2| \tag{97}$$
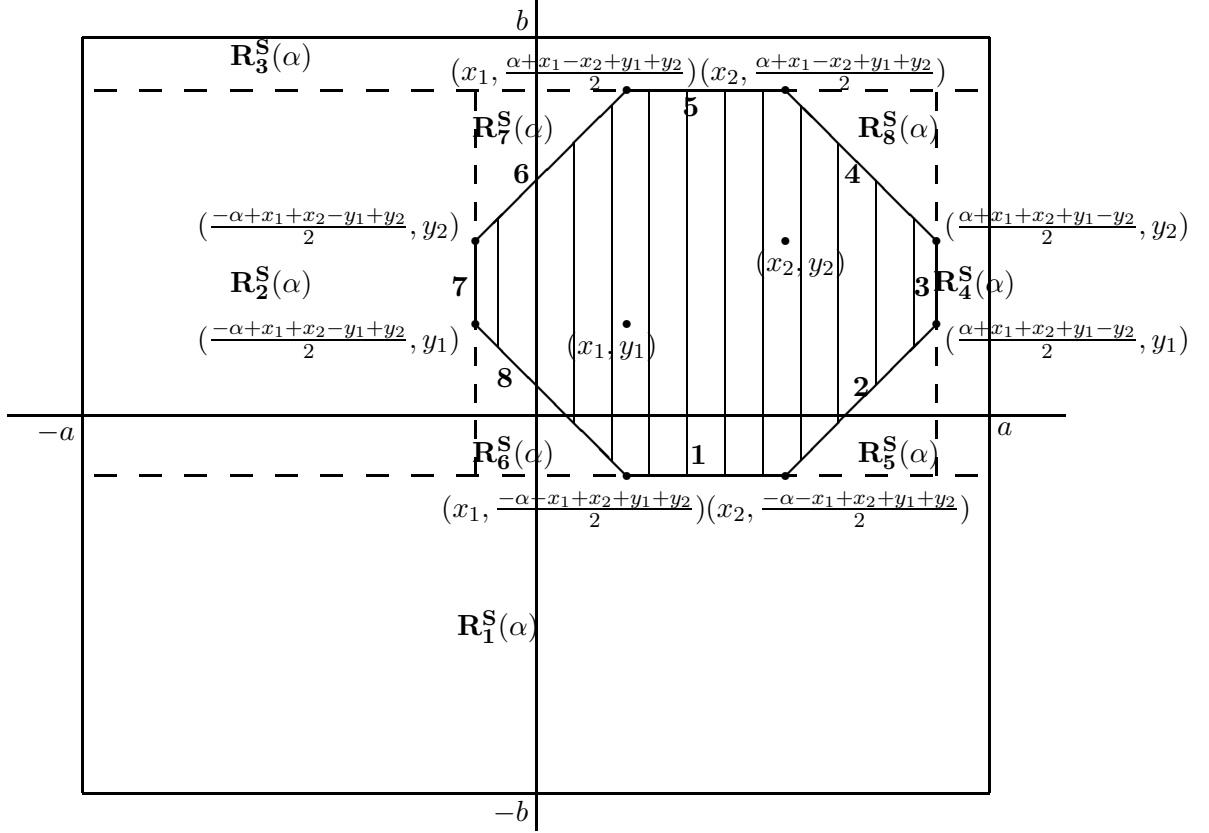
Figure 8: Example of the region in which location of a terminal provides a smaller linehaul distance than $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$ if (x_1, y_1) and (x_2, y_2) are in the same quadrant.

Then, in general, with i and j in the same quadrant, it follows that

$$R_1^S(\alpha) = 2a \left(\frac{-\alpha - \underline{x} + \overline{x} + \underline{y} + \overline{y}}{2} + b \right) \mathbb{I}_{\{\alpha < 2b - \underline{x} + \overline{x} + \underline{y} + \overline{y}\}}$$

$$R_2^S(\alpha) = \left(\frac{-\alpha + \underline{x} + \overline{x} - \underline{y} + \overline{y}}{2} + a \right) \left(\left(b \wedge \frac{\alpha + \underline{x} - \overline{x} + \underline{y} + \overline{y}}{2} \right) - \left((-b) \vee \frac{-\alpha - \underline{x} + \overline{x} + \underline{y} + \overline{y}}{2} \right) \right)$$
$$\times \mathbb{I}_{\{\alpha < 2a + \underline{x} + \overline{x} - \underline{y} + \overline{y}\}}$$

$$R_3^S(\alpha) = 2a \left(b - \frac{\alpha + \underline{x} - \overline{x} + \underline{y} + \overline{y}}{2} \right) \mathbb{I}_{\{\alpha < 2b - \underline{x} + \overline{x} - \underline{y} - \overline{y}\}}$$

$$R_4^S(\alpha) = \left(a - \frac{\alpha + \underline{x} + \overline{x} + \underline{y} - \overline{y}}{2} \right) \left(\left(b \wedge \frac{\alpha + \underline{x} - \overline{x} + \underline{y} + \overline{y}}{2} \right) - \left((-b) \vee \frac{-\alpha - \underline{x} + \overline{x} + \underline{y} + \overline{y}}{2} \right) \right)$$
$$\times \mathbb{I}_{\{\alpha < 2a - \underline{x} - \overline{x} - \underline{y} + \overline{y}\}}$$

$$R_5^S(\alpha) = \frac{1}{2} \left(\left(a \wedge \frac{\alpha + \underline{x} + \overline{x} + \underline{y} - \overline{y}}{2} \right) - \left(\overline{x} \vee \left(\frac{\alpha + \underline{x} + \overline{x} - \underline{y} - \overline{y}}{2} - b \right) \right) \right)^2$$
$$\times \mathbb{I}_{\{\alpha < 2(a+b) - \underline{x} - \overline{x} + \underline{y} + \overline{y}\}}$$

$$R_6^S(\alpha) = \frac{1}{2} \left(\left(\underline{x} \wedge \left(b + \frac{-\alpha + \underline{x} + \overline{x} + \underline{y} + \overline{y}}{2} \right) \right) - \left((-a) \vee \frac{-\alpha + \underline{x} + \overline{x} - \underline{y} + \overline{y}}{2} \right) \right)^2$$

$$R_7^S(\alpha) = \frac{1}{2} \left(\left(x_1 \wedge \left(b - \frac{\alpha - \underline{x} - \overline{x} + \underline{y} + \overline{y}}{2} \right) \right) - \left((-a) \vee \frac{-\alpha + \underline{x} + \overline{x} - \underline{y} + \overline{y}}{2} \right) \right)^2$$
$$\times \mathbb{I}_{\{\alpha < 2(a+b) + \underline{x} + \overline{x} - \underline{y} - \overline{y}\}}$$

$$R_8^S(\alpha) = \frac{1}{2} \left(\left(a \wedge \frac{\alpha + \underline{x} + \overline{x} + \underline{y} - \overline{y}}{2} \right) - \left(\overline{x} \vee \left(\frac{\alpha + \underline{x} + \overline{x} + \underline{y} + \overline{y}}{2} - b \right) \right) \right)^2$$
$$\times \mathbb{I}_{\{\alpha < 2(a+b) - \underline{x} - \overline{x} - \underline{y} - \overline{y}\}}.$$

42

Note that in these area calculations one must keep in mind that part of the boundary of the eight-sided region in Figure 8 may be outside the rectangle $[-a, a] \times [-b, b]$. Thus, depending on the boundary of the eight-sided region, some of the areas $R_1^S(\alpha), \ldots, R_8^S(\alpha)$ may be 0. In the expressions for $R_1^S(\alpha), \ldots, R_8^S(\alpha)$ above, this is accomplished with the indicator functions (except for $R_6^S(\alpha)$, because $a > 0$, $b > 0$, and $\underline{x} + \underline{y} > 0$ imply $R_6^S(\alpha) > 0$).

