A Continuous-time Service Network Design and Vehicle Routing Problem

Yun He[1], Mike Hewitt[2], Fabien Lehuédé[3], Juliette Medina[4], Olivier Péton[3]

[1] Nottingham Business School, Nottingham Trent University, Nottingham NG1 4FQ, UK
yun.he@ntu.ac.uk
[2]Department of Information Systems and Supply Chain Management,
Quinlan School of Business, Loyola University Chicago, US
mhewitt3@luc.edu
[3]IMT Atlantique, LS2N, UMR CNRS 6004, Nantes, France,
fabien.lehuede@imt-atlantique.fr, olivier.peton@imt-atlantique.fr
[4] CRC Services, Rueil Malmaison, France,
j.medina@crc-services.com

Abstract

This paper considers the integrated planning of goods transportation through a multi-echelon supply chain consisting of a nationwide network and regional distribution system. The previously studied Service Network Design and Routing Problem considered similar planning decisions, albeit with multiple restrictions regarding the transportation of goods that can eliminate the opportunities for transportation savings. It also does not explicitly model the opportunity to increase vehicle utilization by having vehicles serve multiple purposes within the supply chain. We propose a mathematical model of the problem we consider that is inspired by the operations of an industrial partner. We present an adaptation of the Dynamic Discretization Discovery algorithm to solve this problem and illustrate its computational effectiveness on instances derived from the operations of a retail distribution network in France. Finally, we illustrate the potential savings enabled by solving the proposed model.

1 Introduction

In a retail supply chain, goods are often transported from manufacturing centers or warehouses associated with suppliers to logistics platforms. These platforms in turn serve as supply points for one or more retail stores. The transportation of goods within a retail supply chain is typically triggered by periodic orders to replenish inventory levels at retail stores. Different orders for the same retail store will often involve the same goods, albeit in different quantities. In many countries, much of the transportation of goods between different facilities in a supply chain is done by truck. For example, in France, truck transportation represented a total of 170 billion ton-kilometers in 2019, with 87% of this quantity transported by a for hire or private fleet carrier[1].

[1]Source: Ministère de la Transition Ecologique, France. https://www.statistiques.developpement-durable.gouv.fr/activite-du-transport-routier-de-marchandises-aux-premier-deuxieme-et-troisieme-trimestres-2020

In this paper, we focus on supply chains wherein retail store order quantities are small relative to vehicle capacity, often just a few pallets. Given the small sizes of orders, instead of transporting them directly from supplier site to retail store, they are transported through a series of logistics platforms that act as consolidation centers. We refer to such platforms as hubs. At a hub, goods are unloaded from an inbound vehicle, consolidated with goods associated with other shipments, and loaded onto an outbound vehicle. Ultimately, goods are transported to a terminal hub located in the same region as the final retail store destination. At that terminal hub, goods are again unloaded, and then consolidated with goods that can be delivered to retail stores on the same last-mile delivery route.

Planning activities in such a supply chain configuration can be complex. Order consolidation creates an interdependence between the transportation plans for different orders. Transshipment of goods at hubs requires synchronization between vehicles. As a result, simplifying strategies are often adopted to avoid dealing with these complexities. To reduce the complexity associated with managing the interdependence of orders, single sourcing policies can be adopted. One such policy is that all outbound orders from a supplier are transported to the the same initial hub. Symmetrically, all orders destined for the same retail store must be delivered from the same terminal hub. A disadvantage of such a strategy is that consolidation opportunities may be lost and unnecessary transfer costs may be induced. For example, when orders from a specific supplier to a specific retail store are collectively of a sufficient size, transporting them from the supplier site directly to the retail store may be more economical than to a hub.

To reduce the complexity associated with temporal synchronization, one strategy is to partition the day at a hub into a sequence of phases, with the first phase focused on inbound transportation-related activities, the second consolidation, and the third outbound transportation. An example of such a strategy at a hub is to first receive goods from regional suppliers until noon. Then, from noon until four p.m., goods are consolidated according to the terminal hub associated with their final destination. From four p.m. on, trucks are loaded for dispatch, to arrive at the terminal hub early the next morning. Finally, local transport to clients from the terminal hub can take place in the morning. One advantage of such a strategy is that it decouples inbound and outbound flows, reducing the need to synchronize activities. One disadvantage is that it can cause goods to wait. For example, goods that arrive at a hub after noon may need to wait until the subsequent day for unloading, consolidation, and outbound transportation. Such forced waiting can lead to additional operational costs and reduced consolidation.

From an operations research perspective, planning the transportation of goods through a nation-wide network and regional distribution system is typically decomposed into two levels. Each level of planning is modeled as an optimization problem and the two solved independently. The first level focuses on planning transportation between hubs, which is often referred to as *long-haul* operations, and is typically modeled with a variant of the Service Network Design Problem (SNDP) [Crainic, 2000]. Shipments are routed from origin hubs through a network of hubs to terminal hubs that are the starting point of local delivery routes. Vehicle transportation to support these shipment moves is also planned by the SNDP. Variants of the SNDP typically presume that vehicle transportation costs can be computed on a per-move basis. However, when outsourcing transportation to a third party provider, there may be pricing schemes in place that lead to economies of scale [Nowak et al., 2019]. For example, the price per unit of distance the carrier transports goods may depend on the total distance of the route the carrier executes.

Planning inbound transportation from suppliers to hubs or outbound transportation from a terminal hub to retail stores, often referred to as *short-haul* operations, is often modeled with a variant of the Vehicle Routing Problems with Time Windows (VRPTW) [Desaulniers et al., 2014]. A byproduct of this decomposition is that it necessitates partitioning a vehicle fleet into one set of vehicles that provides long haul transportation and another that provides short haul. Such a partition can reduce vehicle utilization. Dual use of a vehicle, meaning having it provide both long and short haul transportation, may increase its utilization. However, there may also be rules regarding the structure of such dual use routes.

Emerging professional practices [Medina et al., 2019], partially enabled by advances in hardware and software technology, have motivated attempts to integrate SNDP and VRPTW models into the Service Network Design and Routing Problem (SNDRP). The SNDRP considered by [Medina et al., 2019] primarily focused on synchronizing long haul and short haul activities. It retained the presumption that single sourcing policies were in effect. It also did not explicitly model the dual use of vehicles nor transportation pricing schemes that depend on routes instead of individual transportation moves. Thus, in this paper we extend the SNDRP proposed by Medina et al. [2019] in multiple ways. First, we allow for direct deliveries from suppliers to retail stores (also suggested in hub location by Mahmutoğulları and Kara [2015]). Second, we relax single sourcing policies that allocate a supplier to a single hub or a retail store to a single terminal hub. Third, we model the dual use of vehicles for long and short haul transportation integrating various types of available vehicles. Fourth, we provide a model for transfer cost and time at hubs which integrates the fact that a vehicle may unload and/or load goods when visiting a hub. It also integrates the possibility for a vehicle to drop off part of its load at a hub and then continue on to deliver the rest of its load to a customer. Fifth, we model operational considerations associated with outsourced transportation, including limitations on the number of served stops, maximum detour constraints and carrier rates. These considerations are inspired by the operations of an industrial partner.

We believe this paper makes the following contributions. First, it proposes a mathematical model for a new variant of the SNDRP that is more relevant to agile supply chains. This includes enhancements in our model to integrate the previously mentioned extensions of the SNDRP. Second, we adapt the DDD algorithm of Boland et al. [2017a] to solve the proposed variant of the SNDRP. While DDD was proven to be correct for variants of the SNDP that model vehicle movements on arcs, the model proposed in this paper models vehicle movements on routes. For this and other reasons, adapting DDD to this problem necessitates modifying each step of the algorithm. Third, with an extensive computational study based on the operations of a retail distribution network in France, we establish that the DDD-based algorithm computationally outperforms known benchmarks. Finally, by analyzing high-quality solutions to the proposed mathematical model we provide insights into the magnitude and sources of savings its solutions enable.

The paper is organized as follows. Section 2 presents related bibliographical references in service network design and its integration with vehicle routing problems. It also briefly introduces the DDD algorithm and its recent use in optimizing logistics applications. Section 3 presents the logistics network and the truck routes considered in this paper. The modeling of this logistics network by a time- and route- expanded graph is discussed in Section 4. Then, a mathematical model for the SNDRP is introduced in Section 5. Section 6 describes the adaptation of the DDD algorithmic framework to the problem at hand and proves its correctness. A computational study and managerial insights are presented in section 7. Section 8 concludes the papers and proposes possible extensions and research avenues.

2 Literature

The optimization problem considered in this paper integrates two levels of transportation optimization that have traditionally been addressed separately. The first involves optimizing the routing of goods through a distribution network. Such transportation is often referred to as middle-mile logistics. The second involves optimizing routes responsible for the pick-up of goods from supplier sites to bring to a hub as well as the routes responsible for the delivery of goods from a hub to retail stores. Transportation in this second level is often referred to as first-mile (supplier to hub) or last-mile (hub to retail store) logistics. As such, we next discuss the relevant literature for each of these levels. Note that we do not discuss the strategic design aspects of the network related to hub location or supplier/retail store assignment to hubs in this paper. One reason is that we do not presume the need for a fixed assignment of hubs to stores. Another reason is that we integrate temporal aspects to clearly identify consolidation opportunities at a tactical or operational level. On hub location, we recommend the surveys of Campbell and O'Kelly [2012] and Contreras and O'Kelly [2019] and, the recent paper of Yıldız et al. [2021] which integrates routes that serve several hubs.

Powell and Sheffi [1989] presents one of the earliest examples of optimizing middle-mile logistics operations in the literature. That paper refers to the routing of goods through a terminal network in order to facilitate consolidation and high vehicle utilization as *load planning*. More generally, this optimization problem is referred to as the Service Network Design Problem (SNDP) [Crainic, 2000]. While prevalent for truck-based carriers, the SNDP is also relevant to multi-modal transportation [SteadieSeifi et al., 2014]. As noted by Wolfinger et al. [2018], the problems in this domain are typically considered in the context of tactical planning. As a result, variants of the SNDP rarely design detailed routes for vehicles. Instead, vehicles have often been modeled as resources whose flows should balance at terminals [Pedersen et al., 2009, Jarrah et al., 2009, Erera et al., 2013a]. More precise models have modeled vehicle movements with cyclic itineraries [Andersen et al., 2009b,a, Crainic et al., 2016, 2018, Hewitt et al., 2019] , however they are related to resource management considerations, they cannot explicitly capture the transfer of one commodity from one vehicle to the other and, they do not consider savings (e.g. in transfer cost) related to the dual use of vehicles. The design of routes, including the schedules of drivers, trailers and tractors for a given load plan is proposed in Erera et al. [2013b]. The integration of load planning and vehicle routing is studied in Medina et al. [2019], albeit with severe limits imposed on how goods can flow through the network and no recognition that the same vehicle could support both long-haul moves and local distribution routes. Medina et al. [2019] illustrate that integrating network optimization and vehicle routing can yield significant savings in transportation costs.

The consolidation of goods associated with different orders for joint transportation within a network requires synchronizing their transportation both geographically and temporally. This temporal synchronization is typically captured in the Scheduled Service Network Design Problem (SSNDP) by formulating an optimization problem on a time-expanded network. The creation of such a network necessitates a discretization of time. Boland et al. [2019] observe that while a finer discretization of time can lead to higher quality solutions, it also yields instances of the SSNDP that are much harder to solve. To address this issue, Boland et al. [2017a] propose the *Dynamic Discretization Discovery* (DDD) algorithmic framework, in which a time expanded network is generated dynamically as opposed to statically. This approach is enhanced and extended in Hewitt [2019] and Marshall et al. [2021] for the SSNDP. Hewitt [2022] adapts DDD to a variant of the SSNDP wherein a small set of shipments may be selected for late delivery if the savings potential is significant. Medina et al. [2019] proposes a DDD-based algorithm as one of the first approaches to solving the SNDRP while Belieres et al. [2021] proposes a heuristic based on DDD to optimize flows through a multi-echelon supply chain serving restaurants. Finally, DDD has been extended to solve variants of the Traveling Salesman Problem with Time Windows [Boland et al., 2017b], including when travel times are time-dependent [Vu et al., 2020].

Regarding first and last-mile logistics, the problem considered in this paper involves designing routes that should be synchronized in time and space with cross-docking operations at hubs. Hence, from the perspective of vehicle routing point, the SNDRP is related to the Vehicle Routing Problem with Cross-Docking (VRP-CD Lee et al. [2006]). This problem focuses on designing routes for a set of vehicles, wherein each route can be categorized as either an inbound or outbound route. Inbound routes start at a cross-dock, visit suppliers to pick-up goods, and then end at the same cross-dock. Outbound routes start at a cross-dock with a set of goods to deliver, visits each customer with goods in that set, and then ends at the same cross-dock. At the cross-dock, collected goods are transferred from inbound vehicle to outbound vehicles. The problem was first introduced in Lee et al. [2006] for a situation wherein the time required to transfer goods between vehicles at a cross-dock was the same for all vehicle and fixed in time. The problem is generalized in Wen et al. [2008], which introduces free transfer periods and time windows at suppliers and customers. Wen et al. [2008] proposes a tabu search to solve instances with up to 200 customers. Morais et al. [2014] proposes instances with up to 500 customers and an Iterated Local Search procedure for their solution. Nikolopoulou et al. [2017] and Grangier et al. [2017] propose better-performing solution methods. In much of the literature on the VRPCD, the number of cross-docks is small, often only one. Relatedly, Guastaroba et al. [2016] surveys the literature on Vehicle Routing Problems with intermediate facilities. They observe that direct shipments are rarely considered in problems with cross-docking. More generally, the VRPCD can be classified as an integrated routing and scheduling problem [Paraskevopoulos et al., 2017]. It can also be classified as a VRP with Synchronization Drexl [2012].

3 Problem Setting

In this section, we mathematically define the variant of the SNDRP we consider in this paper. Table 4 in Appendix A summarizes the notation defined in this section.

In the SNDRP, we consider a set of commodities, denoted by $\mathcal{K}$, each representing a certain quantity of goods available at a supply point at a certain time, and has to be delivered to a demand point before a deadline. The set of pick-up points is denoted by $\mathcal{P}$. The set of delivery or demand points is denoted by $\mathcal{D}$. Associated with each commodity $k \in \mathcal{K}$ is a pick-up point, $o_k \in \mathcal{P}$, an earliest time, e_k, when it is available at that pick-up point, a delivery point, $d_k \in \mathcal{D}$, and a latest time, l_k, at which it must be delivered to the delivery point. Also associated with commodity k is a quantity of goods, q_k, that is quoted in pallets. Particularly, the set of commodities that must be delivered to point $i \in \mathcal{D}$ is denoted by $\mathcal{K}_i$. Note that all the goods of one commodity must be transported in the same truck, i.e. they can not be split.

