This article was downloaded by: [216.73.217.81] On: 24 June 2026, At: 01:13
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Transportation Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Intermediate Facilities in Freight Transportation Planning: A Survey

G. Guastaroba, M. G. Speranza, D. Vigo

To cite this article:
G. Guastaroba, M. G. Speranza, D. Vigo (2016) Intermediate Facilities in Freight Transportation Planning: A Survey. Transportation Science 50(3):763-789. https://doi.org/10.1287/trsc.2015.0631

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

TRANSPORTATION SCIENCE
Vol. 50, No. 3, August 2016, pp. 763–789
ISSN 0041-1655 (print) | ISSN 1526-5447 (online)

http://dx.doi.org/10.1287/trsc.2015.0631

Intermediate Facilities in Freight Transportation Planning: A Survey

G. Guastaroba, M. G. Speranza
Department of Economics and Management, University of Brescia, 25121 Brescia, Italy
{gianfranco.guastaroba@unibs.it, grazia.speranza@unibs.it}

D. Vigo
Department of Electrical, Electronic, and Information Engineering, University of Bologna, 40136 Bologna, Italy,
daniele.vigo@unibo.it

Consolidation of freight and merging operations are essential for transportation companies to reduce costs and improve the level of service provided to customers. Such operations take place in intermediate facilities or terminals located between the origins and the destinations of freight. This survey reviews the main contributions from the operations research literature on freight transportation planning problems where the presence of intermediate facilities has a strong impact on the cost of the system and on how goods are delivered. In particular, we focus on the tactical planning issues arising in this context. We have identified three classes of problems with intermediate facilities: vehicle routing problems, transshipment problems, and service network design problems. For each class of problems we provide an overview of the main problem variants, survey the methods used for their solution, and indicate open research directions.

Keywords: literature review; freight transportation planning; vehicle routing problems; transshipment problems; service network design problems
History: Received: January 2015; revisions received: March 2015, April 2015; accepted: April 2015. Published online in *Articles in Advance* January 28, 2016.

1. Introduction

An efficient and cost-effective distribution network is a main component of a supply chain and a key issue for gaining relevant competitive advantages in most industrial contexts. In their book Chopra and Meindl (2009) point out that distribution-related costs make up about 10.5% of the U.S. economy and about 20% of the cost of manufacturing, whereas Musa, Arnaout, and Jung (2010) highlight that 30% of an item price is incurred in the distribution process. It is therefore not surprising that such topics have attracted a massive amount of research aimed at developing innovative optimization models and solution algorithms capable of providing decision makers with effective support tools.

In general, such activities concern the transport of raw materials and products between a set of *origins* (e.g., suppliers or production plants) and a set of *destinations* (e.g., wholesalers, retailers, or final customers), possibly through a set of *intermediate facilities*. Each of the three former groups of entities represents a *stage* (also called layer or tier) of a supply chain. Supply chains may involve different types of intermediate facilities to be visited in sequence (e.g., distribution centers and cross-docks). In such cases, each type of intermediate facility represents a different stage of the supply chain. The transportation of goods occurs between a pair of stages, which represents one level of the distribution network and is often referred to as an *echelon*. Intermediate facilities play an important role in distribution networks, being the main attractors of inbound flows from different origins and the source of outbound flows destined for several destinations. Indeed, even though transporting freight through an intermediate facility increases the distance and time traveled, when compared with a direct delivery from its origin to its destination, economies of scale can be achieved by *consolidating* the freight with other shipments, so that the total cost of the system is reduced by improving the loading of the vehicles. In this situation, direct deliveries should be utilized, for instance, when time restrictions are tight, the distance between origin and destination is short, or the size of the shipment fills the capacity of the vehicle. In other situations direct deliveries are not possible, and freight *must compulsorily move* through an intermediate facility. This typically happens when physical or legal restrictions impose that the mode of transport (or merely the vehicle) used to move the freight out of its origin has to be different from the one used to carry out its delivery. Examples of these situations can be found in the context of city logistics and multimodal transportation.

A large number of different optimization problems arise within the former broad application area, ranging from the choice of the most appropriate service model for specific goods (e.g., direct shipping or cross-docking), the location and sizing decisions for the intermediate facilities, to the definition of routing and scheduling decisions for the fleet of vehicles and the associated crews. These problems may be classified into three classes according to the planning horizon. Specifically, we can identify *strategic decision problems* in which long-term decisions, such as facility location, are considered; *tactical decision problems*, which include midterm decisions, such as how to route the freight through the networks, what type of service to provide, or how to plan maintenance services; finally, *operational decision problems*, which refer to short-term decisions, such as the implementation and adjustment of schedules for services and crews.

1.1. Scope of the Survey

The primary purpose of this paper is to present a state-of-the-art review of the main contributions from the operations research literature on freight transportation planning problems within tactical level decisions. The papers reviewed focus on the intermediate facilities and their impact on freight transportation. Some aspects of operational problems strictly connected with tactical decisions, such as those related to the management of cross-docking terminals, are also mentioned. Whereas the first papers on this class of problems date back to the 1980s (a noncomprehensive list includes Hall 1987a, b; Klincewicz 1990, and Leung, Magnanti, and Singhal 1990), this area has received increasing attention from the research community in the past decade and such contributions deserve to be classified and analyzed.

More formally, the class of problems surveyed in the present paper can be described in graph theoretic terms as follows. Let $G = (N, E \cup A)$ be a weighted graph. Set $N = O \cup H \cup D$ is a set of vertices, where $O = \{1, \ldots, |O|\}$ represents a set of origins, $H = \{|O|+1, \ldots, |O|+|H|\}$ denotes a set of intermediate facilities, and, finally, $D = \{|O|+|H|+1, \ldots, |O|+|H|+|D|\}$ is a set of destinations. Origins, intermediate facilities, and destinations are located at known and fixed locations. Depending on the operations that can be carried out and on the problem variant under investigation, the intermediate facilities are called hubs, cross-docks, platforms, satellites, consolidation centers, or distribution centers. In the following, we use the term intermediate facility when referring to a general setting, whereas we use the most frequently adopted term, if any, when referring to a specific problem variant. Set E is the set of (undirected) edges $\{i, j\}$ such that $i < j$ and $i, j \in N$. Similarly, set A is the set of (directed) arcs (i, j) such that $i, j \in N$. Set E or set A may be empty. Most frequently, some restrictions on the links connecting pairs of vertices are imposed, making graph G not complete. Further details are provided in the following. A nonnegative length or cost c_{ij} is associated with each edge $\{i, j\} \in E$ and each arc $(i, j) \in A$, and represents the travel cost, or sometimes the travel time, to move from vertex i to vertex j. The values c_{ij} satisfy the triangle inequality. A demand is given for each destination $d \in D$ that must be served with the freight available at one or more origins in O. The freight may or must be delivered by a fleet of vehicles through an intermediate facility in H. This survey provides an overview on problems calling for the optimal distribution of goods from origins to destinations, possibly through a set of intermediate facilities where storage, merge, consolidation, or transshipment operations may take place.

1.2. Surveys on Related Problems

The class of problems considered in this survey is closely related to that of hub location problems. Hub location is a fertile research area that blends location decisions and network design aspects. It has a variety of natural applications in long-distance transportation (e.g., air passenger, parcel and postal delivery, and freight transport), and also in other areas such as telecommunication network design. Campbell and O'Kelly (2012, p. 153) defined hub location problems as those that "involve the location of hub facilities through which flows (e.g., of passengers or freight) are to be routed from origins to destinations (e.g., cities)." The similarities between hub location and the class of problems surveyed here are indeed strong, since also in the former problems the hubs are facilities that serve as switching, transshipment, and sorting points, and where flows having different origins are consolidated to take advantage of economies of scale (e.g., see Alumur and Kara 2008). Nevertheless, the key distinguishing features (see again Campbell and O'Kelly 2012) in hub location are the location decisions, and the objective that usually depends on the location of hubs in addition to the routing of the flows. Such features often make the problem of designing hub networks a long-term strategic decision, as pointed out also in Contreras (2015). For further details, we refer the interested reader to one of the several specialized surveys on hub location. In particular, Alumur and Kara (2008) highlighted the strict connections between the literature on hub location and location problems, and classify the contributions on the former class of problems. More recently, Kara and Taner (2011) introduced a taxonomy for hub location problems and reviewed theoretical as well as application oriented studies. The survey by Campbell and O'Kelly (2012) focused on the discussion of the early motivations for analyzing hub location problems, described the early research on this class of problems, provided an overview of some of the current related research, and indicated possible research directions. The survey by Farahani et al. (2013) considered the papers on hub location that appeared after the publication of the review by Alumur and Kara (2008). It contains an overview of the mathematical models and solution methods available, and some case studies illustrating related real-world applications. Finally, Contreras (2015) described key features, assumptions, and properties that are commonly considered in hub location problems, providing some insights on their modeling implications. The author also reviewed some mathematical formulations and solution algorithms, and described the main developments and recent trends in hub location theory.

Wieberneit (2008) and Crainic (2000) wrote related surveys with a focus on service network design problems in freight transportation, one of the classes of problems surveyed in the present paper. Additionally, the recent surveys by Agustina, Lee, and Piplani (2010); Stephan and Boysen (2011); and Van Belle, Valckenaers, and Cattrysse (2012) considered the optimization problems arising with the presence of a cross-dock in the distribution network, such as location of the terminals, terminal layout design, vehicle scheduling for inbound and outbound flows, and dock-door assignment problems. The vehicle scheduling literature was also reviewed in Boysen and Fliedner (2010). Finally, Crainic and Kim (2007) and SteadieSeifi et al. (2014) provided a broader analysis with a focus on multimodal freight transportation.

1.3. Classification and Terminology Used

We classified the family of problems covered in this survey into the following three classes. We consider the following:

(I) *Intermediate facilities in vehicle routing problems*. This class includes variants of the classic vehicle routing problems where one or more intermediate facilities are formally included in the problem definition.

(II) *Intermediate facilities in transshipment problems*. This class includes problems that are closely related to the transshipment problem, an extension of the widely known transportation problem that includes one or more intermediate facilities.

(III) *Intermediate facilities in service network design problems*. This class includes problems that are motivated by applications where the use of intermediate facilities is necessary to achieve economies of scale. They also include some additional issues that are not considered in the previous two classes, such as on which route to provide service, what type of service to use, and how often to offer service on each route.

Within each of the three classes above, we cluster the papers with similar characteristics into groups. Obviously, some papers could be included in several groups. In these cases, we included them in the group that we believe best describes the problem under consideration. We distinguish the following groups of features:

(i) Type of network:

Pure network. Direct deliveries from origins to destinations are not considered, i.e., all shipments are transferred through an intermediate facility.

Hybrid network. Direct deliveries are allowed, i.e., freight can either move through an intermediate facility or directly from origin to destination.

(ii) Number of intermediate facilities:

Single facility. Only one intermediate facility is present.

Multifacility. Multiple intermediate facilities are available.

(iii) Origin-destination structure:

One-to-one. A set of origin-destination pairs is given such that freight available at a given origin has to be delivered to a specific destination.

One-to-many. Only one origin is present that serves several destinations.

Many-to-many. The demand of a commodity associated with each destination can be served with the freight available at several origins.

(iv) Number of commodities:

Single commodity. Only one type of commodity moves over the network.

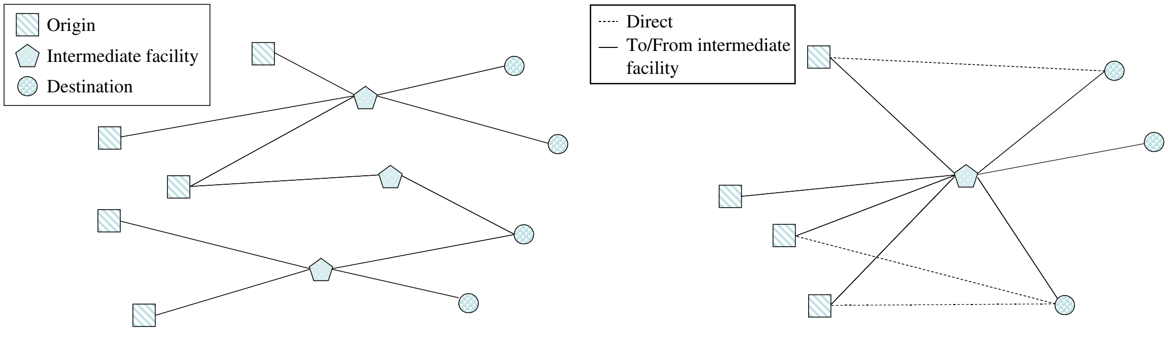
Multicommodity. Several commodities are explicitly considered and must move over the same network.

Some additional characteristics are possibly indicated for each paper. In particular, we highlight the presence of capacity restrictions (typically associated with the vehicles or the intermediate facilities), the possibility to store goods at the intermediate facility, and the inclusion of time restrictions (such as time windows constraints or limits on the level of service to provide).

Figures 1(a) and 1(b) show two illustrative examples displaying a pure network comprising multiple intermediate facilities and a hybrid network including only one intermediate facility, respectively. To keep the pictures simple, links among origins, intermediate facilities, and destinations are not shown. These types of links are permitted only in some classes of problems surveyed in the following (in particular, in the class of vehicle routing problems with intermediate facilities).

1.4. Structure of the Paper

The structure of the paper is as follows. Section 2 provides an overview of different types of intermediate facilities and describes the related application areas. Section 3 gives a survey of the contributions on vehicle routing problems with intermediate facilities. Section 4 examines the literature related to the transshipment problem, whereas the papers on intermediate facilities in service network design problems are reviewed in §5. Finally, §6 draws some conclusions and discusses future research directions.



(a) A pure network with multifacility

(b) A hybrid network with single facility

Figure 1 (Color online) Illustrative Examples of Networks with Intermediate Facilities

Table 1 summarizes the main abbreviations used in the paper to denote optimization models and solution algorithms.

2. The Role of the Intermediate Facilities

This section provides a description of the different types of intermediate facilities considered in the surveyed papers. We provide an overview of the operations that can be carried out at these intermediate facilities, we describe the distribution networks and, since operations research is primarily an application-oriented discipline, we highlight the most appropriate application areas.

