A Decision Aid Algorithm for Long-Haul Parcel Transportation based on Hierarchical Network Structure

Camille Gras, Nathalie Herr, Alantha Newman

A Decision Aid Algorithm for Long-Haul Parcel Transportation based on Hierarchical Network Structure

Camille Gras[a], Nathalie Herr[b], Alantha Newman[a]

[a]Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, 38000 Grenoble, France; [b]Probayes, 38330, Montbonnot, France

ARTICLE HISTORY
Compiled November 14, 2022

Abstract
With the explosion of e-commerce, optimizing parcel transportation has become increasingly important. We study the long-haul stage of parcel transportation which takes place between sorting centers and delivery depots and is performed on a two-level hierarchical network. In our case study, we describe the application framework of this industrial problem faced by a French postal company: There are two vehicle types which must be balanced over the network on a daily basis, and there are two possible sorting points for each parcel, which allows a better consolidation of parcels. These industrial constraints are formalized in the Long-Haul Parcel Transportation Problem (LHPTP). We present a Mixed Integer Linear Program (MILP) and a hierarchical algorithm with aggregation of demands which uses the MILP as a subroutine. We perform numerical experiments on large-size datasets provided by a postal company, which consist of approximately 2500 demands on a network of 225 sites. These tests enable the tuning of certain parameters resulting in a tailored heuristic for the LHPTP. Our algorithm can serve as a decision aid tool for transportation managers to build daily transportation plans, modeled on solutions produced given daily demand forecasts, and can also be used to improve the network design.

KEYWORDS
Optimization; Network Design; Long-Haul Transportation; Parcel Transportation; Hierarchical Network

1. Introduction

In 2020, over two billion people purchased goods or services online (Statista, 2021). Each year, this number increases, generating a greater need for efficient transportation of goods, which provides motivation for parcel transportation companies to constantly adapt and optimize their transportation service networks. The parcel transportation process is composed of four steps: collection, long-haul transportation, distribution and delivery (Sebastian, 2012). The optimization of the long-haul transportation on the road network is addressed in this work, which extends a preliminary conference paper (Gras et al., 2021) and PhD thesis (Gras, 2021).

Long-haul transportation is defined as intercity transportation (Crainic, 2003). Other stages of parcel transportation include how parcels reach an initial sorting center

Institute of Engineering Univ. Grenoble Alpes
CONTACT C. G. Author. Email: camille.gras@grenoble-inp.org

(the first-mile collection problem addressed by Wang and Huang 2020), how parcels are delivered from depots to post offices (the distribution problem addressed by Timperio et al. 2020), and how parcels are sent from post offices to final clients (the last-mile delivery problem addressed by Jiang et al. 2020). These different stages can be solved separately, and we only address the long-haul stage of parcel transportation in this paper. As long-haul parcel transportation addresses transportation of large volumes of parcels over long distances, a better use of vehicles and logistics operations such as sorting leads to improved consolidation and economies of scale.

In this work, the Long-Haul Parcel Transportation Problem (LHPTP) is formally defined. We now give a general overview of the problem framework before outlining our contributions. The input of the LHPTP consists of transportation demands, each represented by a triple consisting of an origin, a destination and a number of parcels, and the road network composed of possible origin, destination and intermediate sites and links between them. The objective of the problem is to minimize the costs (transportation costs, operation costs, personal costs and vehicle costs) for the postal company while delivering all of the demands. Here, we consider a predesigned road network for a French postal system of around 225 sites as the application platform for our algorithm. The network has nodes representing two sets of sites, sorting centers and delivery depots, within a two-level (inner and outer level) hierarchical structure which has a hub-and-spoke configuration on each level of the network. The inner level of the network is composed of sorting centers, which can be inner-hubs. These inner-hubs are points at which an extra consolidation of demands with different origins and/or destinations can be applied. The outer level of the network is composed of sorting centers and delivery depots. We call the sorting centers outer-hubs as they are hubs for their corresponding delivery depots. Over this two-level network, demands have to be routed from sorting centers to delivery depots directly or through at most two other sorting centers. Parcel transportation is performed with two types of vehicles: trucks with one or two containers. The number of vehicles (and containers) in the network is unlimited. The output of the LHPTP is a transportation plan for routing the (projected) demands that is meant to be applied on a daily basis over the course of some time period (a year, for instance).

Our contribution is threefold. Firstly the LHPTP, a new mathematical optimization problem, is introduced and formulated. We present a precise definition corresponding to the problem of long-haul transportation on a two-level network with a sorting operation which arises in our case study on a French postal network. We then show how to formulate it as a path-based Mixed Integer Linear Program (MILP) which mixes paths and flows. The second contribution is a solution approach: We present an algorithm, the Hierarchical Aggregation of Demands (HAD) Algorithm to solve the LHPTP. This algorithm is used to find a solution to realistic problem instances from our case study which have been provided by transportation managers and consist of likely daily demand distributions. To solve the LHPTP, the HAD Algorithm exploits the hierarchical structure of the postal network to design good transportation plans by applying a divide-and-conquer approach that yields tractable subproblems small enough to be solved optimally by MILP. This combination of divide-and-conquer and MILP appears to be rare in the case of service network design problems. Additionally, we were inspired by the idea of Baumung and Gündüz (2015) of sending large demands directly and consolidating residual demands. To apply this idea to our approach, we choose to directly send (bypassing all sorting operations) trucks whose load is more than the truck filling rate threshold, rather than only the trucks that are fully filled. We test various possible truck filling rate thresholds in order to measure the impact of this threshold on the quality of the solutions in terms of cost savings and fill rates. The HAD Algorithm produces good globally optimized solutions for the model datasets from our case study faster than solving the MILP with a solver.The third contribution shows how the algorithm can be used as a decision aid tool by transportation managers in two key ways: (i) to design daily transportation plans based on simulations based on various scenarios and (ii) to modify/set certain parameters of the transportation network to achieve better performance. Firstly, the HAD Algorithm is a tool designated for transportation managers to determine a tactical daily strategy to use on a midterm (e.g., yearly) basis. The datasets used for the tests are projections for the future. The HAD Algorithm allows to simulate various scenarios (e.g., daily demand forecasts). The objective is to avoid running the algorithm every day as the upheaval of routing the parcels according to a different strategy from one day to the next would be too great. Rather, the algorithm provides solutions for the model problem which serve as a guide for the transportation managers who make the final decisions about the on-the-ground strategy for the long-haul routing of parcels. In other words, the transportation managers can use the algorithm as a black-box decision aid tool to design transportation plans. The second way of using the HAD Algorithm is to simulate what happens when various network parameters are modified (e.g., choice of intermediate sorting sites) to see which one will result in the cheapest transportation solution. Using the simulations, we test different ratios of sorting cost versus transportation cost to see how it impacts the optimal set of inner-hubs that should be included in the network.

The remainder of this paper is organized as follows. In Section 2, we present a review of related work in order to position the LHPTP in the literature and compare it to other transportation problems. In Section 3, the application framework from which the LHPTP is derived is described and the LHPTP is defined. We also present a mathematical formulation for the problem. Next, in Section 4, an algorithm which solves the LHPTP via the aggregation of demands and the MILP is presented: the HAD Algorithm. In Section 5, we present the test environment and datasets for numerical experiments. The results obtained by running the HAD Algorithm on these large real datasets are discussed in Section 6. In this section, we also discuss how to use the algorithm to optimally set parameters such as the truck filling rate threshold. In Section 7, we consider how transportation managers can use the HAD Algorithm as a decision aid tool to make decisions about some important aspects of the transportation network such as the selection of inner-hubs. Finally, we conclude with some perspectives on further applications in Section 8.