To finish the expected linehaul distance calculations, it follows from (93) and (96) that we must calculate

$$
\int_{|x_2 - x_1| + |y_2 - y_1|}^{|x_1| + |x_2| + |y_1| + |y_2|} P\left[\lambda_{i,1,j} > \alpha \mid \lambda_{i,1,j} < \lambda_{i,0,j}\right]^\ell \, d\alpha
$$

$$
= \int_{|x_2 - x_1| + |y_2 - y_1|}^{|x_1| + |x_2| + |y_1| + |y_2|} \left[\frac{\sum_{k=1}^8 R_k^S(\alpha) - \sum_{k=1}^8 R_k^S}{4ab - \sum_{k=1}^8 R_k^S}\right]^\ell \, d\alpha. \tag{98}
$$

To calculate (98), we write $\sum_{k=1}^8 R_k^S(\alpha)$ as a piecewise polynomial in α of degree at most two. We will choose $A(\alpha), B(\alpha), C_1(\alpha)$, and C_2 such that

$$
A(\alpha)\alpha^2 + B(\alpha)\alpha + C_1(\alpha) \quad = \quad \sum_{k=1}^8 R_k^S(\alpha) \quad \text{and} \quad C_2 \quad = \quad \sum_{k=1}^8 R_k^S.
$$

Considering the interval over which each indicator function in the expressions for $R_1^S(\alpha), \ldots, R_8^S(\alpha)$, is 0, the interval $[|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|]$ can be partitioned into intervals over which $A(\alpha)$, $B(\alpha)$, and $C_1(\alpha)$ are constant. Also, for $C(\alpha) := C_1(\alpha) - C_2$,

$$
\left[\sum_{k=1}^8 R_k^S(\alpha) - \sum_{k=1}^8 R_k^S\right]^\ell \quad = \quad \left[A(\alpha)\alpha^2 + B(\alpha)\alpha + C(\alpha)\right]^\ell
$$

$$
= \quad \sum_{\left\{\substack{\ell_1, \ell_2, \ell_3 \, : \\ \ell_1 + \ell_2 + \ell_3 = \ell}\right\}} \frac{\ell!}{\ell_1! \ell_2! \ell_3!} \left[A(\alpha)\alpha^2\right]^{\ell_1} \left[B(\alpha)\alpha\right]^{\ell_2} \left[C(\alpha)\right]^{\ell_3}
$$

$$
= \quad \sum_{\left\{\substack{\ell_1, \ell_2, \ell_3 \, : \\ \ell_1 + \ell_2 + \ell_3 = \ell}\right\}} \frac{\ell!}{\ell_1! \ell_2! \ell_3!} A(\alpha)^{\ell_1} B(\alpha)^{\ell_2} C(\alpha)^{\ell_3} \alpha^{2\ell_1 + \ell_2},
$$

and thus it follows from (98) that

$$
\int_{|x_2 - x_1| + |y_2 - y_1|}^{|x_1| + |x_2| + |y_1| + |y_2|} P\left[\lambda_{i,1,j} > \alpha \mid \lambda_{i,1,j} < \lambda_{i,0,j}\right]^\ell \, d\alpha
$$

$$
= \left[\frac{1}{4ab - \sum_{k=1}^8 R_k^S}\right]^\ell \int_{|x_2 - x_1| + |y_2 - y_1|}^{|x_1| + |x_2| + |y_1| + |y_2|} \sum_{\left\{\substack{\ell_1, \ell_2, \ell_3 \, : \\ \ell_1 + \ell_2 + \ell_3 = \ell}\right\}} \frac{\ell!}{\ell_1! \ell_2! \ell_3!} A(\alpha)^{\ell_1} B(\alpha)^{\ell_2} C(\alpha)^{\ell_3} \alpha^{2\ell_1 + \ell_2} \, d\alpha. \tag{99}
$$

Calculating (99) is easy when the shaded region in Figure 8 is inside the rectangle $[-a, a] \times [-b, b]$ for all $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$. In this case, $A(\alpha)$, $B(\alpha)$, and $C_1(\alpha)$ are constant for all $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$. For example, suppose that $a = b$ and

$$
(x_1, y_1) \quad = \quad \left(\frac{1}{4}a, \frac{1}{4}a\right) \quad \text{and} \quad (x_2, y_2) \quad = \quad \left(\frac{1}{2}a, \frac{1}{2}a\right).
$$

Then, $|x_2 - x_1| + |y_2 - y_1| = a/2$ and $|x_1| + |x_2| + |y_1| + |y_2| = 3a/2$, and, for all $\alpha \in (a/2, 3a/2)$,

$$
A(\alpha) \quad = \quad -\frac{1}{2}, \quad B(\alpha) \quad = \quad 0, \quad \text{and} \quad C_1(\alpha) \quad = \quad \frac{65}{16}a^2.
$$

43

Furthermore, $\sum_{k=1}^{8} R_k^S = 25a^2/8$. Suppose, for example, that $\ell = 1$. Then, (99) is easily calculated as

$$\int_{a/2}^{3a/2} P\left[\lambda_{i,1,j} > \alpha \mid \lambda_{i,1,j} < \lambda_{i,0,j}\right] d\alpha = \frac{1}{4a^2 - \frac{25}{8}a^2} \int_{a/2}^{3a/2} \left[A(\alpha)\alpha^2 + B(\alpha)\alpha + C(\alpha)\right] d\alpha$$

$$= \frac{1}{\frac{7}{8}a^2} \int_{a/2}^{3a/2} \left[-\frac{1}{2}\alpha^2 + \left(\frac{65}{16}a^2 - \frac{25}{8}a^2\right)\right] d\alpha$$

$$= \frac{2}{21}a.$$

Note that when $\alpha = 3a/2$, then $R_3^S(\alpha) = R_4^S(\alpha) = 0$. In particular, when $\alpha = 3a/2$, the lines 5 and 3 are exactly on the boundary of the rectangle $[-a, a] \times [-a, a]$. Consider a different choice of (x_1, y_1) and (x_2, y_2), for example,

$$(x_1, y_1) = \left(\frac{1}{4}a, \frac{1}{4}a\right) \quad \text{and} \quad (x_2, y_2) = \left(\frac{3}{4}a, \frac{3}{4}a\right).$$