Networks in which the intermediate facility between origins and destinations acts purely as a transfer location where no, or little, inventory is held are called *cross-docking networks*. WalMart (Stalk, Evans, and Shulman 1992), CompUSA (Gentry 2005), and Nabisco (Gümüş and Bookbinder 2004) have used cross-docking. Within such network settings, products arrive on inbound vehicles to the receiving docks of the cross-docking terminal. The goods are unloaded from the vehicles, sorted by destination, and consolidated with other orders. They are then loaded as rapidly as possible into outbound vehicles at the shipping docks and leave the cross-dock to be delivered to their respective destination. Figure 2(a) shows a schematic example of the operations performed in a typical cross-dock. Implementing a cross-docking network is effective when replenishment lots for the destinations served by a cross-dock are large enough to achieve economies of scale on inbound transportation (Chopra and Meindl 2009). The key to success is to keep the dwell time spent in the cross-dock as short as possible (Apte and Viswanathan 2000). A major benefit is that cross-docks do not need long-term inventory holding capacity (products often stay in cross-docks for less than 12 hours (Simchi-Levi, Kaminsky, and Simchi-Levi 2007), and sometimes less than one hour (Wen et al. 2009)) and products flow faster in the supply chain. Cross-docking also saves on handling costs since products do not have to be moved in and out of storage. However, successful cross-docking requires a significant increase in coordination and synchronization efforts between the inbound and outbound shipments. Additionally,

Table 1 A Summary of the Main Abbreviations Used in the Paper to Denote Optimization Models and Solution Algorithms

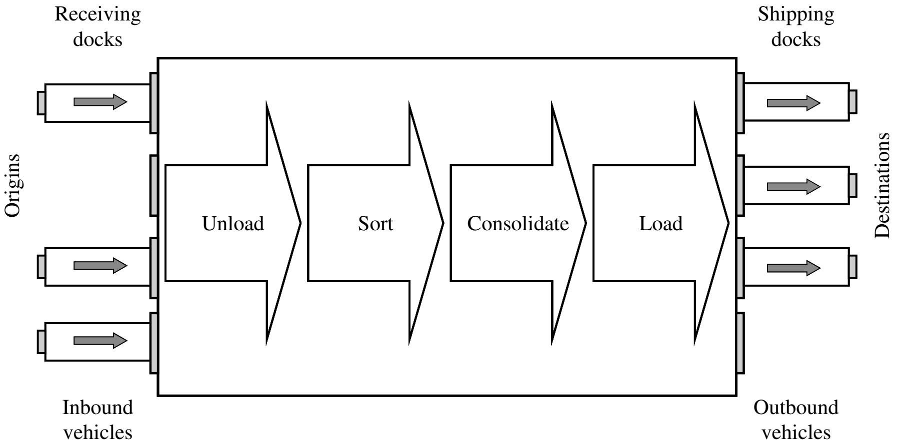
Optimization model	Heuristic algorithm	Exact algorithm
BLP: Binary linear programming BNLP: Binary nonlinear programming ILP: Integer linear programming LP: Linear programming MILP: Mixed-integer linear programming NLIP: Nonlinear integer programming NLP: Nonlinear programming	ACO: Ant colony optimization ALNS: Adaptive large neighborhood search AMP: Adaptive memory procedure AMR: Adaptive multi-restart GA: Genetic algorithm GRASP: Greedy randomized adaptive search procedure ILS: Iterated local search LDB: Lagrangean dual-based LRB: Lagrangean relaxation based LS: Local search MH: Matheuristic MSH: Multi-start heuristic PSH: Problem specific heuristic SA: Simulated annealing SSH: Slope scaling heuristic TS: Tabu search VNS: Variable neighborhood search	B&B: Branch and bound B&C: Branch and cut B&C&P: Branch and cut and price B&P: Branch and price IE: Implicit enumeration LR: Lagrangean relaxation PSE: Problem specific exact

it requires significant start-up investments related to the implementation of advanced information systems that have to link origins, cross-docks, and destinations and to the construction of a fast and responsive transportation system. Typical applications arise in the transportation of perishable products (they need to preserve quality and freshness), frozen food and other refrigerated products (no intermediate storage inside the uncooled cross-dock is allowed, e.g., see Boysen 2010), such as groceries, agricultural (e.g., see Petersen and Røpke 2011), and pharmaceutical products; products with a short life cycle (e.g., the fashion application considered in Hu, Zhao, and Choi 2013); and is in general appropriate when the product demand is stable and when the stock-out unit cost is small (e.g., see Lee, Jung, and Lee 2006). WalMart is extensively cited as an example of a successful usage of a cross-docking network (according to Simchi-Levi, Kaminsky, and Simchi-Levi 2007, WalMart delivers about 85% of its goods using cross-docks).

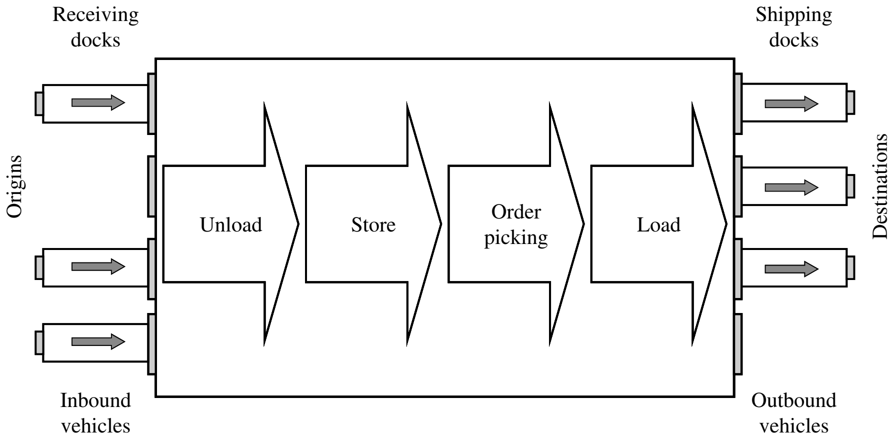
Strictly related to cross-docks are the so-called *in-transit merge centers*. The use of this type of intermediate facility is beneficial when orders from each destination relate to products made up of different components, and these components are available at different origins. In an in-transit merge center, the different components of the same order are then combined together so that the final customer receives only a single delivery. Hewlett Packard, Dell, and Cisco have used this model (see Ala-Risku, Kärkkäinen, and Holmström 2003). Compared to the operations carried out in a cross-dock, the main difference is the additional restriction that consolidation operations cannot begin before all of the components belonging to the same order have arrived at the intermediate facility. Consequently, the main issue arising in this context is how to coordinate shipments of the components from different origins to the in-transit merge center so that their arrivals are synchronized and they can be bundled and shipped as soon as possible to their destination. The requirement of consolidation implies that direct deliveries from origin to destination are not considered, i.e., networks with in-transit merge centers can only take the form of pure networks. The use of in-transit merge centers is best suited to high-value items or when the number of origins is limited. Indeed, when the number of origins is too large, in-transit merge networks can be very difficult to coordinate and implement. Kärkkäinen, Ala-Risku, and Holmström (2003) mentioned that high-tech industries are the prime candidates for the use of in-transit merge networks. In effect, products sold by high-tech companies are often customized and several variants are usually offered for the same product. With an in-transit merge center, it is possible to postpone the assembly of customer-specific variants and thus reduce the need to store multiple configurations. Another related application is the delivery of medical equipment composed of several modules (e.g., see Song, Hsu, and Cheung 2008). In this application, each origin produces certain types of medical devices that represent the modules of a medical equipment. At the destination, these modules are assembled into a single piece of equipment. All modules are required to arrive at their destination in one single shipment so as to minimize the interruption of medical services.

Networks in which inventory is mainly stored in intermediate warehouses, called distribution centers (DCs), which are possibly owned by third parties or distributors, are called *distributor storage networks*. Amazon, W.W. Grainger, and McMaster-Carr (e.g., see Chopra and Meindl 2009) have used such a model. In this type of intermediate facility, inbound vehicles arrive at the receiving dock where freight is unloaded, possibly sorted, and then stored. When the freight has to be delivered to its destination, an operator performs order picking, goods are loaded onto outbound vehicles at the shipping dock and leave the DC. Figure 2(b) shows a schematic example of the operations performed in a typical DC. Distributor storage, as opposed to *manufacturer storage*, is well suited when cheap modes of transportation (e.g., truckload, TL) are used for shipments from origins to DCs, whereas products from DCs to destinations are sent in smaller lots. In these situations, economies of scale can be achieved by having loads on the inbound side much larger than those on the outbound side. For example, WalMart has also used DCs when goods were coming from overseas suppliers to the DC in large shipments, and were, eventually, sent to the stores served by the DC in small lots (see Chopra and Meindl 2009). Furthermore, distribution storage is beneficial for products with high demand. For instance, Amazon and W.W. Grainger hold both medium- and fast-moving items at their DCs, whereas they stock slow-moving items in their factories (see Chopra and Meindl 2009). In this network setting, the response time is usually better than in manufacturer storage, since the intermediate warehouses are, on average, located closer to the destinations and can consolidate the entire order of a customer before shipping. Response time is usually worse than in the case of retail storage. The negative aspects are related to a higher level of inventory required by distributor storage than manufacturer storage, since distributors usually aggregate demand uncertainty at a lower level of the supply chain than manufacturers.

Most of the papers surveyed in the following sections refer to cross-docking networks. Indeed, storage capacities at the intermediate facilities are usually neglected or very limited. Despite the practical relevance of in-transit merge centers and distributor storage networks, only a few papers study their use.



(a) Typical freight flow in a cross-dock terminal



(b) Typical freight flow in a distribution center

Figure 2 **Illustrations of Typical Freight Flows in an Intermediate Facility**

3. Intermediate Facilities in Vehicle Routing Problems

The vehicle routing problem is one of the most important and investigated class of combinatorial optimization problems. It calls for the determination of an optimal set of routes to be performed by a fleet of vehicles to serve a given set of customers (see Toth and Vigo 2002, 2014). In this section we review the main contributions that address problems where one or more intermediate facilities are included in the problem definition. We have identified two main classes of problems: two-echelon vehicle routing problems and pickup and delivery problems with cross-docks. In each of these two classes, we provide a description of a basic problem and illustrate the most studied variants in the class. At the end of the section we point out open research directions.

Table 2 presents an overview of the main problem characteristics of the papers surveyed in this section, whereas Table 3 provides some details on the optimization models and solution methods proposed in each paper. A reference to the solver used to solve the optimization model, if any, is also reported.

3.1. Two-Echelon Vehicle Routing Problems

The class of two-echelon vehicle routing problems (2E-VRPs) has recently received some attention in the literature. The research efforts have concentrated, until now, on the study of a basic member of the family called the two-echelon capacitated vehicle routing problem (hereafter referred to as 2E-CVRP). The literature on the 2E-CVRP is also surveyed in Cuda, Guastaroba, and Speranza (2015). The 2E-CVRP can be defined as follows. Consider a weighted undirected graph $G=(N,E)$. The set $N=\{0\}\cup H\cup D$ is the set of vertices, where vertex 0 represents the single (uncapacitated) origin, $H=\{1,\dots,|H|\}$ is the set of intermediate facilities, and $D=\{|H|+1,\dots,|H|+|D|\}$ is the set of destinations. In the context of 2E-VRPs, the intermediate facilities are called *satellites*. The set $E=E^1\cup E^2$ is the set of edges $\{i,j\}$ such that edge set $E^1=\{\{i,j\}\colon i<j,\ i,j\in\{0\}\cup H\}$

Table 2 **Intermediate Facilities in Vehicle Routing Problems: A Summary of the Main Problem Characteristics**

Reference	Type netw.	No. of I. F.[a]	OD struct.[b]	No. of comm.[c]	Capacity[d]	Storage I. F.	Time restriction[e]	Additional charact.[f]
2E-VRP								
Crainic, Ricciardi, and Storchi (2009)	Pure	M	1-1	MC	H, V		HTW	HV
Perboli, Tadei, and Vigo (2011)	Pure	M	1-M	SC	H, V			
Perboli, Tadei, and Masoero (2010)	Pure	M	1-M	SC	H, V			
Crainic et al. (2011)	Pure	M	1-M	SC	H, V			
Crainic et al. (2013)	Pure	M	1-M	SC	H, V			
Hemmelmayr, Cordeau, and Crainic (2012)	Pure	M	1-M	SC	H, V			
Jepsen, Spoorendonk, and Røpke (2013)	Pure	M	1-M	SC	H, V			
Santos, da Cunha, and Mateus (2013)	Pure	M	1-M	SC	H, V			
Santos, Mateus, and da Cunha (2015)	Pure	M	1-M	SC	H, V			
Baldacci et al. (2013)	Pure	M	1-M	SC	H, V			
PDPCD								
Lee, Jung, and Lee (2006)	Pure	S	M-M	SC	V		PH, SAH	ST
Liao, Lin, and Shih (2010)	Pure	S	M-M	SC	V		PH, SAH	ST
Dondo, Méndez, and Cerdá (2009)	Hybrid	M	M-M	MC	V	✓	HTW, PH	HV, ST Splitting allowed
Dondo, Méndez, and Cerdá (2011)	Hybrid	M	M-M	MC	V	✓	HTW, PH	HV, ST Splitting allowed
Wen et al. (2009)	Pure	S	1-1	SC	V		HTW, PH	ST
Tarantilis (2013)	Pure	S	1-1	SC	V		HTW, PH	ST
Morais, Mateus, and Noronha (2014)	Pure	S	1-1	SC	V		HTW, PH	ST
Dondo and Cerdá (2013)	Pure	S	1-1	SC	V		HTW, PH	ST
Santos, Mateus, and da Cunha (2011a)	Pure	S	1-1	SC	V			
Santos, Mateus, and da Cunha (2011b)	Pure	S	1-1	SC	V			
Liu, Li, and Chan (2003)	Hybrid	S	1-1	SC	V			Unlimited fleet
Petersen and Røpke (2011)	Hybrid	S	1-1	SC	V		HTW	ST
Santos, Mateus, and da Cunha (2013)	Hybrid	S	1-1	SC	V			
Mitrović-Minić and Laporte (2006)	Hybrid	M	1-1	SC			HTW	ST
Bouros et al. (2011)	Hybrid	M	1-1	SC				

[a] S, single facility; M, multifacility.
[b] 1-1, one-to-one; 1-M, one-to-many; M-M, many-to-many.
[c] SC, single commodity; MC, multicommodity.
[d] H, intermediate facilities; V, vehicles.
[e] HTW, hard time windows; PH, planning horizon; SAH, simultaneous arrivals to the intermediate facility.
[f] HV, heterogeneous vehicles; ST, service times.

includes the edges connecting the origin to the satellites, as well as those connecting pairs of satellites. The edge set $E^2 = \{\{i, j\}: i < j,\ i, j \in H \cup D,\ \{i, j\} \notin H \times H\}$ comprises the edges connecting the satellites to the destinations, as well as those connecting pairs of destinations. The freight from origin 0 to the respective destinations is delivered through the satellites in H. Two fleets of vehicles are available to transport freight, one per echelon. Specifically, a limited fleet of K^1 capacitated vehicles is located at the origin. These vehicles all have the same capacity Q^1, and each vehicle starts its route at the origin, visits one or more satellites in H, and then returns to 0 (i.e., each first echelon route comprises only edges in E^1). The delivery of freight from the satellites to the destinations is performed by a limited fleet of K^2 capacitated vehicles starting and finishing their routes at a given satellite in set H (i.e., each second echelon route comprises only edges in E^2). The latter vehicles all have the same capacity Q^2. Each satellite is associated with a given capacity $K^2_{\max}$ defined as the maximum number of second echelon routes that can start from it. Each destination $d \in D$ demands r_d units of freight. Finally, a handling cost l_h for loading/unloading operations is given for each satellite $h \in H$. The 2E-CVRP aims at determining an optimal set of routes at both echelons such that the demand of all destinations is satisfied, satellites and vehicle capacity constraints are not violated, and the total distribution cost is minimized. The total distribution cost is the sum of two components: the routing cost at both echelons, and the handling cost at the satellites.

Figure 3 depicts an example of a feasible solution for a 2E-CVRP. Dashed lines are used to represent the first echelon routes, while the second echelon routes are depicted as solid lines.

The main issue to address when modeling a 2E-VRP is how to connect the flows of freight moving between the two echelons and, in particular, how to manage the dependency of the second level from the first one. Furthermore, note that, if convenient, some satellites may be left unused, and if an assignment

Table 3 **Intermediate Facilities in Vehicle Routing Problems: A Summary of the Optimization Models and Solution Algorithms**

Reference	Optimization model	Heuristic algorithm	Exact algorithm
2E-VRP			
Crainic, Ricciardi, and Storchi (2009)	BLP		
Perboli, Tadei, and Vigo (2011)	MILP	MH	B&C
Perboli, Tadei, and Masoero (2010)			B&C
Crainic et al. (2011)		MSH	
Crainic et al. (2013)		GRASP	
Hemmelmayr, Cordeau, and Crainic (2012)		ALNS	
Jepsen, Spoorendonk, and Røpke (2013)	MILP		B&C
Santos, da Cunha, and Mateus (2013)	MILP		B&P
Santos, Mateus, and da Cunha (2015)	MILP		B&C&P
Baldacci et al. (2013)	ILP		PSE
PDPCD			
Lee, Jung, and Lee (2006)	MILP	TS	
Liao, Lin, and Shih (2010)		TS	
Dondo, Méndez, and Cerdá (2009)	MILP		CPLEX
Dondo, Méndez, and Cerdá (2011)	MILP		Gurobi
Wen et al. (2009)	MILP	AMP-TS	
Tarantilis (2013)		AMR-TS	
Morais, Mateus, and Noronha (2014)		ILS	
Dondo and Cerdá (2013)	MILP	PSH	Solver (?)
Santos, Mateus, and da Cunha (2011a)	BLP		B&P
Santos, Mateus, and da Cunha (2011b)	BLP		B&P
Liu, Li, and Chan (2003)		LS	
Petersen and Røpke (2011)		Parallel ALNS	
Santos, Mateus, and da Cunha (2013)	BLP		B&P
Mitrović-Minić and Laporte (2006)		2-step PSH	
Bouros et al. (2011)		—*Dynamic*—	

of the customers to each satellite is given, then the 2E-CVRP reduces to at most $(1+|H|)$ CVRPs, i.e., one for the first echelon, and at most $|H|$ for the second echelon (e.g., see Perboli, Tadei, and Vigo 2011).