2. Literature Review

In a hierarchical network, the set of sites of the same type (inner-hubs, sorting centers, depots) forms a layer. Each pair of layers constitutes one level of the hierarchical network and is sometimes referred to as an echelon. Two-level networks are a special case of multi-level networks. Surveys on two-level problems have been written by Gonzalez-Feliu (2011) and Cuda, Guastaroba, and Speranza (2015). The most studied two-level problems are the Vehicle Routing Problem (Rahmani, Cherif-Khettaf, and Oulamara, 2016) and the more general Inventory Routing Problem (Farias, Hadj-Hamou, and Yugma, 2021). Hub-and-spoke network are commonly used in the design of long-haul transportation networks. Bryan and O'Kelly (1999) provide a formal definition of a hub-and-spoke network as a network whose links either begin or end at a hub, the other extremities of the links being the spokes, and also provide a comprehensive review of such networks in transportation. O'Kelly (1998) analyzes transportation hub-and-spoke networks and explains why it is sometimes useful to add direct links to them. A hub-and-spoke network in which there are possibilities to bypass the inner-hubs with direct links is called a hybrid hub-and-spoke network (Zäpfel and Wasner, 2002; Zhang, Wu, and Liu, 2007; Lee and Moon, 2014) as opposed to a pure hub-and-spoke network. Zhang, Wu, and Liu (2007) study a hybrid hub-and-spoke network, with demands from sorting center [2] to sorting center; there are three types of vehicles, and they discuss the issue of choosing direct paths versus using hubs.

Zäpfel and Wasner (2002) study the planning and optimization of hybrid hub-and-spoke transportation on the case of parcel delivery in Austria using an MILP. In this problem, there is the same type of vehicle fleet as in the LHPTP, but there is a one-level network, since their demands are from depot to depot. Meisen (2015) works on parcel transportation in Germany (33 hubs and 200 depots). He introduces an MILP, which is an arc-based formulation, in which the next day delivery is the objective of the optimization process. Lee and Moon (2014) study a hybrid hub-and-spoke network in which they select hubs via an MILP which simultaneously solves a hub location problem and optimizes the transportation, on the case of Korea Post. They focus on optimizing the transportation on the inner level of their network and do not address the transportation on the outer level between sorting centers and depots. Zhang, Payan, and Mavris (2021) optimize the long-haul transportation aspect of courier service network efficiency in the USA on a hub-and-spoke network with consolidation. They present an Integer Linear Program (ILP) to enhance consolidation for courier service network design. Their two-level network has a limited fleet to manage and combines air and road transportation.

The Hub-and-Spoke Network Design (HSND) problem, sometimes called the hub network design problem (see O'Kelly and Miller 1994) and originally formulated by O'Kelly (1987) as a quadratic integer programming model, involves four steps: (i) optimizing the locations of the hubs; (ii) linking non-hub sites to hubs; (iii) creating links between the hubs; and (iv) routing flows through this network. In the LHPTP, the focus is on the last three steps.

Lin and Chen (2004) give a definition of a *hierarchical hub-and-spoke network*: it is a network in which each site is assigned to exactly one hub. It is similar to our network but does not contain any inner-hubs. They address the time-constrained hierarchical hub-and-spoke network design problem as they consider simultaneously the time constraint and the design of routes, fleet size and schedules. Their case study is on the ground-exclusive network of an express common carrier in Taiwan. A review of the state of the art of HSND has been presented by O'Kelly (1998) and other HSND problems have been studied by Meng and Wang (2011) and Lin and Chen (2008). Hu, Askin, and Hu (2019) study a hub relay network design problem for long-haul transportation and optimize the driver routes. Baumung and Gündüz (2015) studied an HSND problem for parcel distribution tested on the data of Australia Post (from 10 to 50 sorting centers). Their heuristic consists in sending trucks fully filled with parcels for the same destination (sorting center) until there are not enough parcels at the origin sites to fill a truck. The remaining parcels are called *residual volumes* and these volumes are consolidated into trucks. They define a problem for each set of demands (the main volumes and the residual volumes) and each problem is solved separately. The main volumes are sent in fully filled trucks from sorting center to sorting center and the routing of the remaining parcels, which constitutes a so-called Residual Volumes Hub

[2] A sorting center refers to a site where parcels are (re)sorted and dispatched in containers.

Location Problem, is optimized with an MILP. Note that they consider one vehicle type, one level of network (consisting of sorting centers and hubs) and they do not mention balancing the empty vehicles.

The closest problem to the LHPTP in the literature is the problem studied by Cohn et al. (2007). They study parcel delivery on the US freight transportation network (for UPS), which is modeled as a one level network. They present an arc-based model on a time-space network which they solve via an integer linear program with a non-linear cost structure. Indeed, as parcels with the same origin and destination must use a single path, their model corresponds to an integer linear program.

While most of the key aspects of the LHPTP appear in the parcel transportation literature, they have each been studied separately and not considered together. For example, in the LHPTP there is more than one vehicle type (as in Lee and Moon 2014; Zäpfel and Wasner 2002; Zhang, Wu, and Liu 2007; Zhang, Payan, and Mavris 2021), and there is a two-level network (as in Lee and Moon 2014; Zhang, Payan, and Mavris 2021). The costs of parcel long-haul transportation are minimized on both levels simultaneously (unlike Lee and Moon 2014 which optimizes only the core level). A key feature of the LHPTP is that there is no tour that visits a sorting center and all its associated delivery depots due to bulk transportation requirements (see Section 3.1 for more details). In previous work, this transportation stage often involves a tour (Zäpfel and Wasner, 2002). In the sorting centers, sorting capacity is not considered (unlike Lee and Moon 2014; Baumung and Gündüz 2015) as these sites have appropriately-sized sorting capacity. We simultaneously optimize the long-haul transportation and the balance of vehicles over the course of day on a two-level network with splittable demands which is an original aspect in the scope of parcel transportation to the best of our knowledge.

Moreover, the models presented in the literature related to the LHPTP do not present a systematic approach to the general problem of long-haul parcel transportation. One of the main differences, for example, between previous work and our model is that other works treat the long-haul stage of parcel delivery as a single level; there is very little treatment of two-level networks with logistics operations on parcels permitted between the levels. Furthermore, in a two-level network, some issues arise which do not apply to a simpler model; for instance, balancing vehicles between intermediate sites (i.e., not origin/destination sites).

The LHPTP is actually a family of problems based on various parameters (e.g., allowed logistic operations, limited fleet versus unlimited fleet, vehicle balancing constraints, vehicle types, arc capacity, etc.). In the routing literature, there is no single model that can be used to accurately represent this family of problems.

3. The Long-Haul Parcel Transportation Problem

The LHPTP is basically a Service Network Design problem (Crainic, 2000) with Hub-and-Spoke structure. It has additional distinct properties and strong industrial constraints such as the fact that demands have both fixed origin and destination and the sorting operation has a cost per parcel. This prevents us from simply applying an existing model proposed for another Service Network Design problem in the literature. After the presentation of the application framework of the problem, we introduce a specific MILP for the LHPTP, which can be adapted to solve other long-haul parcel transportation problems (e.g., with different vehicle types, without vehicle balancing, or with sorting capacity constraints, etc.).

3.1. *Application Framework and Problem Description*

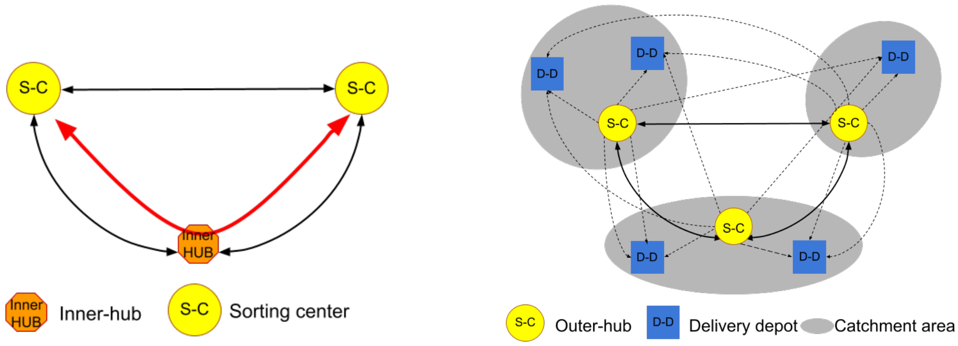
The problem of long-haul parcel transportation via trucks with one or two containers is formulated based on a case study on a French national postal network composed of around 225 sites with around 2500 demands being routed each day between these sites. The *sites* are divided into two sets: *sorting centers* and *delivery depots*. Parcels and vehicles start at sorting centers, while delivery depots are the destination of parcels. In the datasets provided by the postal company and used for the tests, there are on average 17 sorting centers and 208 delivery depots. Some of the sorting centers are also inner-hubs selected by the transportation managers.