Then $R_3^S(\alpha) = R_4^S(\alpha) = 0$ for α in a nonempty portion of the interval

$$(|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|) = (a, 2a),$$

which causes $A(\alpha)$, $B(\alpha)$, and $C_1(\alpha)$ to be piecewise constant with multiple pieces (rather than constant) over the interval $(|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$. Specifically, $R_3^S(\alpha) = R_4^S(\alpha) = 0$ for all $\alpha \in [3a/2, 2a)$. Furthermore,

$$A(\alpha) = \begin{cases} -\frac{1}{2} & \text{if} \quad a < \alpha < \frac{3}{2}a \\ 0 & \text{if} \quad \frac{3}{2}a \leq \alpha < 2a \end{cases}$$

$$B(\alpha) = \begin{cases} 0 & \text{if} \quad a < \alpha < \frac{3}{2}a \\ -a & \text{if} \quad \frac{3}{2}a \leq \alpha < 2a \end{cases}$$

$$C_1(\alpha) = \begin{cases} \frac{17}{4}a^2 & \text{if} \quad a < \alpha < \frac{3}{2}a \\ \frac{37}{8}a^2 & \text{if} \quad \frac{3}{2}a \leq \alpha < 2a \end{cases},$$

and $\sum_{k=1}^{8} R_k^S = 21a^2/8$. When $\ell = 1$, (99) becomes

$$\int_a^{2a} P\left[\lambda_{i,1,j} > \alpha \mid \lambda_{i,1,j} < \lambda_{i,0,j}\right] d\alpha$$

$$= \frac{1}{4a^2 - \frac{21}{8}a^2} \int_a^{2a} \left[A(\alpha)\alpha^2 + B(\alpha)\alpha + C(\alpha)\right] d\alpha$$

$$= \frac{8}{11}\frac{1}{a^2} \left(\int_a^{3a/2} \left[-\frac{1}{2}\alpha^2 + \left(\frac{17}{4}a^2 - \frac{21}{8}a^2\right)\right] d\alpha + \int_{3a/2}^{2a} \left[-a\alpha + \left(\frac{37}{8}a^2 - \frac{21}{8}a^2\right)\right] d\alpha\right)$$

$$= \frac{13}{33}a.$$

In general, integrating (99) requires partitioning the interval $(|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$ into the subintervals over which $A(\alpha)$, $B(\alpha)$, and $C(\alpha)$ are constant. The intervals of α-values over which lines 1–8 are inside or outside the rectangle $[-a, a] \times [-b, b]$, determine the intervals of α-values over which the triangles $R_5^S(\alpha)$, $R_7^S(\alpha)$, and $R_8^S(\alpha)$ have positive areas, and the intervals of α-values over which $A(\alpha)$, $B(\alpha)$, and $C_1(\alpha)$ are constant, from which the calculation of (99) follows directly. We first consider the cases determined by which of the lines 1, 3, 5, and 7 in Figure 8 are outside the rectangle $[-a, a] \times [-b, b]$, and then, if required, consider further subcases determined by which triangles have positive area. Fortunately, dependencies imply that many of the $2^4 \times 2^3$ conceivable cases cannot occur. For example, examination of Figure 8 shows that if line 3 is inside the rectangle so that $(\alpha + x_1 + x_2 + y_1 - y_2)/2 < a$, then $\alpha < 2a - x_1 - x_2 - y_1 + y_2 \leq 2a + x_1 + x_2 - y_1 + y_2$, which implies that line 7 is also inside the rectangle. We show only the cases that can occur.

44

8.1.1 Calculating the piecewise constant functions $A(\alpha)$, $B(\alpha)$, and $C(\alpha)$

For easy reference to Figure 8, and without loss of generality, assume that $x_1 = \underline{x}$, $x_2 = \overline{x}$, $y_1 = \underline{y}$, and $y_2 = \overline{y}$. Note that in this case $R_6^S(\alpha) > 0$ for all $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$. The following cases can occur:

1. Lines 1, 3, 5, and 7 are inside the rectangle, that is,

$$
\begin{aligned}
\alpha \quad \in \quad & [|x_2 - x_1| + |y_2 - y_1|, \ (2a + x_1 + x_2 - y_1 + y_2) \wedge (2a - x_1 - x_2 - y_1 + y_2) \\
& \wedge (2b - x_1 + x_2 + y_1 + y_2) \wedge (2b - x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_5^S(\alpha) > 0$, $R_7^S(\alpha) > 0$, and $R_8^S(\alpha) > 0$. Then,

$$
\begin{aligned}
A(\alpha) &= -\frac{1}{2} \\
B(\alpha) &= 0 \\
C_1(\alpha) &= 4ab + \frac{1}{2}(x_2 - x_1)^2 + \frac{1}{2}(y_2 - y_1)^2 .
\end{aligned}
$$

2. Lines 1, 5, and 7 are inside the rectangle, but line 3 is outside, that is,

$$
\begin{aligned}
\alpha \quad \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2), \\
& (2a + x_1 + x_2 - y_1 + y_2) \wedge (2b - x_1 + x_2 + y_1 + y_2) \wedge (2b - x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_5^S(\alpha) > 0$, $R_7^S(\alpha) > 0$, and $R_8^S(\alpha) > 0$. Then,

$$
\begin{aligned}
A(\alpha) &= -\frac{1}{4} \\
B(\alpha) &= \frac{1}{2}(x_1 + x_2) - a \\
C_1(\alpha) &= a(a + 4b - x_1 - x_2) + \frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{4}(x_2 - x_1)^2 + \frac{1}{4}(y_2 - y_1)^2
\end{aligned}
$$

3. Lines 1, 3, and 7 are inside the rectangle, but line 5 is outside, that is,

$$
\begin{aligned}
\alpha \quad \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2b - x_1 + x_2 - y_1 - y_2), \\
& (2a + x_1 + x_2 - y_1 + y_2) \wedge (2a - x_1 - x_2 - y_1 + y_2) \wedge (2b - x_1 + x_2 + y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_5^S(\alpha) > 0$, $R_7^S(\alpha) > 0$, and $R_8^S(\alpha) > 0$. Then,

$$
\begin{aligned}
A(\alpha) &= -\frac{1}{4} \\
B(\alpha) &= \frac{1}{2}(y_1 + y_2) - b \\
C_1(\alpha) &= b(4a + b - y_1 - y_2) + \frac{1}{4}(x_2 - x_1)^2 + \frac{1}{4}(y_2 - y_1)^2 + \frac{1}{2}(y_1^2 + y_2^2)
\end{aligned}
$$

4. Lines 1 and 7 are inside the rectangle, but lines 3 and 5 are outside, that is,

$$
\begin{aligned}
\alpha \quad \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2), \\
& (2a + x_1 + x_2 - y_1 + y_2) \wedge (2b - x_1 + x_2 + y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_5^S(\alpha) > 0$ and $R_7^S(\alpha) > 0$. The following two subcases can occur:

(a) Part of line 4 is inside the rectangle, that is,

$$
\begin{aligned}
\alpha \quad \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2), \\
& (2a + x_1 + x_2 - y_1 + y_2) \wedge (2b - x_1 + x_2 + y_1 + y_2) \wedge (2(a + b) - x_1 - x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_8^S(\alpha) > 0$. Then,

$$A(\alpha) = 0$$
$$B(\alpha) = -(a+b) + \frac{1}{2}(x_1 + x_2 + y_1 + y_2)$$
$$C_1(\alpha) = (a+b)^2 + 2ab - a(x_1 + x_2) - b(y_1 + y_2) + \frac{1}{2}\left(x_1^2 + x_2^2 + y_1^2 + y_2^2\right).$$