Commodities originating from distinct pick-up points can be consolidated at intermediate facilities that we refer to as *hubs*. The set of hubs is denoted by $\mathcal{H}$. Hubs can also serve as storage points for commodities while they are waiting for transshipment to an outbound vehicle. Associated with hub $h \in \mathcal{H}$ is a per-pallet transfer cost, c_h, and a transfer time, σ_h, that is independent of the number of pallets transferred. The storage and transfer capacities of hubs are assumed to be unlimited.

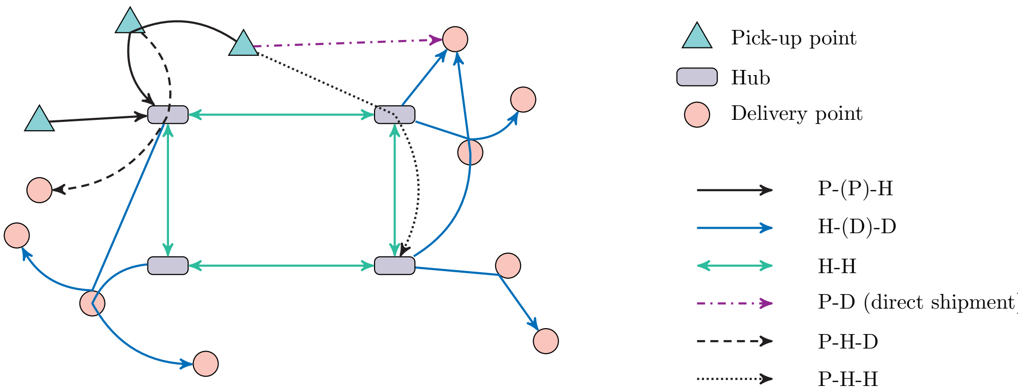


Figure 1: Transportation network of commodities

The physical transportation network is modeled by a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ with node set $\mathcal{N}$ and directed arc set $\mathcal{A}$. Each node $i \in \mathcal{N}$ can be a pick-up point, a delivery point, or a hub. Physical transportation between locations $i, j \in \mathcal{N}$ is modeled by arc $(i, j) \in \mathcal{A}$. Associated with arc $(i, j) \in \mathcal{A}$ is a travel time τ_{ij}.

The transportation of commodities is outsourced to a set of third-party carriers, each of which operates their own fleet of vehicles on routes. Carriers accept some flexibility by integrating one intermediate stop in a route between its origin and its destination but this comes with some limitations and requests. These include constraints on: the maximum distance between visited points, the maximum detour generated by an intermediate stop and, the maximum number of driving hours in a route. Each carrier has its own pricing strategies for pick-ups and deliveries in different regions of the network. We presume that through negotiations with carriers a set of potential routes, $\mathcal{R}$, in the physical transportation network is known. Each route $r \in \mathcal{R}$ visits a sequence of nodes $\mathcal{N}^r \subset \mathcal{N}$ by traversing a sequence of arcs $\mathcal{A}^r \subset \mathcal{A}$. Routes do *not* start and end at the same node, different from the common setting in the VRP. Each route $r \in \mathcal{R}$ requires the use of one vehicle of capacity Q_r, at cost c_r, which can include a fixed vehicle usage cost, stopover costs at locations visited in the network, and transportation costs.

Potential routes must observe one pattern from a predefined set. Namely, with P, H and D denoting pick-up points, hubs and delivery points, the potential patterns are: P-H, P-D, H-D, H-H, P-H-D, P-P-H, P-H-H, H-D-D. Note a route cannot start with a delivery point or end at a pick-up point. As part of the negotiation with carriers, the number of locations visited by a route is limited to at most three. Relatedly, due to the limitation of storage time at hubs and the high cost of transfer activities, transfers between hubs are limited to once per route and the maximum number of hubs visited by a route is limited to two. All routes which are infeasible related to the carrier limitations are removed from $\mathcal{R}$.

Figure 1 illustrates an example of the type of physical transportation network we consider in this paper. The network is composed of pick-up points, delivery points, and hubs. The figure also illustrates different types of routes we consider for transporting goods through the network. More precisely, the figure illustrates inbound pick-up routes of the form P-(P)-H that link pick-up points to hubs, outbound delivery routes of the form H-(D)-D that link hubs with delivery points, inter-hub routes of the form H-H that transport products between hubs, and direct shipment routes of the form P-D that transport goods directly from their pick-up point to their delivery point. It also illustrates pick-up and inter-hub routes of the form P-H-H and pick-up-and-delivery routes of the form P-H-D, both considered as dual use of vehicles. We recall that a hub need not be dedicated to certain pick-up or delivery points, so one delivery points can be visited from several hubs as illustrated in Figure 1.

The Service Network Design and Routing Problem (SNDRP) consists of planning the pick-up and delivery of each commodity by selecting and scheduling routes and assigning trucks to each route, such that the total transportation and transfer costs are minimized, while satisfying time and capacity related constraints. In the next section we present a network-based model of the SNDRP.

4 Network Models

The aim of this section is to present a network-based model of the SNDRP. To do so, we first introduce the concept of a *route-expanded network* which models pre-defined transportation routes along with commodity transfers at hubs in the physical network. Then, to model the scheduling of activities, we introduce the concept of a *route-time-expanded network*, which is the basis of the mathematical model presented in Section 5 and the DDD algorithm in Section 6.

The notation introduced in this section is summarized in Appendix A, with Table 5 presenting notation for the route-expanded network and Table 6 for the route-time-expanded network.

4.1 Modeling motivation

The route model is motivated mainly by the industrial practice. As mentioned previously, the transportation of goods is outsourced to third-party carriers, each operating their own fleet of vehicles. This modeling allows us to aggregate into routes all the information about how goods are transported by each carrier, including various cost structures, routing duration and operational constraints. This modeling also enables us to track, at each hub, when transfer activities happen and how much they cost.

In this way, our model explicitly considers the route(s) that each commodity travels on from its origin to its destination, and the timing of the route(s). This is unlike the models we are aware of in the literature which model the transportation moves a commodity takes and the need for sufficient vehicle capacity on each move. These models do not explicitly identify the vehicle in which a commodity travels. By modeling at this level of detail, our model can exactly capture when transfers occur, a requirement of the problem we consider which is not accurately captured by existing models.

4.2 Route-expanded network

The *route-expanded network* is a representation of the physical network plus additional information about routing and transfer activities. It is denoted by $\mathcal{G}_{\mathcal{R}} = (\mathcal{N}_{\mathcal{R}}, \mathcal{A}_{\mathcal{R}})$. There are three types of nodes in the node set $\mathcal{N}_{\mathcal{R}}$.

1. Route copies, denoted by (i, r), that model the combination of each route $r \in \mathcal{R}$ and physical node $i \in \mathcal{N}$ (pick-up, delivery points or hubs) in route r.
2. Hub nodes, denoted by $(h, -1)$, for each hub $h \in \mathcal{H}$ that are used to model transfers. The value -1 indicates that this node does not belong to any route.

3. Origin and destination nodes for each commodity $k \in \mathcal{K}$, denoted by $(o_k, -1)$ and $(d_k, -1)$ that are used to model the routing of commodities. The value -1 indicates that these nodes do not belong to any route.

The arc set $\mathcal{A}_{\mathcal{R}}$ in this network consists of three types of arcs.

1. Routing arcs, denoted by $((i, r), (j, r))$, for each route $r \in \mathcal{R}$ and each arc $(i, j) \in \mathcal{A}$ traversed by route r. These arcs model that arc (i, j) is included in route r.

2. Transfer arcs, denoted by $((h, r), (h, -1))$ or $((h, -1), (h, r))$ for each hub $h \in \mathcal{H}$ and each route $r \in \mathcal{R}$ visiting h. Arc $((h, r), (h, -1))$ models goods transferred from route r to hub h while arc $((h, -1), (h, r))$ models goods transferred from the hub to route r. Transfers are explained in more detail in Section 4.3.

3. Pick-up and delivery arcs. For each commodity $k \in \mathcal{K}$, arcs of the form $((o_k, -1), (o_k, r))$ model the pick-up of commodity k at its origin by route r. Arcs of the form $((d_k, r), (d_k, -1))$ model the delivery of commodity k to its destination by route r.

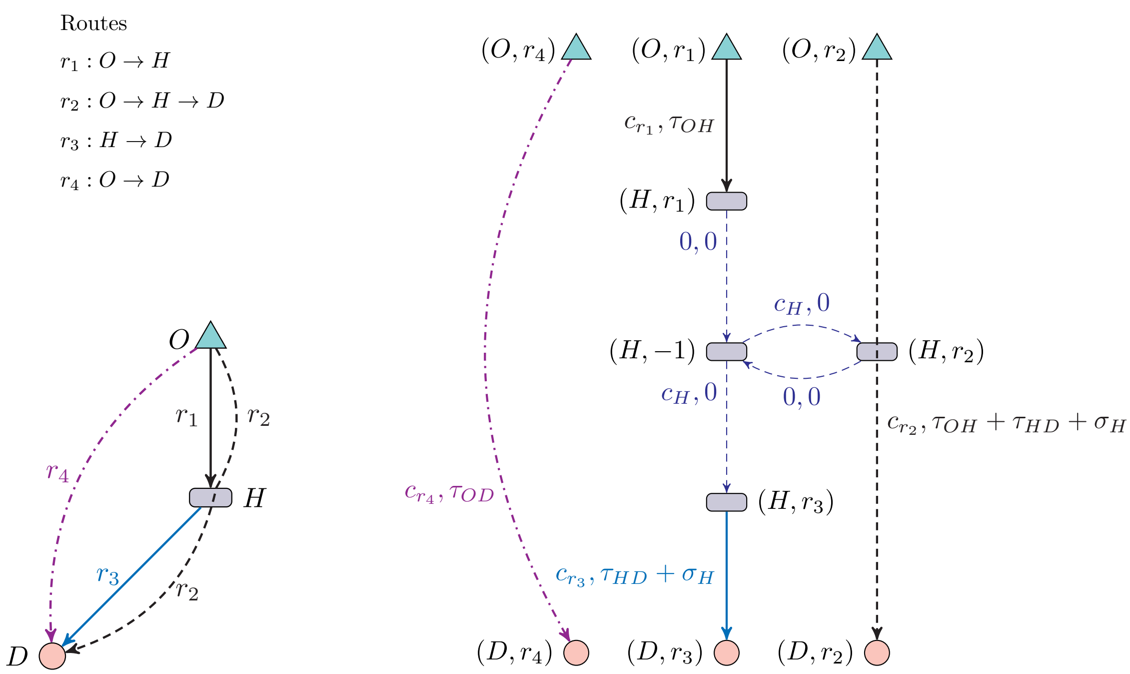
The routing and transfer activities of commodities can thus be modeled in the route-expanded network. Example 1 shows a simple physical network and the corresponding route-expanded network. In the next section, we focus on transfers in the route-expanded network.

Example 1. *Figure 2a illustrates a physical network with 1 pick-up point O, 1 delivery point D and 1 hub H. There are 4 predefined routes. Inbound pick-up or outbound delivery routes* $O \to H$ *or* $H \to D$ *are solid lines. Pick-up-and-delivery routes such as* $O \to H \to D$ *and direct shipment* $O \to D$ *are dashed.*

The associated route-expanded network (without commodity origin/destination nodes) is shown in Figure 2b, with corresponding route cost (left) and total travel time (right) for each route. Transfer arcs are dashed lines.

4.3 Modeling transfers with a route-expanded network

We model transfer cost of commodities at hubs by associating the hub variable cost to the commodities that leave the hub. Hence, no cost or duration is associated with commodities entering a hub. We also presume that the transfer of commodities between vehicles requires a fixed duration of time. The transfer of commodities is modeled in the route-expanded network with the *star structure* defined in Karsten et al. [2015]. Figure 3 illustrates such a structure for hub node h of Example 1.



(a) Physical network (b) Route-expanded network

Figure 2: Example with 1 pick-up point, 1 delivery point and 1 hub: node O in the physical network on the left is passed by routes r_1, r_2, r_4, so in the route-expanded network, nodes $(O, r_1), (O, r_2)$ and (O, r_4) are created, similarly for nodes H and D. For route r_1, routing arc $((O, r_1), (H, r_1))$ is created, similarly for routes r_2, r_3 and r_4. To model the transfer in the hub H, as explained in Section 4.3, an additional central hub node $(H, -1)$ is created, as well as transfer arcs $((H, r_1), (H, -1)), ((H, -1), (H, r_2)), ((H, r_2), (H, -1))$ and $((H, -1), (H, r_3))$. Transfer cost c_H is associated with arcs leaving the central hub node, $((H, -1), (H, r_2))$ and $((H, -1, H, r_3))$. Transfer duration σ_H is installed onto routes that leave the hub, in this example, routes r_2 and r_3.

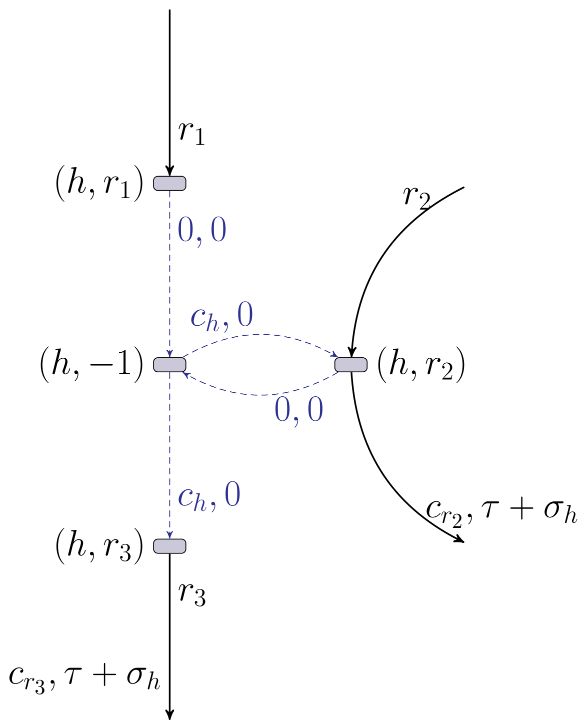


Figure 3: Star structure for modeling transfers

When route r ends at hub h, as r_1 does in Figure 3, then there is an arc from node (h, r) in the route-expanded network to the hub node $(h, -1)$ with cost 0. If h is an intermediate node in route r, like r_2 in Figure 3, then there are bidirectional arcs linking the nodes $(h, -1)$ and (h, r). The transfer cost, c_h, is associated with the outgoing arc $((h, -1), (h, r))$ from the hub node $(h, -1)$. If route r starts at hub h, like in r_3 in Figure 3, then there is an arc from the hub node $(h, -1)$ to the node (h, r) with cost c_h.

