A survey on two-echelon routing problems

R. Cuda(1)  G. Guastaroba(2)  M.G. Speranza(2)

(1) Department of Management, Economics and Quantitative Methods, via dei Caniana 2
   University of Bergamo, Bergamo, Italy
   rosario.cuda@unibg.it

(2) Department of Economics and Management, C.da S. Chiara 50
   University of Brescia, Brescia, Italy
   {guastaro, speranza}@eco.unibs.it

Abstract

The delivery of freight from its origin to its destination is often managed through one or more intermediate facilities where storing, merging and consolidation activities are performed. This type of distribution systems is commonly called multi-echelon, where each echelon refers to one level of the distribution network. Multi-echelon distribution systems are often considered by public administrations when implementing their transportation and traffic planning strategies as well as by private companies in their distribution networks. City logistics and multi-modal transportation systems are the most cited examples of multi-echelon distribution systems. Two-echelon distribution systems are a special case of multi-echelon systems where the distribution network comprises two levels. This latter type of distribution systems has inspired an ever growing body of literature in the last years. This paper provides an overview of two-echelon distribution systems where routes are present at both levels. We consider three classes of problems: the two-echelon location routing problem, the two-echelon vehicle routing problem, and the truck and trailer routing problem. For each class we provide an introduction and survey the foremost related papers appeared in the operations research literature.

Keywords: Literature Review, Two-Echelon Distribution Systems, Two-Echelon Location Routing Problems, Two-Echelon Vehicle Routing Problems, Truck and Trailer Routing Problems

1 Introduction

Transportation of freight occurs from the freight origin (e.g., the supplier or the production plant) to its destination (e.g., the wholesaler, the retailer or the final customer). A supply chain may be seen as composed by stages (also called layers or tiers). Transportation occurs between each pair of stages. Each pair of stages represents one level of the distribution network and is usually referred to as an echelon. Freight transportation is a key driver for many companies since it affects considerably both the product costs and the customer experience. Chopra and Meindl [9] highlight that distribution-related costs make up about 10.5% of the US economy and about 20% of the cost of manufacturing. It
is therefore not surprising that such topic has attracted considerable efforts from the operations research community aimed at developing effective optimization models and solution algorithms capable to provide support to decision makers.

Freight transportation can be broadly categorized into two classes according to the presence of one or more intermediate facilities. Direct shipping takes place when freight is delivered directly from its origin to its destination. Conversely, indirect shipping takes place when freight, or part of the freight, is moved through some intermediate facilities (e.g., cross-docks or distribution centers) before reaching its destination. Two-echelon distribution systems are a special case of multi-echelon systems where the network is composed of two echelons. In this case, after leaving its origin, freight is first delivered to an intermediate facility where storage, merging, consolidation or transshipment operations take place. The freight is then moved from the intermediate facility towards its destination. Given this framework, the flow of freight in one echelon must be coordinated with that in the other echelon. As a consequence, routing problems arising in two-echelon distribution systems cannot be merely decomposed into two sub-problems and then solved separately.

We define as two-echelon routing problems a class of problems that study how to optimally route freight in two-echelon distribution systems, possibly considering also location decisions.

**Area covered.** In recent years, a considerable number of papers focusing on two-echelon routing problems have been published. Some of them tackle variants of the same basic problem, while others propose different solution methods for the same problem. Two-echelon routing problems can be classified according to the type of decisions involved. We consider:

- **strategic planning decisions**: they include decisions concerning the infrastructure of the network, typically the number and the location of the facilities;

- **tactical planning decisions**: they include the routing of freight through the network and the allocation of customers to the intermediate facilities.

This survey aims at providing a classification and a systematic overview of the foremost contributions in the operations research literature on two-echelon routing problems. We survey the literature dealing with two-echelon routing problems where strategic and/or tactical planning decisions are taken into consideration. In particular, we consider the following three classes of two-echelon routing problems.

We refer to the Two-Echelon Location Routing Problems (2E-LRPs, hereafter) when the problem definition involves both strategic and tactical planning decisions, and routes are present at both echelons. Specifically, in a 2E-LRP goods available at different origins (called depots or, sometimes, platforms) have to be delivered to the respective destinations moving mandatorily through intermediate facilities called satellites. An opening cost is associated with each depot and each satellite. The depots, as well as the satellites, to be opened have to be selected from a set of possible depot (satellite) locations.