(b) Line 4 is outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a+b) - x_1 - x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (2b - x_1 + x_2 + y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_8^S(\alpha) = 0$. Then,

$$A(\alpha) = -\frac{1}{8}$$
$$B(\alpha) = -\frac{1}{2}(a+b) + \frac{1}{4}(x_1 + x_2 + y_1 + y_2)$$
$$C_1(\alpha) = \frac{1}{2}(a+b)^2 + 2ab + \frac{1}{2}(b-a)(x_1 + x_2 - y_1 - y_2).$$

5. Lines 1 and 5 are inside the rectangle, but lines 3 and 7 are outside, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2),$$
$$(2b - x_1 + x_2 + y_1 + y_2) \wedge (2b - x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_5^S(\alpha) > 0$, $R_7^S(\alpha) > 0$, and $R_8^S(\alpha) > 0$. Then,

$$A(\alpha) = 0$$
$$B(\alpha) = -2a$$
$$C_1(\alpha) = 2a(a + 2b) + x_1^2 + x_2^2.$$

6. Lines 3 and 7 are inside the rectangle, but lines 1 and 5 are outside, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (2a - x_1 - x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_5^S(\alpha) > 0$, $R_7^S(\alpha) > 0$, and $R_8^S(\alpha) > 0$. Then,

$$A(\alpha) = 0$$
$$B(\alpha) = -2b$$
$$C_1(\alpha) = 2b(2a + b) + y_1^2 + y_2^2.$$

7. Line 7 is inside the rectangle, but lines 1, 3, and 5 are outside, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7^S(\alpha) > 0$. The following three subcases can occur:

(a) Parts of lines 2 and 4 are inside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (2(a+b) - x_1 - x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

46

which implies that $R_5^S(\alpha) > 0$ and $R_8^S(\alpha) > 0$. Then,

$$A(\alpha) = \frac{1}{4}$$

$$B(\alpha) = -\frac{3}{2}a - 2b + \frac{1}{2}(x_1 + x_2) - \frac{1}{4}(y_1 + y_2)$$

$$C_1(\alpha) = a^2 + 4ab + 2b^2 - a(x_1 + x_2) + \frac{1}{4}(x_1 + x_2)^2 + \frac{3}{4}(y_1 + y_2)^2 - y_1 y_2.$$

(b) Part of line 2 is inside the rectangle and line 4 is outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 + y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a+b) - x_1 - x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (2(a+b) - x_1 - x_2 + y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_5^S(\alpha) > 0$ and $R_8^S(\alpha) = 0$. Then,

$$A(\alpha) = \frac{1}{8}$$

$$B(\alpha) = -\frac{1}{2}a - \frac{3}{2}b + \frac{1}{4}(x_1 + x_2 - y_1 - y_2)$$

$$C_1(\alpha) = (a+b)^2 + 2ab - a(x_1 + x_2) + (a+b)(y_1 + y_2) - \frac{3}{4}(x_1^2 + x_2^2) - \frac{1}{4}(y_1^2 + y_2^2)$$
$$+ (x_1 + x_2 - y_1 - y_2)^2.$$

(c) Lines 2 and 4 are outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 + y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a+b) - x_1 - x_2 + y_1 + y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_5^S(\alpha) = R_8^S(\alpha) = 0$. Then,

$$A(\alpha) = 0$$

$$B(\alpha) = -b$$

$$C_1(\alpha) = b(2a + b + x_1 + x_2) + \frac{1}{2}(y_1^2 + y_2^2).$$

8. Line 1 is inside the rectangle, but lines 3, 5 and 7 are outside, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2b - x_1 + x_2 + y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_5^S(\alpha) > 0$. The following three subcases can occur:

(a) Parts of lines 4 and 6 are inside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2b - x_1 + x_2 + y_1 + y_2) \wedge (2(a+b) - x_1 - x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7^S(\alpha) > 0$ and $R_8^S(\alpha) > 0$. Then,

$$A(\alpha) = \frac{1}{4}$$

$$B(\alpha) = -2a - b + \frac{1}{2}(y_1 + y_2)$$

$$C_1(\alpha) = 2a^2 + 4ab + b^2 - b(y_1 + y_2) + \frac{3}{4}(x_1 + x_2)^2 + \frac{1}{4}(y_1 + y_2)^2 + \frac{1}{2}x_1 x_2.$$

47

(b) Part of line 6 is inside the rectangle and line 4 is outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) - x_1 - x_2 - y_1 - y_2),$$
$$(2b - x_1 + x_2 + y_1 + y_2) \wedge (2(a + b) + x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7^S(\alpha) > 0$ and $R_8^S(\alpha) = 0$. Then,

$$A(\alpha) = \frac{1}{8}$$
$$B(\alpha) = -\frac{3}{2}a - \frac{1}{2}b + \frac{1}{4}(-x_1 - x_2 + y_1 + y_2)$$
$$C_1(\alpha) = \frac{3}{2}a^2 + 3ab + \frac{1}{2}b^2 + \frac{1}{2}(a + b)(x_1 + x_2) + \frac{1}{2}(a - b)(y_1 + y_2)$$
$$+ \frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{8}(x_1 + x_2 - y_1 - y_2)^2.$$

(c) Lines 4 and 6 are outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) + x_1 + x_2 - y_1 - y_2),$$
$$(2b - x_1 + x_2 + y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7^S(\alpha) = R_8^S(\alpha) = 0$. Then,

$$A(\alpha) = 0$$
$$B(\alpha) = -a$$
$$C_1(\alpha) = a(a + 2b + y_1 + y_2) + \frac{1}{2}(x_1^2 + x_2^2).$$

9. Lines 1, 3, 5, and 7 are outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2), |x_1| + |x_2| + |y_1| + |y_2|)$$

The following five subcases can occur:

(a) Parts of lines 2, 4, and 6 are inside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2(a + b) - x_1 - x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_5^S(\alpha) > 0$, $R_7^S(\alpha) > 0$, and $R_8^S(\alpha) > 0$. Then,

$$A(\alpha) = \frac{1}{2}$$
$$B(\alpha) = -2(a + b)$$
$$C_1(\alpha) = 2(a + b)^2 + \frac{1}{2}(x_1^2 + x_2^2 + y_1^2 + y_2^2) + x_1 x_2 + y_1 y_2.$$

(b) Parts of lines 2 and 6 are inside the rectangle and line 4 is outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) - x_1 - x_2 - y_1 - y_2),$$
$$(2(a + b) + x_1 + x_2 - y_1 - y_2) \wedge (2(a + b) - x_1 - x_2 + y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_5^S(\alpha) > 0$, $R_7^S(\alpha) > 0$, and $R_8^S(\alpha) = 0$. Then,

$$
\begin{aligned}
A(\alpha) &= \frac{3}{8} \\
B(\alpha) &= -\frac{3}{2}a - \frac{3}{2}b - \frac{1}{4}(x_1 + x_2 + y_1 + y_2) \\
C_1(\alpha) &= \frac{3}{2}(a + b) + \frac{3}{4}(x_1 + x_2 + y_1 + y_2).
\end{aligned}
$$