Regarding the time required for transferring commodities, all arcs linking the hub node $(h, -1)$ and route nodes (h, r) have duration 0. The transfer duration is included in the travel duration of the arc associated with the route movement to the next location. Namely, arc $((h, r), (j, r))$ has travel time $\tau_{hj} + \sigma_h$, where node j is the first location visited in route r after hub h. We note that by construction this route-expanded network guarantees that the transfer of commodities does not happen at pick-up and delivery points.

4.4 Route-Time-Expanded Network

In this section, we discuss the network model, which we refer to as a *route-time-expanded network*, used to represent the temporal aspect of the problem we consider. We note that one assumption of the problem we consider is that there is no en-route waiting time. This assumption is based on the operations of the industrial partner that inspired this research. A consequence of this assumption is that once the start time of a route is determined the visit time at each node in that route is determined. The choice of start time, and subsequent node visit times, is modeled with a route-time-expanded network, $\mathcal{G}_{\mathcal{R},\mathcal{T}} = (\mathcal{N}_{\mathcal{R},\mathcal{T}}, \mathcal{A}_{\mathcal{R},\mathcal{T}})$, that is based on copying each node in the route-expanded network at different instants in time. The network $\mathcal{G}_{\mathcal{R},\mathcal{T}}$ is based on the set of time points $\mathcal{T} = \bigcup_{(i,r)\in\mathcal{N}_{\mathcal{R}}} \mathcal{T}_{(i,r)}$, wherein each set $\mathcal{T}_{(i,r)}$ consists of time points during which route r can visit node i.

The node set $\mathcal{N}_{\mathcal{R},\mathcal{T}}$ of the route-time-expanded network $\mathcal{G}_{\mathcal{R},\mathcal{T}} = (\mathcal{N}_{\mathcal{R},\mathcal{T}}, \mathcal{A}_{\mathcal{R},\mathcal{T}})$ consists of the sets of nodes $\mathcal{N}_{OD}$, $\mathcal{N}_{\Omega}$ and $\mathcal{N}_{\mathcal{H}}$, which we next define.

1. Nodes in the set $\mathcal{N}_{OD}$ represent times at which commodities can be picked up or delivered. Specifically, $\forall k \in \mathcal{K}$, $\mathcal{N}_{OD}$ contains the nodes $(o_k, -1, e_k)$ and $(d_k, -1, l_k)$.

2. Nodes in the set $\mathcal{N}_{\Omega}$ represent times at which routes visit physical locations. Specifically, the set $\mathcal{N}_{\Omega}$ contains nodes of the form (i, r, t) with route $r \in \mathcal{R}$, physical location $i \in \mathcal{N}^r$, and an instant of time $t \in \mathcal{T}_{(i,r)}$. We refer to these nodes as *timed route copies* in the route-time-expanded network.

3. Nodes in the set $\mathcal{N}_{\mathcal{H}}$ model activities at hubs at different times. Specifically, the set $\mathcal{N}_{\mathcal{H}}$ contains nodes of the form $(h, -1, t)$ with $h \in \mathcal{H}$ and $t \in \mathcal{T}$. These nodes are called *central hub nodes* in the route-time-expanded network.

Each node in set $\mathcal{N}_{\mathcal{R},\mathcal{T}}$ can thus be denoted by a triplet (i, r, t) with $r \in \mathcal{R} \cup \{-1\}$, $i \in \mathcal{N}^r \cup \mathcal{H}$ and $t \in \mathcal{T}_{(i,r)}$. Regarding the duration of activities, recall that τ_{ij} is the travel duration of physical arc (i, j), while σ_h represents the transfer duration at hub $h \in \mathcal{H}$. Given these quantities the duration, $\hat{\tau}_{ij}$, is computed as follows:

$$\hat{\tau}_{ij} = \begin{cases} \tau_{ij} & \text{if } i \text{ is not a hub} \\ \tau_{ij} + \sigma_i & \text{otherwise} \end{cases} \quad (1)$$

The arc set $\mathcal{A}_{\mathcal{R},\mathcal{T}}$ of the route-time-expanded network $\mathcal{G}_{\mathcal{R},\mathcal{T}} = (\mathcal{N}_{\mathcal{R},\mathcal{T}}, \mathcal{A}_{\mathcal{R},\mathcal{T}})$ consists of the sets of arcs $\mathcal{A}_p, \mathcal{A}_f, \mathcal{A}_l, \mathcal{A}_h$, which we next define.

1. Arcs in the set $\mathcal{A}_p$ represent transportation. Specifically, the set $\mathcal{A}_p$ contains arcs of the form $((i, r, t_i), (j, r, t_i + \hat{\tau}_{ij}))$ presuming that route $r \in \mathcal{R}$ visits nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$. As en-route waiting is not allowed, if route r departs on arc (i, j) at node t_i then it must arrive at node j at time $t_i + \hat{\tau}_{ij}$.

2. Arcs in the set $\mathcal{A}_f$ represent transferring commodities from a route to a hub or vice-versa. Specifically, the set $\mathcal{A}_f$ contains arcs of the form $((h, r, t), (h, -1, t'))$ or $((h, -1, t), (h, r, t')), t \in \mathcal{T}_{(h,r)}$, with $t \leq t'$, presuming route $r \in \mathcal{R}$ visits hub $h \in \mathcal{H}$.

3. Arcs in the set $\mathcal{A}_l$ are linking arcs representing a route picking up or delivering a commodity. Specifically, the set $\mathcal{A}_l$ contains arcs of the form $((o_k, -1, e_k), (o_k, r, t)), t \in [e_k, l_k]$ or $((d_k, r', t'), (d_k, -1, l_k)), t' \in [e_k, l_k]$ for each commodity $k \in \mathcal{K}$.

4. Arcs in the set $\mathcal{A}_h$ represent holding commodities being held at a hub. Specifically, the set $\mathcal{A}_h$ contains arcs of the form $((h, -1, t^h_m), (h, -1, t^h_{m+1}))$ with $h \in \mathcal{H}$ and t^h_m, t^h_{m+1} two consecutive time points modeled for hub h.

For each physical route $r \in \mathcal{R}$, we let $\Omega(r)$ denote the set of its timed copies. Each timed route copy $\omega \in \Omega(r)$ is defined by route r and starting time t. Also associated with ω is the binary parameter $a^{\omega t_1 t_2}_{i_1 i_2}$, which equals 1 if ω goes from node $i_1 \in \mathcal{N}^r$ at time t_1 to node $i_2 \in \mathcal{N}^r$ at time t_2. Let $\Omega = \bigcup_{r\in\mathcal{R}} \Omega(r)$ be the set of all timed route copies. We define a *complete route-time-expanded network* as follows.

Definition 1. *A* complete route-time-expanded network $\mathcal{G}^C_{\mathcal{R},\mathcal{T}}$, *is a route-time-expanded network based upon a route set Ω that contains all timed routes that start and end within a pre-defined planning horizon.*

Example 2. *Figure 4 illustrates an example of the route-time-expanded network of the physical network as in Figure 2a with two commodities. Suppose that the duration $\tau_{OH} = 1$, $\tau_{HD} = 1$, $\tau_{OD} = 2$ and $\sigma_H = 1$. Also suppose that commodity 1 is available for pick-up at its origin, O, at time 0 and is due for delivery at its destination, D, at time 5. Commodity 2 is also available for pick-up at O, albeit at time 1 and is to be delivered to D at time 4. Figure 4 illustrates all nodes in set $\mathcal{N}_{\mathcal{R},\mathcal{T}}$, but only a subset of arcs in the set $\mathcal{A}_{\mathcal{R},\mathcal{T}}$.*

5 Mathematical Model for the continuous time SNDRP

In this section, we present a mathematical model of the problem we consider that is formulated on a route-time-expanded network. Regarding decision variables, for each arc $((i_1, r_1, t_1), (i_2, r_2, t_2)) \in \mathcal{A}_{\mathcal{R},\mathcal{T}}$ and each commodity $k \in \mathcal{K}$, we let the binary variable $x^k_{(i_1,r_1,t_1),(i_2,r_2,t_2)}$ indicate whether commodity k travels on arc $((i_1, r_1, t_1), (i_2, r_2, t_2))$. For each timed copy $\omega \in \Omega(r)$ of route $r \in \mathcal{R}$, we let the integer variable z_ω indicate the number of vehicles that travel on the route timed copy ω in the solution. The mathematical model is as follows.

$$\begin{aligned}\min z = \quad & \sum_{r\in\mathcal{R}} \sum_{\omega\in\Omega_r} c_r z_\omega \\ & + \sum_{k\in K} \sum_{h\in\mathcal{H}} \sum_{((h,-1,t),(h,r,t'))\in\mathcal{A}_f} c_h\, q_k\, x^k_{(h,-1,t),(h,r,t')}\end{aligned}$$

subject to

$$\sum_{((o_k,-1,e_k),(o_k,r,t))\in\mathcal{A}_l} x^k_{(o_k,-1,e_k),(o_k,r,t)} = 1 \quad \forall k \in \mathcal{K} \tag{2}$$

$$\sum_{((d_k,r,t),(d_k,-1,l_k))\in\mathcal{A}_l} x^k_{(d_k,r,t),(d_k,-1,l_k)} = 1 \quad \forall k \in \mathcal{K} \tag{3}$$

$$\sum_{((i_1,r_1,t_1),(i_2,r_2,t_2))\in\mathcal{A}_{\mathcal{R},\mathcal{T}}} x^k_{(i_1,r_1,t_1),(i_2,r_2,t_2)} = \sum_{((i_2,r_2,t_2),(i_1,r_1,t_1))\in\mathcal{A}_{\mathcal{R},\mathcal{T}}} x^k_{(i_2,r_2,t_2),(i_1,r_1,t_1)} \quad \forall k \in \mathcal{K}, \forall (i_1, r_1, t_1) \in \mathcal{N}_\Omega \tag{4}$$

$$x^k_{((h,r,t),(h,-1,t'))\in\mathcal{A}_f} \leq \sum_{((i,r,t-\hat{\tau}_{ih}),(h,r,t))\in\mathcal{A}_p} x^k_{((i,r,t-\hat{\tau}_{ih}),(h,r,t))}$$

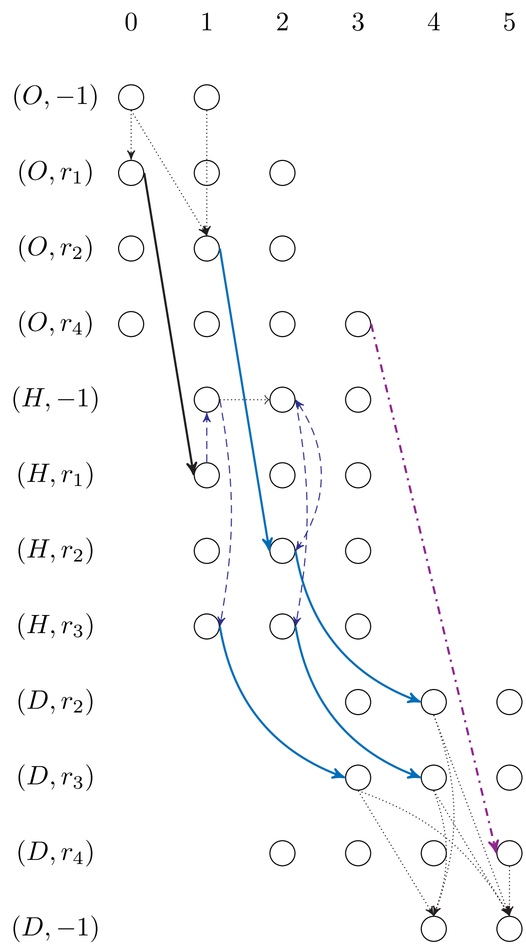


Figure 4: Example of a route-time-expanded network of the physical network in Figure 2a with two commodities

$$\forall k \in \mathcal{K}, \forall h \in \mathcal{H}, \forall((h,r,t),(h,-1,t')) \in \mathcal{A}_f \qquad (5)$$

$$x^k_{((h,-1,t'),(h,r,t)) \in \mathcal{A}_f} \leq \sum_{((h,r,t),(i,r,t+\hat{\tau}_{hi})) \in \mathcal{A}_p} x^k_{((h,r,t),(i,r,t+\hat{\tau}_{hi}))}$$

$$\forall k \in \mathcal{K}, \forall h \in \mathcal{H}, \forall((h,-1,t'),(h,r,t)) \in \mathcal{A}_f \qquad (6)$$

$$\sum_{k \in K} q_k x^k_{(i_1,r,t_1),(i_2,r,t_2)} \leq Q_r a^{\omega t_1 t_2}_{i_1 i_2} z_\omega \quad \forall r \in \mathcal{R},\ \forall \omega \in \Omega(r),\ \forall((i_1,r,t_1),(i_2,r,t_2)) \in \mathcal{A}_p \qquad (7)$$

$$x^k_{(i_1,r_1,t_1),(i_2,r_2,t_2)} \in \{0,1\} \quad \forall((i_1,r_1,t_1),(i_2,r_2,t_2)) \in \mathcal{A}_{\mathcal{R},\mathcal{T}}, \forall k \in \mathcal{K}, \qquad (8)$$

$$z_\omega \in \mathbb{N} \quad \forall r \in \mathcal{R},\ \forall \omega \in \Omega(r) \qquad (9)$$

The objective function consists of two terms. The first term is the sum of fixed route costs while the second corresponds to the sum of variable transfer costs for all commodities and all hubs. Constraints (2) and (3) ensure that each commodity is picked up and delivered. Constraints (4) are flow conservation constraints for commodities. Constraints (5) ensure that if there is a commodity enters a hub node, then there is at least one physical route going into that same hub. Constraints (6) are symmetric, albeit for a commodity departing a hub node. Constraints (7) are the capacity constraints on transportation arcs. The variables and their domains are defined by Constraints (8)–(9).

We define the Continuous Time Service Network Design and Routing Problem (CTSNDRP) as follows.

Definition 2. *The* Continuous Time Service Network Design and Routing Problem *(CTSNDRP) is the mathematical model presented above formulated on a complete route-time-expanded network $\mathcal{G}_{\mathcal{R},\mathcal{T}}$.*

6 The CTSNDRP Dynamic Discretization Discovery Algorithm

In this section, we adapt the Dynamic Discretization Discovery (DDD) algorithmic framework of Boland et al. [2017a] to solve the CTSNDRP and name it *DDD for CTSNDRP*. We prove that by applying DDD properly, the CTSNDRP can be solved exactly.