The *parcels* are physical objects which must be routed from a designated origin sorting center to a specified delivery depot. Each container has the same volume and can transport a large number of parcels, and each parcel is assumed to be the same size, which is a very small size compared to the volume. A *demand* is constituted of all the parcels which share the same origin and destination and is represented by the triple consisting of an origin, a destination and a number of parcels. A daily demand forecast (the same each day) is given as input to the LHPTP. (In our case study, several demand forecasts that model "average days" are provided by transportation managers.) The objective is to design a transportation plan to deliver all the demands at a minimum cost (see objective function 1a). The cost of a solution is composed of the costs of the logistics operations (e.g., sorting) and costs of transportation (e.g., fuel, vehicles and drivers' salary, etc.).

In this paper, only one logistics operation is incorporated: the *sorting*. It occurs in sorting centers. After a sorting, all parcels in the same container share the same intermediate destination on their operational path. *Consolidation* is when demands destined for different final sites are put in the same container. The parcels are shipped in bulk in containers. As a container cannot be partially unloaded, it transports parcels to a single site, which is either the final delivery depot or an intermediate sorting center for each of these parcels. This means when consolidation between demands occurs, there is always a sorting operation which comes afterwards: sorting centers are breakbulk terminals (Crainic, 2003).

The parcels are transported on a road network composed of two hierarchically nested hub-and-spoke networks (see Figure 1). The *inner level* of this two-level network is composed of sorting centers. A sorting center on an operational path between two sorting centers is called an *inner-hub*. The *outer level* of the network is made of sorting centers and delivery depots. Indeed, each delivery depot is affiliated with a sorting center called an outer-hub, usually the closest (with respect to travel time). A *catchment area* is the area which contains the delivery depots affiliated with a sorting center, including this sorting center. The mapping of the catchment areas is created with the experience of transportation managers. This mapping is used in the current operational strategy which handles the parcels regionally. The parcels are sent to the catchment area containing their delivery depot, sorted in their corresponding outer-hub and sent to their delivery depot to be transported in the last-mile portion.

The inner network concerns the sorting centers. It is a hybrid hub-and-spoke network as shown in Figure 1a. Parcels can either be sent directly from their origin sorting center to the sorting center corresponding to their destination before they reach their delivery depots. Or they can be consolidated at an inner-hub and then sorted at a sorting center before they reach their delivery depots. Consolidation of demands allows to better fill trucks and reduce the number of trucks required, thus it results in costs savings. Figure 1b shows the outer network, composed of sorting centers and delivery depots



(a) Inner hybrid hub-and-spoke network. (b) Outer hybrid hub-and-spoke network.

Figure 1. Caption: The two levels of a two-level hybrid hub-and-spoke network.
Figure 1. Alt Text: Two pictures of directed graphs to depict the two levels of two-level hybrid hub-and-spoke network, one picture corresponding to each level. The first graph has three nodes representing three sorting centers, one of which is a inner hub that can be used to route parcels between the other two sorting centers. The second graph uses edges to show the connection between the delivery depots (represented by blue squares) to their respective sorting centers (represented by yellow circles). The grey areas depict catchment areas.

(the final destinations). These delivery depots (D-D in Figure 1b) are the spokes of the hybrid hub-and-spoke network in which the outer-hubs are the associated sorting centers. Thus, the key point of Figure 1 is that parcels can be consolidated once or twice, depending on whether or not they use an inner-hub.

Two types of *vehicles* transport the parcels on the network: trucks with one container and trucks with two containers (also called twin trailers). The number of vehicles and the number of containers are unlimited. These quantities can be determined appropriately in the optimization process to suit the requirements. All vehicles are trucks, and the travel time for a fixed path is the same for all of them. The vehicle type (i.e., one or two containers) affects only its capacity and cost.

Links are used by vehicles to go from one site to another in the network (they can also be viewed as arcs in a graph). An *operational link* is defined as a link from one site to another site associated with a vehicle type and a specified time period. The cost of an operational link is the cost for the vehicle to traverse the link. An *operational path* is defined as a set of consecutive operational links between sites in which each site is associated with an operation (sorting or delivery) performed on a parcel flow at a time slot. The cost of a path is the cost of the sortings on this path which depends on the number of parcels sorted. Each parcel can either be (i) sent on a direct path from its origin sorting center to a delivery depot, or (ii) it undergoes a single sorting at a sorting center (not necessarily an inner-hub), or (iii) it undergoes two sortings, the first one at an inner-hub and the last one at the sorting center to which its delivery depot is associated (see Figure 2). In the LHPTP, each demand can be split over multiple operational paths in the transportation plan: this is called *disaggregate shipping* (Leung, Magnanti, and Singhal, 1990).

The parcels are required to be delivered in one or two days (respectively called D+1 and D+2 deliveries). Thus a transportation plan includes demands from the previous day that require two days to reach their destination. This transportation plan is sometimes called daily transportation plan. We remark that to design a daily transportation plan, it is necessary to ensure that enough vehicles and containers are available each day on each site to accommodate all the parcels. This is guaranteed by enforcing the number of outgoing and incoming vehicles to be equal for each site

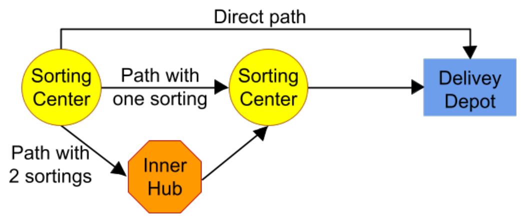


Figure 2. Caption: The three possible types of operational paths for routing parcels.
Figure 2. Alt Text: A picture of a directed graph illustrating the three types of operational paths. The graph has four vertices denoted by two yellow circles (sorting centers), one blue square (delivery depot) and one inner hub (orange octagon). The three path types are (i) a direct link from a yellow circle to the blue square, (ii) a path from one yellow circle to the yellow circle ending at the blue square, and (iii) a path from a yellow circle to orange octagon to the other yellow circle ending at the blue square.

in the course of a day. To do this, we use balance constraints on vehicles and more specifically, on vehicle types (i.e., that in and out flow are equal at each site). This sometimes requires the transport or *repositioning* of empty vehicles.

The LHPTP can now be defined. The input of this optimization problem consists of:

- A transportation network with sites, their types and the distances between them;
- Demands between these sites (from sorting centers to delivery depots in our case);
- Operation and transportation costs for each vehicle type;
- Capacities of vehicles.

The output of the LHPTP is a *transportation plan* for the input demands which consists of the operational path(s) for each demand and the number of vehicles with one or two containers on each operational link. A solution for the LHPTP is required to deliver all the parcels in a given network while respecting the vehicle capacity and balance constraints (discussed next). The objective is to minimize the sorting and transportation costs.

3.2. *Mathematical Formulation*

The MILP formulation for the LHPTP is presented here. Since the definition of the problem is based on possible operational paths, we propose a path-based model. This model is designed in a way which allows to easily monitor the possible paths proposed to the solver, especially when it comes to removing and/or adding some path types.