(c) Part of line 6 is inside the rectangle and lines 2 and 4 are outside the rectangle, that is,

$$
\begin{aligned}
\alpha \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2) \\
& \vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) - x_1 - x_2 + y_1 + y_2), \\
& (2(a + b) + x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_7^S(\alpha) > 0$ and $R_5^S(\alpha) = R_8^S(\alpha) = 0$. Then,

$$
\begin{aligned}
A(\alpha) &= \frac{1}{4} \\
B(\alpha) &= -a - b - \frac{1}{2}x_1 - \frac{1}{2}x_2 \\
C_1(\alpha) &= \left(a + b + \frac{1}{2}(x_1 + x_2)\right)^2 + \frac{1}{4}(y_1 + y_2)^2.
\end{aligned}
$$

(d) Part of line 2 is inside the rectangle and lines 4 and 6 are outside the rectangle, that is,

$$
\begin{aligned}
\alpha \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2) \\
& \vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) + x_1 + x_2 - y_1 - y_2), \\
& (2(a + b) - x_1 - x_2 + y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_5^S(\alpha) > 0$ and $R_7^S(\alpha) = R_8^S(\alpha) = 0$. Then,

$$
\begin{aligned}
A(\alpha) &= \frac{1}{4} \\
B(\alpha) &= -a - b - \frac{1}{2}y_1 - \frac{1}{2}y_2 \\
C_1(\alpha) &= \left(a + b + \frac{1}{2}(y_1 + y_2)\right)^2 + \frac{1}{4}(x_1 + x_2)^2.
\end{aligned}
$$

(e) Lines 2, 4, and 6 are outside the rectangle, that is,

$$
\begin{aligned}
\alpha \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2a - x_1 - x_2 - y_1 + y_2) \\
& \vee (2b - x_1 + x_2 + y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) + x_1 + x_2 - y_1 - y_2) \\
& \vee (2(a + b) - x_1 - x_2 + y_1 + y_2), |x_1| + |x_2| + |y_1| + |y_2|)
\end{aligned}
$$

which implies that $R_5^S(\alpha) = R_7^S(\alpha) = R_8^S(\alpha) = 0$. Then,

$$
\begin{aligned}
A(\alpha) &= \frac{1}{8} \\
B(\alpha) &= -\frac{1}{2}(a + b) - \frac{1}{4}(x_1 + x_2 + y_1 + y_2) \\
C_1(\alpha) &= \frac{1}{2}(a + b)(a + b + x_1 + x_2 + y_1 + y_2) \\
& \quad + \frac{1}{8}(x_1 + x_2 + y_1 + y_2)^2.
\end{aligned}
$$

49

8.2 Origin and Destination Located in Adjacent Quadrants

First, suppose that $x_1 \leq 0 \leq x_2$ and $0 \leq y_1 \leq y_2$. Then i is in quadrant II and j is in quadrant I, as in Figure 9. For $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$, let

$$
R_1^A(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ -a \leq x \leq a \text{ and } -b \leq y \leq \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2}\right\}\right)
$$

$$
R_2^A(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} -a \leq x \leq \frac{-\alpha + x_1 + x_2 - y_1 + y_2}{2} \\ \text{and } \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2} \leq y \leq \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \end{array}\right\}\right)
$$

$$
R_3^A(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ -a \leq x \leq a \text{ and } \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \leq y \leq b\right\}\right)
$$

$$
R_4^A(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} \frac{\alpha + x_1 + x_2 + y_1 - y_2}{2} \leq x \leq a \\ \text{and } \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2} \leq y \leq \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \end{array}\right\}\right)
$$

$$
R_5^A(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} 2(x - y) \geq \alpha + x_1 + x_2 - y_1 - y_2 \text{ and } x_2 \leq x \leq \frac{\alpha + x_1 + x_2 + y_1 - y_2}{2} \\ \text{and } \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2} \leq y \leq y_1 \end{array}\right\}\right)
$$

$$
R_6^A(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} 2(x + y) \geq -\alpha + x_1 + x_2 + y_1 + y_2 \text{ and } \frac{-\alpha + x_1 + x_2 - y_1 + y_2}{2} \leq x \leq x_1 \\ \text{and } \frac{-\alpha - x_1 + x_2 + y_1 + y_2}{2} \leq y \leq y_1 \end{array}\right\}\right)
$$

$$
R_7^A(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} 2(y - x) \geq \alpha - x_1 - x_2 + y_1 + y_2 \text{ and } \frac{-\alpha + x_1 + x_2 - y_1 + y_2}{2} \leq x \leq x_1 \\ \text{and } y_2 \leq y \leq \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \end{array}\right\}\right)
$$

$$
R_8^A(\alpha) \quad := \quad \text{Area}\left(\left\{(x,y) \ : \ \begin{array}{l} 2(x + y) \geq \alpha + x_1 + x_2 + y_1 + y_2 \text{ and } x_2 \leq x \leq \frac{\alpha + x_1 + x_2 + y_1 - y_2}{2} \\ \text{and } y_2 \leq y \leq \frac{\alpha + x_1 - x_2 + y_1 + y_2}{2} \end{array}\right\}\right).
$$

Figure 9 shows the regions defining $R_1^A(\alpha), \ldots, R_8^A(\alpha)$ for a value of $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$. The lines 1–8 in Figure 9 form the boundaries of the regions defining $R_1^A(\alpha), \ldots, R_8^A(\alpha)$.

For the special case with $\alpha = \lambda_{i,0,j} = |x_1| + |x_2| + |y_1| + |y_2|)$, let

$$
R_k^A \quad := \quad R_k^A(\lambda_{i,0,j}).
$$

Then, the probabilities in (94) and (95) are given by

$$
P\left[\lambda_{i,1,j} < \lambda_{i,0,j}\right] \quad = \quad \frac{4ab - \sum_{k=1}^{8} R_k^A}{4ab}
$$

and

$$
P\left[\lambda_{i,1,j} > \alpha \mid \lambda_{i,1,j} < \lambda_{i,0,j}\right] \quad = \quad \frac{\sum_{k=1}^{8} R_k^A(\alpha) - \sum_{k=1}^{8} R_k^A}{4ab - \sum_{k=1}^{8} R_k^A} \tag{100}
$$

Next, we allow i and j to be in any adjacent quadrants, and we allow $x_1 > x_2$ or $y_1 > y_2$. It is easy to see how to choose a new coordinate system with the same origin and scale as the original coordinate system and with axes coinciding with the axes of the original coordinate system, so that j will be in the new quadrant I and i will be in the new quadrant II. Furthermore, using Figure 9, it is easy to observe that in the new coordinate system, the areas $R_1^A(\alpha), \ldots, R_8^A(\alpha)$ depend only on the minimum and maximum of the distances of i and j from the new x axis. Hence, if x_1 and x_2 have opposite signs and y_1 and y_2 have the same sign, then let

$$
\underline{x} := -|x_1|, \quad \overline{x} := |x_2|, \quad \underline{y} := |y_1| \wedge |y_2|, \quad \overline{y} := |y_1| \vee |y_2|, \quad a' := a, \quad \text{and} \quad b' := b \tag{101}
$$