We refer to the Two-Echelon Vehicle Routing Problems (2E-VRPs, from now on) when the problem definition involves only tactical planning decisions, and the routing is present at both echelons. In a 2E-VRP the set of depot and the set of satellites to use is given, and no cost is associated with the use of any depot and any satellite.
Finally, we consider the Truck and Trailer Routing Problems (TTRPs, henceforth). In a TTRP freight transportation is managed by means of a set of trucks and trailers with the following restrictions. A subset of customers have to be served by a truck alone, whereas the remaining customers can be served either by a complete vehicle (i.e., a truck pulling a trailer) or by a truck alone. The nature of the TTRPs is different from the above two classes of problems. However, TTRPs are two-echelon routing problems since in a feasible solution a two-level route may be present with the following characteristics: the first level route is traveled by a complete vehicle, whereas the second level route, starting and ending at a vertex visited in the first level tour, is traveled by the truck alone.

**Applications.** Due to the many real-life problems that can be modeled as two-echelon distribution systems, an increasing number of examples of design and implementation of this type of distribution system appears in the literature. We mention, among other applications, city logistics, multi-modal transportation, postal and parcel delivery, milk collection, press and grocery distribution.

City logistics is probably the most frequently cited application. Crainic et al. [18] claim that “city logistics aims to reduce the nuisances associated to freight transportation in urban areas while supporting their economic and social development”. Indeed, freight transportation in urban areas is one of the main reasons of congestion, disorder, pollution emissions and noises. Implementing a two-echelon distribution system could be an effective response to these issues. In such systems each satellite corresponds to a facility located, usually, in the outskirts of the city where large trucks are allowed to arrive and where goods headed to different destinations are unloaded, sorted and consolidated. Goods are then loaded onto smaller and environment-friendly (also called eco-friendly) vehicles that are allowed to travel in the city center and can serve the final customers. Several papers cited in this survey are related to this particular application.

Although multi-modal transportation is not as cited as city logistics, it represents a relevant application of freight distribution systems involving two or more echelons. In recent years the number of intermodal logistic centers in Central and South-West European countries has increased significantly (e.g., see [28]). As an example we mention the ship-road multi-modal distribution system (e.g., see [27]) where the freight travels from the supplier to a satellite by ship (i.e., the first echelon) and then is loaded onto a truck that delivers it to its final destination (i.e., the second echelon).

**Surveys for related problems and structure of the paper.** Among the related classes of problems we mention the Location Routing Problem (LRP) and the Vehicle Routing Problem (VRP). For an overview on LRPs we refer the interested reader to the surveys by Nagy and Salhi [40] and, more recently, by Prodhon and Prins [48], whereas the paper by Laporte [33] provides a summary of the most important studies on the VRP. Two-echelon freight transportation optimization problems are analyzed in González Feliu [23] that aims at identifying the main concepts and issues. Papers dealing with the presence of intermediate facilities in distribution networks are surveyed in Guastaroba et al. [24]. In the latter survey the focus is mostly put on the role of the intermediate facilities in service network design problems and more on the service network design aspects than on the routing aspect. A partial overlapping concerns the 2E-VRPs that are covered in both papers. Multi-echelon issues that are not explicitly related to the routing of freight, and therefore not included in this survey, are considered, among others, in Pirkul and
Jayaraman [46], Tragantalerngsak et al. [53], Marín and Pelegrín [38], Crainic et al. [17]. Moreover, papers that consider the routing of freight but assume the presence of more than two echelons, and therefore are not included in this survey, are considered, among others, in Ambrosino and Scutellá [1] and Hamidi et al. [25].