Let us define some notations:

- D is the set of demands. A demand is a triple (s_d, t_d, v_d), where s_d is the origin, t_d is the destination and v_d is the volume (i.e., number of parcels to be routed from s_d to t_d).
- S is the set of all the sites (i.e., sorting centers and delivery depots),
- L_{veh} is the set of operational links between sites whose element $l_{i,j}$ is the link from site i to site j. $L_{veh} = L_{1c} \cup L_{2c}$ where L_{1c} is the set of operational links for vehicles with one container and L_{2c} is the set of operational links for vehicles with two containers,
- P_d is the set of possible operational paths for the demand $d \in D$,
- P_d^l is the set of possible operational paths using the operational link $l \in L_{veh}$ for demand $d \in D$,
- V is the set of vehicle types.

There are two types of variables, for parcel flows and vehicle flows:

- $x_p^d \in [0,1]$, with $d \in D$ and $p \in P_d$, represent parcel flows. It is the percentage of a demand d using an operational path p.
- $y_l \in \mathbb{N}$, with $l \in L_{veh}$, represent vehicle flows. They are integers which represent the number of vehicles of the type veh on each operational link l.

Let us also denote:

- c_p the cost of using the operational path p, (associated with the demand sharing origin and destination with p),
- c_l the cost of using the operational link l for a vehicle type,
- C_{1c} the capacity of vehicles with one container, C_{2c} the capacity of vehicles with two containers (i.e., $C_{2c} = 2 \times C_{1c}$).

With this notation, our model, denoted *LHPTP-MILP*, can be described as follows.

$$\min \sum_{d \in D} \sum_{p \in P_d} c_p x_p^d + \sum_{veh \in V} \sum_{l \in L_{veh}} c_l y_l \tag{1a}$$

s.t.:

$$\forall\, d \in D, \sum_{p \in P_d} x_p^d = 1 \tag{1b}$$

$$\begin{aligned} \forall\, l \in L_{1c}, \sum_{d \in D} \sum_{p \in P_d^l} v_d x_p^d \leqslant y_l \cdot C_{1c} \\ \forall\, l \in L_{2c}, \sum_{d \in D} \sum_{p \in P_d^l} v_d x_p^d \leqslant y_l \cdot C_{2c} \end{aligned} \tag{1c}$$

$$\begin{aligned} \forall\, s \in S, \sum_{i \in S} \sum_{l_{i,s} \in L_{1c}} y_{l_{i,s}} = \sum_{j \in S} \sum_{l_{s,j} \in L_{1c}} y_{l_{s,j}} \\ \forall\, s \in S, \sum_{i \in S} \sum_{l_{i,s} \in L_{2c}} y_{l_{i,s}} = \sum_{j \in S} \sum_{l_{s,j} \in L_{2c}} y_{l_{s,j}} \end{aligned} \tag{1d}$$

with:

$$x_p^d \in [0,1] \tag{1e}$$

$$y_l \in \mathbb{N} \tag{1f}$$

The LHPTP-MILP is composed of the objective (1a) and three major types of constraints: the delivery constraints, the capacity constraints and the design-balance constraints. The delivery constraints (1b) declare that all the parcels must be delivered. The link capacity constraints (1c) associate the number of vehicles and the number of parcels on each operational link. Finally, the design-balance constraints (see 1d) balance trucks between sites on a daily basis. Since it takes the vehicle type into account, this results in a balancing of containers. Note that we do not require that a vehicle be returned to its origin. (We use MILP* to denote the LHPTP-MILP without the design-balance constraints (1d), since we will not need them for some subproblems.)

A container can carry parcels with different origins and destinations; thus an optimal solution might not route each demand on its shortest path. Therefore the LHPTP is not a shortest path problem due to the usage of consolidation. Indeed, the usage of consolidation, which is a crucial tool used to reduce costs, makes the problem computationally difficult. Different sorting centers and different time slots can be used for the sortings. Furthermore, the possibility of disaggregate shipping adds to the complexity of the problem and this results in a combinatorial explosion on the number of possible operational paths, even though the path length is bounded. This makes it impossible to find an optimal solution with the MILP on realistic sized datasets (with around 225 sites and 2500 demands) in a reasonable time frame.

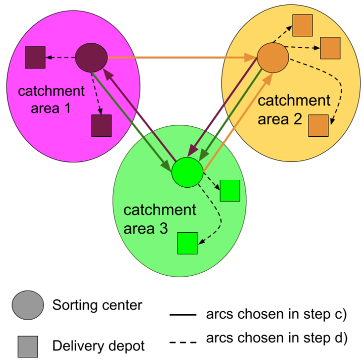
4. Hierarchical Aggregation of Demands Algorithm

As seen in the previous section, the LHPTP can be formulated as an MILP, and an optimal solution for this MILP is an optimal solution for the LHPTP. However, this MILP cannot be solved in reasonable time when considering realistic sized datasets. Thus, we propose the HAD Algorithm, which divides the problem into smaller subproblems, each of which can usually be solved optimally via the MILP or by other means. Then, it combines the solutions to these subproblems to make a final solution of good quality.

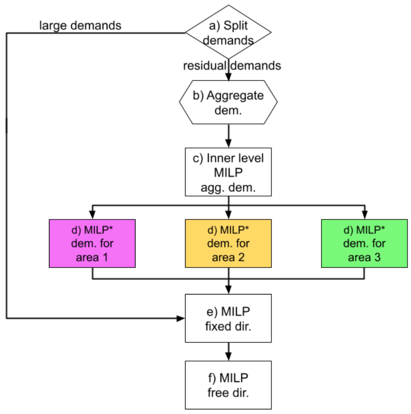
The idea of the proposed algorithm is first to find an optimal transportation plan for the inner level of the network (solid arcs and the sites they connect in Figure 3a), which involves choosing which inner-hubs to use. Then this transportation plan is extended to the outer level of the network (dotted arcs and the sites they connect in Figure 3a), and in the last steps of the algorithm, these solutions are combined and refined to obtain a transportation plan for the whole network. Figure 3b gives an overview of the steps of the HAD algorithm.

In the LHPTP, the goal is to transport the demands from sorting centers to delivery depots. Hence, for the inner subproblem, we create aggregate demands from sorting centers to sorting centers in order to dissociate the two levels. An aggregate demand is the sum of demands from a sorting center to all the corresponding delivery depots of another (destination) sorting center. It is possible to aggregate the demands as the last sorting has to be done in the sorting center of the catchment area of the delivery depot of destination. However, not all demands are included in these aggregate demands. We consider also some direct paths which might be selected in an optimal solution. The LHPTP has two levels which can both be bypassed by these direct paths, and a heuristic based on a filling rate threshold σ is used to optimize this decision. A different method (namely, the MILP) determines how many sorting are used for parcels routed on the inner level (not via direct paths). Indeed, as the optimization on the inner level with the MILP reaches a 1% gap, there is no need for a heuristic to decide if an aggregate demand (from sorting to sorting center) needs to be consolidated in an inner-hub or not.

The input of the HAD Algorithm is the input of the optimization problem plus the truck filling rate threshold σ, where $\sigma \in [0, 1]$. We emphasize that determining the value of σ is key to having a low cost global solution. Here, this value is found empirically by testing different thresholds (see Section 6). The output of the HAD Algorithm is the transportation plan defined in Section 3. We define the capacity of the smallest



(a) Example of network with 3 catchment areas.



(b) Flow chart for the HAD Algorithm.