Before reviewing the papers that introduced optimization models or solution algorithms for the 2E-CVRP, it is worth highlighting that some studies investigated the potential benefits achievable using a delivery strategy that implements a 2E-CVRP compared to a single-echelon strategy based on solving a CVRP. Crainic et al. (2010) reported on some experimental studies conducted to evaluate the impact of some instance modifications (such as the depot location, the number of satellites, and the customer distribution) on the total distribution cost of a 2E-CVRP. In addition, Crainic et al. (2012) reported a study on the impact that the inclusion in the travel costs of components other than the distance (e.g., fixed costs for using a link, or operational and environmental costs) may have on the optimal solution of a 2E-CVRP. Both studies indicate that substantial benefits can be achieved by applying a delivery strategy that uses a 2E-CVRP instead of one implementing a CVRP.

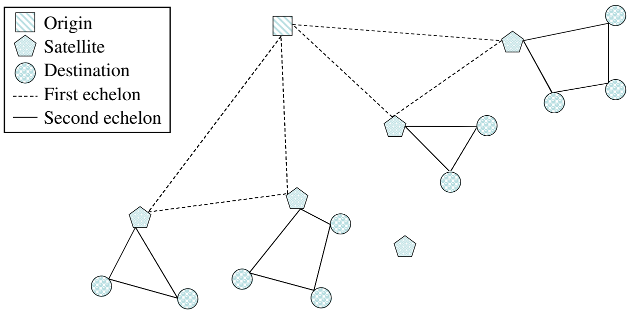


Figure 3 **(Color online) An Example of a Feasible Solution for a 2E-CVRP**

Two-echelon distribution systems are often encountered in city logistics. The paper by Crainic, Ricciardi, and Storchi (2009) is often cited as the one that inspired the research on 2E-VRPs, even if the latter term appeared only later. In particular, Crainic, Ricciardi, and Storchi (2009) considered the following two-echelon distribution network in a city logistics context. The freight is available at some origins called, in the city logistics terminology, city distribution centers (CDCs). At a CDC, freight is loaded onto so-called urban vehicles. Each urban vehicle receives a departure time and a route, travels to one or several satellites, unloads, and then leaves the system or moves to a CDC where it waits for its next departure. At the satellites, freight is unloaded from urban vehicles, consolidated, and then loaded onto so-called city freighters, which are vehicles suitable for moving within dense city zones.

Each city freighter performs a route starting from a depot, loads freight at a satellite, and then serves the designated destinations before traveling to another satellite or finishing its route at a depot, possibly different from where it started. Both fleets are composed of heterogeneous vehicles. The authors formulated the problem as a binary linear programming (BLP) model, introduced and analyzed some variants, and described possible heuristics, but did not report computational results.

Perboli, Tadei, and Vigo (2011) provided a formal introduction of the family of 2E-VRPs and, in particular, of the 2E-CVRP described above. The authors proposed a mixed-integer linear programming (MILP) formulation and, to strengthen its linear programming (LP) relaxation, introduced two families of valid inequalities derived from the CVRP literature. They presented two matheuristics for the 2E-CVRP based on their MILP model, and a simple branch-and-cut (B&C) algorithm. Computational experiments were conducted on instances derived from the benchmark CVRP test problems introduced in Christofides and Eilon (1969), and on a set taken from Crainic et al. (2010). Crainic et al. (2011) introduced a family of multi-start heuristics (MSH) for the 2E-CVRP based on the idea of separating the first and second echelon routing problems, and then iteratively solving the two resulting routing subproblems. The performance of the heuristics was compared with the matheuristics described in Perboli, Tadei, and Vigo (2011) and the lower bounds reported in Perboli, Tadei, and Masoero (2010). The same authors used the idea of decomposing the problem also in Crainic et al. (2013), where a greedy randomized adaptive search procedure (GRASP) combined with a path relinking procedure was introduced to solve the 2E-CVRP.

The current best performing heuristic for the 2E-CVRP is the adaptive large neighborhood search (ALNS) of Hemmelmayr, Cordeau, and Crainic (2012), a metaheuristic framework that yielded an excellent performance for several routing problems (see Pisinger and Røpke 2007). In Hemmelmayr, Cordeau, and Crainic (2012) the authors adapted to the 2E-CVRP some existing neighborhood search operators originally introduced in Pisinger and Røpke (2007) for the CVRP, and introduced new ones that exploit the structure of the 2E-CVRP. They also showed that the one-echelon location-routing problem (see Prodhon and Prins 2014 for a recent survey) can be reduced to a special case of the 2E-CVRP in which routing is performed only in the second echelon. Computational experiments for the 2E-CVRP were performed on several sets of benchmark instances, and a new set of large-scale instances generated from test problems originally introduced in Prins, Prodhon, and Wolfler Calvo (2004) for a location-routing problem. When possible, the performance of the ALNS was compared with the best known solutions reported in Crainic et al. (2011); Perboli, Tadei, and Masoero (2010); and Perboli, Tadei, and Vigo (2011). Computational results show that the ALNS heuristic improves several best known solutions.

In addition to the above-mentioned B&C algorithms designed in Perboli, Tadei, and Vigo (2011) and Perboli, Tadei, and Masoero (2010), other exact algorithms for the 2E-CVRP are described in the following papers. Jepsen, Spoorendonk, and Røpke (2013) showed, by means of an example, that the optimization model contained in Perboli, Tadei, and Vigo (2011) for the 2E-CVRP may not provide correct upper bounds when more than two satellites are selected in the solution. Therefore, the authors designed a different MILP formulation that overcomes this limitation but is highly symmetric and tends to provide poor lower bounds. As a consequence, the authors derived an alternative MILP model, which is a relaxation for the 2E-CVRP but eliminates symmetry and provides better lower bounds. Since this relaxation provides lower bounds but not necessarily feasible solutions for the 2E-CVRP, the authors devised a feasibility test and a specialized branching scheme to obtain feasible integer solutions. They developed a B&C algorithm to solve the 2E-CVRP using the specialized branching rule. Santos, da Cunha, and Mateus (2013) formulated the 2E-CVRP as an MILP model that uses, as main decision variables, binary variables indexed by vehicles and associated with each feasible first and second echelon route. They designed two branch-and-price (B&P) algorithms to solve the optimization model. The computational results of the B&P algorithms were compared with those of the B&C algorithms tested in Perboli, Tadei, and Vigo (2011) and Perboli, Tadei, and Masoero (2010). According to Santos, Mateus, and da Cunha (2015), the presence of binary variables indexed by vehicles makes the MILP formulation of Santos, da Cunha, and Mateus (2013) symmetric. To overcome this drawback, Santos, Mateus, and da Cunha (2015) introduced a new MILP formulation that uses integer variables to count the number of vehicles traveling a given first echelon route, and binary variables indexed only by second echelon routes. Several classes of valid inequalities derived from the literature on the traveling salesman problem and the CVRP were used to strengthen the formulation, and a branch-and-cut-and-price (B&C&P) algorithm was designed to solve the problem. They provided a comparison with the B&C of Perboli, Tadei, and Vigo (2011) and the B&P tested in Santos, da Cunha, and Mateus (2013).

Baldacci et al. (2013) developed the current state-of-the-art exact algorithm for the 2E-CVRP based on an integer linear programming (ILP) formulation that was used to derive both continuous and integer relaxations. The authors devised a lower bounding procedure based on dynamic programming, a dual ascent method, and an exact algorithm based on decomposing the 2E-CVRP into a limited set of multidepot CVRPs with side constraints. Computational results were given for benchmark instances (namely, those considered in Jepsen, Spoorendonk, and Røpke 2013 and a subset of those tested by Hemmelmayr, Cordeau, and Crainic 2012) and a new set of instances. Whenever possible, the performance of the exact method was compared with the B&C algorithms introduced in Perboli, Tadei, and Vigo (2011) and Jepsen, Spoorendonk, and Røpke (2013). The comparison showed that the algorithm designed by Baldacci et al. (2013) outperforms the competing exact algorithms in terms of size of the instances solved, the number of instances solved to optimality, and computing times.

3.2. Pickup and Delivery Problems with Cross-Docks

Pickup and delivery problems (PDPs) constitute a class of vehicle routing problems in which goods (or people) have to be collected at an origin (pickup location) and then transported to a destination (delivery location). This class of problems inspires an ever growing body of literature (see the recent surveys by Berbeglia et al. 2007; Berbeglia, Cordeau, and Laporte 2010, and Parragh, Doerner, and Hartl 2008).

The classification based on the origin-destination structure we presented in §1 is often applied to PDPs (e.g., see the aforementioned survey by Berbeglia et al. 2007). In particular, PDPs with a many-to-many structure are characterized by the presence of several origins and destinations for each commodity, i.e., pickup and delivery locations are unpaired. Conversely, in one-to-one problems a set of requests is given. Each request is identified by an origin-destination pair and the size of the load to be picked up at the origin and delivered to its destination. The usual assumption in one-to-one PDPs is that the entire request must be served by the same vehicle, i.e., the same vehicle visits both the origin and the destination. This is frequently called the *pairing constraint*.

Pickup and delivery problems with cross-docks (PDPCDs) extend classical PDPs by introducing the opportunity for a vehicle to drop or pickup loads at an intermediate facility, called *cross-dock* (or, sometimes, transshipment point) in this context. The introduction of this type of facility may lead to a more efficient use of the vehicle capacities but increases the complexity of the problem considerably. The main issue in a PDPCD is how to model the operations carried out at the cross-dock and, in particular, how to model the flows of freight moving from one vehicle (that picked up the freight from its origin) to a different vehicle (that will deliver the freight to its destination). In PDPCDs the same vehicle fleet performs both pickup and delivery operations. Additionally, the presence of a handling cost incurred (or time spent) for each unit of load moved, as well as other side constraints, prevent the problem from being decomposed into two independent vehicle routing problems.

In this section, we first review the papers dealing with many-to-many PDPCDs, and those studying one-to-one variants. Whereas the early papers on this problem relate to the many-to-many case, most of the attention has later focused on the one-to-one case. This can probably be explained by considering that many-to-many problems, both for classical PDPs and PDPCDs, are rarely encountered in practice.

3.2.1. Many-to-Many Pickup and Delivery Problems with Cross-Docks. The basic prototype of the class of many-to-many PDPCDs with single facility, single commodity, and defined on a pure network, can be defined as follows. Let $G = (N, E)$ be a weighted undirected graph. Set $N = O \cup \{0\} \cup D$ is the set of vertices, where $O = \{1, \ldots, |O|\}$ is the set origins, vertex 0 represents the unique (uncapacitated) cross-dock present, and $D = \{|O|+1, \ldots, |O|+|D|\}$ is the set of destinations. The set $E = E^1 \cup E^2$ is the set of edges $\{i, j\}$ such that edge set $E^1 = \{\{i, j\}: i < j,\ i, j \in O \cup \{0\}\}$ includes the edges connecting the origins to the cross-dock, and pairs of origins. The edge set $E^2 = \{\{i, j\}: i < j,\ i, j \in \{0\} \cup D\}$ is made up of the edges connecting the cross-dock to the destinations, as well as those connecting pairs of destinations. A given load q_o is available at each origin $o \in O$, whereas a demand r_d is associated with each destination $d \in D$. A limited fleet of K homogeneous vehicles is available at the cross-dock to perform pickup and delivery operations. Each vehicle k has a capacity Q. A vehicle used for pickup operations starts its route in 0, visits some origins where it picks up a load, and then returns to 0. Similarly, a vehicle used for delivery operations starts its route in 0, visits some destinations, and then returns to 0. The cross-dock is the depot for the vehicles. Split pickups or deliveries are usually not allowed.

Lee, Jung, and Lee (2006) examined a variant of the basic problem in which all of the vehicles are constrained to arrive simultaneously at the cross-dock once they conclude their pickup routes. This limitation was introduced to allow goods to be moved immediately from inbound to outbound vehicles without storing any inventory at the cross-dock. A *service time* was given for each vertex of the network to represent the time spent visiting that node. Furthermore, each vehicle must return to the cross-dock before the end of a scheduled *planning horizon*. The objective was to determine the optimal number of vehicles, their routes, their schedules, and their arrival times at the cross-dock, while minimizing the sum of transportation costs and fixed costs of vehicles. The problem was cast as an MILP model and solved by means of a tabu search (TS) heuristic with a very simple neighborhood structure.

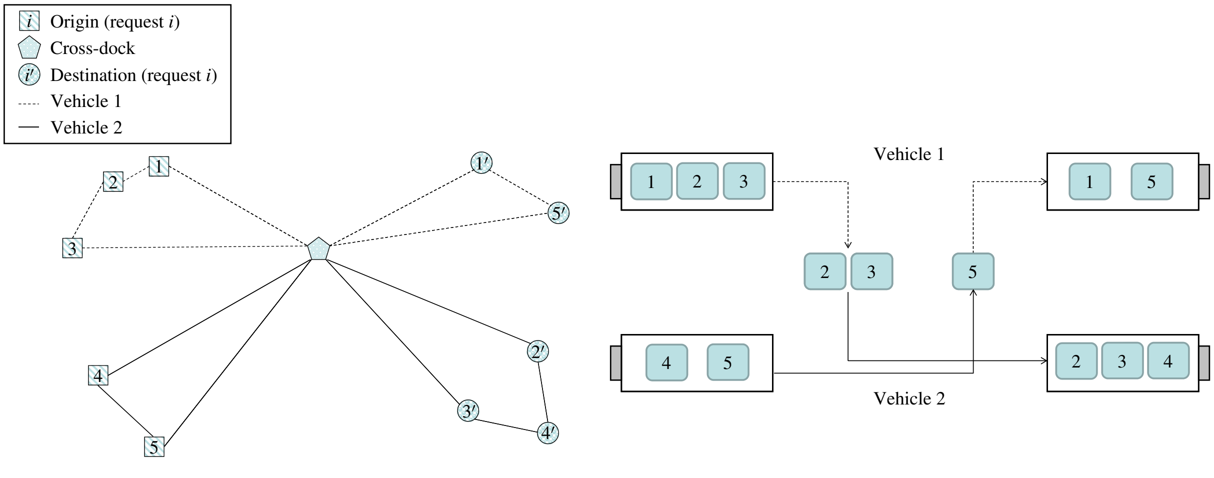
Computational experiments were performed on a set of randomly generated instances. The main drawback of this heuristic is that the number of vehicles used in the initial solution remains fixed in the rest of the execution. This limitation was overcome in the TS algorithm of Liao, Lin, and Shih (2010) who conducted some computational experiments on a new set of randomly generated instances. Their results showed that the new TS algorithm outperforms the previous one both in terms of solution quality and computing time.