<table>
<thead>
<tr>
<th>Optimization Model</th>
<th>Heuristic and Exact Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILP: Integer Linear Programming</td>
<td>ALNS: Adaptive Large Neighborhood Search</td>
</tr>
<tr>
<td>MILP: Mixed Integer Linear Programming</td>
<td>B&amp;C: Branch-and-Cut</td>
</tr>
<tr>
<td></td>
<td>B&amp;P: Branch-and-Price</td>
</tr>
<tr>
<td></td>
<td>GRASP: Greedy Randomized Adaptive Search Procedure</td>
</tr>
<tr>
<td></td>
<td>ILS: Iterated Local Search</td>
</tr>
<tr>
<td></td>
<td>SA: Simulated Annealing</td>
</tr>
<tr>
<td></td>
<td>TS: Tabu Search</td>
</tr>
<tr>
<td></td>
<td>VND: Variable Neighborhood Descent</td>
</tr>
<tr>
<td></td>
<td>VNS: Variable Neighborhood Search</td>
</tr>
</tbody>
</table>

Table 1: A summary of the abbreviations used in the paper.

The structure of the paper is as follows. Section 2 is devoted to the 2E-LRPs. The 2E-VRPs are considered in Section 3, while the TTRPs are surveyed in Section 4. Finally, Section 5 draws some concluding remarks and outline possible future research directions.

A summary of the abbreviations used in the paper can be found in Table 1.

2 Two-Echelon Location Routing Problems

In this section, we first provide a general description of the class of 2E-LRPs, then describe the Capacitated 2E-LRP, which is the simplest and most studied member of the class, and, finally, review the foremost papers tackling a problem belonging to this class.

2.1 Problem Class Definition

We consider a two-echelon distribution network composed of three disjoint sets of vertices corresponding to potential locations for the depots (i.e., the origins), potential locations for the satellites (i.e., the intermediate facilities), and the customers (i.e., the destinations), respectively. The customers are situated at known and fixed locations. On the other hand, the location of the depots and the satellites to use is not determined a priori. To keep the exposition clear, in this section we refer to potential locations for the depots (satellites) shortly as the depots (satellites). We will highlight the cases when location decisions are not involved for the depots or the satellites.
The above-mentioned distribution network can be decomposed into two echelons. The first echelon comprises the links between the depots and the satellites, and those connecting pairs of satellites. The second echelon connects the satellites and the customers, and also includes the links between pairs of customers. An opening cost is associated with each depot and each satellite. A demand is given for each customer that must be served with freight available at one or more depots. The freight has to be delivered to the customers through the satellites, compulsorily. Along the distribution network, freight transportation is performed by two different fleets of vehicles, one at each echelon. Vehicles belonging to the first echelon are referred to as primary vehicles, whereas those in the second echelon are called secondary vehicles. 2E-LRPs aim at locating a number of depots and/or satellites among candidate sites and determining a set of delivery routes, at both echelons, such that the total system cost is minimized.

In the following, we classify 2E-LRPs according to the notation introduced in Laporte [31] and later modified in Crainic et al. [19] to explicitly consider location decisions. Particularly, Laporte [31] introduces the following notation to synthetically represent multi-echelon location routing problems: $\lambda/M_1/\ldots/M_{\lambda-1}$, where $\lambda$ is the number of stages, and $M_i$ is the type of distribution mode between stages $i$ and $i+1$. If only return trips (i.e., trips to and from a single vertex) are allowed between stages $i$ and $i+1$, then $M_i = R$. If trips leaving stage $i$ may be routes (i.e., trips through several vertices), then $M_i = T$. Crainic et al. [19] suggest to mark with an overline the distribution mode identifier (i.e., $R$ or $T$) if location decisions have to be taken at stage $i$. Given this notation, we consider as 2E-LRPs $3/\overline{T}/T$, $3/T/\overline{T}$ and $3/\overline{T}/T$ problems, i.e., problems where routes may be present at both echelons, and location decisions must be taken for at least one echelon. We highlight that few authors (e.g., Nikbakhsh and Zegordi [43] and Dalfard et al. [20]) call 2E-LRP a problem where routes are not allowed on the first echelon. These papers are not covered in this survey. If location decisions have not to be taken for any echelon, the problem reduces to a 2E-VRP analyzed in Section 3 and corresponding to a $3/T/T$ problem.