Figure 3. Caption: A juxtaposition of the topological and chronological evolution of the HAD Algorithm. **Figure 3. Alt Text:** The first figure shows three colored circles, each corresponding to a catchment area. The fuschia and green circles each contain two (darker) squares and one circle corresponding to its two delivery depots and sorting center, which are connected by dotted lines. The orange circle contains three squares and one circle. The second figure is a flow chart with the six steps of the HAD algorithm, where each square corresponds to solving an MILP. Step d of the flow chart is depicted with fuschia, green and orange squares, and each color corresponds to a subproblem in the first figure.

vehicle and denote it C. (Note that in our case $C = C_{1c}$, i.e., C is the capacity of a container.) The pseudocode for the HAD Algorithm is presented in Algorithm 1. The stages of the HAD Algorithm (1) are now detailed.

a) Splitting Demand Volumes into Two Sets

The first step consists of splitting the demand volumes into two sets. The first set of demands is the large demands, which are those whose volume is above a given threshold $\sigma \cdot C$. The remaining set of demands is called the residual demands (Baumung and Gündüz, 2015). Define $k := \lceil v_d/C \rceil$ with v_d being the volume of demand d. For each demand d, if $v_d \leq \sigma \cdot C$, then demand d is a residual demand. Otherwise, if $v_d \in [k \cdot \sigma \cdot C, k \cdot C]$, then d is a large demand as it is possible to use k containers. In the last case, $v_d \in [(k-1) \cdot C, k \cdot \sigma \cdot C]$, the demand d is divided into twin demands: A large demand of volume $(k-1) \cdot C$ and a residual demand of volume $v_d - (k-1) \cdot C$. Note that sometimes $k \cdot \sigma \cdot C < (k-1) \cdot C$ and in this case, v_d is a large demand.

For instance, suppose the given threshold is 60% of the volume of a container, which is 1000. And suppose there is a demand whose volume of 1100 is a bit larger than the capacity of a container. Then this demand is split into one large demand with volume 1000 and one residual demand with volume 100. In the set of residual demands, there are demands from sorting centers to delivery depots which are smaller than the given threshold. Large demands are unique in their set but they can have a twin residual demand.

The two sets of demands are handled differently in the algorithm. The large demands are set aside and possibly routed directly at the end of the algorithm (see Steps e and f of the HAD Algorithm). The residual demands are routed according to the solution of MILP (see Equation 1 and Steps b to d). Note that the routing of the large demands is determined only after determining the routing of the residual demands: Either they are routed directly, or they are combined and sent with the residual demands if there is leftover capacity.

Algorithm 1: Hierarchical Aggregation of Demands Algorithm

Input: Transportation network with sites (and their types) and distances between them; demands between these sites; operations and transportation costs; a truck filling rate threshold σ and vehicle capacity C.
Output: Transportation plan for the demands.

a) Split demands

for each demand d **do**
  if $v_d < \sigma \cdot C$ **then Add** d to the set of residual demands
  else $k = \lceil v_d/C \rceil$
    if $k \cdot \sigma \cdot C \leq v_d \leq k \cdot C$ **then Add** d to the set of large demands
    else Split d into a large demand of volume $(k-1)C$ and a residual demand (twin demands)

b) Aggregate residual demands

for each residual demand d **do**
  Aggregate it with the demands with the same associated sorting center as destination

c) Solve the aggregate subproblem (inner level)

Solve the MILP which is made only of operational links between sorting centers

d) Extend for each catchment area (outer level)

for each one of the n catchment areas **do**
  Solve MILP* (MILP without constraints 1d) to route the demands towards their catchment area

e) Add the solutions of all residual demands, large demands (on direct paths) and integrate the repositioning of empty vehicles (global level)

Solve the MILP with all the residual demands and with only the chosen operational paths to optimize the vehicle flow. The large demands are enforced to use a direct paths. (for up to one hour)

f) Optimize the large demands

Solve the MILP with all the demands, the solution of the previous MILP fixed and the option for the large demands to follow a direct path or an already chosen operational path. (for up to one hour)

b) Aggregating the Residual Demands

In order to optimize the routing on the two levels of the network separately, we need to have demands on the inner level of the network and demands for both levels. To achieve this, we force the last sorting for a parcel to be done in the sorting center associated with the delivery depot of its destination. This requirement allows to aggregate residual demands which are sent to a common outer-hub.

The aggregate demands represent the addition of the residual demands from a single origin to all the corresponding delivery depots of each sorting center. An aggregate demand is denoted d_{agg}. The set of delivery depots in the catchment area of the sorting center j is denoted S^j. The volume of the aggregate demand from the initial sorting center i to the sorting center j is $v^{i,j}_{d_{agg}} = \sum\limits_{d: s_d = i, t_d \in S^j} v_d$. When the origin and destination are clear from context, this volume can be denoted $v_{d_{agg}}$.

For instance, one can consider two sorting centers A and B (see Figure 4). All residual demands originated from A and destined for a delivery depot in the catchment area of B are aggregated. This results in an aggregate demand from A to B whose volume is the sum of the volumes of these residual demands. Note that all aggregate demands are truncated demands which end up in a sorting center (whereas original demands end in a delivery depot).

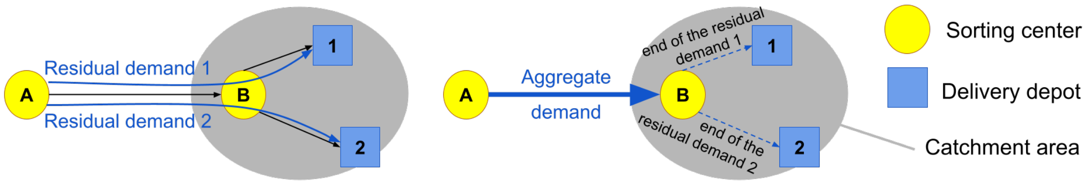


Figure 4. Caption: Aggregation of demands by catchment area.
Figure 4. Alt Text: A picture showing two sorting centers (A and B), two delivery depots and one catchment area containing B and the two delivery depots. The left-hand and right-hand figures show the demands before and after aggregation, respectively.

c) Finding the Routing for Aggregate (Residual) Demands Between Sorting Centers

Considering only aggregate (residual) demands from sorting center to sorting center, the problem size is reduced as there are fewer possible destinations. Now the same problem of routing parcels is solved on a much smaller network of around 20 sites. Thus, it is possible to use the MILP (1). Note that this formulation includes the design-balance constraint (1d). The MILP applied on the network composed of the sorting centers and the aggregate demands provides a transportation plan for these truncated demands.

d) Extending the Routing for Each Catchment Area

At this point, at least one operational path(s) has been chosen for each aggregate demand. In this step, the aggregate demands are disaggregated and turned back into residual demands ending in delivery depots. One or more operational path has been chosen for each group of residual demands from its origin to the sorting center associated with their destination and these paths have to be completed so that each parcel reaches its final destination. Two reasons makes this non-trivial: (i) paths between sorting centers can be shared by several aggregate demands, and (ii) paths between sorting centers reserved for each aggregate demand can be shared by several residual demands.

(i) If in Step c, two aggregate demands were sharing a vehicle on an operational link, and one of these aggregate demands is split over more than one operational paths (for example, if 20% of an aggregate demand A is sent on an operational path which shares an operational link with the operational path selected for an aggregate demand B), then the parcels for one of these two demands (demand A in this example) should not completely use the vehicle capacity allowed on this operational link. It could happen as the catchment areas are solved separately and that would be a problem in the next steps. Thus parcels need to be enforced to use the operational links chosen for each aggregate demand by the MILP of Step c with respect of the capacities fixed in this MILP (to avoid mixing the catchment areas).

(ii) The global capacity of the paths chosen for the aggregate demands has to be reassigned and it has to be shared between residual demands as it was in Step c.

Indeed, the capacity reserved on each path may be used by several residual demands. For example, let us consider an aggregate demand A which is disaggregated in residual demands A_1 and A_2. If in Step c the paths p_a and p_b have been chosen for A, the residual demand A_1 might use both paths p_a and p_b while A_2 only uses the path p_b. In this case, the capacity reserved for the path p_b must be shared.