Dondo, Méndez, and Cerdá (2009) considered a multicommodity PDPCD with several side constraints. The demand of each destination may include several products, while each product is available at a subset of origins and DCs. A given demand for each product is associated with every DC, representing a target inventory to achieve for that commodity. The network is hybrid and the destinations must be served within hard time windows. A set of heterogeneous vehicles is available, each of which having some attributes, including a fixed cost for usage, an average travel speed, and volume and weight capacity requirements. Service time at each vertex was modeled as the sum of two components: a vertex-dependent fixed time for preparation, and a variable part proportional to the unloaded or loaded quantity. Load splitting was allowed, i.e., several vehicles may stop at the same origin (destination) to perform pickup (delivery) operations. The problem was formulated as an MILP, which was solved by CPLEX. The authors presented the results for five sample instances based on Spanish town locations. Dondo, Méndez, and Cerdá (2011) studied an extension of this problem, which allows some intermediate facility to act as a cross-dock. The authors introduced an MILP formulation that generalizes the model introduced in Dondo, Méndez, and Cerdá (2009). The formulation comprises two objective functions that were treated as primary and secondary targets in their solution approach. Indeed, an MILP model that minimizes the total fixed and variable transportation cost was first solved. Then, an LP problem was obtained by fixing the binary variables to their optimal values and replacing the objective function with the minimization of the total travel time. Computational results were reported for five sample instances solved with Gurobi. These test instances were derived from those used in Dondo, Méndez, and Cerdá (2009).

3.2.2. One-to-One Pickup and Delivery Problems with Cross-Docks. The main problem of the class of one-to-one PDPCDs with single facility, single commodity, and defined on a pure network, can be described as follows. Let $\mathcal{R}$ be a set of requests, and $G=(N,E)$ be a weighted undirected graph. Set $N=O\cup\{0\}\cup D$ is the set of vertices, where $O=\{1,\ldots,|\mathcal{R}|\}$ is the set of origins, vertex 0 represents the unique (uncapacitated) cross-dock facility, and $D=\{1',\ldots,|\mathcal{R}|'\}$ is the set of destinations. For each request $i\in\mathcal{R}$, a load of size q_i has to be transported from its origin i to its destination i'. Edge set E and the fleet of K vehicles are defined as in §3.2.1. Split pickups or deliveries are usually not allowed. The assumption that distinguishes the one-to-one PDPCD from its corresponding PDP is that two vehicles can now handle the same request through the use of the cross-dock. In this situation, one vehicle collects the freight at its pickup location and drops it at the cross-dock, and a different vehicle loads the freight and carries out its delivery. Consequently, a pickup request is unloaded at the cross-dock only if its delivery is performed by another vehicle.

Figure 4 depicts an example of a feasible solution for a one-to-one PDPCD. Figure 4(a) shows the pickup and delivery routes for two vehicles. Dashed lines are used to represent the pickup and delivery routes traveled by vehicle 1. The corresponding routes traveled by vehicle 2 are depicted as solid lines. Figure 4(b) depicts the associated load exchanges between the two vehicles. For instance, requests 2 and 3 are collected by vehicle 1, and their delivery is carried out by vehicle 2.

PDPCDs with Time Windows. Several researchers studied members of the family of one-to-one PDPCDs that include time windows restrictions. The first contribution is due to Wen et al. (2009) and considers a PDPCD where a time window was also associated with the cross-dock to model the presence of a specific planning horizon. Additionally, the authors introduced a service time for both unloading and loading operations carried out at the cross-dock. The service time was modeled as the sum of two components: a fixed time for preparation, and a variable part that is proportional to the unloaded/loaded size. The problem stemmed from an application faced by a Danish company that dealt with collecting orders from some suppliers and dispatching them to supermarkets via a cross-dock. The authors proposed an MILP formulation that aims at minimizing the total distance traveled, subject to several constraints, which include those devoted to model the operations carried out by each vehicle at the cross-dock. The authors pointed out that relaxing the latter constraints, the problem essentially reduces to two independent CVRPs with time windows. The sum of the optimal costs of the two latter problems therefore yields a lower bound on the optimal solution value. The problem was solved by means of a TS heuristic embedded within an adaptive memory procedure (AMP-TS). Computational experiments were conducted on data sets generated from real data provided by a Danish logistics consultancy and show that the AMP-TS provided good quality solutions within reasonable computing times. Tarantilis (2013) introduced a TS algorithm embedded within an adaptive multi-restart framework (AMR-TS) for the PDPCD introduced in Wen et al. (2009), which is also capable of solving a number of variants



(a) Pickup and delivery routes

(b) Requests change vehicles at the cross-dock

Figure 4 (Color online) An Example of a Feasible Solution for a One-to-One PDPCD

with some adjustments. Computational experiments conducted on the same set of instances show that the AMR-TS algorithm improves several of the solutions reported in Wen et al. (2009) (only best solution costs obtained over different runs are reported). The first variant considered by Tarantilis (2013) assumes the presence of two different fleets of vehicles: the first used to perform only pickup operations, the second to carry out deliveries. This assumption implies that the entire freight collected by a pickup vehicle must be unloaded at the cross-dock and then loaded onto a delivery vehicle. Computational results show that the effect of introducing the two different fleets of vehicles on the total cost of the solution is relatively small. The second variant studied is the open version of the problem, i.e., vehicles are not based at the cross-dock, and pickup operations may start directly from any of the origins and delivery operations may finish after visiting the last destination of an open route. Morais, Mateus, and Noronha (2014) designed three iterated local search (ILS) heuristics for the PDPCD introduced in Wen et al. (2009). The best performing of these ILS heuristics maintains a set of elite solutions, which are used in a restart procedure, and has an intensification phase that is based on the solution of a set partitioning formulation for the problem of identifying which pickup and, separately, which delivery routes to select. Computational experiments were reported for the set of instances introduced in Wen et al. (2009) and a set of randomly generated instances of larger size produced by the authors from benchmark test problems available for the CVRP with time windows. The results show that the aforementioned ILS algorithm outperforms the AMP-TS designed in Wen et al. (2009), and yields better results than the AMR-TS developed in Tarantilis (2013) solving the larger-scale instances, whereas worse results are achieved on the smaller-scale ones.

Dondo and Cerdá (2013) extended the PDPCD introduced in Wen et al. (2009) by associating a service time also with each origin and each destination, in addition to the cross-dock. The authors cast their problem as an MILP model, and considered two alternative objective functions: the first one minimizes the total routing cost, whereas the second minimizes the time the last outbound vehicle returns to the cross-dock (i.e., the makespan). Because the size of the MILP formulation increases sharply with the number of requests, the authors considered a modification of the MILP model incorporating a set of constraints mimicking the sweep heuristic (see Gillett and Miller 1974) to assign requests to vehicles. Consequently, the resulting formulation provides heuristic solutions for the original problem. The authors compared the solutions found solving the latter formulation with those obtained with the MILP model without the additional constraints on small-scale instances generated by them. Their results show that the MILP sweep-based formulation finds good solutions within acceptable computing times. Additional results are given for the MILP sweep-based formulation solving larger-scale instances, both including and removing the time windows constraints, and also using both objective functions.

PDPCDs with Handling Costs. A handling cost is sometimes incurred when loading or unloading operations are carried out at a cross-dock. Santos, Mateus, and da Cunha (2011a) studied a PDPCD with a handling cost that depends on the vehicle that is loaded or unloaded and on the specific request. The problem is cast as a route-based BLP model that minimizes the total cost, equal to the sum of transportation and handling costs. The authors presented a B&P algorithm and conducted computational experiments on 50-request instances extracted from those of Wen et al. (2009). Santos, Mateus, and da Cunha (2011b) claimed that the BLP model formulated in Santos, Mateus, and da Cunha (2011a) suffers from large symmetries because of the presence of route-related binary variables indexed by the vehicles. To overcome this drawback, the authors introduced a novel BLP formulation along with a new B&P algorithm. Computational results confirm that the new B&P algorithm outperforms the exact method tested in Santos, Mateus, and da Cunha (2011a) in terms of duality gaps achieved within a given time limit.

PDPCDs with Hybrid Networks. A strong assumption made in most of the PDPCDs surveyed in this section is that all vehicles stop at the cross-dock during their route, even if no request has to be moved to a different vehicle for its delivery. Clearly, solutions of this type are not cost-effective when the origin and the associated destination of a given request are geographically close. Nevertheless, only a few authors have investigated variants where this restriction is removed. Liu, Li, and Chan (2003) were among the first to study a basic one-to-one PDPCD where vehicles are not forced to stop at the cross-dock after completing their pickup routes. Therefore, feasible solutions may include two types of routes: pure pickup and delivery routes (i.e., no intermediate stop at the cross-dock is performed) and routes that stop at the cross-dock to move loads from an inbound vehicle to an outbound vehicle. The only other characteristic of the problem studied in Liu, Li, and Chan (2003) that differs from the basic one-to-one PDPCD defined above is that in the former paper the authors assumed that the number of vehicles available is unlimited. The authors designed a local search (LS) heuristic. The basic idea of the heuristic is to partition the set of requests into two subsets: those served using direct deliveries and those served via the cross-dock. Given this partition, a set of CVRPs is solved by means of the classical Clarke and Wright (1964) heuristic for the VRP. The local search part of the algorithm consists of moving requests from one subset of requests to the other. Computational experiments were carried out on a set of randomly generated instances and aimed at providing a comparison, in terms of total distance traveled, between implementing a hybrid network, a pure network, or a distribution network that allows direct shipments only. The results show that adopting a hybrid network may save on average around 10% of the total traveling distance compared with either of the other two distribution systems.

Petersen and Røpke (2011) considered the PDPCD faced by a major Danish transporter of flowers with activities reaching most of northern and central Europe. In this problem, each request consists of a specific amount of flower containers to be picked up from an origin (i.e., gardeners' greenhouses) and then delivered to a destination (i.e., florists and supermarkets). If a request is served through the cross-dock, repacking and consolidation operations imply a cost per container handled. A service time for loading and unloading operations at the cross-dock as well as at each origin and destination is introduced as a stepwise linear function of the number of loaded or unloaded containers. The cross-dock, as well as each origin and destination, must be visited within a hard time window. The objective of the problem is to minimize the sum of transportation and handling costs. The authors designed a parallel ALNS. Computational experiments were performed on real-world data that comprise up to approximately 1,000 requests. The results indicate that it is beneficial to serve many requests via the cross-dock.

Santos, Mateus, and da Cunha (2013) extended their previous research (see Santos, Mateus, and da Cunha 2011a, b) by allowing direct deliveries from origins to destinations. This variant is formulated as a BLP and a B&P algorithm is designed for its solution. Computational experiments were carried out comparing the solutions for the PDPCD using a hybrid network with those for the PDPCD adopting a pure network (the latter solutions were obtained by means of an improved version of the B&P algorithm designed in Santos, Mateus, and da Cunha 2011b). The results show that on the tested instances, the average cost reduction due to the introduction of direct deliveries is approximately 3.3%.

PDPs with Transfers. In the PDP with transfers, requests can be moved from a vehicle to another at given locations, called transshipment points, and not necessarily at a cross-dock. The main difference between a cross-dock and a transshipment point is that whereas the former is the starting and ending vertex of each route, the latter can be any given location visited by a pair of vehicles and where goods can be moved from one vehicle to another. Mitrović-Minić and Laporte (2006) presented an empirical study on the usefulness of transshipment points in the context of courier companies serving large geographic areas. Indeed, in this application context companies often allow transshipments between vehicles to enable drivers to remain close to their home area, and the vehicle capacity can be neglected (loads are letters or small parcels). The authors also assumed that each location must be visited within a hard time window and is associated with a fixed service time. The authors designed a two-step heuristic to solve the problem. Computational experiments performed on randomly generated instances show that the introduction of transshipment points may significantly reduce the total distance traveled. Bouros et al. (2011) studied a dynamic PDP with transshipments. In this problem, a set of routes, called the static plan, is given a priori. Then, additional requests arrive at arbitrary times and the static plan must be modified to accommodate them. The objective is to minimize a combination of two costs. The first cost measures the delay, with respect to the static plan, because of serving the new requests. The second cost gives a measure of the promptness of satisfying the new requests. The authors solved the problem by means of a heuristic based on identifying, for each new request, the shortest path from the origin to the destination. The performance of this algorithm was compared with that of a heuristic combining an insertion algorithm with a TS mechanism. Computational experiments conducted on randomly generated instances show that the first heuristic is able to find solutions significantly faster than the second algorithm, with slightly worse solution costs, especially for the large-scale instances.

3.3. Research Directions

Since basically all of the contributions cited on 2E-VRPs relate to the basic 2E-CVRP, a promising research avenue is the study of other variants that include more realistic features. Several interesting variants are strongly related to the definition of multiechelon networks for freight movements in a city logistics context, a quickly emerging area of applications that is still relatively unexplored. In these problems, vehicles may perform several trips during the work day that use the same or different facilities. Furthermore, the service may have to be planned over a time horizon of several days because of strong variations in the demand patterns, thus making multitrip and multiperiod features particularly relevant since they may change considerably the network design and operations strategies. We also believe that it is worth investigating extensions of the 2E-CVRP where hard or soft time windows restrictions are associated with the arrival or the departure of vehicles to the satellites, or the origins and the destinations. We would also like to bring to the readers' attention all of the aspects related to operations synchronization and the possible consideration of multiple depots, i.e., where goods are available at more than one origin, and also the 2E-CVRP with a hybrid network, i.e., where direct deliveries are allowed if they are cost-effective or help to satisfy time constraints. Finally, a new and interesting research area is related to the use of environmentally friendly vehicles in urban freight distribution, such as electric or hybrid ones, which introduces a number of special characteristics to be addressed.

Another natural extension is the generalization of these problems to more than two echelons. An illustrative example is a network where DCs are positioned on the stage closer to the origins, and satellites are located on the stage closer to the destinations. The freight available at the origins is first delivered to a DC where it is temporarily stored. Goods are then transported from the DCs to the satellites, and then to their destinations. This class of problems represents a broader vision of the supply chain than the basic 2E-CVRP, which concentrates on the last two echelons of the chain. On the other hand, it raises several issues that complicate the problem considerably.

As far as the literature on PDPCDs is concerned, promising research directions include the analysis of some variants that have received little or no attention. As mentioned above, despite many-to-many PDPCDs represent a relatively unexplored research area, these problems are rarely encountered in practice. On the other hand, possible research directions worth pursuing include extensions of some of the variants reviewed above to their multifacility versions. In these problems many cross-docks are available, each one possibly associated with a given fleet size, handling cost, and service time. Furthermore, PDPCDs defined on hybrid networks have been rarely studied. Another interesting variant includes constraints imposing the synchronized arrivals and departures of vehicles at the cross-docks. Such constraints may impose, for instance, that once a vehicle has unloaded its freight at a cross-dock, those goods are immediately loaded onto another vehicle for delivery. The introduction of this type of constraint is, for example, of paramount importance when modeling the delivery of perishable goods to retail shops in urban areas. Finally, from an algorithmic perspective, an open research area is the design of exact methods that have currently been considered only for some basic variants of PDPCDs.

In both classes of problems a very limited number of studies assume the presence of DCs as intermediate facilities. Indeed, in almost all of the contributions cited the intermediate facilities are assumed to act as cross-docks, as defined in §2, where storage opportunities are either forbidden, very limited, or even ignored. Nevertheless, the inclusion of DCs requires the definition of additional elements, such as their capacity, storage costs, or a target inventory level to achieve at the end of a planning horizon. Finally, stochastic and dynamic counterparts of the problems surveyed in this section have not received any attention.