The idea of the DDD algorithm is to solve the continuous time problem on a smaller partially defined network first, under the proof that the solution on the partial network is a relaxation of the original problem. Then the partial network is completed gradually by some refinement procedure, so that the solution on the partial network can be converged to the optimal solution of the original problem defined on the complete network.

Compared to the DDD algorithm, our algorithm operates on a partial route-time-expanded network, which is different from time-expanded network in Boland et al. [2017a]. In this network, only some of the transfer arcs and linking arcs have a travel or transfer duration that may be shorter than the real duration or even be negative. More precisely, we suppose that:

- travel times of transportation arcs are fixed to their exact duration;
- transfer arcs from a timed route copy of a hub node to the associated central hub node have 0 duration;

- transfer arcs from a central hub node to a timed route copy of a hub node can have a shorter duration;
- linking arcs from a timed origin node to a timed route copy of the same origin can have a shorter duration;
- linking arcs from a timed route copy of a delivery node to the associated timed destination node can only have positive or zero duration, because delivery can never be later than the latest delivery time.

Under these assumptions, we define the partial route-time-expanded network and the Service Network Design and Routing (SNDR) problem on this network.

Definition 3. *A* partial route-time-expanded network $\mathcal{G}'_{\mathcal{R},\mathcal{T}'} = (\mathcal{A}'_{\mathcal{R},\mathcal{T}'}, \mathcal{N}'_{\mathcal{R},\mathcal{T}'})$ *is a route-time-expanded network defined with a subset of timed routes $\Omega' \subset \Omega$, where time points $\mathcal{T}'$ contains only a subset of the time points in $\mathcal{T}$ in the complete network.*

In the following, notations for the partial network are the same as for the complete network with an apostrophe ($'$). The mathematical model defined on a partially expanded network $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$ is the same as defined in Section 5 and is denoted by SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$).

6.1 Properties

In this section, we prove that the solution of SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$) is a relaxation of CTSNDRP under the following properties of the partial route-time expanded network $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$.

Property 1. *For each route $r \in \mathcal{R}$, if, according to all commodity available and due times as well as the travel duration on the route, its earliest starting time to carry any commodity is a_r, then there exists a timed copy of $r \in \Omega'(r)$ that starts at time a_r.*

Indeed, we can compute the earliest and latest starting time a_r^k and b_r^k of each route r for carrying each commodity k according to the commodity available and due times e_k and l_k and the travel duration on the route. In this way, we can set up a compatible list of routes for carrying each commodity based on travel times. This is implemented as a pre-processing procedure to and will be explained in detail in Section 6.2.1.

Property 2. *For each commodity $k \in \mathcal{K}$, if route $r \in \mathcal{R}$ visits its origin o_k at time t at the latest, then as long as the commodity available time $e_k \leq t$, for any timed copy of $\omega \in \Omega'(r)$ that passes o_k at a time $t_o \leq t$, there is a linking arc $((o_k, -1, e_k), (o_k, r, t_o))$. If route $r \in \mathcal{R}$ visits its destination d_k at time $t_d \leq l_k$, then there is a linking arc $((d_k, r, t_d), (d_k, -1, l_k))$.*

Property 3. *Every route $r \in \mathcal{R}$ has at least a timed copy in the time expanded network.*

If a route does not have a timed copy in the time expanded network, this route cannot carry any commodity according to Property 1 and Property 2. Then the route is not useful for transportation and should be eliminated from the set of routes $\mathcal{R}$.

Property 4. *For each hub $h \in \mathcal{H}$ visited by route $r \in \mathcal{R}$, if there exists a transfer arc $((h, r), (h, -1))$ in the route-expanded network, then every timed copy $(h, r, t) \in \mathcal{N}'_{\mathcal{R},\mathcal{T}'}$ is linked to node $(h, -1, t) \in \mathcal{N}'_{\mathcal{R},\mathcal{T}'}$ by arc $((h, r, t), (h, -1, t)) \in \mathcal{A}'_f$.*

This property states that every incoming timed route copy visiting a hub is linked to a central hub node at exactly the same time when the hub is visited by this route.

Property 5. *For each hub $h \in \mathcal{H}$, if there exists a transfer arc $((h,-1),(h,r))$ in the route-expanded network and a central hub node at time t, and if the latest time of route r to visit hub h to serve any commodity is not earlier than t, then there exists an arc $((h,-1,t),(h,r,t')) \in \mathcal{A}'_f$ with $t' \leq t$.*

This property allows early outgoing of a hub by a timed route copy before the actual visiting time t under the condition that the hub is visited before the latest visiting time of that route.

Property 6. *If arc $((h,-1,t),(h,r,t')) \in \mathcal{A}'_f$, then there does not exist a node $(h,r,t'') \in \mathcal{N}'_{\mathcal{R},\mathcal{T}'}$ with $t' < t'' \leq t$.*

This property states that for a fixed route, the outgoing arc of a central hub copy can only be linked to the latest timed route copy that is earlier than the actual transfer time.

Property 7. *For each hub $h \in \mathcal{H}$, if it is visited by a sequence of time points $\{t_1, t_2, \ldots, t_n\}$, then there is a holding arc $((h,-1,t_i),(h,-1,t_{i+1}))$ between every two consecutive time points t_i and t_{i+1}.*

Given a partially route-time-expanded network satisfying the above properties, we prove the following theorem.

Theorem 1. *Let $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$ be a partially route-time-expanded network that satisfies Properties 1–7. Then the SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$) is a relaxation of the CTSNDRP.*

Proof. Taking an optimal solution to the CTSNDRP, we demonstrate that it can be converted to a solution to SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$) defined on a partially route-time expanded network $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$ with a cost not larger than that of the continuous time solution. In the following, let $S^* = (x^*, z^*)$ be an optimal solution to the CTSNDRP and let $\underline{S} = (\underline{x}, \underline{z})$ be the converted solution to the SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$).

For this, we need to first prove that after conversion, (1) the same set of physical routes is used for the transportation of each commodity and, (2) for each timed route copy used in the continuous time solution S^*, consolidations of commodities stay the same in the converted solution $\underline{S}$.

Taking an optimal solution $S^* = (x^*, z^*)$ to the CTSNDRP defined on the complete route-time-expanded network $\mathcal{G}_{\mathcal{R},\mathcal{T}}$, let us introduce the following sets. The set of timed route copies used in S^* is defined by $\Omega^* = \{\omega \in \Omega | z^*_\omega > 0\}$. The set of timed route copies used for the transportation of commodity $k \in \mathcal{K}$ is $\Omega^*_k = \{\omega \in \Omega | z^*_\omega > 0, k \in \mathcal{K}_\omega\}$. The set of commodities transported on each timed copy $\omega \in \Omega(r)$ of route r is defined by

$$\mathcal{K}_\omega = \{k \in \mathcal{K} | \exists((i_1,r,t_1),(i_2,r,t_2)) \in \mathcal{A}^\omega_p, x^{*k}_{(i_1,r,t_1),(i_2,rt_2)} = 1\}.$$

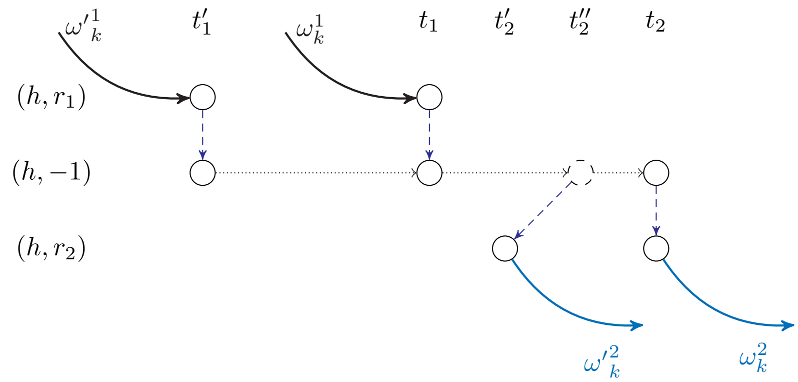
For each commodity $k \in \mathcal{K}$, its path in the completely expanded network forms the set

$$\mathcal{A}^*_k = \{((i_1,r_1,t_1),(i_2,r_2,t_2)) \in \mathcal{A}_\Omega | x^{*k}_{(i_1,r_1,t_1),(i_2,r_2,t_2)} = 1\}.$$

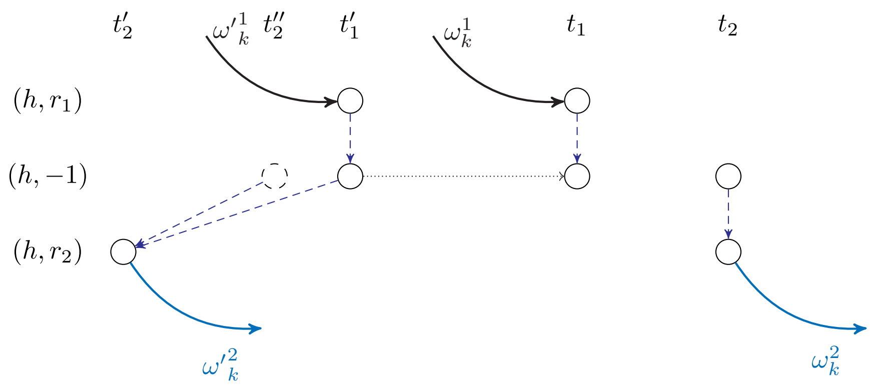
First, let us prove that each timed route copy in Ω^* in the completely expanded network can be mapped to one timed copy of the same route in the partially expanded network.

Take a timed route copy $\omega \in \Omega*$, suppose ω is a timed copy of route $r \in \mathcal{R}$. According to Property 3, r has at least a timed copy in the partially expanded network. According to Property 1, there exist timed copies of r starting no later than ω. Take the latest one ω' that starts no later than ω, and set $\underline{z}_\omega = z^*_{\omega'}$ in solution $\underline{S}$. In the following, ω' is called the *shifted timed route copy* of ω in the partially expanded network.

Now, for each commodity $k \in \mathcal{K}$, let us prove that it can be carried by the same set of routes in the partially expanded network. For this, we distinguish two cases:



(a) $t'_1 \leq t''_2$



(b) $t'_1 > t''_2$

Figure 5: Case 2: $|\underline{\Omega}_k| \geq 2$

<u>Case 1:</u> $|\Omega^*_k| = 1$. In this case, k is transported by only 1 route, noted r with timed copy ω_k i.e. there is no transfer in the transportation of commodity k. Take the shifted timed route copy ω'_k of ω_k in the partially expanded network. Since ω'_k starts no later than ω_k, we are sure that commodity k can be delivered in time by using ω'_k. According to property 2, ω'_k is linked to both origin and destination timed copies of k. So it is valid to use ω'_k for the delivery of commodity k in the solution $\underline{S}$ of SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$).

<u>Case 2:</u> $|\underline{\Omega}_k| \geq 2$. In this case, k is transported by at least two routes and there must be at least one transfer in the path of commodity k. Let us prove the case with one transfer in the path of k and proceed other cases by induction.

Suppose that the transfer takes place at hub h. It is visited at time t_1 and t_2 ($t_1 \leq t_2$) by two different timed copies ω^1_k and ω^2_k of routes r_1 and r_2, respectively, as illustrated in Figure 5.

Let $\omega'^1_k, \omega'^2_k \in \Omega'(r)$ the corresponding shifted timed copies in the partial network, which visit hub h at time t'_1 and t'_2, respectively. Since r_1 does not start with h and r_2 does not end with h, according to the definition of star transfer structure in Section 4.3 and Properties 4–5, nodes $(h, -1, t'_1)$ and $(h, -1, t''_2)$ with $t'_2 \le t''_2$, as well as transfer arcs $((h, r_1, t'_1), (h, -1, t'_1))$ and $((h, -1, t''_2), (h, r_2, t'_2))$ exist in the partially expanded network.

If $t'_1 \le t''_2$, since waiting at hub h is always allowed according to Property 7, node $(h, -1, t'_1)$ is connected to node $(h, -1, t''_2)$ via waiting arcs. Otherwise ($t'_1 > t''_2$), we have $t'_2 \le t''_2 < t'_1 \le t_1 \le t_2$. Let us prove that arc $((h, -1, t'_1), (h, r_2, t'_2))$ exists. By Property 5, there exists an arc $((h, -1, t'_1), (h, r_2, t'))$ with $t' \le t'_1 \le t_1 < t_2$. By Property 6, $t' \ge t'_2$. Since ${\omega'}^2_k$ is the latest timed copy of route r_2 not later than ω^2_k in the partially expanded network, $t' \le t'_2$. So $t' = t'_2$, arc $((h, -1, t'_1), (h, r_2, t'_2))$ exists in the partially expanded network.

The same commodity path can thus be kept using the same but shifted set of routes in the converted solution $\underline{S}$.

The case with more than one transfers can be easily proved by induction.

Suppose that the same set of physical routes is used to carry commodity k and there are $n-1$ transfers in the path of k in the continuous time solution. For $i \in \{1, 2, \ldots, n-1\}$, h_i is the hub where the i-th transfer takes place, t_i is the time that h is visited by route r_i with timed copy ω^i_k ($t_1 \le t_2 \le \ldots \le t_{n-1}$), and ω'_i the shifted timed route copy with t'_i the shifted visiting time. Given a commodity path with n transfers, suppose that the n-th transfer takes place at hub h_n, which is visited by route r_n at time t_n with timed copy ω^n_k. Let ${\omega'}^n_k$ the shifted timed route copy which visit h_n at time t'_n. Compare t'_n with t'_{n-1} and proceed in the same way as in the case with one transfer.

Second, let us prove that each timed copy can keep the same set of commodities in the converted solution $\underline{S}$.

For each timed route copy $\omega \in \Omega(r)$ of $r \in \mathcal{R}$, after the conversion, each $k \in \mathcal{K}_\omega$ is carried by shifted timed route copy $\omega' \in \Omega'(r)$ starting not later than ω. According to the rule for setting z, the number of timed route copies used in $\underline{S}$ cannot be more than in S^*. Thus, it exists a feasible solution where all $k \in \mathcal{K}_\omega$ is carried by shifted timed copy $\omega' \in \mathcal{G}'_{\mathcal{R},\mathcal{T}'}$ of ω. And the same consolidation can be kept in solution $\underline{S}$. So there exists a feasible solution in the partially expanded network with a cost not higher than S^*. As a result, the optimal solution in the partially expanded network has a cost not larger than that of the solution in the complete network.

In conclusion, the SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$) defined on a partially expanded network $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$ satisfying properties 1–7 is a relaxation of the CTSNDRP. $\square$

6.2 Algorithm

The DDD algorithm can thus be adapted for solving the CTSNDRP. The general scheme of the algorithm is presented by Algorithm 2. Example 3 in Section 6.3 presents how the algorithm works.