Otherwise, if x_1 and x_2 have the same sign and y_1 and y_2 have opposite signs, then let

$$
\underline{x} := -|y_1|, \quad \overline{x} := |y_2|, \quad \underline{y} := |x_1| \wedge |x_2|, \quad \overline{y} := |x_1| \vee |x_2|, \quad a' := b, \quad \text{and} \quad b' := a \tag{102}
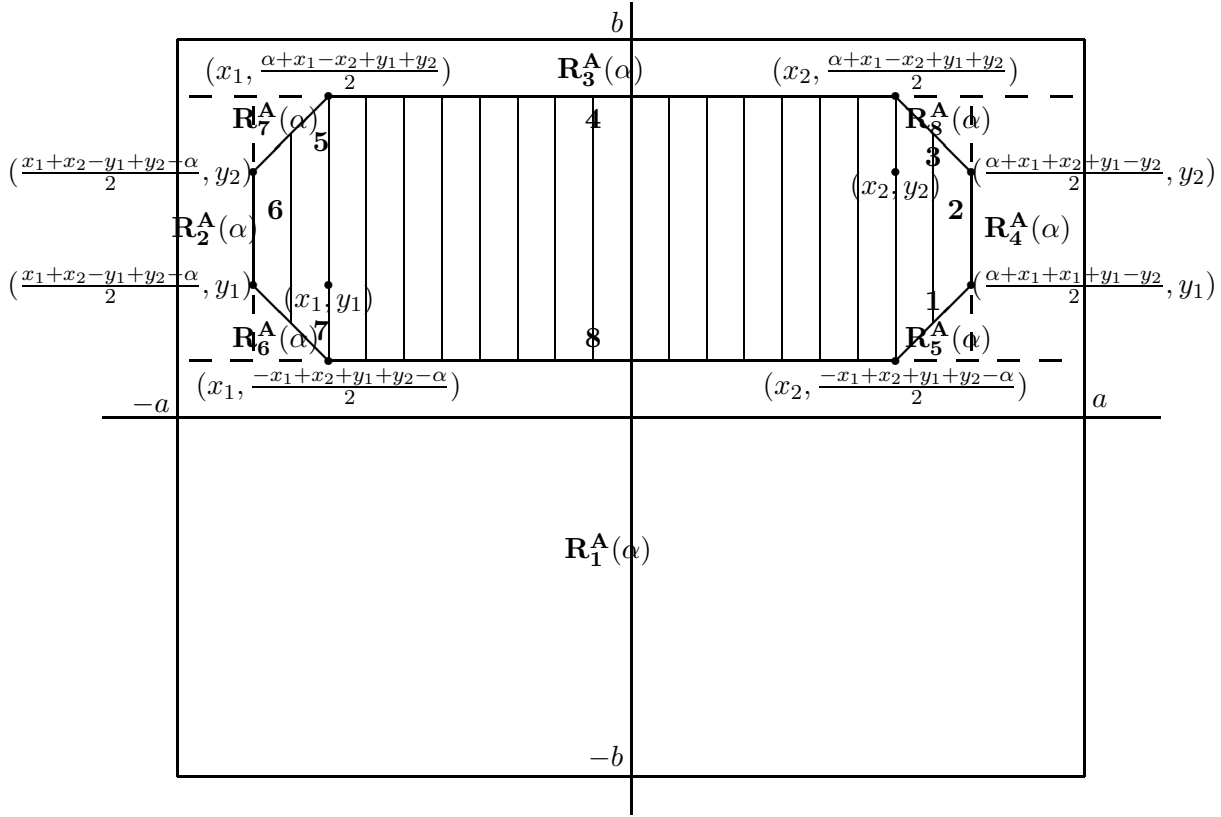$$

Figure 9: Example of the region in which location of a terminal provides a smaller L_1 linehaul distance than $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$ if (x_1, y_1) and (x_2, y_2) are in adjacent quadrants.

Then

$$R_1^A(\alpha) = -a'\alpha + a'(-\underline{x} + \overline{x} + \underline{y} + \overline{y}) + 2a'b'$$

$$R_2^A(\alpha) = \left(\left(\left(b' \wedge \frac{\alpha + \underline{x} - \overline{x} + \underline{y} + \overline{y}}{2}\right) - \frac{-\alpha - \underline{x} + \overline{x} + \underline{y} + \overline{y}}{2}\right)\left(a' + \frac{-\alpha + \underline{x} + \overline{x} - \underline{y} - \overline{y}}{2}\right)\right.$$
$$\times \mathbb{I}_{\{\alpha < 2a' + \underline{x} + \overline{x} - \underline{y} + \overline{y}\}}$$

$$R_3^A(\alpha) = 2a'\left(b' - \frac{\alpha + \underline{x} - \overline{x} + \underline{y} + \overline{y}}{2}\right)\mathbb{I}_{\{\alpha < 2b' - \underline{x} + \overline{x} - \underline{y} - \overline{y}\}}$$

$$R_4^A(\alpha) = \left(a' - \frac{\alpha + \underline{x} + \overline{x} + \underline{y} - \overline{y}}{2}\right)\left(\left(b' \wedge \frac{\alpha + \underline{x} - \overline{x} + \underline{y} + \overline{y}}{2}\right) - \frac{-\alpha - \underline{x} + \overline{x} + \underline{y} + \overline{y}}{2}\right)$$
$$\times \mathbb{I}_{\{\alpha < 2a' - \underline{x} - \overline{x} - \underline{y} + \overline{y}\}}$$

$$R_5^A(\alpha) = \frac{1}{2}\left(\left(a' \wedge \frac{\alpha + \underline{x} + \overline{x} + \underline{y} - \overline{y}}{2}\right) - \overline{x}\right)^2$$

$$R_6^A(\alpha) = \frac{1}{2}\left(\underline{x} - \left((-a') \vee \frac{-\alpha + \underline{x} + \overline{x} - \underline{y} + \overline{y}}{2}\right)\right)^2$$

$$R_7^A(\alpha) = \frac{1}{2}\left(\left(\underline{x} \wedge \left(b' - \frac{\alpha - \underline{x} - \overline{x} + \underline{y} + \overline{y}}{2}\right)\right) - \left((-a') \vee \frac{-\alpha + \underline{x} + \overline{x} - \underline{y} + \overline{y}}{2}\right)\right)^2$$
$$\times \mathbb{I}_{\{\alpha < 2(a' + b') + \underline{x} + \overline{x} - \underline{y} - \overline{y}\}}$$

$$R_8^A(\alpha) = \frac{1}{2}\left(\left(a' \wedge \frac{\alpha + \underline{x} + \overline{x} + \underline{y} - \overline{y}}{2}\right) - \left(\overline{x} \vee \left(\frac{\alpha + \underline{x} + \overline{x} + \underline{y} + \overline{y}}{2} - b'\right)\right)\right)^2$$
$$\times \mathbb{I}_{\{\alpha < 2(a' + b') - \underline{x} - \overline{x} - \underline{y} - \overline{y}\}}$$

51

Similar to Section 8.1, we write $\sum_{k=1}^{8} R_k^A(\alpha)$ as a piecewise polynomial in α of degree at most two, and we choose $A(\alpha), B(\alpha), C_1(\alpha)$, and C_2 such that

$$A(\alpha)\alpha^2 + B(\alpha)\alpha + C_1(\alpha) \quad = \quad \sum_{k=1}^{8} R_k^A(\alpha) \quad \text{and} \quad C_2 \quad = \quad \sum_{k=1}^{8} R_k^A.$$