4. Intermediate Facilities in Transshipment Problems

The classical *transshipment problem* is an extension, originally introduced by Orden (1956), of the standard transportation problem where, besides origins and destinations, we are given a set of additional vertices corresponding to transshipment centers. More formally, the transshipment problem can be defined as follows (e.g., see Bazaraa, Jarvis, and Sherali 2010). Let $G=(N,A)$ be a weighted directed graph. Set $N=O\cup H\cup D$ is a set of vertices, where $O=\{1,\ldots,|O|\}$ represents a set of origins, $H=\{|O|+1,\ldots,|O|+|H|\}$ denotes a set of intermediate facilities (henceforth referred to as *transshipment centers*), and, finally, $D=\{|O|+|H|+1,\ldots,|O|+|H|+|D|\}$ is a set of destinations. Set $A=A^1\cup A^2\cup A^3$ is the set of arcs (i,j) such that arc set $A^1=\{(i,j)\colon i\in O,\ j\in H\}$ includes the arcs

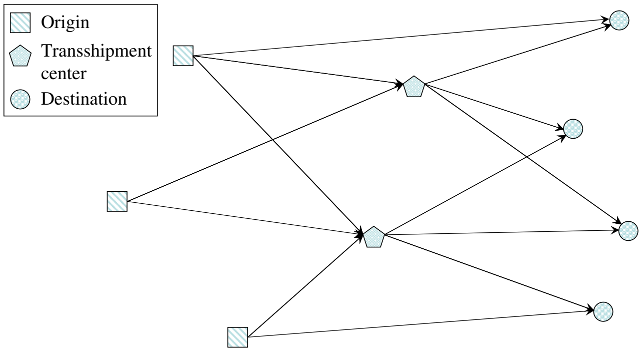


Figure 5 (Color online) An Example of a Feasible Solution for a Transshipment Problem

connecting origins to transshipment centers. Arc set $A^2 = \{(i, j): i \in H, j \in D\}$ comprises the arcs connecting transshipment centers to destinations. Finally, arc set $A^3 = \{(i, j): i \in O, j \in D\}$ contains the arcs linking origins with destinations. Note that if $A^3 = \varnothing$, then the network is pure. Associated with each origin $o \in O$ there is a given supply of q_o units of a commodity, whereas each destination $d \in D$ demands r_d units of the commodity. The problem calls for the determination of an optimal shipping plan from origins to destinations that minimizes the total transportation cost. It is worth mentioning that the classical transshipment problem can be converted into a transportation problem (e.g., see Bazaraa, Jarvis, and Sherali 2010). Additionally, the above transshipment problem models the flows of a single commodity over a network. Multicommodity variants arise when several commodities, differentiated by their physical characteristics, must move over the same network. Figure 5 illustrates an example of a feasible solution for a transshipment problem.

All papers discussed in this section consider a problem closely related to the classical transshipment problem. We have classified the papers mainly according to the assumption made on the origin-destination structure. We first review the papers that assume a many-to-many structure, followed by those considering one-to-one and one-to-many transshipment problems. Finally, we comment on the papers dealing with in-transit merge centers that, because of the particular role of this intermediate facility, we decided to keep separate from the others. At the end of the section we point out open research directions.

Table 4 presents an overview of the main problem characteristics of the papers surveyed in this section, whereas Table 5 provides some details on the optimization models and solution methods proposed in each paper. A reference to the solver used to solve the optimization model, if any, is also reported.

4.1. Many-to-Many Transshipment Problems

To our knowledge, Lim et al. (2005) were the first to extend the classical transshipment problem by introducing some characteristics arising from cross-dock networks. Specifically, the authors studied a problem defined on a pure network with multiple facilities where a hard time window is associated with each origin and destination. Every transshipment center has a maximum inventory holding capacity and an associated storage cost. Additionally, the authors assumed

Table 4 Intermediate Facilities in Transshipment Problems: A Summary of the Main Problem Characteristics

Reference	Type netw.	No. of I. F.[a]	OD struct.[b]	No. of comm.[c]	Capacity[d]	Storage I. F.	Time restriction[e]	Additional charact.
Lim et al. (2005)	Pure	M	M-M	SC	H, S	✓	FxdS, FlbS, HTW	
Miao et al. (2012)	Pure	M	M-M	SC	H, S	✓	FxdS, HTW, PSTW	Penalties
Miao, Fu, and Yang (2012)	Pure	M	M-M	SC	H, S	✓	FxdS, HTW	Penalties
Chen et al. (2006)	Pure	M	M-M	MC	H	✓	HTW, PH	
Marjani, Moattar Husseini, and Karimi (2012)	Pure	M	M-M	MC	H	✓	HTW, PH, STW	
Li et al. (2006)	Hybrid	M	M-M	SC	V	✓	HTW	
Ma et al. (2011)	Hybrid	M	M-M	SC	V	✓	HTW	
Lapierre, Ruiz, and Soriano (2004)	Hybrid	M	1-1	SC				
Musa, Arnaout, and Jung (2010)	Hybrid	M	1-1	SC	V			
Üster and Agrahari (2010)	Hybrid	M	1-1	SC	V			
Zäpfel and Wasner (2002)	Hybrid	S	1-1	SC	V			
Berman and Wang (2006)	Hybrid	M	1-1	MC	V			Pipeline and dest. inventory
Ali and O'Connor (2010)	Pure	M	1-M	MC	V			
Croxton, Gendron, and Magnanti (2003)	Pure	M	M-M	MC	H	✓		
Song, Hsu, and Cheung (2008)	Pure	S	1-1	SC		✓	DD, FxdS, RT	

[a] S, single facility; M, multifacility.
[b] 1-1, one-to-one; 1-M, one-to-many; M-M, many-to-many.
[c] SC, single commodity; MC, multicommodity.
[d] H, intermediate facilities; S, schedules; V, vehicles.
[e] DD, due dates; FlbS, flexible schedules; FxdS, fixed schedules; HTW, hard time windows; PH, planning horizon; PSTW, preferred service time windows; RT, release times; STW, soft time windows.

Table 5 Intermediate Facilities in Transshipment Problems: A Summary of the Optimization Models and Solution Algorithms

Reference	Optimization model	Heuristic algorithm	Exact algorithm
Lim et al. (2005)			
Miao et al. (2012)	ILP	Adaptive GA, Adaptive TS	CPLEX
Miao, Fu, and Yang (2012)	ILP	Hybrid GA	CPLEX
Chen et al. (2006)	ILP	SA, TS, Hybrid SA + TS	
Marjani, Moattar Husseini, and Karimi (2012)	Bi-Obj. ILP	TS, VNS, Hybrid SA	
Li et al. (2006)	ILP	2-step PSH	CPLEX
Ma et al. (2011)	ILP	2-step PSH	CPLEX
Lapierre, Ruiz, and Soriano (2004)	BNLP	TS, VNS, Hybrid TS + VNS	CPLEX
Musa, Arnaout, and Jung (2010)	ILP	ACO	LINGO
Üster and Agrahari (2010)	BLP	PSH	CPLEX
Zäpfel and Wasner (2002)	MILP	PSH	
Berman and Wang (2006)	BNLP	LRB	B&B
Ali and O'Connor (2010)	MILP		CPLEX
Croxton, Gendron, and Magnanti (2003)	ILP		PSE
Song, Hsu, and Cheung (2008)	Discrete NLP	LDB	CPLEX

that a set of shipping schedules is provided by some transportation companies. The schedules can be fixed or flexible. Indeed, schedules are fixed when departure and arrival times cannot be modified, such as schedules for railway and airline transportation services. On the other hand, schedules are flexible when departure and arrival times can vary in time; the typical example is a schedule for truck transportation. A travel cost and a capacity, among other attributes, are associated with each schedule. The objective of the problem is to determine an optimal set of shipping schedules minimizing the sum of travel and storage costs without violating time windows or the capacity constraints. The problem is analyzed under different assumptions concerning the number of shipments that can depart from an origin, or that can arrive at a destination. The authors analyzed three cases: (i) the case where multiple shipments can leave the same origin and multiple deliveries can arrive at a given destination; (ii) the case where only a single shipment per origin and a single delivery per destination are allowed (which is typical when the setup cost for each shipment is relatively high); and (iii) a combination of the two previous cases. They also identified, among such cases, those that are polynomially solvable and proved that the remaining cases are $\mathcal{NP}$-hard. No optimization models or solution algorithms were proposed. Miao et al. (2012) analyzed a variant of the single shipment-single delivery case. In this variant, preferred service and hard time windows are associated with each destination with the following properties. If the destination is not served within its preferred service time interval (but within its hard time windows), then a penalty is incurred. A higher penalty is paid if the destination cannot be served at all, because, for instance, no transportation schedule is available to carry out the delivery within its hard time window. No penalty is incurred if the destination is visited within its preferred service time window. The problem is formulated as an ILP model that determines an optimal set of schedules minimizing the total cost given by the sum of transportation, inventory storage, and penalty costs. The authors implemented two types of metaheuristics, namely, an adaptive TS heuristic and an adaptive genetic algorithm (GA). In both algorithms the search space is explored by means of a variable neighborhood search (VNS) algorithm. Computational experiments are performed on two sets of randomly generated instances. The results show that both heuristics perform better than CPLEX when terminated after a time threshold, and that the adaptive TS heuristic outperforms the adaptive GA on large-scale instances, whereas the opposite is true for small-scale instances. Miao, Fu, and Yang (2012) analyzed a variant similar to the latter problem. By contrast to the problem studied in Miao et al. (2012), these authors did not consider preferred service time windows for destinations, but associated a penalty with each origin. This penalty is incurred whenever a shipment available at an origin cannot be shipped out within its hard time window. The authors provided an ILP formulation of the problem drawing on the optimization model introduced in Miao et al. (2012), and designed a hybrid GA that integrates a greedy approach with a VNS algorithm. Computational experiments were carried out on randomly generated instances.

Chen et al. (2006) studied a variant of the transshipment problem where hard time windows are associated with origins and destinations, and a maximum storage capacity as well as a storage cost are given for each transshipment center. This multicommodity problem is defined on a pure multifacility network. The problem is formulated as an ILP model that minimizes the sum of transportation and storage costs within a planning horizon. Three heuristics were designed to solve the problem: a TS heuristic, a simulated annealing (SA) heuristic, and a hybrid metaheuristic in which tabu lists are embedded within an SA framework. Computational experiments, conducted on a set of randomly generated instances, show that the TS algorithm outperforms the two other heuristics. Marjani, Moattar Husseini, and Karimi (2012) extended the above problem as follows. Shipments between transshipment centers are allowed, initial and final inventories at the intermediate facilities are possible (whereas in Chen et al. 2006 they are constrained to be zero), and late deliveries are allowed for a subset of the destinations. The problem is formulated as a bi-objective ILP. One objective minimizes the total cost consisting of transportation and storage costs, whereas the second one minimizes the total tardiness (i.e., a measure of the delay for the late shipments). The authors developed three heuristics to solve the problem: a TS, a VNS, and a hybrid SA. The initial solution for all of the heuristics is constructed by a procedure that first assigns origins to destinations, and then identifies a transshipment center for each origin-destination pair using a heuristic. Computational experiments were performed on randomly generated instances and show the key role of the heuristic used to compute the initial solution. The results also indicate that the VNS outperforms the other two algorithms.

In all of the above papers the number of vehicles traveling on each arc of the network is not an explicit decision variable. Li et al. (2006) studied a variant of the transshipment problem that explicitly considers the number of vehicles traveling on each arc of a hybrid network. An unlimited fleet of homogeneous and capacitated vehicles is available to deliver the freight. A setup cost is associated with each arc, and is incurred whenever a vehicle traverses it. A storage cost is associated with each (uncapacitated) transshipment center. The resulting problem is cast as an ILP model that minimizes the total cost given by the sum of transportation, setup, and storage costs, while satisfying hard time window and vehicle capacity constraints. The authors implemented a two-step solution approach based on the idea of splitting the demand into full TL and less-than-truckload (LTL) plans. In the first step, a network flow algorithm is used to find the TL plan. Hence, an initial LTL plan is determined with the remaining shipments by means of a greedy algorithm. This plan is then improved by means of a GA in the second step. Computational experiments that were run on randomly generated instances show that the heuristic is faster than CPLEX. The above problem was also studied in Ma et al. (2011) who developed a two-step heuristic similar to the algorithm implemented in Li et al. (2006). The main difference between the two heuristics is that in the first one the second step consists of using a squeaky wheel optimization heuristic (e.g., see Joslin and Clements 1999) instead of a GA. In the computational experiments the authors compared the aforementioned heuristic with a two-step solution approach quite similar to that designed in Li et al. (2006). The effectiveness of the two algorithms was validated by comparing their results with CPLEX solving the ILP model provided in Li et al. (2006). The heuristic that uses the squeaky wheel optimization is on average faster, whereas the GA usually produces better quality solutions.

4.2. One-to-One Transshipment Problems

Some authors have studied variants of the transshipment problem that assume a one-to-one origin-destination structure. Lapierre, Ruiz, and Soriano (2004) considered such a structure for a transshipment problem defined on a hybrid network. Additionally, the authors explicitly included real-world handling rates into the problem definition. In particular, each request has two attributes: weight and density (defined as weight per cubic foot). The cost of delivering a request directly from its origin to its destination is obtained from a table by selecting the value corresponding to the request weight and density. The computation of the costs of shipments moving through transshipment centers is more complicated since these depend on the set of requests involved and, more precisely, on the total weight and average density of the requests moving on the same arc from an origin to a transshipment center, and from the center to a destination. The authors introduced a binary nonlinear programming (BNLP) formulation to minimize the total transportation cost subject to linear constraints. They designed a hybrid algorithm combining TS and VNS whose performance was validated on instances generated using real-world cost structures published by a U.S. carrier and random demands. They compared the solutions found by the hybrid algorithm either to the optimal ones or to lower bounds provided by CPLEX. Finally, the performance of the hybrid metaheuristic was found to be superior to both the TS and the VNS algorithms.

Musa, Arnaout, and Jung (2010) considered a problem in which an unlimited fleet of homogeneous and capacitated vehicles is available to serve requests over a hybrid network. The authors developed an ILP formulation for the problem that minimizes the total transportation cost, and designed an ant colony optimization (ACO) algorithm for its solution. The performance of the ACO algorithm was assessed by comparing its solutions with those found by the LINGO solver within a maximum computing time. Computational results show that the ACO algorithm outperforms LINGO, especially on large-scale instances.

Üster and Agrahari (2010) studied a transshipment problem defined on a hybrid network. In this problem a request is either delivered directly, or has to move through two transshipment centers. Such problems model situations in which requests coming from different origins are first collected at a consolidation center and subsequently, transferred to a deconsolidation center using a cheap mode of transportation (e.g., TL transportation). The requests are finally distributed from the deconsolidation centers to their destinations. The authors assumed that LTL transportation is used to move a request from its origin to a consolidation center, from a deconsolidation center to its destination, and from its origin directly to its destination. Conversely, loads from consolidation centers to deconsolidation centers are assumed to be shipped using a TL mode by means of a limited fleet of homogeneous and capacitated vehicles. The authors constructed a BLP formulation for this problem that minimizes the total transportation cost, i.e., the sum of transportation costs for direct shipments and through intermediate facilities. They also designed three heuristics based on different types of neighborhood functions and branching schemes. The first heuristic uses a deterministic branching scheme; the second one adopts a probabilistic branching scheme based on the concept, borrowed from SA algorithms, of acceptance probability functions; the last one is a TS algorithm. Computational experiments were conducted on randomly generated instances. The results indicate that, on average, the heuristic with probabilistic branching finds good quality solutions and outperforms the other approaches in terms of computing times.

Zäpfel and Wasner (2002) examined the problem faced by a medium-size Austrian parcel delivery provider that guarantees 24-hour service completion. The network used by the company is pure and is made up of one central transshipment center and 10 terminals, each terminal acting both as an origin and a destination for the parcels. The main goal of this paper is to provide evidence of the effectiveness of using a hybrid network instead of the pure network currently adopted by the company. In their study, the authors made the assumption that the freight can be moved by means of two vehicle types, with one or two containers. These vehicle types have different capacities and traveling costs. The authors modeled the problem as an MILP to minimize the total travel cost, and they designed a heuristic for its solution. A possible extension where the problem is combined with that of determining the routes for the pickup and delivery of the parcels was described along with an MILP formulation, but no solution algorithm was given.