To take into account these two points, the following new constraint is added to the MILP in which $P^{ext}_{p_{agg}}$ denotes the set of operational paths extending the aggregate path p_{agg}, which is a path associated to the aggregate demand d_{agg}. We recall that any operational path p is associated to one specific demand d, which can be denoted as $d(p)$. The term $x^{d_{agg}}_{p_{agg}}$ is defined analogously as x^d_p in Section 3.2. It represents the rate of the parcel flow of the demand d_{agg} on the path p_{agg}.

$$\forall\, x^{d_{agg}}_{p_{agg}},\ \sum_{p \in P^{ext}_{p_{agg}}} \sum_{d=d(p)} v_d x^d_p \leq v_{d_{agg}} x^{d_{agg}}_{p_{agg}}. \tag{2}$$

This problem is solved separately for each catchment area. For each set of residual demands, whose final destination belongs to the same catchment area, MILP* (i. e., the MILP without constraint 1d) is used to extend the operational path (from Step c) from final sorting center to final destination. For each subproblem, the network of all sorting centers (network considered in the previous step) and delivery depots of the relevant catchment area is considered. The variables are restricted so that the operational paths output by the MILP* solution follow the operational paths already selected in the previous step. MILP* does not include the design-balance constraint, in order to optimize globally the repositioning of empty vehicles (and not by catchment area). Recall that the design-balance constraint means that the number of outgoing and incoming trucks must be equal for each site in the course of a day (e.g., it ensures that each vehicle is returned to some sorting center).

At the end of this step, operational paths for all the residual demands have been computed. But the algorithm is not finished yet, since these operational paths are in several transportation plans, one for each catchment area, and the objective is to design a single global transportation plan.

e) Merging the Residual Demands Solutions with the Large Demands on Direct Paths and Integrating the Repositioning of Empty Vehicles

In this step, an MILP is run to select direct paths for the large demands and simultaneously merge these paths with the paths for the residual demands from each catchment area (chosen in Step d) while respecting the repositioning of empty vehicles. We fix in the MILP the operational paths variables for parcels (the x^d_p variables in the MILP) and the variables representing the number of vehicles on each operational link (the y_l variables in the MILP) accumulated in the previous steps. Indeed, the previous steps have optimized the inner level and these operational links are not recomputed. The operational links which specify how parcels end up in their respective delivery depot are also not recomputed. Indeed, one catchment area (or the inner level) is optimized in each subproblem, thus these operational links cannot be used in two separate subproblems and there is no need to recompute these operational links to properly merge the solutions of the subproblems.

We now have a global solution which routes all the demands. In this step, the repositioning of empty vehicles from delivery depots to sorting centers is optimized as the design-balance constraint of the MILP is activated. Two reasons motivate this choice: It allows first to optimize simultaneously the repositioning of empty vehicles and the direct paths and secondly to have a transportation plan which answers the original problem which can be compared with the one obtained at the end of the next step. Note that, there is more than one direct path option for each demand (because of the two types of vehicles). Thus there is an optimization done by the MILP solver for the large demands even if they are allowed to use only direct paths. At the end of this step, there is a global solution built from the solutions to the subproblems solved in the previous steps, with the large demands sent on direct paths.

f) Optimizing Operational Paths for the Large Demands

In this step, the variables corresponding to the solution obtained at the previous step for the residual demands are fixed, and the large demands are allowed to follow either a direct path or the operational path used by their residual twin demand when there is enough "leftover" space.

5. Test Environment and Datasets

The HAD Algorithm is implemented on a Linux server with 32 CPU and 150 Gbytes of RAM along with CPLEX 12.8 solver. Six realistic datasets provided by a postal company are used. These datasets represent six different configurations of the network with logistics sites spread across mainland France. A dataset consists of:

- a set of sites with their location, type and the travel times between each pair of sites;
- a set of demands between these sites;
- costs: kilometric cost and the sorting costs.

In these datasets, the number of sites ranges from 154 to 292 and the number of demands ranges from 2312 to 3627. The kilometric cost for a vehicle is three times the cost to sort one parcel in a sorting center. More information about the datasets can be found in Gras (2021).

The MILP solver is allowed to run for up to one hour to solve each subproblem (inner level subproblem, subproblems to extend to each catchment area and merging subproblems). The solver can stop before the one hour of computation if it reaches a gap of 0.01%. This is nearly always the case for the subproblem addressing the extension to the catchment areas (Step d), which usually takes less than one minute to reach this gap. As it is possible to fix nearly all the variables before entering the merging stage of the algorithm, it requires only a few seconds to reach this 0.01% gap. Thus, the total computing time is much less than the maximum time allowed. The only subproblem which systematically takes one hour of computation is the inner level problem. The time to build the models for each subproblem and to save the results are not taken into account as they are negligible compared to the execution time of the algorithm.

It is effective to divide instances from the case study into smaller problems which can be solved optimally with exact methods. It allows for solving larger instances of the problem as the HAD Algorithm yields solutions more quickly and with lower cost than those generated by directly solving the global MILP using only a solver. Next, we discuss how the results from the experiments can be used to guide the choice of parameter settings for the HAD Algorithm.

6. Results on Algorithm Performance and Fill Rate

The goal of this section is to test the performance of the HAD Algorithm. An input for the algorithm is the truck filling rate threshold σ. Thus, in these experiments, whose results are presented in Table 1, we test different values for σ, which is the threshold for splitting demand volumes into large demands and residual ones. Note that the truck filling rate threshold is used to designate demands as large or residual (where the former are sent on direct paths), while actual fill rate is used to denote the average capacity of trucks used over the whole network. All values of σ from 10% to 100% are tested (with a step of 10, i.e., 20%, 30%, etc.) on each of the six datasets. Therefore Table 1 is based on 60 runs of the HAD Algorithm, 10 runs for each dataset. Each row gives the minimum, average and maximum values of the respective results over the six datasets for a fixed truck filling rate threshold. In addition, the first two rows of Table 1 correspond to simulations solving the global MILP without any heuristic (rows “none" in Table 1). Now we explain what each of the seven columns in Table 1 shows.

The first column states the truck filling rate threshold σ. The second column shows the computing time (min, avg, max) of the complete algorithm over the six datasets. The third column, called Sol fixed dir. (solution with fixed direct paths), shows the cost of the solution found at the end of Step e of the HAD Algorithm. The fourth column, called Built solution, presents the cost of the final solution which is the output of the MILP for the first two rows, and the output at the end of Step f for the last 10 rows. The next column presents the gap which represents the distance to the best lower bound computed by the solver for an exact resolution after a six-hour run (first rows of Table 1). The sixth and seventh columns give the fill rate not including and including (respectively) the empty vehicles. Note that in these experiments, each MILP run within the HAD Algorithm have a one hour time limit and all sorting centers are considered as inner-hubs.

Table 1. Comparison of the truck filling rate thresholds

Thres.	Time (h)			Sol fixed dir. e)			Built sol f)			Gap (%)			Act. fill rate (no empty trucks) (%)			Global fill rate (%)		
(%)	min	**avg**	max	min	**avg**	max	min	**avg**	max	min	**avg**	max	min	**avg**	max	min	**avg**	max
none	1	**1**	1	—			377	**477**	807	11.6	**21.5**	58.7	47.5	**66.5**	72.2	31.6	**41.0**	43.7
none	6	**6**	6	—			377	**406**	426	10.2	**13.7**	17.4	70.0	**71.3**	73.2	42.7	**43.4**	44.4
100	1.0	**1.1**	1.3	376	**415**	444	376	**414**	442	13.4	**15.5**	16.7	82.6	**84.0**	86.0	52.4	**54.1**	55.7
90	1.0	**1.1**	1.2	375	**411**	437	375	**411**	436	13.2	**14.7**	15.9	82.1	**84.2**	86.4	52.0	**54.0**	55.5
80	1.0	**1.2**	1.2	374	**406**	429	374	**406**	429	12.2	**13.8**	14.9	81.6	**83.9**	85.9	51.4	**53.4**	54.8
70	1.0	**1.1**	1.2	369	**405**	437	369	**405**	435	11.8	**13.5**	15.1	80.2	**82.4**	84.9	50.8	**52.1**	53.6
60	1.0	**1.1**	1.2	367	**399**	419	367	**399**	419	10.1	**12.2**	13.9	79.5	**81.8**	83.6	49.3	**50.9**	52.0
50	1.1	**1.2**	1.3	374	**401**	419	374	**401**	419	10.0	**12.7**	14.3	76.7	**78.7**	80.3	47.4	**48.3**	48.7
40	1.1	**1.2**	1.5	374	**406**	426	374	**406**	426	11.5	**13.8**	15.7	75.4	**76.0**	77.1	45.5	**45.7**	46.4
30	1.0	**1.2**	1.5	393	**421**	435	393	**421**	435	13.3	**16.9**	19.1	70.1	**70.9**	72.7	41.1	**41.6**	42.1
20	1.0	**1.3**	1.5	436	**463**	480	436	**463**	480	19.9	**24.5**	27.7	60.1	**61.5**	63.3	34.1	**34.7**	35.4
10	1.1	**1.3**	1.5	559	**583**	627	559	**583**	627	33.0	**39.9**	44.6	42.7	**45.6**	49.4	23.0	**24.6**	26.6