8.2.1 Calculating the piecewise constant functions $A(\alpha)$, $B(\alpha)$, and $C(\alpha)$

For easy reference to Figure 9, and without loss of generality, assume that $x_1 = \underline{x}$, $x_2 = \overline{x}$, $y_1 = \underline{y}$, $y_2 = \overline{y}$, $a = a'$, and $b = b'$. Note that in this case $R_5^A(\alpha) > 0$ and $R_6^A(\alpha) > 0$ for all $\alpha \in (|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$. As in Section 8.1, when partitioning the interval $(|x_2 - x_1| + |y_2 - y_1|, |x_1| + |x_2| + |y_1| + |y_2|)$ into the subintervals over which $A(\alpha)$, $B(\alpha)$, and $C(\alpha)$ are constant, we show only the cases that can occur, namely the following cases:

1. Lines 2, 4, and 6 are inside the rectangle, that is,

$$\begin{aligned}
\alpha \quad \in \quad & [|x_2 - x_1| + |y_2 - y_1|, \\
& (2a + x_1 + x_2 - y_1 + y_2) \wedge (2b - x_1 + x_2 - y_1 - y_2) \wedge (2a - x_1 - x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}$$

which implies that $R_7^A(\alpha) > 0$ and $R_8^A(\alpha) > 0$. Then,

$$\begin{aligned}
A(\alpha) \quad &= \quad -\frac{1}{2} \\
B(\alpha) \quad &= \quad 0 \\
C_1(\alpha) \quad &= \quad 4ab + \frac{1}{2}\left(x_1^2 + x_2^2 + y_1^2 + y_2^2\right) - x_1 x_2 - y_1 y_2.
\end{aligned}$$

2. Lines 2 and 4 are inside the rectangle, but line 6 is outside, that is,

$$\begin{aligned}
\alpha \quad \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2), \\
& (2b - x_1 + x_2 - y_1 - y_2) \wedge (2a - x_1 - x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}$$

which implies that $R_7^A(\alpha) > 0$ and $R_8^A(\alpha) > 0$. Then,

$$\begin{aligned}
A(\alpha) \quad &= \quad -\frac{1}{4} \\
B(\alpha) \quad &= \quad -a - \frac{1}{2}(x_1 + x_2) \\
C_1(\alpha) \quad &= \quad a(a + 4b + x_1 + x_2) + \frac{3}{4}\left(x_1^2 + x_2^2\right) + \frac{1}{4}\left(y_2 - y_1\right)^2 - \frac{1}{2}x_1 x_2.
\end{aligned}$$

3. Lines 4 and 6 are inside the rectangle, but line 2 is outside, that is,

$$\begin{aligned}
\alpha \quad \in \quad & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2), \\
& (2a + x_1 + x_2 - y_1 + y_2) \wedge (2b - x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}$$

which implies that $R_7^A(\alpha) > 0$ and $R_8^A(\alpha) > 0$. Then,

$$\begin{aligned}
A(\alpha) \quad &= \quad -\frac{1}{4} \\
B(\alpha) \quad &= \quad -a + \frac{1}{2}(x_1 + x_2) \\
C_1(\alpha) \quad &= \quad a(a + 4b - x_1 - x_2) + \frac{3}{4}\left(x_1^2 + x_2^2\right) + \frac{1}{4}\left(y_2 - y_1\right)^2 - \frac{1}{2}x_1 x_2.
\end{aligned}$$

52

4. Lines 2 and 6 are inside the rectangle, but line 4 is outside, that is,

$$\alpha \quad \in \quad [(|x_2 - x_1| + |y_2 - y_1|) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (2a - x_1 - x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7^A(\alpha) > 0$ and $R_8^A(\alpha) > 0$. Then,

$$A(\alpha) \quad = \quad -\frac{1}{4}$$
$$B(\alpha) \quad = \quad -b + \frac{1}{2}(y_1 + y_2)$$
$$C_1(\alpha) \quad = \quad b(4a + b - y_1 - y_2) + \frac{1}{4}(x_2 - x_1)^2 + \frac{3}{4}(y_1^2 + y_2^2) - \frac{1}{2}y_1 y_2.$$

5. Line 2 is inside the rectangle, but lines 4 and 6 are outside, that is,

$$\alpha \quad \in \quad [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2a - x_1 - x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_8(\alpha) > 0$. The following two subcases can occur:

(a) Part of line 5 is inside the rectangle, that is,

$$\alpha \quad \in \quad [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2a - x_1 - x_2 - y_1 + y_2) \wedge (2(a + b) + x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7(\alpha) > 0$. Then,

$$A(\alpha) \quad = \quad 0$$
$$B(\alpha) \quad = \quad -(a + b) - \frac{1}{2}(x_1 + x_2) + \frac{1}{2}(y_1 + y_2)$$
$$C_1(\alpha) \quad = \quad a^2 + b^2 + 4ab + a(x_1 + x_2) - b(y_1 + y_2) + \frac{1}{2}(x_1^2 + x_2^2 + y_1^2 + y_2^2).$$

(b) Line 5 is outside the rectangle, that is,

$$\alpha \quad \in \quad [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a + x_1 + x_2 - y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) + x_1 + x_2 - y_1 - y_2),$$
$$(2a - x_1 - x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7(\alpha) = 0$. Then,

$$A(\alpha) \quad = \quad -\frac{1}{8}$$
$$B(\alpha) \quad = \quad -\frac{1}{2}(a + b) - \frac{1}{4}(x_1 + x_2) + \frac{1}{4}(y_1 + y_2)$$
$$C_1(\alpha) \quad = \quad \frac{1}{2}(a^2 + b^2) + 3ab + \frac{1}{2}(a - b)(x_1 + x_2 + y_1 + y_2)$$
$$+ \frac{3}{8}(x_1^2 + x_2^2 + y_1^2 + y_2^2) + \frac{1}{4}(-x_1 x_2 + x_1 y_1 + x_1 y_2 + x_2 y_1 + x_2 y_2 - y_1 y_2).$$

6. Line 6 is inside the rectangle, but lines 2 and 4 are outside, that is,

$$\alpha \quad \in \quad [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7(\alpha) > 0$. The following two subcases can occur:

53

(a) Part of line 3 is inside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (2(a+b) - x_1 - x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_8(\alpha) > 0$. Then,

$$A(\alpha) = 0$$
$$B(\alpha) = -(a+b) + \frac{1}{2}(x_1 + x_2 + y_1 + y_2)$$
$$C_1(\alpha) = 4ab + a^2 + b^2 - a(x_1 + x_2) - b(y_1 + y_2) + \frac{1}{2}(x_1^2 + x_2^2 + y_1^2 + y_2^2).$$

(b) Line 3 is outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2)$$
$$\vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a+b) - x_1 - x_2 - y_1 - y_2),$$
$$(2a + x_1 + x_2 - y_1 + y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_8(\alpha) = 0$. Then,

$$A(\alpha) = -\frac{1}{8}$$
$$B(\alpha) = -\frac{1}{2}(a+b) + \frac{1}{4}(x_1 + x_2) + \frac{1}{4}(y_1 + y_2)$$
$$C_1(\alpha) = 3ab + \frac{1}{2}(a^2 + b^2) + \frac{1}{2}(b - a)(x_1 + x_2 - y_1 - y_2)$$
$$+ \frac{1}{2}(x_1^2 + x_2^2 + y_1^2 + y_2^2) - \frac{1}{8}(x_1 + x_2 + y_1 + y_2)^2.$$