Berman and Wang (2006) considered a multicommodity variant in which requests are defined by an origin-destination-commodity vector and by the size of the load. This problem aims at determining a distribution strategy in a hybrid network that minimizes the sum of transportation, pipeline inventory, and storage costs (the latter cost is introduced only for the destinations). Pipeline inventory cost can be regarded as a cost incurred for goods that have left the warehouses but have not yet arrived at their destination, and are therefore still within the firm distribution chain. An unlimited fleet of homogeneous and capacitated TL vehicles is available to transport the commodities. The authors designed a BNLP formulation with a nonlinear objective function, subject to linear constraints. A heuristic and an exact algorithm are implemented to solve the problem. The heuristic is based on a Lagrangean relaxation (LR) of the nonlinear formulation, and on a greedy algorithm to compute an initial feasible solution and an upper bound. The exact algorithm is a B&B algorithm that uses the LR to obtain a lower bound for each subproblem. The proposed algorithms were tested on randomly generated instances of different sizes.

4.3. One-to-Many Transshipment Problems

We are aware of only one paper in which the origin-destination structure is assumed to be one-to-many. Specifically, Ali and O'Connor (2010) carried out a computational study to investigate the computational tractability of the following one-to-many transshipment problem. The network considered is pure and comprises only one origin where several commodities are available to serve the demand of a set of destinations. The commodities are heterogeneous in terms of their unit volume, and are delivered using two fleets of capacitated vehicles. The vehicles belonging to the first fleet carry the freight from the origin to a transshipment center, and those in the second fleet transport the products from transshipment centers to their destinations. The vehicles are homogeneous within each fleet, and a fixed cost is incurred for every vehicle used. The problem is cast as an MILP that minimizes the sum of the vehicle fixed costs. The authors introduced two families of valid inequalities to strengthen their formulation. Computational experiments were conducted by solving with CPLEX a set of instances generated by the authors modifying, among other elements, the number and location of the destinations, as well as the quantity of a commodity demanded by each destination. The tested instances consist of up to 32 transshipment centers and 2,400 destinations. The results show the substantial contribution provided by the two families of valid inequalities in terms of duality gaps and number of instances solved within a given time limit.

4.4. In-Transit Merge Centers

Despite the practical relevance of in-transit merge centers, we are aware of only a few studies that focus on their use. Croxton, Gendron, and Magnanti (2003) were the first to analyze such networks from the perspective of optimizing tactical planning issues. They studied a multicommodity and multiperiod problem in which a given demand for a final product is associated with each destination in each time period. The final product is obtained by consolidating at an in-transit merge center diverse components available in unlimited quantities at different origins. In-transit merge centers have a volume capacity and may hold short-term inventory when this is cost-effective. The authors assumed that the components and the final products can be transported by means of four modes (namely, small package, LTL, TL, and airplane), each one with its own cost structure that is piecewise linear. The authors introduced an ILP formulation that tries to capture the piecewise linear structure of the cost functions, and use disaggregation techniques to derive a hierarchy of LP relaxations. To solve these LP relaxations the authors developed a cutting-plane procedure with some embedded rounding heuristics to obtain feasible solutions, which are subsequently improved by B&B. Computational experiments were performed using instances based on a real application arising in the U.S. computer industry.

Song, Hsu, and Cheung (2008) studied the problem of determining an optimal delivery of medical equipment composed of several modules in the Asia-Pacific region. The components become available at each origin at given release times, and are consolidated into a single shipment to be delivered to the in-transit merge center. At the center, all products heading to the same destination are consolidated into a single shipment that is sent before a given deadline. Consolidation at the center can be as early as possible (i.e., the shipment is consolidated immediately after all of the orders headed to the same destination have arrived) or as late as possible (i.e., the shipment is consolidated immediately before it is delivered to the destination), depending on the customer requirements and the different storage costs at the center for consolidated and unconsolidated shipments. Freight deliveries are constrained by fixed transportation schedules, making it possible to consider only a limited number of feasible transportation times for each shipment. The problem is formulated as a discrete nonlinear programming (NLP) model where the main decision variables belong to given discrete sets that represent, generally speaking, the set of feasible arrival (departure) times to (from) the in-transit merge center. The authors defined a Lagrangean dual-based (LDB) heuristic to solve the problem based on the observation that the Lagrangean dual of the model can be optimally solved as a linear problem, hence allowing the fast computation of a lower bound, and a procedure can be applied to its optimal solution to compute an upper bound on the primal problem. The performance of the algorithm is assessed by comparing the best upper bound found by the heuristic with the best between the lower bound provided by the Lagrangean dual and that returned by CPLEX, solving an MILP formulation of the problem, within a given time threshold.

4.5. Research Directions

Most papers working with a many-to-many structure assume the presence of a pure network and consider only one type of commodity. Consequently, a relatively unexplored research avenue is the study of more realistic variants, for instance those considering hybrid networks, or multicommodity counterparts where handling costs at the transshipment centers may depend on the commodity. Furthermore, other more realistic variants are those that consider the number of vehicles moving over the network as an explicit decision variable.

The majority of the papers considering a one-to-one origin-destination structure include the presence of a single commodity moving over a hybrid network. By contrast to the papers assuming a many-to-many structure, here the vehicles are usually considered to be homogeneous. Hence, a possible extension is to study the heterogeneous case, arising for instance when vehicles are differentiated in terms of traveling cost, time, or capacity. This extension is particularly relevant at a strategic level when fleet sizing and composition should be addressed, and would reflect the fact that an important decision faced by the management of transportation companies is to choose the type of vehicle to use on each route. In addition, no author has considered other characteristics that appear to be quite frequent and relevant in practical problems, such as the presence of time restrictions for the service or the consideration of a storage capability for transshipment centers. Finally, very few contributions deal with in-transit merge centers, in spite of their practical importance.

More generally, each paper surveyed in this section addresses a specific variant of the problem. We think that an important research direction in this area is represented by the definition and analysis of some basic common variants that capture its main features (note that the basic transshipment problem can be easily solved using a network simplex algorithm, see Bazaraa, Jarvis, and Sherali 2010). Possible examples are variants with only time windows or in which the number of vehicles is an explicit decision variable. Additionally, although most of the papers consider multifacility problems, the possibility of hauling shipments between two transshipment centers is always neglected. It would be interesting to investigate the magnitude of the possible savings that can be achieved by introducing shipments between these intermediate facilities. Furthermore, in most papers the transportation cost is defined per unit of product, whereas the fare structures published by carriers are often more complicated and may depend on several shipment attributes (e.g., weight and volume) or may be described by nonlinear function of the size of the shipment transported. The paper by Lapierre, Ruiz, and Soriano (2004) may represent the starting point for this line of research.

As far as the solution methods are considered, most of the papers proposed a heuristic algorithm, whereas very few of them designed exact methods, which, therefore, represent another open research direction. Finally, to our knowledge stochastic and dynamic counterparts of the problems surveyed in this section have received no attention, whereas multiperiod variants are covered only marginally.

5. Intermediate Facilities in Service Network Design Problems

Service network design problems simultaneously incorporate two types of major decisions (e.g., see Crainic and Kim 2007). The first type concerns the definition of the service network itself, i.e., the selection of the routes (in terms of origins, destinations, intermediate stops, and physical routes) over which a service is offered, as well as the attributes of each service (such as the mode of transportation to use and the frequency of each service). The second type of decision involves the specification of how the freight moves from its origin to its destination, i.e., the determination of the flows on the services and through the nodes of the service network. This class of problems arises in many application contexts related to freight transportation, typically involving trips over relatively long distances (see, e.g., the surveys of Wieberneit 2008 and Crainic 2000). Consequently, it is not our aim to provide a comprehensive discussion on the subject, but we will instead concentrate on those papers that focus on the intermediate facilities.

In this class of problems, the network consists of two sets of vertices representing terminals and intermediate facilities, respectively. Furthermore, the origin-destination structure is one-to-one, and each request is defined by an origin-destination pair, the size of freight to be hauled from the origin to the destination, and, usually, a level of service. In practice, each terminal has an associated service area, so that all loads originating within the assigned area are picked up and consolidated at the terminal. Additionally, the freight that must be delivered to a given local area is first dispatched to the associated terminal, and eventually delivered to the final customer. Consequently, each terminal is the origin of several requests and, at the same time, the destination of some other requests. The problems considered in the following papers focus on the movements of freight among terminals and intermediate facilities, i.e., pickup and delivery operations within the local areas are neglected. In addition to being locations where consolidation operations are carried out, intermediate facilities are sometimes origins or destinations for some requests. Finally, connections between intermediate facilities are usually possible.

This section is structured as follows. First, we review the papers addressing the design of service networks for time-definite freight common carriers. We then survey the contributions related to the problems faced by companies providing express shipment deliveries, and, subsequently, we review the papers on LTL carriers. Finally, the section ends with some directions for further research.

Depending on the application, terminals and intermediate facilities need specific equipment and layout, and are identified with different names. Indeed, terminals are usually called *centers* in the applications related to time-definite freight common carriers, *airstops* or *gateways* in those concerning express shipment deliveries, or *end-of-lines* in the LTL transport industry. Intermediate facilities are referred to as *hubs* in the first and second type of applications, and as *breakbulks* in the third one.

Table 6 presents an overview of the main problem characteristics of the papers surveyed in this section, whereas Table 7 provides some details on the optimization models and solution methods proposed in each paper. A reference to the solver used to solve the optimization model, if any, is also reported.

5.1. Time-Definite Freight Common Carriers

Freight common carriers provide door-to-door pickup and delivery services for small shipments, offering their services to the general public. They typically haul freight according to defined and published routes, time schedules, and rate tables. The term *time-definite* indicates that these companies provide guaranteed on-time delivery. Examples of companies in this industrial segment are UPS, FedEx, and DHL (see Lin and Chen 2008).

Service network design problems faced by these type of carriers in East Asia were studied in a number of papers by a research group in Taiwan. The basic problem they studied can be sketched as follows. Once all shipments to be delivered are collected at an origin center, they are consolidated and then loaded onto a fleet of capacitated vehicles. The vehicles are dispatched to various hubs for further consolidation. The loads are then moved from hubs to the respective destination centers, to be delivered to the final customers. Different types of vehicles can be used. For instance, trucks with trailers can be used when the distance to be traveled is short (e.g., from a center to a hub), whereas aircraft can be adopted for long-haul travels (e.g., between two hubs). As far as the network structure is considered, each node (both terminals and hubs) has a handling cost, a service time (proportional to the size of freight handled), and a capacity. A traveling cost, traveling time, and capacity are associated with each arc, where the capacity represents the total available capacity of all of the vehicles assigned to the arc. In its basic version,

Table 6 **Intermediate Facilities in Service Network Design Problems: A Summary of the Main Problem Characteristics**

Reference	Type netw.	No. of I. F.[a]	OD struct.[b]	No. of comm.[c]	Capacity[d]	Time restriction[e]	Additional charact.[f]
Lin (2001)	Pure	M	1-1	SC	A, H, T, V	LoS	ST
Lin and Wu (2001)	Pure	M	1-1	SC	A, H, T, V	LoS	ST
Lin and Chen (2004)	Pure	M	1-1	SC		LoS	ST
Lin and Chen (2008)	Hybrid	M	1-1	SC	A, H, T, V	LoS	ST
Lin (2004)	Pure	M	1-1	SC	A, H, T, V	LoS	ST
Lin and Lin (2007)	Pure	M	1-1	SC	A, H, T, V	LoS	ST
Barnhart and Schneur (1996)	Hybrid	S	1-1	SC	V	LoS, HTW	HV
Kim et al. (1999)	Pure	M	1-1	SC	A, H, V	LoS, HTW	HV
Barnhart et al. (2002)	Pure	M	1-1	SC	A, H, V	LoS, HTW	HV
Armacost, Barnhart, and Ware (2002)	Pure	M	1-1	SC	A, H, V	LoS, HTW	HV
Powell and Sheffi (1983)	Pure	M	1-1	SC	A, V	LoS	
Powell (1986)	Pure	M	1-1	SC	A, V	LoS	
Powell and Sheffi (1989)	Pure	M	1-1	SC	A, V	LoS	
Jarrah, Johnson, and Neubert (2009)	Hybrid	M	1-1	SC	A, V	LoS	
Cheung and Muralidharan (1999)	Hybrid	M	1-1	SC	V	LoS	
Cheung and Muralidharan (2000)	Pure	M	1-1	SC	V	LoS	

[a] S, single facility; M, multifacility.
[b] 1-1, one-to-one.
[c] SC, single commodity.
[d] A, arcs; H, intermediate facilities; T, terminals; V, vehicles.
[e] HTW, hard time windows; LoS, level of service.
[f] HV, heterogeneous vehicles; ST, service times.

this problem aims at minimizing the sum of traveling and handling costs, while not violating the capacity limitations and some further operational restrictions required by the application under study. Among the latter restrictions, constraints that guarantee the level of service are always present, whereas constraints imposing that there must be a directed in-tree rooted at any destination are usually included (they ensure that in any feasible solution only one path exists between each origin-destination pair).

Table 7 **Intermediate Facilities in Service Network Design Problems: A Summary of the Optimization Models and Solution Algorithms**

Reference	Optimization model	Heuristic algorithm	Exact algorithm
Lin (2001)	BLP		LR, IE-ϵ
Lin and Wu (2001)	ILP		IE-ϵ
Lin and Chen (2004)	BLP		IE
Lin and Chen (2008)	BLP		IE
Lin (2004)	—*Stochastic programming*—		
Lin and Lin (2007)	—*Stochastic programming*—		
Barnhart and Schneur (1996)	BLP	2-step PSH	
Kim et al. (1999)	MILP	PSH	
Barnhart et al. (2002)	MILP	PSH	
Armacost, Barnhart, and Ware (2002)	BLP		CPLEX
Powell and Sheffi (1983)	NLIP	LS	
Powell (1986)		LS	
Powell and Sheffi (1989)	NLIP	LS	
Jarrah, Johnson, and Neubert (2009)	MILP	SSH	
Cheung and Muralidharan (1999)		—*Simulation*—	
Cheung and Muralidharan (2000)		—*Dynamic*—	

Lin (2001) studied the operations carried out by a common carrier in Taiwan. In this problem, the carrier owns a fleet of trucks pulling trailers to move freight over a pure network. The problem is cast as a BLP model, and two exact algorithms are designed. The first algorithm is based on a LR of the optimization model, whereas the second one is an implicit enumeration algorithm with ϵ-optimality (IE-ϵ), where ϵ-optimality refers to a tolerance for an optimality condition (in particular, $\epsilon = 0$ requires the proven optimal solution). Computational experiments were performed using real data and show that the IE-ϵ algorithm outperforms the LR algorithm, especially for large networks. The author assumed that the vehicles assigned to each arc of the graph are known a priori, and did not explicitly consider any issue related to the balancing of the vehicles over the network. Such issues are included in the problem considered by Lin and Wu (2001) where the number of vehicles assigned to each arc is a decision variable. As a consequence, the capacity associated with each arc is not known a priori but is an output of the optimization process. Furthermore, the movements of empty vehicles must be taken into consideration to balance the number of vehicles entering and exiting each node. The study conducted by Lin and Wu (2001) stems from the observation that carriers usually run their delivery operations once a day. Hence, the authors studied the potential impact that the introduction of multiple frequency delivery operations may have on the size of the fleet of vehicles. The problem is formulated as an ILP model, and an IE-ϵ algorithm is designed for its solution. Computational experiments were conducted by considering two instances where the results related to the use of a single frequency delivery operations were compared to those based on a double frequency. The results show that running delivery operations twice may yield some savings compared with a single frequency, primarily because this allows the use of a smaller fleet.