### 2.2 The Capacitated 2E-LRP

This subsection is devoted to the Capacitated 2E-LRP, hereafter referred to as the 2E-CLRP, that is the basic and most studied problem among the 2E-LRPs. In the 2E-CLRP both primary and secondary vehicles are capacitated, homogeneous within the same echelon and a fixed cost of usage is paid for each vehicle routed. Both fleets of vehicles are assumed to be unlimited. A capacity limit for each depot as well as each satellite, representing the maximum amount of freight that can be handled in the facility, is also considered. Each open satellite has to be visited by exactly one primary vehicle. Similarly, each customer has to be served by exactly one secondary vehicle that starts from an open satellite (i.e., customer demand cannot be split).

The 2E-CLRP aims at finding the optimal set of location sites for the depots and the satellites as well as the optimal set of vehicle routes that satisfy the customer demands and do not violate the capacity constraints. The total cost, to be minimized, is given by the sum of the fixed costs for opening the facilities, the usage cost of the vehicles, and the routing costs. Under the notation mentioned in the previous subsection, the 2E-CLRP is a $3/\overline{T}/T$ problem.
Figure 1: An example of a feasible solution for a 2E-CLRP.

Figure 1 shows an example of a feasible solution for a 2E-CLRP. The squares represent the depots, the pentagons are the satellites, while the circles are the customers. The routes belonging to the first echelon are represented as dashed lines, whereas the routes belonging to the second echelon are depicted as solid lines.

More formally, the 2E-CLRP may be described as the following problem. Let us consider a weighted undirected graph $G = (N, E)$. Set $N = D \cup S \cup C$ is the set of vertices, where $D = \{1, \ldots, |D|\}$ represents the set of $|D|$ depots ($|A|$ denotes the cardinality of set $A$), $S = \{|D| + 1, \ldots, |D| + |S|\}$ is the set of $|S|$ satellites, and $C = \{|D| + |S| + 1, \ldots, |D| + |S| + |C|\}$ is the customer set. These three sets of vertices are pairwise disjoint. Set $E = E^1 \cup E^2$ is the set of edges $\{i, j\}$ such that edge set $E^1 = \{\{i, j\} : i < j, i, j \in D \cup S, \{i, j\} \notin D \times D\}$ includes the edges connecting the depots to the satellites, as well as those connecting pairs of satellites. Edge set $E^2 = \{\{i, j\} : i < j, i, j \in S \cup C, \{i, j\} \notin S \times S\}$ comprises the edges connecting the satellites to the customers, as well as those connecting pairs of customers.

An opening cost $f_i$ and a capacity $k_i$ are given for each depot and each satellite $i \in D \cup S$. Each customer $c \in C$ has a known and deterministic demand $d_c$ to be served by exactly one vehicle. A fleet of homogeneous primary vehicles with a capacity $Q^1$ is shared by the depots, while a set of homogeneous secondary vehicles with a capacity $Q^2$ is shared by the satellites. A fixed cost $f^1$ is paid for each primary vehicle routed, whereas for the use of each secondary vehicle a fixed cost $f^2$ is paid.

To give some further insights into the problem definition, we provide a route-based mathematical formulation that needs the introduction of the following notation. Let $R^1$ be a set of primary routes. Each primary route $r \in R^1$ starts from a depot $d \in D$, visits one or several satellites in $S$, and ends at $d$. Similarly, we denote as $R^2$ the set of secondary routes.
routes, each one starting from a satellite $s \in S$, visiting one or several customers in $C$, and ending at $s$. Note that primary routes traverse only edges in $E^1$, while secondary routes traverse only edges in $E^2$. Given a primary route $r \in \mathcal{R}^1$ and a satellite $s \in S$, let $\alpha_{rs} \in \{0,1\}$ be a binary parameter equal to 1 if satellite $s$ is visited in route $r$, and 0 otherwise. On the other hand, given a secondary route $r \in \mathcal{R}^2$ and a customer $c \in C$, let $\beta_{rc} \in \{0,1\}$ be a binary parameter equal to 1 if customer $c$ is visited in route $r$, and 0 otherwise. The cost of each route $r \in \mathcal{R}^1 \cup \mathcal{R}^2$ is $p_r$. Finally, given a secondary route $r \in \mathcal{R}^2$, let $d_r = \sum_{c \in C : c \in r} d_c \leq Q^2$ denote the total demand of customers visited. Hence, each secondary route $r \in \mathcal{R}^2$ does not violate the vehicle capacity by construction.