7. Line 4 is inside the rectangle, but lines 2 and 6 are outside, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2a + x_1 + x_2 - y_1 + y_2),$$
$$(2b - x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

which implies that $R_7(\alpha) > 0$ and $R_8(\alpha) > 0$. Then,

$$A(\alpha) = 0$$
$$B(\alpha) = -2a$$
$$C_1(\alpha) = 4ab + 2a^2 + x_1^2 + x_2^2.$$

8. Lines 2, 4, and 6 are outside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$|x_1| + |x_2| + |y_1| + |y_2|)$$

The following four subcases can occur:

(a) Parts of lines 3 and 5 are inside the rectangle, that is,

$$\alpha \in [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2a + x_1 + x_2 - y_1 + y_2) \vee (2b - x_1 + x_2 - y_1 - y_2),$$
$$(2(a+b) + x_1 + x_2 - y_1 - y_2) \wedge (2(a+b) - x_1 - x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))$$

54

which implies that $R_7(\alpha) > 0$ and $R_8(\alpha) > 0$. Then,

$$
A(\alpha) = \frac{1}{4}
$$

$$
B(\alpha) = -2a - b + \frac{1}{2}(y_1 + y_2)
$$

$$
C_1(\alpha) = 2a^2 + b^2 + 4ab - b(y_1 + y_2) + \frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{4}(x_1 + x_2)^2 + \frac{1}{4}(y_1 + y_2)^2
$$

(b) Part of line 5 is inside the rectangle and line 3 is outside the rectangle, that is,

$$
\begin{aligned}
\alpha \in \ & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2a + x_1 + x_2 - y_1 + y_2) \\
& \vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) - x_1 - x_2 - y_1 - y_2), \\
& (2(a + b) + x_1 + x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_7(\alpha) > 0$ and $R_8(\alpha) = 0$. Then,

$$
A(\alpha) = \frac{1}{8}
$$

$$
B(\alpha) = -\frac{3}{2}a - \frac{1}{2}b - \frac{1}{4}(x_1 + x_2 - y_1 - y_2)
$$

$$
\begin{aligned}
C_1(\alpha) = \ & \frac{3}{2}a^2 + \frac{1}{2}b^2 + 3ab + \frac{1}{2}a(x_1 + x_2 + y_1 + y_2) + \frac{1}{2}b(x_1 + x_2 - y_1 - y_2) \\
& + \frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{8}(x_1 + x_2 - y_1 - y_2)^2.
\end{aligned}
$$

(c) Part of line 3 is inside the rectangle and line 5 is outside the rectangle, that is,

$$
\begin{aligned}
\alpha \in \ & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2a + x_1 + x_2 - y_1 + y_2) \\
& \vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) + x_1 + x_2 - y_1 - y_2), \\
& (2(a + b) - x_1 - x_2 - y_1 - y_2) \wedge (|x_1| + |x_2| + |y_1| + |y_2|))
\end{aligned}
$$

which implies that $R_7(\alpha) = 0$ and $R_8(\alpha) > 0$. Then,

$$
A(\alpha) = \frac{1}{8}
$$

$$
B(\alpha) = -\frac{3}{2}a - \frac{1}{2}b + \frac{1}{4}(x_1 + x_2 + y_1 + y_2)
$$

$$
\begin{aligned}
C_1(\alpha) = \ & \frac{3}{2}a^2 + \frac{1}{2}b^2 + 3ab - \frac{1}{2}a(x_1 + x_2 - y_1 - y_2) - \frac{1}{2}b(x_1 + x_2 + y_1 + y_2) \\
& + \frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{8}(x_1 + x_2 + y_1 + y_2)^2.
\end{aligned}
$$

(d) Lines 3 and 5 are outside the rectangle, that is,

$$
\begin{aligned}
\alpha \in \ & [(|x_2 - x_1| + |y_2 - y_1|) \vee (2a - x_1 - x_2 - y_1 + y_2) \vee (2a + x_1 + x_2 - y_1 + y_2) \\
& \vee (2b - x_1 + x_2 - y_1 - y_2) \vee (2(a + b) + x_1 + x_2 - y_1 - y_2) \vee (2(a + b) - x_1 - x_2 - y_1 - y_2), \\
& |x_1| + |x_2| + |y_1| + |y_2|)
\end{aligned}
$$

which implies that $R_7(\alpha) = 0$ and $R_8(\alpha) = 0$. Then,

$$
A(\alpha) = 0
$$

$$
B(\alpha) = -a
$$

$$
C_1(\alpha) = a^2 + 2ab + a(y_1 + y_2) + \frac{1}{2}(x_1^2 + x_2^2).
$$

9 Online Appendix

9.1 Computational Results for MILP formulation

For very small problem instances, the mixed integer linear program (3) can be solved with available software. Table 3 shows CPU times for some small instances obtained with CPlex 9.0. All instances were randomly generated, with a single scenario each.

Number of Origins	4	5	6
Number of Destinations	4	5	6
Number of Candidate Terminals	3	3	3
Instance 1	62.62	182.23	43164.04
Instance 2	308.35	36895.92	22475.01
Instance 3	31.65	692.00	Stop after 187643.12
Instance 4	184.23	5002.46	Stop after 115830.61
Instance 5	490.47	1873.43	Stop after 379172.34
Mean	215.46	8929.21	N/A
Standard Deviation	188.67	15745.65	N/A

Table 3: CPU time in seconds for MILP formulation

9.2 Data Sets

The data sets used for the results reported in Sections 7 and 9.1 can be found at
http://www.scl.gatech.edu/research/casestudies/. The distances d_{ij} between pairs of points i and j were given by the least great-circle distances in miles between the pairs of points. Also, for the purpose of the continuous approximation, (latitude, longitude) coordinates were converted to Cartesian (x, y) coordinates according to the Albers Equal Area Conic Projection Method. The L_1 metric between pairs of (x, y) coordinates were multiplied with a factor of 3150, which gives approximately the least great-circle distance in miles between the pair of points. All scenarios were assigned equal weights $p(\omega)$. In addition, the following parameters were used:

Fixed cost per terminal per time period $c_m = \$10000$.

Transportation cost per vehicle-mile $= \$1$.

Cost per time period for each vehicle based at each terminal $C_v = \$100$.

Cost for each vehicle that is used during a time period $c_v = \$100$.

Vehicle capacity used in detailed vehicle routing calculations $Q_v = 3000\text{ft}^3$.

Vehicle capacity used in continuous approximation calculations $Q_v = 2900\text{ft}^3$.

β-coefficient for approximating detour distance $\beta = 2$.

The rectangle $[-a, a] \times [-b, b]$ containing the terminals was chosen as follows: Let (x_i, y_i) denote the coordinates of origin or destination i. Let $\overline{x} := \max_{i \in \mathcal{O} \cup \mathcal{D}} x_i$, $\underline{x} := \min_{i \in \mathcal{O} \cup \mathcal{D}} x_i$, $\overline{y} := \max_{i \in \mathcal{O} \cup \mathcal{D}} y_i$, $\underline{y} := \min_{i \in \mathcal{O} \cup \mathcal{D}} y_i$. Then $a := (\overline{x} - \underline{x})/2$ and $b := (\overline{y} - \underline{y})/2$.