First, the earliest and latest starting time of each route for serving each commodity is obtained by pre-processing the network with respect to commodity available and due times as well as travel duration on each route. Then, an initial partial route-time-expanded network ${\mathcal{G}'}_{\mathcal{R},\mathcal{T}'}$ that satisfies properties 1–7 is created (Algorithm 3).

Then, SNDR(${\mathcal{G}'}_{\mathcal{R},\mathcal{T}'}$) is solved under a time limit T. Its solution $\overline{S}$ is passed to a MIP presented in Section 6.2.3, which reschedules the timed route copies and checks whether any new timed route copies need to be added to the partially expanded network ${\mathcal{G}'}_{\mathcal{R},\mathcal{T}'}$. The MIP gives us a set of timed route copies Ω_Δ that are to be added in the partially expanded network ${\mathcal{G}'}_{\mathcal{R},\mathcal{T}'}$. Meanwhile, a linear programming (LP) model presented in Section 6.2.4 is solved to convert $\overline{S}$ to a feasible solution S of the CTSNDRP.

By adding these missing timed route copies to Ω' and by restoring all properties in the partially expanded network, $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$ is updated. Then, a new instance of the SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$) is created and the algorithm iterates until the stop condition is reached. This stop condition is either the optimal solution of the CTSNDRP or the attainment of a global time limit.

Different components of the algorithm are explained in detail in this section.

Algorithm 2 DDD FOR CTSNDRP

Require: Physical network $\mathcal{G} = \{\mathcal{N}, \mathcal{A}\}$, route set $\mathcal{R}$, commodity set $\mathcal{K}$
Preprocessing
Create a partially route-time expanded network $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$ (CREATE-INITIAL)
while stop condition not attained **do**
  solve SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$) under a time limit T
  get solution $\underline{S}$ of the SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$)
  solve converting LP to get an upper bound $\overline{S}$ of the CTSNDRP
  solve rescheduling MIP and get set Ω_Δ of the timed route copies to be added to $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$
  if Ω_Δ is empty **then**
    if optimality gap is smaller than a certain criteria **then**
      solved, output the converted solution
    else
      increase time limit T
    end if
  else
    set $\Omega' \leftarrow \Omega_\Delta \cup \Omega'$
    for $\omega = (r, t) \in \Omega_\Delta$ **do**
      REFINE(r, t) and save the timed route copies of hubs if there exist any
    end for
    Refine transfer to timed route copies of hubs
    RESTORE network $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$ so that properties 2–6 are satisfied
  end if
end while

6.2.1 Preprocessing

Preprocessing computes the departure time window of each route in relation to the earliest available and latest delivery time of each commodity.

To compute the earliest departure time of each route, we implement a Dijkstra algorithm on the route-expanded network. For each commodity k, starting by setting the visit time to its origin node to the earliest available time e_k, we update the arrival time of k to each hub node reachable from o_k by adding the travel duration and keep the earliest time. We repeat the process until all the routes reaching the destination node d_k have been tested. Then we obtain the earliest departure time a_r^k of each route r which can be used to serve k in the network. The earliest departure time of each route r is then the earliest among all commodities $a_r = \min_{k \in \mathcal{K}} a_r^k$, which is the earliest departure time that a route can be used to serve any commodities in the network.

The latest departure time of each route can be obtained in a similar way. For each commodity k, instead of starting from o_k at time e_k, we start from its destination node d_k at the due time l_k

and go backward. We update the arrival time of each hub node leading to d_k by subtracting the travel duration and keep the latest time and proceed in the similar way as calculating the earliest departure time. The latest departure time of each route $b_r = \max_{k\in\mathcal{K}} b_r^k$is the latest time that a route can depart to serve any commodities in the network.

Then, in initialization, routes that start at their earliest departure time are added to the partial network. In this way, the initial route-time-expanded network contains more feasible timed copies, which results in fewer timed copies added in future iterations and better initial bounds. The latest departure times are later used in refine and restore to make sure that we don't introduce any infeasible connections between routes in the partially expanded network.

6.2.2 Initialization

The initial route-time-expanded network contains the origin and destination nodes of each commodity, the earliest timed copy of each physical route and the corresponding central hub nodes with holding, linking and transfer arcs, so that properties 1–7 are satisfied.

Algorithm 3 CREATE-INITIAL

Require: Route set $\mathcal{R}$
  add origin/destination nodes into $\mathcal{N}_{\mathcal{R},\mathcal{T}'}$
  for all $r \in \mathcal{R}$ **do**
    add one timed copy that starts at the earliest starting time
    add linking arcs if needed
    add central hub node if needed
  end for
  add holding and transfer arcs

6.2.3 Identification of missing timed route copies

In each iteration, after solving SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$) on partially expanded network $\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$, solution $\underline{S}$ is passed to a MIP to detect missing timed copies. In this MIP, the path of each commodity, the set of used timed route copies and the plan of consolidations are obtained from solution $\underline{S}$ and are fixed as parameters. The variables represent violations of any time constraints. The objective is to minimize the total number of arcs that are allowed to be too short.

Let $\Pi_k = \{n_0, n_1, n_2, \ldots, n_{|\Pi_k|}\}$ the path of each commodity $k \in \mathcal{K}$ in solution $\underline{S}$ omitting central hub nodes, with $n_0 = (o_k, -1, e_k)$, $n_{|\Pi_k|} = (d_k, -1, l_k)$ and each node n_p with index $p \in \{1, 2, \ldots, |\Pi_k| - 1\}$ is a route node. From the path Π_k of each commodity k, one can obtain the earliest and latest departure time e_p^k and l_p^k of each node n_p in this path. The earliest visit time of each node n_p in the path of commodity k is calculated by

$$\forall p \in \{1, 2, \ldots, |\Pi_k| - 1\}, \;\; e_p^k = e_k + \sum_{m=1}^{m=p} \hat{\tau}_{n_{m-1}n_m}.$$

The latest time of each node n_p in the path of commodity k is calculated by

$$\forall p \in \{1, 2, \ldots, |\Pi_k| - 1\}, \;\; l_p^k = l_k - \sum_{m=1}^{m=p} \hat{\tau}_{n_{|\Pi_k|-m-1}n_{|\Pi_k|-m}}.$$

For each timed route copy $\omega \in \Omega'$ used in solution $\underline{S}$, let $\underline{s}_\omega$ be its start time. The set of commodities transported by ω is denoted by $\mathcal{K}_\omega$. For each hub $h \in \mathcal{H}$, let $\mathcal{N}'_{\Omega',h}$ be the node set of its timed route copies. Since no en-route waiting is allowed, parameter u_r^n represents the cumulative travel duration on route r from the start node of r to the node n.

Since the travel time inside each route and the commodity due times are always respected by construction, the only possible violations of time constraints occur on either the pick-up or the transfer of each commodity k.

Let variables ϵ indicate time constraint violations of solution $\underline{S}$ applied to CTSNDRP. For each commodity $k \in \mathcal{K}$ and for each pick-up arc (n_0, n_1) with $n_0 = (o_k, -1, e_k)$ and $n_1 = (o_k, r, t)$, binary variable $\epsilon_{0,1}^k$ equals 1 if the pick-up is planned earlier than the earliest available time e_k of commodity k. For each commodity $k \in \mathcal{K}$, if two consecutive nodes n_p and n_{p+1} in its path Π_k correspond to timed route copies of a hub h in two different routes (i.e. $n_p = (h, r_1, t_1) \in \mathcal{N}'_{\Omega',h}$ and $n_{p+1} = (h, r_2, t_2) \in \mathcal{N}'_{\Omega',h}$ with $r_1 \neq r_2$), binary variable $\epsilon_{p,p+1}^k$ equals 1 if the transfer is allowed to be shorter than the real transfer duration at hub h. To make the model simpler, for each commodity $k \in \mathcal{K}$, for each index $p \in \{1, \dots, |\Pi_k|\}$, intermediate continuous variable δ_p^k is for departure time at node $n_p \in \mathcal{N}'_{\Omega'}$ in commodity path Π_k.

The objective (10) is to minimize the total number of violated time constraints. If the objective value is not zero, non-zero ϵ values indicate timed route copies that need to be rescheduled in the partially expanded network. The second term of the objective (10) is weighted by the value p to detect violations near the start of a commodity path first.

$$\min \sum_{k \in \mathcal{K}} \epsilon_{0,1}^k + \sum_{k \in \mathcal{K}} \sum_{\substack{p \in \{1,2,\dots,|\Pi_k|-2\} \\ \exists h \in \mathcal{H}, n_p, n_{p+1} \in \mathcal{N}'_{\Omega',h}}} p \epsilon_{p,p+1}^k \tag{10}$$

$$\delta_1^k \geq e_k(1 - \epsilon_{0,1}^k) \quad \forall k \in \mathcal{K} \tag{11}$$

$$l_k \geq \delta_{|\Pi_k|-1}^k \quad \forall k \in \mathcal{K} \tag{12}$$

$$\delta_{p+1}^k - \delta_p^k + (t_1 - t_2)\epsilon_{p,p+1}^k \geq 0, \\ \forall k \in \mathcal{K}, \forall p \in \{1, 2, \dots, |\Pi_k| - 2\}, n_p = (h, r_1, t_1), n_{p+1} = (h, r_2, t_2) \in \mathcal{N}'_{\Omega',h} \tag{13}$$

$$\delta_p^k = \underline{s}_\omega + u_r^i, \quad \forall r \in \mathcal{R}, \forall \omega \in \Omega'(r), \forall k \in \mathcal{K}_\omega, \forall n_p = (i, r, t) \in \Pi_k \tag{14}$$

$$\delta_p^k \geq 0 \quad \forall p \in \{1, 2, \dots, |\Pi_k| - 1\}, \forall k \in \mathcal{K} \tag{15}$$

$$\epsilon_{p,p+1}^k \in \{0, 1\} \quad \forall k \in \mathcal{K}, \forall p \in \{0, 1, \dots, |\Pi_k| - 1\} (n_p, n_{p+1}) \in \mathcal{A}'_l \cup \mathcal{A}'_f \tag{16}$$

Constraints (11) and (12) are for the earliest available and latest due time of each commodity. Constraints (13) are for the transfer duration on the path of each commodity. Constraints (14) set the departure time of each node. Constraints (15) and (16) define the variable domains.

From the non-zero ϵ values in the solution of the MIP model, missing timed route copies are identified by a route r with a new starting time t. Route r is identified from node n_{p+1} of a non-zero $\epsilon_{p,p+1}^k$ of commodity k. If n_{p+1} is an origin node, the new starting time is set to be the earliest starting time of the route visiting n_p to serve k. If n_p and n_{p+1} are hubs, the new starting time of the route r visiting n_{p+1} is set to visit n_{p+1} at time δ_p^k or the latest starting time of the route r to serve k. If a timed route copy in solution $\underline{S}$ is scheduled before its earliest visit time given by e_p^k, then a new copy is added to visit n_p at e_p^k.

6.2.4 Conversion to a solution to CTSNDRP

From a solution $\underline{S}$ to SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$), a feasible solution $\overline{S}$ to CTSNDR can be obtained by solving an LP. The LP fixes commodity paths using the same physical routes as in the solution $\underline{S}$, while making variable the starting time of each route and the consolidation plan.

The same notations are used for the commodity paths as in Section 6.2.3. There are three sets of variables. For each commodity $k \in \mathcal{K}$, for each $p \in \{1, \ldots, |\Pi_k|\}$, continuous variable δ_p^k represents the departure time at node $n_p \in \mathcal{N}'_{\Omega'}$ in the commodity path Π_k. For each commodity $k \in \mathcal{K}$, for each timed copy $\omega \in \Omega(r)_k$ of route r used by k, continuous variable s_ω^k is for the starting time of copy ω when used to transport commodity k. For each timed route copy $\omega \in \Omega'$ shared by at least two commodities, for each pair of commodities $k_1, k_2 \in \mathcal{K}_\omega$, continuous variable $\epsilon_\omega^{k_1,k_2}$ indicates the difference of start times of ω when carrying k_1 and $k2$.

The objective (17) is to minimize the total difference of starting times of timed route copies when used by different commodities, so that the consolidation plan given by solution $\underline{S}$ can be kept as much as possible.

$$\min \sum_{\omega \in \Omega} \sum_{k_1,k_2 \in \mathcal{K}_\omega} \epsilon_\omega^{k_1,k_2} \tag{17}$$

$$\begin{aligned}
\delta_1^k &\geq e_k && \forall k \in \mathcal{K} && (18)\\
l_k &\geq \delta_{|\Pi_k|-1}^k && \forall k \in \mathcal{K} && (19)\\
\delta_{p+1}^k - \delta_p^k &\geq 0 && \forall k \in \mathcal{K}, \forall p \in \{1,2,\ldots,|\Pi_k|-2\}\ n_p, n_{p+1} \in \mathcal{N}'_{\omega',h} && (20)\\
\delta_p^k &= s_\omega^k + u_r^i && \forall r \in \mathcal{R}, \forall \omega \in \Omega(r), \forall k \in \mathcal{K}_\omega, \forall n_p = (i,r,t) \in \Pi_k && (21)\\
s_\omega^{k_1} - s_\omega^{k_2} &\leq \epsilon_\omega^{k_1,k_2} && \forall \omega \in \Omega, \forall k_1, k_2 \in \mathcal{K}_\omega && (22)\\
s_\omega^{k_2} - s_\omega^{k_1} &\leq \epsilon_\omega^{k_1,k_2} && \forall \omega \in \Omega, \forall k_1, k_2 \in \mathcal{K}_\omega && (23)\\
\delta_n^k &\geq 0 && \forall k \in \mathcal{K}, \forall n \in \Pi_k && (24)\\
\epsilon_\omega^{k_1 k_2} &\geq 0 && \forall k_1, k_2 \in \mathcal{K}_\omega, k_1 < k_2, \forall \omega \in \Omega && (25)\\
s_\omega^k &\geq 0 && \forall k \in \mathcal{K}, \forall \omega \in \Omega_k && (26)
\end{aligned}$$

Constraints (18) and (19) are for the earliest available and latest due time of each commodity. Constraints (20) make sure that transfer duration is respected. Constraints (21) ensure that travel duration is respected. Constraints (22) and (23) computes the difference of start times of the same timed copy when used by different commodities. Constraints (24)–(26) are variable domains.

If a solution with zero objective is found by the LP, then solution $\underline{S}$ on the partial network can be converted to one on the complete network with the same route starting times and consolidation plans. In this case, if the solution $\underline{S}$ is optimal for SNDR($\mathcal{G}'_{\mathcal{R},\mathcal{T}'}$), then an optimal solution for CTSNDRP is found.