The route-based formulation uses the following sets of variables. The facility location binary variable $y_d$, with $d \in D$, takes value 1 if and only if depot $d$ is opened. Likewise, binary variable $y_s$, with $s \in S$, takes value 1 if and only if satellite $s$ is used in the solution. Furthermore, binary variable $x_r$, with $r \in \mathcal{R}^1 \cup \mathcal{R}^2$, takes value 1 if and only if route $r$ belongs to the solution. Finally, let $q_{rs} \geq 0$, with $r \in \mathcal{R}^1, s \in S$, be a nonnegative flow variable that gives the load of the vehicle, possibly zero, on $r$ that has to be delivered to satellite $s$.

The 2E-CLRP can be cast as the following Mixed Integer Linear Programming (MILP) model:

\[
\begin{align*}
\text{minimize} & \quad \sum_{d \in D} f_d y_d + \sum_{s \in S} f_s y_s + \sum_{r \in \mathcal{R}^1} (f^1 + p_r) x_r + \sum_{r \in \mathcal{R}^2} (f^2 + p_r) x_r \\
\text{subject to} & \quad \sum_{s \in S} q_{rs} \leq Q^1 x_r \quad r \in \mathcal{R}^1 \\
& \quad \sum_{r \in \mathcal{R}^1} \sum_{d \in D} q_{rs} \leq k_d y_d \quad d \in D \\
& \quad \sum_{r \in \mathcal{R}^1} q_{rs} \leq k_s y_s \quad s \in S \\
& \quad \sum_{r \in \mathcal{R}^1} \alpha_{rs} x_r = y_s \quad s \in S \\
& \quad \sum_{r \in \mathcal{R}^1} q_{rs} = \sum_{r \in \mathcal{R}^2} d_r x_r \quad s \in S \\
& \quad \sum_{r \in \mathcal{R}^2} \beta_{rc} x_r = 1 \quad c \in C \\
& \quad y_d \in \{0,1\} \quad d \in D \\
& \quad y_s \in \{0,1\} \quad s \in S \\
& \quad x_r \in \{0,1\} \quad r \in \mathcal{R}^1 \cup \mathcal{R}^2 \\
& \quad q_{rs} \geq 0 \quad r \in \mathcal{R}^1, s \in S.
\end{align*}
\]
Figure 2: A summary of the papers on 2E-LRPs.

The 2E-CLRP is \( \mathcal{NP} \)-hard as it generalizes other known \( \mathcal{NP} \)-hard problems, such as the Capacitated LRP (CLRP, henceforth) and the Capacitated VRP (CVRP, from now on).

2.3 Literature Review

This section is organized as follows. We first provide a broad overview of the literature on 2E-LRPs, and then give some details on each of the papers reviewed. We report on some of the early papers on 2E-LRPs, and we will concentrate on the latest advances of the related literature.

Most of the early papers on 2E-LRPs consider location decisions at only one stage,
usually the second one. Indeed, Jacobsen and Madsen [29], Madsen [37], Laporte [31] and Laporte and Nobert [32] study a $3/T/T$ problem. Only recently researchers have focused on 2E-LRPs where location decisions involve two stages. The 2E-CLRP is the most studied member of the class of 2E-LRPs. It is tackled in Boccia et al. [5], Crainic et al. [19], Contardo et al. [11], and Schwengerer et al. [52]. A $3/T/T$ problem consisting in a special case of the 2E-CLRP with a Single Depot whose position is known a priori ($2E-CLRPSD$, from now on) has been recently investigated in Nguyen et al. [41] and Nguyen et al. [42]. It is also worth highlighting that several mathematical formulations and optimization algorithms proposed in the literature are extensions of optimization models and algorithms originally proposed for related problems, especially LRP, Capacitated Facility Location Problem (CFLP) and VRP. Most of the papers propose a heuristic algorithm. To the best of our knowledge, the best performing heuristic and exact algorithms for the 2E-CLRP are the Adaptive Large Neighborhood Search (ALNS) and the Branch-and-Cut (B&C), respectively, introduced in [11].