A model commonly used in the air transportation industry is based on moving all traffic to and from terminals along "spokes" that are connected to a hub. The associated network is termed the hub-and-spoke network (HSN). Lin, Lin, and Lin (2003) provided a classification and detailed description of the HSNs in the context of air express common carriers. In particular, they identified three basic types of HSNs, namely, a pure HSN, an HSN with pickup stopovers and feeders, and an HSN with center-to-center directs. In a pure HSN, aircraft connect centers with hubs and pairs of hubs. The second type of network is often adopted when the size of the loads to be picked up is relatively small. In these situations, small aircraft, called feeders, stop over a sequence of centers to pick up additional freight before arriving at an air transit center where the freight from the feeders is unloaded and reloaded onto larger aircraft destined to a hub for consolidation. Stopovers are also possible for the larger aircraft on the route from an air transit center to a hub. Moreover, stopovers and feeders are usually used when carrying out delivery operations. An HSN with center-to-center direct deliveries is an HSN with stopovers and feeders where shipments can be unloaded during stopovers. Only in the last type of network are direct deliveries between centers allowed, thus resulting in a hybrid network.

Lin and Chen (2004) studied a special type of HSN with feeders and stopovers where centers are clustered and assigned to one "primary" hub through which all shipments must move. Specifically, the freight originating at all centers in a given cluster is first moved to the assigned primary hub, and, similarly, all shipments to be delivered to any center belonging to the cluster must pass through the assigned primary hub. This type of HSN is called hierarchical. Lin and Chen (2004) analyzed its use in the context of time-definite freight common carriers. The authors considered a simplified situation where only fixed and traveling costs for each vehicle used are taken into account. Furthermore, a constraint ensuring the level of service is included, while no capacity or other specific constraints are considered. The problem calls for the determination of the mode used to deliver shipments that minimizes the total cost, given by the sum of the fixed and traveling costs, using both primary feeders/aircraft (i.e., vehicles moving between primary hubs) and secondary feeders/aircraft (i.e., vehicles moving among centers in a cluster, and between a center and its primary hub). The problem is formulated as a BLP and solved by means of an IE algorithm. Lin and Chen (2008) analyzed a general service network design problem for time-definite freight common carriers, capable of capturing and integrating the three basic types of HSNs mentioned above. The problem studied in Lin and Chen (2008) is general and includes, among other elements, the management of a fleet of different types of vehicles, each type with its own capacity, a handling cost and a maximum handling capacity for each node, and vehicle balancing constraints. The authors formulated the problem as a BLP, solved by means of an IE algorithm. Computational experiments were conducted on instances derived from real data provided by Federal Express AsiaOne.

In the papers mentioned above, all problem data are assumed to be deterministic. Some authors have considered that some data and, particularly the demands, are uncertain. Specifically, Lin (2004) considered several of the characteristics of the industrial segment that we mentioned above, and considered that the size of the load associated with each request is a random variable with a discrete distribution. The problem is formulated as a two-stage stochastic programming problem and solved heuristically. A generalization of the latter problem to a multiperiod setting is considered in Lin and Lin (2007). Under the latter assumption, the problem becomes a multistage stochastic programming problem that the authors solved by means of a heuristic.

5.2. Express Shipment Delivery Companies

A similar application was studied in a series of papers by Barnhart and Schneur (1996); Kim et al. (1999); Barnhart et al. (2002); and Armacost, Barnhart, and Ware (2002) for the delivery of express shipments. At its core, the latter problem is akin to that faced by time-definite freight common carriers: the company guarantees that a shipment is picked up and delivered within a specified time interval. Nevertheless, the papers reviewed in the following differ from those cited above in the operational restrictions of the specific application.

Barnhart and Schneur (1996) developed an optimization model and a solution algorithm for the problem faced by an express shipment company. The company hauls shipments over a network comprising several terminals and one hub. These shipments are delivered by using either a limited fleet of ground vehicles and aircraft (of different types) owned by the company, or commercial airlines (e.g., when it was not possible to guarantee the delivery of a shipment with the required level of service). Only a subset of terminals (referred to as airstops) are potentially visited by an aircraft, whereas the others (called centers) can only be visited by a ground vehicle. As a consequence, the movements of freight over this type of network may comprise a combination of ground (e.g., from a center to an airstop) and air services (e.g, from an airstop to the hub with, possibly, some stopovers). The authors formulated the problem as a BLP that includes several restrictions related to the operations of an air service, such as constraints ensuring that the number of aircraft departing from the hub should be spread out over a period of time. The solution approach implemented by the authors is a two-step heuristic. First, the LP relaxation of the model is solved, and then, starting from its optimal solution, the BLP model is restricted to a subset of the decision variables and solved using CPLEX. Computational experiments were reported for a set of instances representing the problem faced by the target express shipment company.

Some papers studied the optimal design of the service network used by UPS. The first attempt to model and solve this problem is reported in Kim et al. (1999). The network used by UPS includes several hubs, and the shipments can enter (or exit) the system only at terminals corresponding to an airport (here called gateway). The company imposed that each request stopped at one terminal at most on the route from its origin terminal to a hub and, similarly, that at most one stop was possible from a hub to its destination terminal. Other restrictions are related to the sorting and landing capacities of each hub, the presence of vehicle balance constraints, and the requirement that there must be at least one service route from each terminal to the UPS major hub, and vice versa. The authors cast the problem as an MILP and designed a heuristic algorithm whose general idea is to decompose the problem into two subproblems. A model was first solved to generate routes that specify the restricted network over which shipments are allowed to move. Then, and given the restricted network, the heuristic determines the optimal movements for the shipments. This latter heuristic was improved in Barnhart et al. (2002) by alternating between the generation of routes and the determination of shipment movements, and by allowing shipment decisions to influence the selection of routes, and vice versa. Furthermore, the authors introduced in Barnhart et al. (2002) a new formulation to determine the optimal shipment movements. Conventional network design optimization models for express shipment deliveries, such as the formulation described in Kim et al. (1999), are difficult to solve, primarily because their LP relaxations yield poor lower bounds. To overcome this drawback, Armacost, Barnhart, and Ware (2002) introduced an improved formulation based on the use of what the authors call “composite variables.” The basic idea of these variables is to represent combinations of aircraft routes that implicitly capture shipment movements. Thus, shipment movement variables are no longer explicitly represented as separate decision variables. The resulting formulation is a BLP model (see also the appendix in Armacost et al. 2004). Computational experiments conducted on the network of UPS show that the LP relaxation of the formulation using the composite variables gives stronger lower bounds than earlier models for the same problem. Finally, Armacost et al. (2004) provided a general account of the collaboration between academics and specialists at the UPS described in the papers just mentioned. In this paper, a particular emphasis is placed on the description of the modeling and computational obstacles encountered, the implementation issues, and the impact on the cost of operating the network by UPS and on how the planners carry out their job.

5.3. Less-Than-Truckload Carriers

LTL carriers are specialized in transporting shipments whose weight usually ranges between a few hundred to a few thousand pounds (e.g., see Jarrah, Johnson, and Neubert 2009). Typically, these shipments cannot fill the entire vehicle capacity on their own. Thus, it is beneficial for carriers to consolidate several shipments using a network that can be sketched as follows (see Powell and Sheffi 1983 for further details). The network is composed of end-of-lines and breakbulks that correspond to terminals and intermediate facilities, respectively, in the general description provided at the beginning of this section. Each end-of-line is usually associated with a “primary” breakbulks. All shipments that originate from (or have to be delivered to) any end-of-line move through its primary breakbulks. Moving from its origin end-of-line to its destination end-of-line, a shipment can stop at more than one breakbulks for further sorting and consolidation. Additionally, since some routes are too long to be handled by one driver, the trip may have to be broken into two or more segments to allow a change of drivers. An additional restriction is that the flows of all shipments destined to a given destination must follow a directed tree. This guarantees that a unique path is selected between every pair of nodes of the network, but also that the path is independent of the origin of the shipment (see Powell and Sheffi 1989). Finally, like the other applications presented in this section, LTL carriers provide a guaranteed level of service, usually defined as the number of weekdays within which a shipment will arrive at its destination.

Powell and Sheffi (1983) were among the first to study the service network design problem in the LTL industry from an operations research perspective. In the nonlinear integer programming (NLIP) model they proposed, the level of service is not explicitly considered, but rather approximated by imposing that if an arc is used in a feasible solution, then a minimum frequency in terms of vehicles per time period must be provided. This guarantees an acceptable level of service since it ensures that a shipment never waits too long for the next vehicle to leave. Furthermore, the authors ignored the repositioning of empty trailers. They designed a local search heuristic, which, after computing an initial solution, applies a sequence of local changes based on adding and removing links. The former study is extended in Powell (1986) where the problem was decomposed hierarchically into a network design problem used to determine on which arcs to provide service, and two subproblems to decide how the freight should be routed and what should be the flows of empty trailers to balance the network. A mathematical formulation for each problem is provided. The solution algorithm is an LS heuristic that adds and removes links taking advantage of the structure of LTL networks. Additionally, after each local change, the freight routing subproblem is reoptimized, and the empty balancing subproblem is optimally solved. The research described in these two papers was carried out within a project concerning the definition of an interactive optimization software to solve the service network design problem faced by an LTL carrier operating in the U.S. The design and implementation of this software is described in detail in Powell and Sheffi (1989). Furthermore, and compared to the problem studied in Powell (1986), in Powell and Sheffi (1989) the presence of full TL shipments, which can influence the movements of empty trailers, is considered explicitly. The software described in Powell and Sheffi (1989) uses a solution approach based on hierarchically decomposing the problem, similar to Powell (1986), but with the inclusion of a subproblem to determine the routing of full TL shipments. The software also combines the use of local changes to the network structure and the optimization of the subproblems with user interaction (e.g., the user can specify regions of the network to examine for possible improvements). Several details on the architecture of the software, the implementation issues, and the impacts in terms of cost reduction and the approach used to do network planning at the target company are also provided.

Jarrah, Johnson, and Neubert (2009) studied the optimal design of service networks for an LTL carrier operating in North America. Compared to the papers mentioned above, the main contributions introduced here are the explicit account of the time dimension when determining freight movements, the inclusion of days-of-week freight fluctuations, and the consideration of the level of service defined explicitly as a maximum number of days between the pickup and the delivery of a shipment (rather than approximating it by imposing a minimum frequency on each arc used). Based on the observation that in an LTL application shipment movements headed to a given destination form a directed tree rooted at that node, the authors cast the problem as a MILP using trees as decision variables. The solution approach implemented is based on a slope scaling heuristic (SSH; see Crainic, Gendron, and Hernu 2004) where the LP relaxation of the MILP is solved by column generation. Several variants were designed along this general idea.

All of the previous authors have considered deterministic algorithms for the problem faced by LTL companies. Some researchers used simulation and dynamic approaches. Specifically, Cheung and Muralidharan (1999) developed a system to simulate the operations of an LTL carrier that considers several operational features, including driver rules and several factors to decide when closing a trailer (such as a minimum weight and volume loaded, or minimum frequencies to be provided). The simulation study was based on randomly generating the arrival times, as well as the sizes, of the shipments. The authors carried out a simulation study aimed at evaluating the impact, on the cost of operating the system and on the level of service provided, of modifying some of the operational features mentioned above. Cheung and Muralidharan (2000) used a dynamic approach to move priority shipments, i.e., shipments that are handled first, for example, because they are likely to arrive late at their destination. This type of shipment represents only a small portion, usually between 5% and 10%, of the total shipments of an LTL carrier. In this paper, priority shipment movements were determined by using real-time information. The problem was approximated by finding dynamic shortest paths over a network with random arc costs.

5.4. Research Directions

The papers reviewed in this section are often related to a specific application and are frequently the results of collaborations between academics and companies. We believe that additional efforts should be devoted to the design of optimization models and solutions methods for more general service network design problems than those currently available in the literature (some examples of general formulations and solution algorithms can be found in Wieberneit 2008 and in the few contributions cited therein). Furthermore, most papers were published before 2005. In the last 10 years, major developments have taken place in the design of algorithms, both exact, such as B&C and B&C&P, and heuristic, such as ALNS, MH, and GA. We expect that these developments, coupled with the improvement of computer capabilities, can make these challenging problems more manageable. Furthermore, most papers considered problems at a national or regional level, whereas our economy is becoming increasingly global, and requires the design of international or even intercontinental service networks often based on the use of multiple modes of transport. This dramatically increases the size of real-world instances and the difficulty of solving them. It also requires the introduction of further operational restrictions and characteristics, for example the presence of several consolidation centers along the same route.

Finally, more research efforts should also be devoted to the design of models for dynamic and stochastic problems.

6. Conclusions

We have focused on the presence of intermediate facilities in freight transportation planning problems. In such facilities, different operations are carried out, including consolidation, merging, transshipment, and storage operations. The detailed study of the role of these facilities has recently received increasing attention, stimulated by the new approaches aimed at designing more efficient and cost-effective systems to deliver freight. Several successful applications have inspired operations researchers to investigate these problems, and to design optimization models and algorithms. The number of contributions on this topic is significant, but the literature is fragmented. This survey has provided a classification of the literature into three classes: intermediate facilities in vehicle routing problems, in transshipment problems, and in service network design problems. For each class of problems we have identified the most important variants, reviewed the main exact and heuristic solution approaches available, and indicated several promising research areas.

Acknowledgments

The authors would like to express their appreciation for the insightful comments and suggestions made by the associate editor who handled this paper.

References

Agustina D, Lee CKM, Piplani R (2010) A review: Mathematical models for cross docking planning. *Internat. J. Engrg. Bus. Management* 2(2):47–54.

Ala-Risku T, Kärkkäinen M, Holmström J (2003) Evaluating the applicability of merge-in-transit. *Internat. J. Logist. Management* 14(2):67–82.

Ali AI, O'Connor DJ (2010) The impact of distribution system characteristics on computational tractability. *Eur. J. Oper. Res.* 200(2):323–333.

Alumur S, Kara BY (2008) Network hub location problems: The state of the art. *Eur. J. Oper. Res.* 190(1):1–21.

Apte UM, Viswanathan S (2000) Effective cross docking for improving distribution efficiencies. *Internat. J. Logist.: Res. Appl.* 3(3):291–302.

Armacost AP, Barnhart C, Ware K (2002) Composite variable formulations for express shipment service network design. *Transportation Sci.* 36(1):1–20.

Armacost AP, Barnhart C, Ware KA, Wilson AM (2004) UPS optimizes its air network. *Interfaces* 34(1):15–25.

Baldacci R, Mingozzi A, Roberti R, Wolfler Calvo R (2013) An exact algorithm for the two-echelon capacitated vehicle routing problem. *Oper. Res.* 61(2):298–314.

Barnhart C, Schneur RR (1996) Air network design for express shipment service. *Oper. Res.* 44(6):852–863.

Barnhart C, Krishnan N, Kim D, Ware K (2002) Network design for express shipment delivery. *Comput. Optim. Appl.* 21(3):239–262.

Bazaraa MS, Jarvis JJ, Sherali HD (2010) *Linear Programming and Network Flows*, 4th ed. (John Wiley & Sons, Hoboken, NJ).

Berbeglia G, Cordeau J-F, Laporte G (2010) Dynamic pickup and delivery problems. *Eur. J. Oper. Res.* 202(1):8–15.

Berbeglia G, Cordeau J-F, Gribkovskaia I, Laporte G (2007) Static pickup and delivery problems: A classification scheme and survey. *TOP* 15(1):1–31.

Berman O, Wang Q (2006) Inbound logistic planning: Minimizing transportation and inventory cost. *Transportation Sci.* 40(3):287–299.