If a solution with strictly positive objective is found by the LP, then some of the consolidations in solution $\underline{S}$ need to be separated, introducing higher costs. An upper bound to CTSNDRP is obtained in this case.

6.2.5 Refine and restore

The algorithm for refining and restoring the partially expanded network is similar to the algorithm proposed by Boland et al. [2017a]. The difference is that routes are manipulated instead of arcs.

In particular, attention needs to be paid when adding transfer arcs so that timed route copies can only be linked to central hub nodes that are visited no later than the latest starting time of the route. This is to make sure that no late delivery is introduced.

The partially expanded network is refined by adding new timed route copies discovered by the MIP defined in 6.2.3 and linking them to the existing network. This includes

- adding new timed route copies and central hub nodes according to Property 4;
- adding linking arcs for origin and destination nodes to maintain Property 2;
- replacing holding arcs by new ones if new central hub nodes are added so that Property 7 is maintained;
- link new central hub to existing timed route copies to keep Property 5.

After the refinement of the partially-expanded network, for each new timed route copies added, unnecessary transfers and linking arcs are removed so that Property 6 is kept.

6.3 Example

The following example illustrates how the adapted DDD algorithm works for solving the CTSNDRP on a very simple network. In this example, the solution on the initial network is a relaxation of the final one, by allowing the two commodities to be consolidated on an early copy of route r_1. The algorithm discovers and removes links to these copies that are too early and add new ones that are needed, to gradually converge to the optimal solution of the problem defined on the complete network.

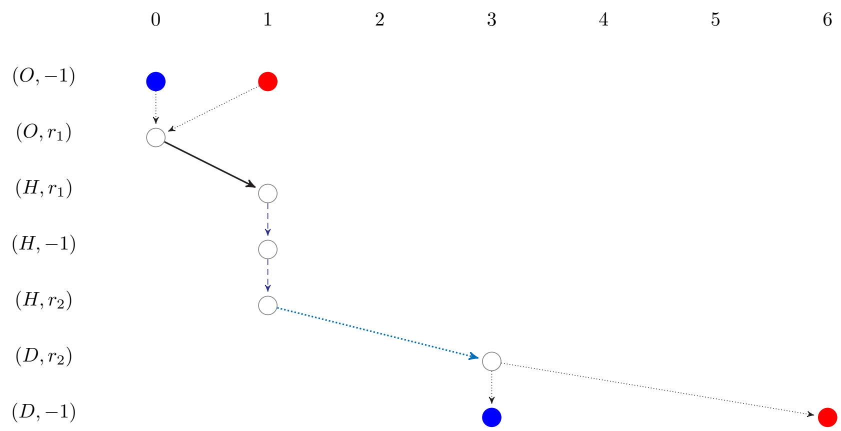
Example 3. *Consider a network with three points, an origin node O, a destination node D and a hub H. Inbound route r_1 goes from O to H, and outbound route r_2 goes from H to D. Two commodities k_1 and k_2 both originate from O and to be delivered to D. k_1 is available at time 1 and due at time 6. k_2 is available at time 0 and due at time 3.*

Figure 6a to 6e show one iteration of the algorithm. The network is initialized with route r_1 starting at time 0, and r_2 starting at time 1 as in Figure 6a. After solving the SNDR on this partially expanded network, the two commodity travels together on the only copies of route r_1 and r_2. The algorithm discovers that the arc linking the origin node of k_1 $(O, -1, 1)$ and the time route copy $(O, r_1, 0)$ is too short (Figure 6b). It proposes to add a copy of r_1 at time 1. The new timed copy, together with the corresponding linking, transfer and holding arcs are then added to the network as in Figures 6c and 6d. The network is restored so that all the properties of the partially expanded network are maintained (Figure 6e).

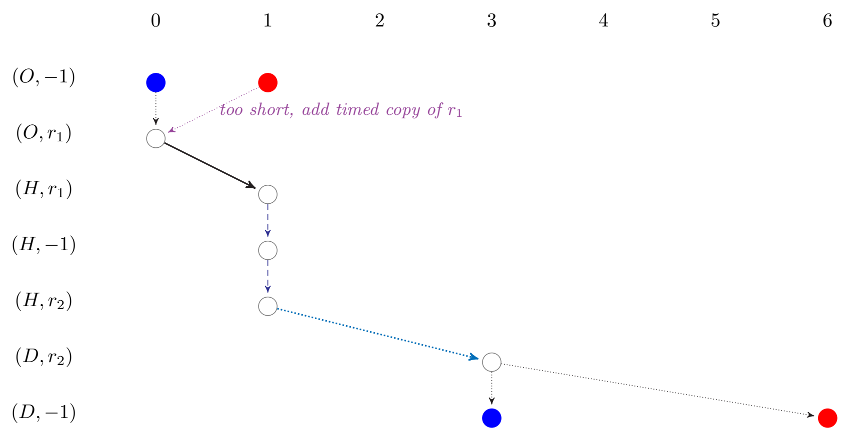
The process is repeated until the optimal solution is found (Figure 6f). The final network expanded is shown in Figure 7, in which the two commodities travel separately.

7 Computational Results

In this section, we computationally study two issues. The first is the performance of the proposed DDD-based algorithm. The second is the savings potential associated with different routing strategies considered by the CTSNDRP (e.g. direct delivery from supplier to retail store). All experiments were conducted with instances generated based on historical data from our industrial

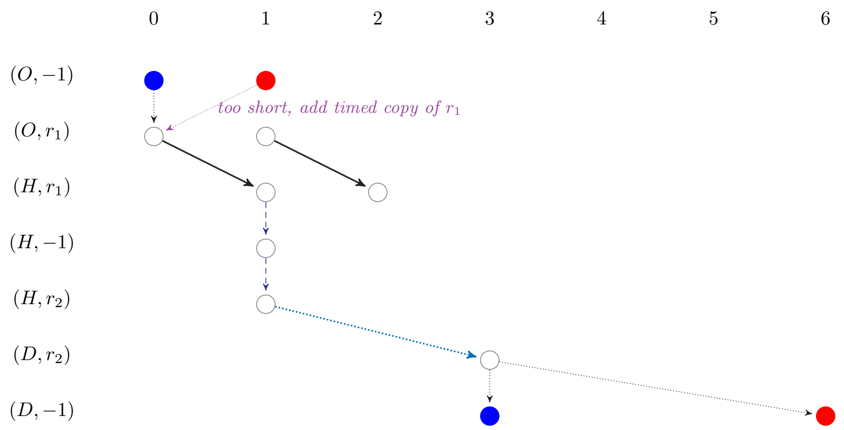


(a) Initialization

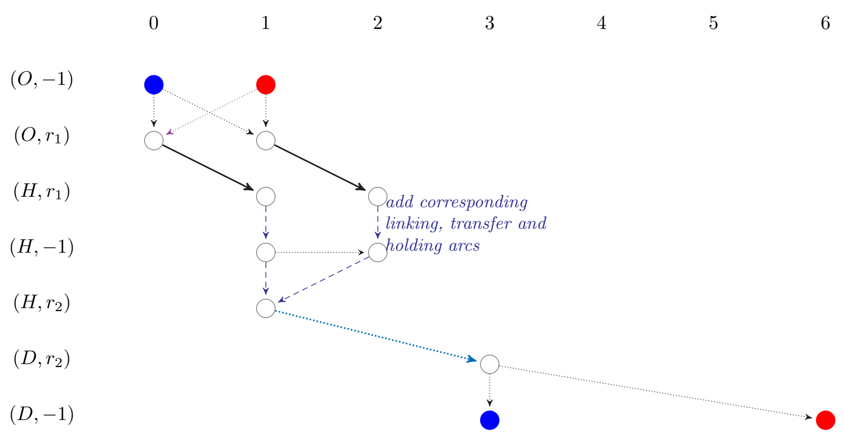


(b) Solve SNDR model, check if the solution can be converted. if yes, solved; otherwise, detect arcs that are too short

Figure 6: Example of DDD for CTSNDRP (Initialization and Discretization Discovery)

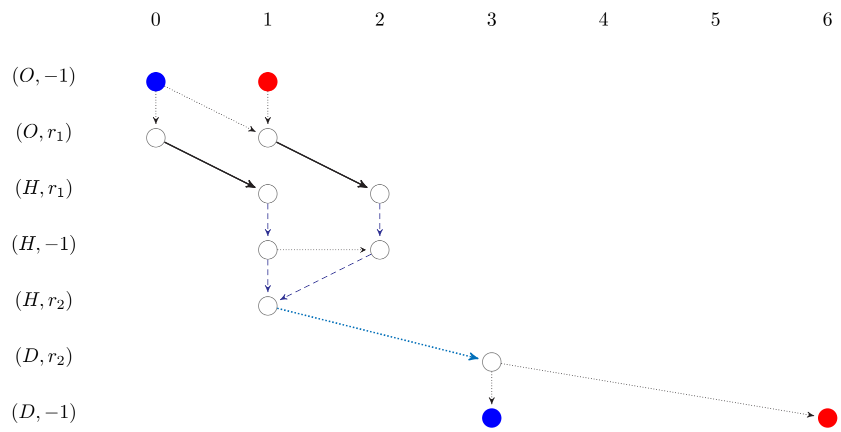


(c) Refine the network, add new timed copies of routes

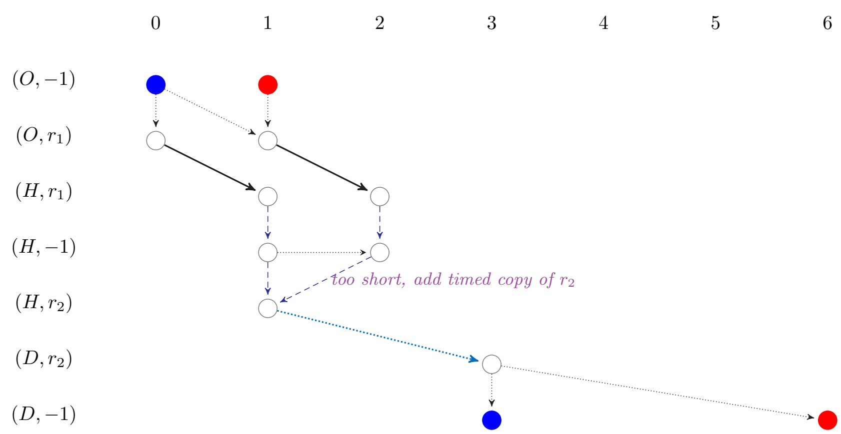


(d) Refine the network, add other arcs

Figure 6: Example of DDD for CTSNDRP (Refine)



(e) Restore the network, remove arcs that are no longer valid



(f) If not solved, repeat 2–4: solve SNDR on the new partially expanded network

Figure 6: Example of DDD for CTSNDRP (Restore and Repeat)

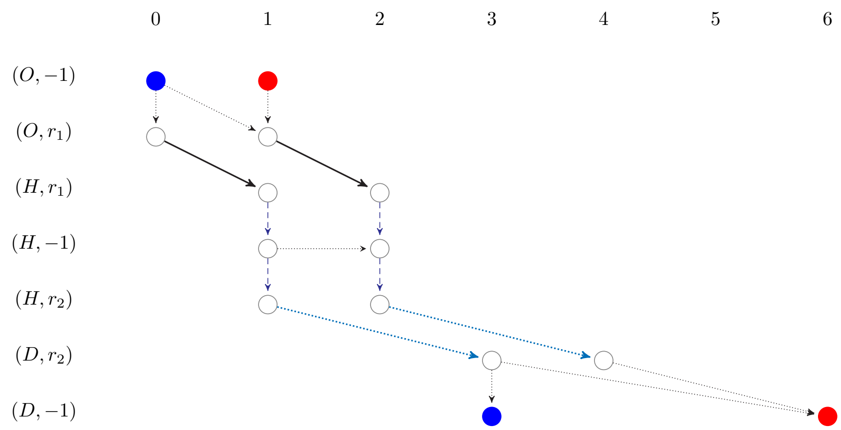


Figure 7: Final network of Example 3

partner. A description of the instance generation process is presented in Section 7.1. A description of the different operational scenarios we considered in order to assess savings potentials is presented in Section 7.2. Section 7.3 benchmarks the performance of DDD with a MIP solver. Finally, Section 7.4 presents an analysis of savings potential associated with different routing strategies.

From a computational perspective, we note that all experiments were conducted on a server running Linux, Ubuntu 20.04.2 LTS, equipped with an Intel Xeon Gold 6230 @ 2.10GHz. The algorithm is implemented in Java and calls the CPLEX 20.1.0 MIP solver with 0.01% relative MIP gap tolerance and all the other parameters left at default values. The Java virtual machine was run with a 2GB memory limit and a time limit of 2 hours.

7.1 Instance generation

The instances used in the computational study represent a network with 6 suppliers, 4 hubs, and 291 geographical delivery points scattered around France. The physical location of each site (supplier, hub, delivery point) is known, as are the time at which each location opens. Furthermore, all carrier trucks are 33 pallets semi-trailers (the standard in Europe). The industrial partner provided cost information regarding specific sets of inbound routes from suppliers to hubs, outbound routes from hubs to delivery points, inter-hub routes, and direct delivery routes. These routes were complemented with other routes that were generated based on the patterns explained in Section 3 and their costs were computed in agreement with information provided by the industrial partner. Lastly, the industrial partner also provided data regarding the cost and duration of transfer activities at hubs.

The historical data represents one year of operation and contains more than 33,000 orders. We model each order as a distinct commodity. To generate instances, we partitioned the historical data by week. Then, for a given week, we randomly sampled orders that originated in that week to define commodities. In the initial historical data, each facility may only be open during certain hours on each day but the earliest available times (e_k) or latest delivery times (l_k) for orders are not provided. These times are computed as follows. The earliest available time of each order e_k is randomly set to a time during which the corresponding supplier is known to be open. The latest due time of each order l_k is set to be the latest delivery date according to an approximation of the travel duration from the supplier to the delivery point and the opening times of the delivery point.

Ultimately, 37 instances were generated. The number of commodities in these instances range from 9 to 102. The quantity of each commodity varies between 1 and 33 pallets with a standard deviation of 10.94, 50% of the commodities have fewer than 6 pallets and 75% of the commodities have fewer than 19 pallets.

7.2 Operational scenarios

To understand the savings potential of the different routing strategies allowed by the CTSNDRP, we define a base scenario, denoted B, which corresponds to practices regularly employed by our industrial partner. Costs incurred in this operational scenario serve as a basis for comparison in our analysis.

In scenario B, each inbound route of type P-H goes from one pick-up point to one hub. There are inter-hub routes of type H-H between each pair of hubs and each inter-hub route contains two hubs. The allocation of delivery points is fixed: each delivery point is allocated to it nearest hub. Each outbound route can visit at most three sites, among which two are delivery points. Outbound routes are of type H-D or H-D-D.