A summary of the main characteristics of the papers included in this section is reported in Figure 2, including some details on the problem type, the optimization model introduced, the solution method proposed (heuristic and/or exact), and the main specific characteristics of the problem studied.

With the exception of the 2E-CLRP, the problems of the 2E-LRP class studied in the literature are rather different from each other. Often only one or very few papers have been published on a specific variant, and only few variants have been studied so far. On the one hand, this makes sometimes impossible to conduct a performance comparison of different solution approaches, while, on the other hand, it indicates that several opportunities for future research directions are open.

Among the first papers on a 2E-LRP, the paper by Jacobsen and Madsen [29] is worth mentioning. Their research is motivated by a real application concerning newspaper distribution in the western part of Denmark. Newspapers that are available at a printing office (i.e., the only depot) have to be delivered to some sale points (i.e., the customers) moving through some transfer points (i.e., the satellites). The delivery has to adhere to a set of constraints regarding vehicle capacities for primary vehicles, maximum duration of secondary routes and latest delivery times to sale points. Furthermore, as no special equipment is required at a satellite, opening costs are not considered for these facilities. The decisions to be made are: the location of the satellites, the primary routes, the secondary routes, and the sequencing of primary routes that gives the order in which the primary vehicles leave. The problem is a $3/T/T$ with only one depot and several side constraints. As the size of the instances involved in their case study is large (the number of customers is around 4500), the authors design three different heuristics to solve the problem. Slight modifications of two of the these heuristics, improving their performance, are proposed in Madsen [37] and applied to the same case study. The aforementioned paper by Laporte [31] provides an overview of multi-echelon location routing problems along with a list of case studies, some of which may be modeled as a 2E-LRP (among them, distribution of consumer goods, rubber collection and newspaper delivery). Laporte [31] (see also Laporte and Nobert [32] for a slightly different problem) shows that, by means of a graph transformation, an asymmetric 2E-LRP may be transformed into a Traveling Salesman Problem (TSP) with additional constraints. Suggestions on possible solution
To the best of our knowledge, the 2E-CLRP has been firstly formalized in Boccia et al. [5]. The authors introduce the problem and present a Tabu Search (TS) heuristic based on algorithms originally designed for LRP5s, namely, the nested approach proposed by Nagy and Salhi [39] and the two-phase iterative approach introduced by Tuzun and Burke [54]. The basic idea of the algorithm is to, firstly, decompose the original problem into two LRP5s and, secondly, decompose each resulting LRP into a CFLP and a Multi-Depot Vehicle Routing Problem (MDVRP). Hence, the TS algorithm consists of two main phases: a location phase, in which the number and the location of the facilities are determined, and a routing phase, in which the routing component is considered and possibly improved. A bottom-up approach is used. Specifically, a second echelon solution is firstly built and then, given that solution, a first echelon solution is computed and optimized. Computational experiments are run for small-scale (up to 4 depots, 10 satellites and 25 customers) and large-scale (up to 5 depots, 20 satellites and 200 customers) instances generated by the authors. Three MILP formulations for the 2E-CLRP are introduced in Crainic et al. [19]. The first, using three-index variables, and the third, adopting one-index variables, are inspired from classical VRP formulations, whereas the second, using two-index variables, is derived from the MDVRP literature. The authors develop an instance generator with the scope of reproducing a schematic representation of a multi-level urban area and to test the effectiveness of the proposed formulations. Computational experiments are conducted solving two of the three optimization models proposed (the third formulation is not considered in the experiments) by means of the XPRESS solver on a set of instances generated by the authors. The mathematical formulations are compared in terms of computing times, lower bounds provided and quality of the solutions. The computational results show that the three-index formulation provides better lower bounds and outperforms the two-index model on medium-scale instances.