From Table 1, we observe that the HAD Algorithm can provide better solutions in five to six times less computation time compared to the LHPTP-MILP when run for six hours without any heuristic (second row with threshold none in Table 1). This holds when the truck filling rate threshold of 60% is used, for which the HAD Algorithm exhibits the best performance in terms of cost. Note that this algorithm provides a solution for the complete instance in reasonable time but might be globally suboptimal because the choice between performing at least one sorting or sending the parcel on a direct path is handled in a heuristic way. Thus, for instance, for a parcel sent directly, some possible operational paths which could be chosen in a globally optimal solution are not considered in the HAD Algorithm. Moreover, the solutions obtained in Step e and Step f are nearly the same. (As a reminder, in Step e, the large demands are forced to use direct paths, while in Step f they can use the path(s) of their twin residual demand.) This shows the natural "greedy" choices made when using direct paths are locally almost optimal, since they are not changed much by Step f.

To apply the HAD Algorithm, we need to determine the setting of the parameter σ which yields the best results. (This can also depend on the settings of other parameters such as, for example, sorting costs.) Table 1 shows the truck filling rate threshold which provides the best results is $60\% \pm 10\%$. When the truck filling rate threshold is 60%, the actual fill rates with and without considering empty vehicles are respectively 50.9% and 81.8%. We refer to the former as *global fill rate* and to the latter as the *actual fill rate*. Slightly higher actual fill rates (like 84.2%) can be obtained with higher filling rate thresholds but at the expense of sorting for more consolidation. One might think that maximizing the actual fill rate of trucks would lead to cost minimization, because when trucks are more filled, fewer trucks are needed, but this is not the case. The results displayed in Table 1 show that the higher actual fill rate does not correspond to the lowest cost solution. In fact, it is profitable to send a truck filled at 60% (seemingly suboptimal) on a direct path as it allows to save on both repositioning of empty vehicles and sorting costs. Note that the optimal truck filling rate threshold computed empirically depends on the datasets, and would have to be recomputed if the HAD Algorithm were to be used on another dataset.

7. Utilizing the HAD Algorithm as a Decision Aid Tool for Inner-Hub Selection

The HAD algorithm is a decision aid tool for making managerial decisions. The transportation managers can use the solution obtained via the algorithm as a guide to create a routing of parcels for an actual input demand profile. If the daily demand profile changes, implementing a completely new solution each day is too expensive in terms of running the algorithm and also in terms of implementing its output (e.g., the cost of upheaval/changing a solution from day to day). Thus, the algorithm is not meant to be run every day (i.e. for all data sets). To employ the solution as a guide, the transportation managers can analyze the key performance indicators of the best solution proposed for the most "similar" data set and use it to make the decision to open new paths (or not) in the strategy applied on the ground.

The HAD Algorithm can also be a useful management tool to make decisions about the design of the postal network. For instance, the choice of inner-hubs is part of the network design, which in turns affects the quality and cost of the transportation plans (through consolidation of demands) that can be produced on the respective network. Here we show how the HAD Algorithm can help transportation managers with the choice of inner-hubs. The sorting centers that are used as inner-hubs by the HAD Algorithm are specified as part of the transportation network given as input for the LHPTP. In the real-life application of the French postal company, the two to four inner-hubs that are used have been chosen at some point by transportation managers for operational and/or technical reasons. In the experiments, we ignore this constraint and assume that all the sorting centers can be inner-hubs. This gives more possibilities for consolidation, which occurs at the inner-hubs. Moreover, we want to confirm if the choice of the inner-hubs made by the transportation managers is the one which results in a lowest cost solution.

In the transportation plans whose costs are presented in Table 1, nearly all the sorting centers are used as inner-hubs. This is because the sorting costs are not very high compared to the kilometric cost and because the demands are small compared to the vehicle capacity. Therefore consolidation of demands is useful as it allows to reduce the number of vehicles used, even if this leads to more sorting operations. This might not be the case with much higher sorting costs and this is tested and discussed next.

Table 2 presents the outcomes when the sorting cost is increased, that is when the cost is 10 or 20 times the cost used in the realistic case study. Each row represents a value for the sorting cost for the six datasets. Hence we present, as in Table 1, the minimum, average and maximum values over the six datasets. The first column of Table 2 gives the sorting cost. The second column shows the computing time of the HAD Algorithm. The third and fourth column present respectively the costs of the solutions of Step e and Step f. The fifth column shows the gap to the best lower bound computed in six hours. The last column presents the number of inner-hubs used by the solution of the HAD Algorithm.

Table 2. Comparison of the sorting costs with $\sigma = 60\%$

Sorting cost	Time (h)	Sol fixed dir. e)	Built sol f)	Gap [3] (%)	Used inner-hubs
	min | **avg** | max	min | **avg** | max	min | **avg** | max	min | **avg** | max	min | **avg** | max
normal	1.0 | **1.1** | 1.2	367| **399** | 419	367| **399** | 419	10.1 | **12.2** | 13.9	13 | **14.7** | 17
10 × normal	0.0 | **0.1** | 0.2	1635 | **1733** | 1806	1634 | **1731** | 1803	78.6 | **79.8** | 80.7	8 | **8.5** | 10
20 × normal	0.0 | **0.1** | 0.2	3030 | **3188** | 3330	3026 | **3182** | 3324	88.3 | **89.0** | 89.6	4 | **5.2** | 7

From Table 2, several observations can be made regarding hub selection and sorting cost. Firstly, Table 2 shows the number of inner-hubs used decreases when the sorting cost increases. Secondly, when the sorting costs increase from normal to 10 × normal, the algorithm runs much faster as operational paths with sorting become too expensive and are ignored. Therefore, many constraints become irrelevant and it simplifies the MILP. Thirdly, we observe that with a sorting cost 20 × normal (last row of Table 2), five hubs are used on average and by very small demands. And notably, these inner-hubs do not include all the inner-hubs chosen by transportation managers in real life. The fourth observation is that with the "normal" costs, all sorting centers are selected as inner-hubs by the HAD Algorithm. This choice of inner-hubs differs from what is currently in use by the French postal company, where not all inner-hubs are used. Thus the HAD Algorithm can be a useful decision support tool for the transportation managers in choosing which inner-hubs to put in use.

8. Conclusion and Future Research Perspectives

This study introduces the LHPTP. We describe the application framework of this industrial transportation problem faced by a French postal company, and then we formally define the optimization problem. The LHPTP concerns the optimization of the long-haul transportation of parcels on a two-level hybrid hub-and-spoke network. There are two types of vehicles to manage according to the number of containers and the daily repositioning of empty vehicles from delivery depot to sorting centers.

[3] Gap to the lower bound of the MILP without heuristic with the same variables.

The problem contains strong industrial constraints and has not been addressed in the literature to the best of our knowledge. We have presented an MILP to solve the LHPTP, but the datasets provided by a postal company are too large for an MILP solver to directly find a good solution in reasonable time.