We compare the costs from solving instances of the CTSNDRP constructed to reflect scenario B with the following operational scenarios. Each of the following scenarios augment Scenario B with additional routing options as follows.

- *Scenario D*: Is *Scenario B* but where we also allow commodities to be shipped directly from their pick-up point to their delivery point using routes of type P-D.
- *Scenario MS (multi-sourcing)*: Is *Scenario B* but where we additionally allow a delivery point to receive goods from any of the existing hubs in the network. In this scenario, outbound routes of type H-D and H-D-D contain all combinations from hubs to delivery points and each route contains at most two delivery points. These routes must also be valid according to carrier constraints.
- *Scenario PDR*: Is *Scenario B* but where routes are allowed to visit different types of facilities. Namely, we consider additional routes of the form P-H-H and P-P-H, and pick-up-and-delivery routes of the form P-H-D.

We also tested combinations of different scenarios, which we note with a plus (+) sign. For example, MS+PDR+D is the combination of scenarios MS, PDR and D, in which there are multiple hubs supplying a single delivery point (as in scenario MS), routes visiting several types of facilities (as in PDR) and direct deliveries (as in D).

7.3 Performance analysis

To benchmark the performance of the adapted DDD algorithm, we constructed a complete network (denoted CN) as follows. Between the preprocessed earliest and latest starting time of each route, the route is copied at each time point. Each commodity origin node is connected to all the route copies that can visit the origin node at the earliest available time of the commodity onward. Similarly, for each route that can visit the destination node of a commodity, there is a route copy that is connected to the destination node no later than the latest due time of the commodity. Transfers at hubs are always allowed, as long as the transfer duration is enough. We solved the resulting instance of the CTSNDRP on CN using CPLEX.

For these experiments, the memory allocated to CPLEX was increased to 6GB when solving the CN. Table 1 compares solving the CN with CPLEX to the DDD-based algorithm. In this table, for each method (CPLEX, DDD), the column nbFeas contains the number of instances for which the method could produce a feasible solution. The column nbOpt contains the number of instances for which the method could produce a provably optimal solution. The column avgGap contains the average optimality gap reported at termination by each method while column avgT contains the average time to termination for each method. Finally, the column UBGap contains the difference in percentage between the objective function value of the DDD solution and the solution produced by CPLEX. We note UBGap is only reported for instances for which CPLEX could produce a feasible solution.

Scenario	CPLEX				DDD				avg
	nbFeas	nbOpt	avgGap	avgT	nbFeas	nbOpt	avgGap	avgT	UBGap
B	34	2	13.9%	6776	37	27	1.2%	2819	3.3%
D	34	4	12.4%	6389	37	28	1.0%	2394	2.9%
MS	28	3	24.9%	6515	37	19	5.5%	4043	19.9%
MS+D	29	3	24.3%	6547	37	19	4.4%	4117	15.8%
PDR	30	3	15.9%	6519	37	17	3.3%	4697	3.2%
PDR+D	29	3	16.2%	6496	37	22	2.9%	3773	4.7%
MS+PDR	20	3	33.2%	6326	37	7	10.4%	6056	48.3%
MS+PDR+D	20	3	24.8%	6261	37	10	10.0%	5548	15.6%
Average	29.25	3.00	20.70%	6,478.63	37.00	18.63	4.84%	4,180.88	14.21%

Table 1: Performance analysis - comparison with CPLEX

We see that the proposed DDD-based algorithm outperforms CPLEX in multiple ways. First, it finds a feasible solution for every instance while CPLEX does not. Second, it produces far more provably optimal solutions. Third, it produces solutions that are of provably higher quality, and in less time. In scenarios MS+PDR and MS+PDR+D, CPLEX could not find a feasible solution for 17 instances because the memory limit was reached. Compared to CPLEX, the UBGap of the most difficult scenario MS+PDR is 48.3%. We conclude from these results that the proposed DDD algorithm outperforms the benchmark CPLEX.

To further improve the performance of the DDD algorithm on an instance reflecting one of the more difficult scenarios, MS+PDR and MS+PDR+D, we first executed DDD for two hours on that instance, but for an easier scenario (PDR, PDR+D) to derive an initial solution. We then executed DDD for two hours for the actual scenario, albeit warm-started with the initial solution from the easier scenario. The results from using this warm start solution are reported in Table 2.

We see that with the warm start, DDD is able to prove optimality for more instances and produce solutions of provably high-quality for instances it can not solve without a warm start.

7.4 Result analysis

In this section, we study the savings potential of solutions derived from solving instances of the proposed model that are constructed to represent different operational scenarios. To do so, we com-

Scenario	Without warm-start					With warm-start				
	nbFeas	nbOpt	avgGap	avgT	UBGap	nbFeas	nbOpt	avgGap	avgT	UBGap
MS+PDR	37	7	10.4%	6056	48.3%	37	11	6.4%	5626	52.6%
MS+PDR+D	37	10	10.0%	5548	15.6%	37	11	6.4%	5313	16.8%
Average	37	8.50	10.20%	5,802.00	31.95%	37	11	6.40%	5,469.50	34.70%

Table 2: Warm-starting DDD with solutions from easier scenarios

pare the cost and characteristics of the solution in different operational scenarios to give managerial insights.

To estimate the savings potential of different routing and consolidation configurations in the network, we compare the cost of instances under scenarios D, MS and PDR with the base scenario B. We calculate the relative gap as the objective value in scenario X minus the objective value in scenario B divided by the objective value in scenario B using the formula below:

$$gap = \frac{\mathrm{obj}_X - \mathrm{obj}_B}{\mathrm{obj}_B}.$$

Table 3 reports the mean, minimum and maximum values of this gap.

Scenario	average reduction	min reduction	max reduction
B vs D	22.0%	12.2%	30.9%
B vs MS	8.4%	4.7%	12.3%
B vs PDR	18.3%	11.4%	27.1%

Table 3: Average, min, max cost reductions achieved individually by the various practices.

We see that allowing more flexibility in the routing strategy can significantly reduce the total cost. Among the three scenarios (D, MS and PDR), allowing direct shipping (scenario D) gives the most reduction of total cost with an average of 22.0%, followed by visiting several types of facilities in the same route (scenario PDR) with an average of 18.3% and lastly by relaxing single sourcing constraints (scenario MS), with an average of 8.4%.

To understand why there are cost reductions, Figure 8 shows the breakdown of total costs by different types of routes under each of the operational scenarios and, Figure 9 shows the number of times that each types of route has been used in the solution. Note that the costs and route counts reported for scenarios MS+PDR and MS+PDR+D in these two figures are those obtained by DDD when warm-started. In Figure 8, the total height of each bar shows the total cost, which is composed of fixed route costs of various types of routes and variable transfer cost of category "Var" in the figure. Similarly, in Figure 9, the total height of each bar shows the total number of routes of all types used in this scenario. In both figures, the total value of each scenario are marked on top of each bar.

In general, direct shipping (P-D) is taken advantage of when allowed, which results in lower total costs and fewer routes used in each category.

Under scenario MS, more commodities are consolidated in the same inbound (P-H) and outbound (H-D or H-D-D) routes, which is reflected by the fact that the number of inbound (P-H) and outbound (H-D or H-D-D) routes stays almost the same in Figure 9, while there is a significant

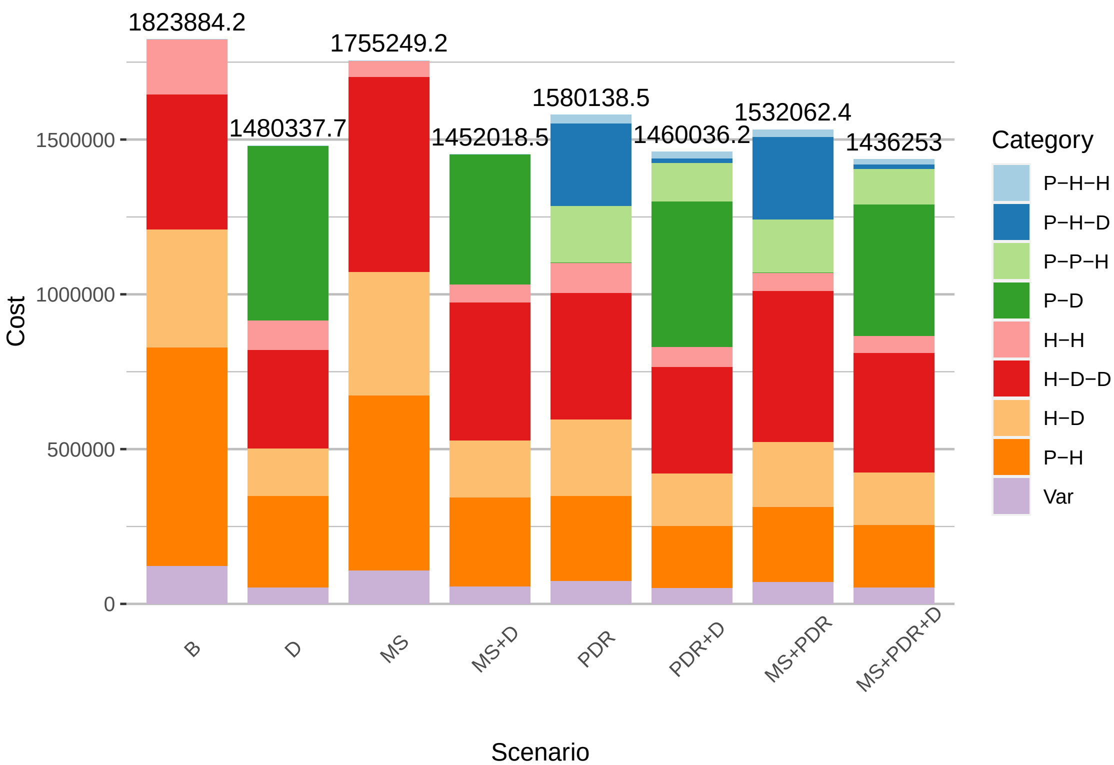


Figure 8: Cost distribution for each scenario

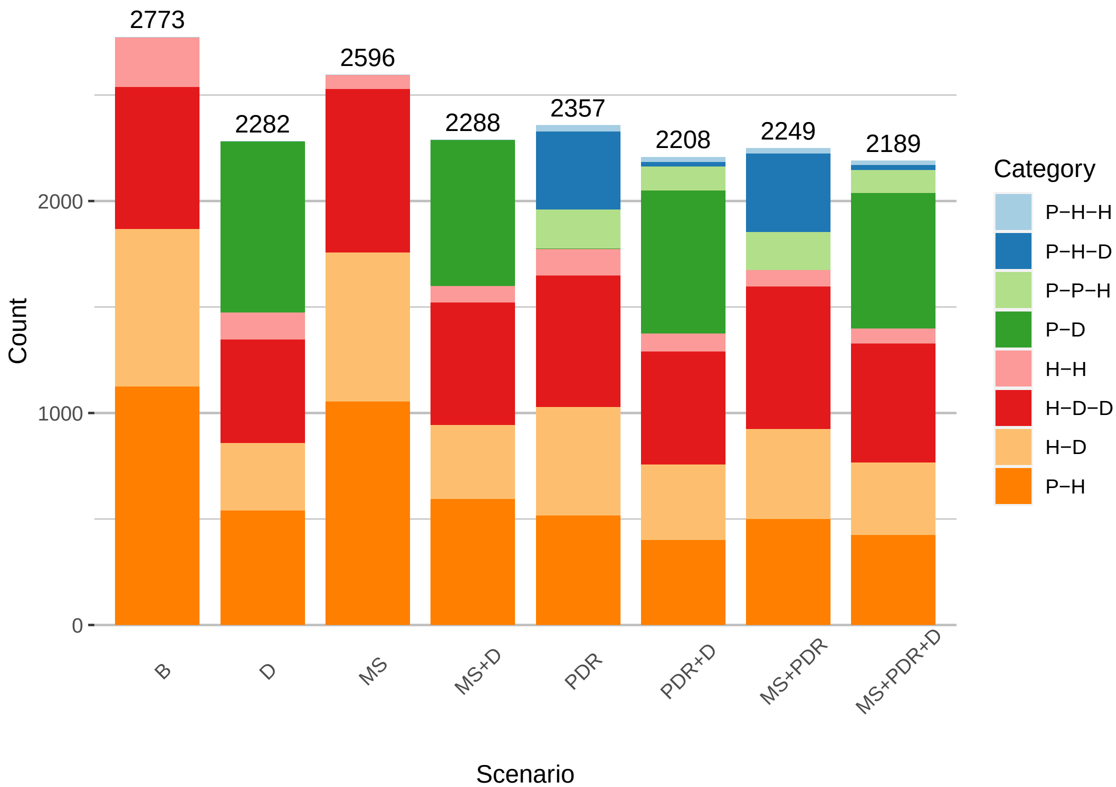


Figure 9: Route usage for each scenario

reduction in inbound (P-H) and inter-hub (H-H) route costs, but an increase in outbound route costs of type H-D-D because some of the commodities might have to travel longer.

In scenarios PDR and MS+PDR, transfers are also included in dual routes of the form P-H-H and P-H-D. As can be seen by Figure 9, there is a significant dual usage of pick-up-and-delivery routes P-H-D in scenarios PDR and MS+PDR. Compared to scenario B, the combined usage of dual routes P-H-H, P-H-D and more flexible inbound routes P-P-H helps to reduce the inbound and outbound (P-H and H-D) route costs, and inter-hub (H-H) route costs, resulting in a reduction in the total costs. There is not much use of dual routes P-H-H, potentially because the quantity of each commodity in the instances is quite small and routes P-H and H-H can better serve these.

The combined scenarios with MS+PDR enable greater savings than the two independent scenarios MS and PDR, but are more difficult to solve than one single scenario, due to huge number of initial routes added to the route-expanded network during the solution. However, Figure 9 shows that the total number of routes used in scenario MS+PDR is still smaller than all the other scenarios, indicating further possible reductions in costs if the performance of the algorithm can be improved even more to deal with large numbers of routes.

The most efficient practice seems to be to allow direct shipping. When direct shipping is allowed, the impact of multiple sourcing and pick-up-and-delivery routes (dual usage) is less. This observation could be due to the fact that all the points in the network are within France, and each two points can be reached within one day of drive. If opportunities of consolidation are limited, it is better to ask for a direct delivery. In case no direct shipping is possible, allowing flexible dual routing options can achieve the same effect.