Bouros P, Sacharidis D, Dalamagas T, Sellis T (2011) Dynamic pickup and delivery with transfers. Pfoser D, Tao Y, Mouratidis K, Nascimento MA, Mokbel M, Shekhar S, Huang Y, eds. *Advances in Spatial and Temporal Databases*, Lecture Notes Comput. Sci., Vol. 6849 (Springer-Verlag, Berlin Heidelberg), 112–129.

Boysen N (2010) Truck scheduling at zero-inventory cross docking terminals. *Comput. Oper. Res.* 37(1):32–41.

Boysen N, Fliedner M (2010) Cross dock scheduling: Classification, literature review and research agenda. *Omega* 38(6):413–422.

Campbell JF, O'Kelly ME (2012) Twenty-five years of hub location research. *Transportation Sci.* 46(2):153–169.

Chen P, Guo Y, Lim A, Rodrigues B (2006) Multiple crossdocks with inventory and time windows. *Comput. Oper. Res.* 33(1):43–63.

Cheung RK, Muralidharan B (1999) Impact of dynamic decision making on hub-and-spoke freight transportation networks. *Ann. Oper. Res.* 87:49–71.

Cheung RK, Muralidharan B (2000) Dynamic routing for priority shipments in LTL service networks. *Transportation Sci.* 34(1): 86–98.

Chopra S, Meindl P (2009) *Supply Chain Management: Strategy, Planning, and Operation*, 4th ed. (Prentice Hall, Boston).

Christofides N, Eilon S (1969) An algorithm for the vehicle-dispatching problem. *Oper. Res. Quart.* 20(3):309–318.

Clarke G, Wright JW (1964) Scheduling of vehicles from a central depot to a number of delivery points. *Oper. Res.* 12(4):568–581.

Contreras I (2015) Hub location problems. Laporte G, Nickel S, Saldanha da Gama F, eds. *Location Science* (Springer-Verlag, Berlin Heidelberg), 311–344.

Crainic TG (2000) Service network design in freight transportation. *Eur. J. Oper. Res.* 122(2):272–288.

Crainic TG, Kim KH (2007) Intermodal transportation. Barnhart C, Laporte G, eds. *Transportation*, Handbooks Oper. Res. Management Sci., Vol. 14 (North-Holland, Amsterdam), 467–537.

Crainic TG, Gendron B, Hernu G (2004) A slope scaling/Lagrangean perturbation heuristic with long-term memory for multicommodity capacitated fixed-charge network design. *J. Heuristics* 10(5):525–545.

Crainic TG, Ricciardi N, Storchi G (2009) Models for evaluating and planning city logistics systems. *Transportation Sci.* 43(4):432–454.

Crainic TG, Mancini S, Perboli G, Tadei R (2010) Two-echelon vehicle routing problem: A satellite location analysis. *Procedia—Soc. Behavioral Sci.* 2(3):5944–5955.

Crainic TG, Mancini S, Perboli G, Tadei R (2011) Multi-start heuristics for the two-echelon vehicle routing problem. Merz P, Hao J-K, eds. *Evolutionary Computation in Combinatorial Optimization*, Lecture Notes Comput. Sci., Vol. 6622 (Springer-Verlag, Berlin Heidelberg), 179–190.

Crainic TG, Mancini S, Perboli G, Tadei R (2012) Impact of generalized travel costs on satellite location in the two-echelon vehicle routing problem. *Procedia—Soc. Behavioral Sci.* 39:195–204.

Crainic TG, Mancini S, Perboli G, Tadei R (2013) A GRASP with path-relinking metaheuristic for the two-echelon vehicle routing problem. Di Gaspero L, Schaerf A, Stützle T, eds. *Advances in Metaheuristics*, Oper. Res./Comput. Sci. Interfaces Series, Vol. 53 (Springer, New York), 113–125.

Croxton KL, Gendron B, Magnanti TL (2003) Models and methods for merge-in-transit operations. *Transportation Sci.* 37(1):1–22.

Cuda R, Guastaroba G, Speranza MG (2015) A survey on two-echelon routing problems. *Comput. Oper. Res.* 55:185–199.

Dondo R, Cerdá J (2013) A sweep-heuristic based formulation for the vehicle routing problem with cross-docking. *Comput. Chemical Engrg.* 48:293–311.

Dondo R, Méndez CA, Cerdá J (2009) Managing distribution in supply chain networks. *Indust. Engrg. Chemistry Res.* 48(22):9961–9978.

Dondo R, Méndez CA, Cerdá J (2011) The multi-echelon vehicle routing problem with cross docking in supply chain management. *Comput. Chemical Engrg.* 35(12):3002–3024.

Farahani RZ, Hekmatfar M, Arabani AB, Nikbakhsh E (2013) Hub location problems: A review of models, classification, solution techniques, and applications. *Comput. Indust. Engrg.* 64(4):1096–1109.

Gentry C (2005) Million-dollar savings. *Chain Store Age* 81(2):54–56.

Gillett BE, Miller LR (1974) A heuristic algorithm for the vehicle-dispatch problem. *Oper. Res.* 22(2):340–349.

Gümüş M, Bookbinder JH (2004) Cross-docking and its implications in location-distribution systems. *J. Bus. Logist.* 25(2):199–228.

Hall RW (1987a) Comparison of strategies for routing shipments through transportation terminals. *Transportation Res. Part A: General* 21(6):421–429.

Hall RW (1987b) Direct versus terminal freight routing on a network with concave costs. *Transportation Res. Part B: Methodological* 21(4):287–298.

Hemmelmayr VC, Cordeau J-F, Crainic TG (2012) An adaptive large neighborhood search heuristic for two-echelon vehicle routing problems arising in city logistics. *Comput. Oper. Res.* 39(12):3215–3228.

Hu Z-H, Zhao Y, Choi T-M (2013) Vehicle routing problem for fashion supply chains with cross-docking. *Math. Problems Engrg.* http://dx.doi.org/10.1155/2013/362980.

Jarrah AI, Johnson EL, Neubert LC (2009) Large-scale, less-than-truckload service network design. *Oper. Res.* 57(3):609–625.

Jepsen M, Spoorendonk S, Røpke S (2013) A branch-and-cut algorithm for the symmetric two-echelon capacitated vehicle routing problem. *Transportation Sci.* 47(1):23–37.

Joslin DE, Clements DP (1999) Squeaky wheel optimization. *J. Artificial Intelligence Res.* 10(5):353–373.

Kara BY, Taner MR (2011) Hub location problems: The location of interacting facilities. Eiselt H, Marianov V, eds. *Foundations of Location Analysis* (Springer, New York), 273–288.

Kärkkäinen M, Ala-Risku T, Holmström J (2003) Increasing customer value and decreasing distribution costs with merge-in-transit. *Internat. J. Physical Distribution Logist. Management* 33(2):132–148.

Kim D, Barnhart C, Ware K, Reinhardt G (1999) Multimodal express package delivery: A service network design application. *Transportation Sci.* 33(4):391–407.

Klincewicz JG (1990) Solving a freight transport problem using facility location techniques. *Oper. Res.* 38(1):99–109.

Lapierre SD, Ruiz AB, Soriano P (2004) Designing distribution networks: Formulations and solution heuristic. *Transportation Sci.* 38(2):174–187.

Lee YH, Jung JW, Lee KM (2006) Vehicle routing scheduling for cross-docking in the supply chain. *Comput. Indust. Engrg.* 51(2):247–256.

Leung JMY, Magnanti TL, Singhal V (1990) Routing in point-to-point delivery systems: Formulations and solution heuristics. *Transportation Sci.* 24(4):245–260.

Li X, Lim A, Miao Z, Rodrigues B (2006) Reducing transportation costs in distribution networks. Ali M, Dapoigny R, eds. *Advances in Applied Artificial Intelligence*, Lecture Notes Comput. Sci., Vol. 4031 (Springer-Verlag, Berlin Heidelberg), 1138–1148.

Liao CJ, Lin Y, Shih SC (2010) Vehicle routing with cross-docking in the supply chain. *Expert Systems Appl.* 37(10):6868–6873.

Lim A, Miao Z, Rodrigues B, Xu Z (2005) Transshipment through crossdocks with inventory and time windows. *Naval Res. Logist.* 52(8):724–733.

Lin C-C (2001) The freight routing problem of time-definite freight delivery common carriers. *Transportation Res. Part B: Methodological* 35(6):525–547.

Lin C-C (2004) Load planning with uncertain demands for time-definite freight common carriers. *Transportation Res. Record: J. Transportation Res. Board* 1873:17–24.

Lin C-C, Chen SH (2004) The hierarchical network design problem for time-definite express common carriers. *Transportation Res. Part B: Methodological* 38(3):271–283.

Lin C-C, Chen SH (2008) An integral constrained generalized hub-and-spoke network design problem. *Transportation Res. Part E: Logist. Transportation Rev.* 44(6):986–1003.

Lin C-C, Lin JS-J (2007) The multistage stochastic integer load planning problem. *Transportation Res. Part E: Logist. Transportation Rev.* 43(2):143–156.

Lin C-C, Wu Y-C (2001) The multiple frequency delivery operations in the time-definite delivery industry. *J. Oper. Res. Soc.* 52(11):1215–1224.

Lin C-C, Lin Y-J, Lin D-Y (2003) The economic effects of center-to-center directs on hub-and-spoke networks for air express common carriers. *J. Air Transport Management* 9(4):255–265.

Liu J, Li CL, Chan CY (2003) Mixed truck delivery systems with both hub-and-spoke and direct shipment. *Transportation Res. Part E: Logist. Transportation Rev.* 39(4):325–339.

Ma H, Miao Z, Lim A, Rodrigues B (2011) Crossdocking distribution networks with setup cost and time window constraint. *Omega* 39(1):64–72.

Marjani MR, Moattar Husseini SM, Karimi B (2012) Bi-objective heuristics for multi-item freights distribution planning problem in crossdocking networks. *Internat. J. Advanced Manufacturing Tech.* 58(9–12):1201–1216.

Miao Z, Fu K, Yang F (2012) A hybrid genetic algorithm for the multiple crossdocks problem. *Math. Problems Engrg.* http://dx.doi.org/10.1155/2012/316908.

Miao Z, Yang F, Fu K, Xu D (2012) Transshipment service through crossdocks with both soft and hard time windows. *Ann. Oper. Res.* 192(1):21–47.

Mitrović-Minić S, Laporte G (2006) The pickup and delivery problem with time windows and transshipment. *INFOR—Inform. Systems Oper. Res.* 44(3):217–228.

Morais VWC, Mateus GR, Noronha TF (2014) Iterated local search heuristics for the vehicle routing problem with cross-docking. *Expert Systems Appl.* 41(16):7495–7506.

Musa R, Arnaout JP, Jung H (2010) Ant colony optimization algorithm to solve for the transportation problem of cross-docking network. *Comput. Indust. Engrg.* 59(1):85–92.

Orden A (1956) The transhipment problem. *Management Sci.* 2(3):276–285.

Parragh SN, Doerner KF, Hartl RF (2008) A survey on pickup and delivery problems. Part II: Transportation between pickup and delivery locations. *J. für Betriebswirtschaft* 58(2):81–117.

Perboli G, Tadei R, Masoero F (2010) New families of valid inequalities for the two-echelon vehicle routing problem. *Electronic Notes Discrete Math.* 36:639–646.

Perboli G, Tadei R, Vigo D (2011) The two-echelon capacitated vehicle routing problem: Models and math-based heuristics. *Transportation Sci.* 45(3):364–380.

Petersen H, Røpke S (2011) The pickup and delivery problem with cross-docking opportunity. Böse JW, Hu H, Jahn C, Shi X, Stahlbock R, Voß S, eds. *Computational Logistics*, Lecture Notes Comput. Sci., Vol. 6971 (Springer-Verlag, Berlin Heidelberg), 101–113.

Pisinger D, Røpke S (2007) A general heuristic for vehicle routing problems. *Comput. Oper. Res.* 34(8):2403–2435.

Powell WB (1986) A local improvement heuristic for the design of less-than-truckload motor carrier networks. *Transportation Sci.* 20(4):246–257.

Powell WB, Sheffi Y (1983) The load planning problem of motor carriers: Problem description and a proposed solution approach. *Transportation Res. Part A: General* 17(6):471–480.

Powell WB, Sheffi Y (1989) Design and implementation of an interactive optimization system for network design in the motor carrier industry. *Oper. Res.* 37(1):12–29.

Prins C, Prodhon C, Wolfler Calvo R (2004) Nouveaux algorithmes pour le problème de localisation et routage sous contraintes de capacité. Dolgui A, Dauzère-Pérès S, eds. *Modelisation et simulation pour l'analyse et l'optimisation des systèmes industriels et logistiques (MOSIM 2004)*, Vol. 2 (Lavoisier, Paris), 1115–1122.

Prodhon C, Prins C (2014) A survey of recent research on location-routing problems. *Eur. J. Oper. Res.* 238(1):1–17.

Santos FA, da Cunha AS, Mateus GR (2013) Branch-and-price algorithms for the two-echelon capacitated vehicle routing problem. *Optim. Lett.* 7(7):1537–1547.

Santos FA, Mateus GR, da Cunha AS (2011a) A branch-and-price algorithm for a vehicle routing problem with cross-docking. *Electronic Notes Discrete Math.* 37:249–254.

Santos FA, Mateus GR, da Cunha AS (2011b) A novel column generation algorithm for the vehicle routing problem with cross-docking. Pahl J, Reiners T, Voß S, eds. *Network Optimization*, Lecture Notes Comput. Sci., Vol. 6701 (Springer-Verlag, Berlin Heidelberg), 412–425.

Santos FA, Mateus GR, da Cunha AS (2013) The pickup and delivery problem with cross-docking. *Comput. Oper. Res.* 40(4):1085–1093.

Santos FA, Mateus GR, da Cunha AS (2015) A branch-and-cut-and-price algorithm for the two-echelon capacitated vehicle routing problem. *Transportation Sci.* 49(2):355–368.

Simchi-Levi D, Kaminsky P, Simchi-Levi E (2007) *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies*, 3rd ed. (Irwin/McGraw-Hill, New York).

Song H, Hsu VN, Cheung RK (2008) Distribution coordination between suppliers and customers with a consolidation center. *Oper. Res.* 56(5):1264–1267.

Stalk G, Evans P, Shulman LE (1992) Competing on capabilities: The new rules of corporate strategy. *Harvard Bus. Rev.* 70(2):57–69.

SteadieSeifi M, Dellaert NP, Nuijten W, Van Woensel T, Raoufi R (2014) Multimodal freight transportation planning: A literature review. *Eur. J. Oper. Res.* 233(1):1–15.

Stephan K, Boysen N (2011) Cross-docking. *J. Management Control* 22(1):129–137.

Tarantilis CD (2013) Adaptive multi-restart tabu search algorithm for the vehicle routing problem with cross-docking. *Optim. Lett.* 7(7):1583–1596.

Toth P, Vigo D, eds. (2002) *The Vehicle Routing Problem* (SIAM, Philadelphia).

Toth P, Vigo D, eds. (2014) *Vehicle Routing: Problems, Methods, and Applications*, 2nd ed. (MOS-SIAM, Philadelphia).

Üster H, Agrahari H (2010) An integrated load-planning problem with intermediate consolidated truckload assignments. *IIE Transactions* 42(7):490–513.

Van Belle J, Valckenaers P, Cattrysse D (2012) Cross-docking: State of the art. *Omega* 40(6):827–846.

Wen M, Larsen J, Clausen J, Cordeau J-F, Laporte G (2009) Vehicle routing with cross-docking. *J. Oper. Res. Soc.* 60(12):1708–1718.

Wieberneit N (2008) Service network design for freight transportation: A review. *OR Spectrum* 30(1):77–112.

Zäpfel G, Wasner M (2002) Planning and optimization of hub-and-spoke transportation networks of cooperative third-party logistics providers. *Internat. J. Production Econom.* 78(2):207–220.