To the best of our knowledge, the only exact method available in the literature for the 2E-CLRP is the B&C algorithm proposed in Contardo et al. [11]. In this paper, the authors propose a novel two-index vehicle flow formulation which is the base of a B&C algorithm able to solve small and medium-scale instances to optimality. The mathematical formulation is a MILP model that is strengthened by means of several valid inequalities derived from the papers on the CLRP by Belenguer et al. [4] and by Contardo et al. [10]. The authors introduce also an ALNS algorithm that outperforms the other heuristics proposed for the 2E-CLRP. Both the exact and the heuristic algorithms are based on the idea of decomposing the 2E-CLRP into two LRP5s, one at each echelon. This enables them to apply algorithms proposed for the CLRP at each echelon and then combine the partial solutions to achieve a globally feasible solution. The exact and the heuristic algorithms are tested on three sets of instances: two sets introduced in Nguyen et al. [41] for the 2E-CLRP (see below for further details), and a third set comprising 2E-CLRP instances. The latter set is composed of the large-scale instances introduced in Boccia et al. [5] (also tested in Nguyen et al. [42]) plus a set of small-scale instances generated according to the specifications described in [5]. The performance of the ALNS is compared with that of the heuristics tested in [41] and [42]. The latter heuristics are clearly outperformed by the ALNS algorithm that improved 133 best-known solutions out of 147. On the other hand, the B&C algorithm solved to optimality 75 instances out of 147. Additionally, the
comparison of the two methods shows that the lower bounds obtained by the B&C lie, on average, no further than 3.06% below the solution values found by the ALNS algorithm.

Schwengerer et al. [52] present a Variable Neighborhood Search (VNS) for the 2E-CLRP drawing on a VNS algorithm proposed in [47] for a LRP. Computational results are given for the three sets of instances tested in [11]. The effectiveness of the VNS is validated comparing its performance with that of the heuristics tested in [41], [42] and [11]. The computational results show that the proposed VNS outperforms the heuristics proposed in [41] and [42] but, on average, is outperformed by the ALNS introduced in [11].

Other members of the class of 2E-LRPs are studied by the following authors. Nguyen et al. [41] and Nguyen et al. [42] study the 2E-CLRPSD mentioned above, i.e. a special case of the 2E-CLRP where only one depot is available, and its position is known a priori. The problem can be classified as a $3/T/T$ problem. Nguyen et al. [41] formulate the 2E-CLRPSD with an Integer Linear Programming (ILP) model with two-index decision variables and present four constructive heuristics (three of them with some random elements) and a Greedy Randomized Adaptive Search Procedure (GRASP) complemented by a learning process and a path relinking procedure to solve the problem. The GRASP uses three greedy randomized heuristics to generate trial solutions and two Variable Neighborhood Descent (VND) procedures to improve them. The authors generated two sets of instances for the 2E-CLRPSD that involve up to 10 satellites and 200 customers. Computational results show that the GRASP with learning process and path relinking outperforms the other heuristics considered in their paper. It is worth highlighting that for the randomized heuristics the authors report only the best solution value obtained over some runs (10 or 5, depending on the heuristic), but no information on their average performance is provided. The same authors propose in [42] a new MILP formulation for the 2E-CLRPSD and a multi-start Iterated Local Search (ILS) algorithm with some special features. These special features include an acceptance criterion for children solution (i.e., solutions obtained applying on the incumbent one a random perturbation and, then, a local search) that can be accepted only if their gap from the best-known solution does not exceed a given threshold; two improvement procedures based on the exploration of two neighborhoods of different complexity; and a list storing recent visited solutions inspired to a TS algorithm. The resulting multi-start ILS is then reinforced by a path relinking procedure. The algorithm, along with some simpler and less performing heuristics, is tested on the benchmark instances for the 2E-CLRPSD generated in [41]. Also in this paper, the authors report only the best solution value obtained over some runs, while no information on the average performance is given. No comparison is provided among the performance of the multi-start ILS with path relinking and that of the heuristics used in [41]. We compared the best solution costs reported in the two aforementioned papers. The comparison shows that the multi-start ILS with path relinking introduced in [42] reports a slightly better average solution cost than the GRASP with learning process and path relinking proposed in [41], but it is considerably slower (computational experiments are conducted on the same hardware). Finally, even if the multi-start ILS with path relinking introduced in [42] was initially designed for the 2E-CLRPSD, in the computational experiments section the authors describe how it can be extended to the 2E-CLRP. The resulting algorithm is tested on the subset of large-scale instances introduced in [5]. Computational
results indicate that the multi-start ILS with path relinking performs better than the TS of [5]. Nevertheless, as mentioned above, it is outperformed by the ALNS introduced in [11].