8 Conclusion

In this paper, we presented an integrated service network design and routing problem that models various routing options that have not yet been considered in the literature. The problem is modeled on a route-time expanded network with a star structure for transfers. One advantage of this modeling is that complex routing cost calculations are separated from the decision process, so it is easy to integrate other types of routing options. We presented an adaptation of the Dynamic Discretization Discovery (DDD) algorithm and proved its correctness. We conducted a case study using data derived from the operations of an industrial partner operating a logistics network in France. Numerical results show that flexible routing strategies such as direct shipping or using routes combining pick-up and delivery sites can significantly reduce costs. To a lesser extent, allowing hubs to serve different delivery points can also reduce the total costs.

This work suggests multiple future research directions. Because the decision of routing and timing is closely linked, when adding new timed copies, it is difficult to say which timing is the best. In each iteration, solving the SNDR defined on the partially expanded network is costly in computing time, especially when instances get larger and when routing options are more complicated. This is mainly because the combinations of possible consolidations becomes very large with more orders to deliver and more routes to consider. Currently, we use predefined routes in the solution process and the number of sites visited by each route is limited to three. This number is quite realistic with respect to the professional practices of long-distance transportation. In the future, it might be interesting to integrate routing into the solution process, generating routes as in column generation. During the execution of the DDD algorithm, the partially route-time-expanded network is constructed carefully so that the infeasiblity only comes from the timing of the routes. Future research could investigate whether it is possible to expand the network not only by time but also by route, by allowing infeasible routes initially, dynamically discover infeasible commodity path and adding more routes to remove these infeasible paths during the solution process.

From a practical point of view, this more agile and flexible network organization can bring huge savings to logistics and transport companies. This study can be extended in different ways. It would be interesting to consider other types of costs such as energy or environmental considerations, given the fact that it is very easy to integrate other cost structures with the route model in this study. How to handle the uncertainty met during the transportation process is another topic to study in the future. In practice, a delivery point can order different parts from different suppliers and ask them to be delivered at the same delivery route, how to handle this synchronization requirements with the route-time-expanded network is also a point to consider in the future. Under the development of the Internet of Things technologies, we would have detailed data about the location and the duration of travel of each carrier. In the future, we can also develop a model with route generation and carrier side constraints such as driving working hours, maximum number of sites visited in the route and other driver working regulations. Other applications consider limited capacities at hubs. He et al. [2022] consider a limit on the number of vehicles that can be loaded and unloaded simultaneously at a hub in a city logistics context. They integrate this limit in a DDD algorithm, and this could be further investigated in the SNDRP as defined in our paper.

Appendix A Notation Tables

symbol	meaning
$\mathcal{D}$	the set of delivery points
$\mathcal{P}$	the set of pick-up points
$\mathcal{H}$	the set of hubs
$\mathcal{K}$	the set of commodities
o_k	the pick-up point of commodity k
d_k	the delivery point of commodity k
e_k	the earliest available time of commodity k
l_k	the latest available time of commodity k
q_k	the quantity of commodity k
c_h	transfer cost per unit at hub h
σ_h	transfer duration at hub h
$\mathcal{G}$	the directed graph for the physical network
$\mathcal{N}$	the set of nodes in graph $\mathcal{G}$
$\mathcal{A}$	the set of arcs in graph $\mathcal{G}$
τ_{ij}	travel duration on arc (i, j)
c_{ij}	cost for each unit transported on arc (i, j)
$\mathcal{R}$	the set of pre-defined transportation routes
$\mathcal{N}^r$	the set of nodes traversed by route r
$\mathcal{A}^r$	the set of arcs traversed by route r
Q_r	the capacity in terms of units of products of route r
c_r	the fixed cost of route r

Table 4: Notations for problem settings in Section 3

symbol	meaning
$\mathcal{G}_{\mathcal{R}}$	route-expanded network
$\mathcal{N}_{\mathcal{R}}$	the set of nodes in the route-expanded network
$\mathcal{A}_{\mathcal{R}}$	the set of arcs in the route-expanded network
$(i, r) \in \mathcal{N}_{\mathcal{R}}$	a route copy for route r visiting physical node i
$(h, -1) \in \mathcal{N}_{\mathcal{R}}$	a central hub node for hub h
$(o_k, -1) \in \mathcal{N}_{\mathcal{R}}$	an origin node of commodity k
$(d_k, -1) \in \mathcal{N}_{\mathcal{R}}$	a destination node of commodity k
$((i, r), (j, r)) \in \mathcal{A}_{\mathcal{R}}$	a routing arc from physical node i to j on route r
$((h, r), (h, -1))$ or $((h, -1), (h, r)) \in \mathcal{A}_{\mathcal{R}}$	a transfer arc at hub h
$((o_k, -1), (o_k, r)) \in \mathcal{A}_{\mathcal{R}}$	a pick-up arc for commodity k by route r
$((d_k, r), (d_k, -1)) \in \mathcal{A}_{\mathcal{R}}$	a delivery arc for commodity k by route r

Table 5: Notations for route-expanded network

symbol	meaning
$\mathcal{G}_{\mathcal{R},\mathcal{T}}$	route-time-expanded network
$\mathcal{N}_{\mathcal{R},\mathcal{T}}$	the set of nodes in route-time-expanded network
$\mathcal{A}_{\mathcal{R},\mathcal{T}}$	the set of arcs in route-time-expanded network
$\mathcal{N}_{OD} \subset \mathcal{N}_{\mathcal{R},\mathcal{T}}$	the set of timed copies of origin/destination of each commodity
$\mathcal{N}_{\Omega} \subset \mathcal{N}_{\mathcal{R},\mathcal{T}}$	the set of timed route copies
$\mathcal{N}_{\mathcal{H}} \subset \mathcal{N}_{\mathcal{R},\mathcal{T}}$	the set of timed central hub nodes
$\mathcal{A}_p \subset \mathcal{A}_{\mathcal{R},\mathcal{T}}$	the set of transportation arcs or timed copies of routing arcs
$\mathcal{A}_f \subset \mathcal{A}_{\mathcal{R},\mathcal{T}}$	the set of timed transfer arcs
$\mathcal{A}_l \subset \mathcal{A}_{\mathcal{R},\mathcal{T}}$	the set of linking arcs or timed copies of pick-up and delivery arcs
$\mathcal{A}_h \subset \mathcal{A}_{\mathcal{R},\mathcal{T}}$	the set of holding arcs at hubs
$\mathcal{T}$	the set of time points
$\mathcal{T}_{(i,r)}$	the set of time points of a node (i, r) in the route-expanded network
Ω	the set of timed copies of all timed route copies
$\Omega(r)$	the set of timed copies of route r
$\omega \in \Omega(r)$	a timed copy of route r
$a_{i_1 i_2}^{\omega t_1 t_2}$	binary parameter equal to 1 if ω goes from node i_1 at time t_1 to node i_2 at time t_2
$\hat{\tau}_{ij}$	general travel duration from node i to node j in the route-time-expanded network
(i, r, t)	general representation of a node in the route-time-expanded network

Table 6: Notations in route-time-expanded network

Acknowledgment

This work was supported by the French Agency for Ecological Transition (Agence de la transition écologique (ADEME)). The authors would like to thank the two reviewers for helping improving this paper.

References

J. Andersen, T. G. Crainic, and M. Christiansen. Service network design with asset management: Formulations and comparative analyses. *Transportation Research Part C: Emerging Technologies*, 17(2):197 – 207, 2009a.

J. Andersen, T. G. Crainic, and M. Christiansen. Service network design with management and coordination of multiple fleets. *European Journal of Operational Research*, 193(2):377 – 389, 2009b. ISSN 0377-2217.

S. Belieres, M. Hewitt, N. Jozefowiez, and F. Semet. A time-expanded network reduction matheuristic for the logistics service network design problem. *Transportation Research Part E: Logistics and Transportation Review*, 147:102203, 2021.

N. Boland, M. Hewitt, L. Marshall, and M. Savelsbergh. The continuous-time service network design problem. *Operations Research*, 65(5):1303–1321, 2017a. doi: 10.1287/opre.2017.1624. URL `https://doi.org/10.1287/opre.2017.1624`.

N. Boland, M. Hewitt, D. M. Vu, and M. Savelsbergh. Solving the traveling salesman problem with time windows through dynamically generated time-expanded networks. In *International Conference on AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, pages 254–262. Springer, 2017b.

N. Boland, M. Hewitt, L. Marshall, and M. Savelsbergh. The price of discretizing time: a study in service network design. *EURO Journal on Transportation and Logistics*, 8(2):195–216, June 2019. ISSN 2192-4376. doi: 10.1007/s13676-018-0119-x. URL `https://www.sciencedirect.com/science/article/pii/S2192437620300327`.

J. F. Campbell and M. E. O'Kelly. Twenty-five years of hub location research. *Transportation Science*, 46(2):153–169, 2012.

I. Contreras and M. O'Kelly. *Hub Location Problems*, pages 327–363. Springer International Publishing, Cham, 2019.

T. G. Crainic. Service network design in freight transportation. *European Journal of Operational Research*, 122(2):272–288, 2000.

T. G. Crainic, M. Hewitt, M. Toulouse, and D. Vu. Service network design with resource constraints. *Transportation Science*, 50(4):1380–1393, 2016.

T. G. Crainic, M. Hewitt, M. Toulouse, and D. M. Vu. Scheduled service network design with resource acquisition and management. *EURO Journal on Transportation and Logistics*, 7(3): 277–309, 2018.

G. Desaulniers, O. B. Madsen, and S. Ropke. Chapter 5: The vehicle routing problem with time windows. In *Vehicle Routing: Problems, Methods, and Applications, Second Edition*, pages 119–159. SIAM, 2014.

M. Drexl. Synchronization in vehicle routing-a survey of VRPs with multiple synchronization constraints. *Transportation Science*, 46(3):297–316, 2012.

A. Erera, M. Hewitt, M. Savelsbergh, and Y. Zhang. Improved load plan design through integer programming based local search. *Transportation Science*, 47(3):412–427, 2013a.

A. L. Erera, M. Hewitt, M. W. Savelsbergh, and Y. Zhang. Creating schedules and computing operating costs for LTL load plans. *Computers and Operations Research*, 40(3):691 – 702, 2013b.

P. Grangier, M. Gendreau, F. Lehuédé, and L.-M. Rousseau. A matheuristic based on large neighborhood search for the vehicle routing problem with cross-docking. *Computers & Operations Research*, 84:116–126, 2017.

G. Guastaroba, M. G. Speranza, and D. Vigo. Intermediate Facilities in Freight Transportation Planning: A Survey. *Transportation Science*, 50(3):763–789, Aug. 2016. doi: 10.1287/trsc.2015.0631.

E. He, N. Boland, G. Nemhauser, and M. Savelsbergh. An exact algorithm for the service network design problem with hub capacity constraints. *Networks*, 80(4):572–596, 2022.

M. Hewitt. Enhanced dynamic discretization discovery for the continuous time load plan design problem. *Transportation Science*, 53(6):1731–1750, 2019.

M. Hewitt. The flexible scheduled service network design problem. *Transportation Science*, 2022.

M. Hewitt, T. G. Crainic, M. Nowak, and W. Rei. Scheduled service network design with resource acquisition and management under uncertainty. *Transportation Research Part B: Methodological*, 128:324–343, 2019.

A. I. Jarrah, E. Johnson, and L. C. Neubert. Large-scale, less-than-truckload service network design. *Operations Research*, 57(3):609–625, 2009.

C. V. Karsten, D. Pisinger, S. Ropke, and B. D. Brouer. The time constrained multi-commodity network flow problem and its application to liner shipping network design. *Transportation Research Part E: Logistics and Transportation Review*, 76:122–138, 2015.

Y. H. Lee, J. W. Jung, and K. M. Lee. Vehicle routing scheduling for cross-docking in the supply chain. *Computers and Industrial Engineering*, 51"(2):247 – 256, 2006.

A. I. Mahmutoğulları and B. Y. Kara. Hub Location Problem with Allowed Routing between Nonhub Nodes: Hub Location Problem. *Geographical Analysis*, 47(4):410–430, 2015.

L. Marshall, N. Boland, M. Savelsbergh, and M. Hewitt. Interval-based dynamic discretization discovery for solving the continuous-time service network design problem. *Transportation Science*, 55(1):29–51, 2021.

J. Medina, M. Hewitt, F. Lehuédé, and O. Péton. Integrating long-haul and local transportation planning: the service network design and routing problem. *EURO Journal on Transportation and Logistics*, 8(2):119–145, 2019.

V. W. Morais, G. R. Mateus, and T. F. Noronha. Iterated local search heuristics for the vehicle routing problem with cross-docking. *Expert Systems with Applications*, 41(16):7495–7506, 2014.

A. I. Nikolopoulou, P. P. Repoussis, C. D. Tarantilis, and E. E. Zachariadis. Moving products between location pairs: Cross-docking versus direct-shipping. *European Journal of Operational Research*, 256(3):803–819, Feb. 2017. ISSN 0377-2217. doi: 10.1016/j.ejor.2016.06.053.

M. Nowak, M. Hewitt, and H. Bachour. Mileage bands in freight transportation. *European Journal of Operational Research*, 272(2):549–564, 2019.

D. C. Paraskevopoulos, G. Laporte, P. P. Repoussis, and C. D. Tarantilis. Resource constrained routing and scheduling: Review and research prospects. *European Journal of Operational Research*, 263(3):737–754, 2017.

M. B. Pedersen, T. G. Crainic, and O. B. Madsen. Models and tabu search metaheuristics for service network design with asset-balance requirements. *Transportation Science*, 43(2):158–177, 2009.

W. B. Powell and Y. Sheffi. Or practice—design and implementation of an interactive optimization system for network design in the motor carrier industry. *Operations Research*, 37(1):12–29, 1989.

M. SteadieSeifi, N. P. Dellaert, W. Nuijten, T. Van Woensel, and R. Raoufi. Multimodal freight transportation planning: A literature review. *European journal of operational research*, 233(1): 1–15, 2014.

D. M. Vu, M. Hewitt, N. Boland, and M. Savelsbergh. Dynamic discretization discovery for solving the time-dependent traveling salesman problem with time windows. *Transportation Science*, 54 (3):703–720, 2020.

M. Wen, J. Larsen, J. Clausen, J.-F. Cordeau, and G. Laporte. Vehicle routing with cross-docking. *Journal of the Operational Research Society*, 60(12):1708–1718, 2008.

D. Wolfinger, F. Tricoire, and K. F. Doerner. A matheuristic for a multimodal long haul routing problem. *EURO Journal on Transportation and Logistics*, pages 1–37, 2018.

B. Yıldız, H. Yaman, and O. E. Karaşan. Hub location, routing, and route dimensioning: Strategic and tactical intermodal transportation hub network design. *Transportation Science*, 55(6):1351–1369, 2021.