Accordingly, we present a new algorithm that exploits the two-level structure of the network: the HAD Algorithm, which divides the problem into subproblems solved through MILPs. As the hybrid hub-and-spoke network offers the possibility to use direct paths, this option is allowed for large demands (above a truck filling rate threshold). The residual demands (lower than the truck filling rate threshold) are gathered into aggregate demands which stay on the inner-level of the network. We have tested various truck filling rate thresholds to partition the demands to find which one suits our datasets the best. This constitutes a tailored, adaptive heuristic which provides better quality solutions for the LHPTP. In addition to being a heuristic which solves the LHPTP, the HAD Algorithm is a decision aid tool designed for transportation managers to address anticipated demand forecasts, which allows them to design their networks and transportation plans accordingly. For instance, they can use it to evaluate the network performance resulting from a particular network design decision (concerning the choice of inner-hubs for instance).

One perspective for future work is to study the impact of using different truck filling rate thresholds on different parts of the network or for different types of demands. This threshold could also be based on additional criteria such as distance and transportation costs. Another research direction is further applications of the HAD Algorithm as a decision aid tool. Indeed, the HAD Algorithm can be used to assist in making managerial decisions other than the inner-hub selection, such as which direct paths to open, or how to improve the repositioning of empty vehicles, for example not requiring that they remain in the same catchment area. This would require specifically designed simulations and modifications of the HAD algorithm that could be done as future work.

Acknowledgments

We thank Nicolas Teypaz (Probayes) for ideas and discussion and Sylvain Ducomman (Probayes) for comments.

Disclosure statement

No potential conflict of interest to report.

Data availability statement

The company that provided the data did not agree for their data to be shared publicly, so supporting data is not available. However, the datasets are described comprehensively in Gras (2021).

Funding

This work is funded by ANRT, CIFRE n°2019/0351.

References

Baumung, Martin N., and Halil I. Gündüz. 2015. "Consolidation of Residual Volumes in a Parcel Service Provider's Long-Haul Transportation Network." *Computational Logistics* 437–450.

Bryan, Deborah L, and Morton E O'Kelly. 1999. "Hub-and-spoke networks in air transportation: an analytical review." *Journal of regional science* 39 (2): 275–295.

Cohn, Amy, Sarah Root, Alex Wang, and Douglas Mohr. 2007. "Integration of the load-matching and routing problem with equipment balancing for small package carriers." *Transportation Science* 41 (2): 238–252.

Crainic, Teodor Gabriel. 2000. "Service network design in freight transportation." *European Journal of Operational Research* 122 (2): 272 – 288. http://www.sciencedirect.com/science/article/pii/S0377221799002337.

Crainic, Teodor Gabriel. 2003. "Long-haul freight transportation." In *Handbook of transportation science*, 451–516. Springer.

Cuda, Rosario, Gianfranco Guastaroba, and Maria Grazia Speranza. 2015. "A survey on two-echelon routing problems." *Computers & Operations Research* 55: 185–199.

Farias, Katyanne, Khaled Hadj-Hamou, and Claude Yugma. 2021. "Model and exact solution for a two-echelon inventory routing problem." *International Journal of Production Research* 59 (10): 3109–3132. https://doi.org/10.1080/00207543.2020.1746428.

Gonzalez-Feliu, Jesus. 2011. "Two-echelon freight transport optimisation: unifying concepts via a systematic review." *Working Papers on Operations Management* 2 (1): 18–30.

Gras, Camille. 2021. "The Long-Haul Parcel Transportation Problem : Path-based Models and Divide-and-Conquer Algorithms." Theses, Université Grenoble Alpes [2020-....]. https://tel.archives-ouvertes.fr/tel-03462717.

Gras, Camille, Van-Dat Cung, Nathalie Herr, and Alantha Newman. 2021. "A Hierarchical Network Approach for Long-Haul Parcel Transportation." In *IFIP International Conference on Advances in Production Management Systems*, 168–178. Springer.

Hu, Zhengyang, Ronald G. Askin, and Guiping Hu. 2019. "Hub relay network design for daily driver routes." *International Journal of Production Research* 57 (19): 6130–6145. https://doi.org/10.1080/00207543.2019.1571253.

Jiang, Li, Mohamed Dhiaf, Junfeng Dong, Changyong Liang, and Shuping Zhao. 2020. "A traveling salesman problem with time windows for the last mile delivery in online shopping." *International Journal of Production Research* 58 (16): 5077–5088. https://doi.org/10.1080/00207543.2019.1656842.

Lee, Jeong-Hun, and Ilkyeong Moon. 2014. "A hybrid hub-and-spoke postal logistics network with realistic restrictions: A case study of Korea Post." *Expert systems with applications* 41 (11): 5509–5519.

Leung, Janny M. Y., Thomas L. Magnanti, and Vijay Singhal. 1990. "Routing in Point-to-Point Delivery Systems: Formulations and Solution Heuristics." *Transportation Science* 24 (4): 245–260. https://doi.org/10.1287/trsc.24.4.245.

Lin, Cheng-Chang, and Sheu-Hua Chen. 2004. "The hierarchical network design problem for time-definite express common carriers." *Transportation research part B: methodological* 38 (3): 271–283.

Lin, Cheng-Chang, and Sheu-Hua Chen. 2008. "An integral constrained generalized hub-and-spoke network design problem." *Transportation Research Part E: Logistics and Transportation Review* 44 (6): 986–1003.

Meisen, Matthias. 2015. "Optimizing long-haul transportation considering alternative transportation routes within a parcel distribution network." In *Quantitative Approaches in Logistics and Supply Chain Management*, 129–147. Springer.

Meng, Qiang, and Xinchang Wang. 2011. "Intermodal hub-and-spoke network design: incorporating multiple stakeholders and multi-type containers." *Transportation Research Part B: Methodological* 45 (4): 724–742.

O'Kelly, Morton E. 1987. "A quadratic integer program for the location of interacting hub facilities." *European journal of operational research* 32 (3): 393–404.

O'Kelly, Morton E. 1998. "A geographer's analysis of hub-and-spoke networks." *Journal of transport Geography* 6 (3): 171–186.

O'Kelly, Morton E., and Harvey J. Miller. 1994. "The hub network design problem: A review and synthesis." *Journal of Transport Geography* 2 (1): 31 – 40. `http://www.sciencedirect.com/science/article/pii/0966692394900329`.

Rahmani, Younes, Wahiba Ramdane Cherif-Khettaf, and Ammar Oulamara. 2016. "The two-echelon multi-products location-routing problem with pickup and delivery: formulation and heuristic approaches." *International Journal of Production Research* 54 (4): 999–1019. `https://doi.org/10.1080/00207543.2015.1040127`.

Sebastian, Hans-Jürgen. 2012. "Optimization approaches in the strategic and tactical planning of networks for letter, parcel and freight mail." In *Modeling for Decision Support in Network-Based Services*, 36–61. Springer.

Statista. 2021. "E-commerce worldwide." October. `https://www.statista.com/topics/871/online-shopping/`.

Timperio, Giuseppe, Sunil Tiwari, José Manuel Gaspar Sánchez, Rafael Adrián García Martín, and Robert de Souza. 2020. "Integrated decision support framework for distribution network design." *International Journal of Production Research* 58 (8): 2490–2509. `https://doi.org/10.1080/00207543.2019.1680894`.

Wang, Xin, and George Q. Huang. 2020. "When and How to Share First-Mile Parcel Collection Service." *European Journal of Operational Research* 288.

Zäpfel, Günther, and Michael Wasner. 2002. "Planning and optimization of hub-and-spoke transportation networks of cooperative third-party logistics providers." *International Journal of Production Economics* 78 (2): 207–220.

Zhang, Jian, Yaohua Wu, and Pei Liu. 2007. "Routing Problem for Hybrid Hub-and-spoke Transportation Network: a Case Study of a LTL Carrier." In *2007 IEEE International Conference on Automation and Logistics*, 170–175. IEEE.

Zhang, Wenxin, Alexia P Payan, and Dimitri N Mavris. 2021. "Improving Courier Service Network Efficiency through Consolidations." In *AIAA Scitech 2021 Forum*, 1233.