A summary of the solution methods proposed in each paper on a 2E-LRP along with the maximum size solved is reported in Table 2. \( |D|_{\text{max}} \) denotes the maximum number of depots, \( |S|_{\text{max}} \) the maximum number of satellites, and \( |C|_{\text{max}} \) the maximum number of customers. Note that when in a single paper several algorithms are proposed, we report the best performing one.

### 3 Two-Echelon Vehicle Routing Problems

In this section, we first provide a general description of the class of 2E-VRPs, then describe the Capacitated 2E-VRP, which is the simplest and most studied member of the class, and, finally, review the foremost papers tackling a problem belonging to this class.

#### 3.1 Problem Class Definition

2E-VRPs can be seen as special cases of 2E-LRPs where the location of the depots and the satellites to use is given. Therefore, the definition provided in Subsection 2.1 can be modified accordingly to hold for 2E-VRPs and, to the sake of brevity, we do not describe it here. We here report only the goal of 2E-VRPs.

2E-VRPs aim at finding a set of primary and secondary routes such that the demand of all customers is satisfied, while the total system cost is minimized. According to the notation introduced in Subsection 2.1, 2E-VRPs may be classified as \( 3/T/T \) problems.

Note that in the literature on 2E-VRPs, the depots are sometimes called consolidation centers, whereas the intermediate facilities are referred to as distribution centers or, when such distribution centers are smaller than a depot and/or have only short-term inventory holding capacity, they are called satellite platforms, or, simply, satellites (see [45]). To the sake of simplicity, henceforth we refer to an intermediate facility shortly as satellite.

#### 3.2 The Capacitated 2E-VRP

This section is devoted to the *Capacitated 2E-VRP*, in the following referred to as the *2E-CVRP*, that is the basic and most studied version of the 2E-VRP. In the 2E-CVRP one (uncapacitated) depot and a set of satellites are available at given locations. A limited fleet of primary vehicles is located at the depot, while a given number of secondary vehicles
is shared by the satellites. Each satellite is associated with a given capacity defined as an upper bound on the number of secondary vehicles that can be routed from it (instead of a maximum amount of freight that it can handle as in the 2E-CLRP). Furthermore, both primary and secondary vehicles are capacitated and homogeneous within the same echelon. Direct shipments from the depot to the customers are not allowed. Each satellite can be served by one or more primary vehicles. Each primary vehicle can deliver the freight of one or more customers, as well as can serve more than one satellite in the same route. Conversely, each customer is served by exactly one secondary vehicle, i.e., split deliveries are not allowed on the second echelon. Each satellite is associated with a handling cost paid for loading/unloading operations.

The 2E-CVRP aims at finding a set of routes at both levels such that the demand of all customers is satisfied, the satellites and vehicles capacity constraints are not violated, while the total distribution cost is minimized. The total distribution cost is given by two components: the cost of primary and secondary routes used, and the handling cost at the satellites.

More formally, the 2E-CVRP can be described as the following problem. Let us consider a weighted undirected graph $G = (N, E)$. Set $N = \{0\} \cup S \cup C$ is the set of vertices, where vertex 0 represents the depot, $S = \{1, \ldots, |S|\}$ is the set of $|S|$ satellites, and $C = \{|S| + 1, \ldots, |S| + |C|\}$ is the set of customers. Set $E = E_1 \cup E_2$ is the set of edges $\{i, j\}$ such that edge set $E_1 = \{\{i, j\} : i < j, \ i, j \in \{0\} \cup S\}$ includes the edges connecting the depot to the satellites, as well as those connecting pairs of satellites. Edge set $E_2 = \{\{i, j\} : i < j, \ i, j \in S \cup C, \ \{i, j\} \notin S \times S\}$ comprises the edges connecting the satellites to the customers, as well as those connecting pairs of customers. A fleet of $\nu_1$ homogeneous and capacitated primary vehicles is located at the depot. Each primary vehicle has a capacity $Q_1$ and starts and ends its route at the depot after visiting one or more satellites. A fleet of $\nu_2$ homogeneous and capacitated secondary vehicles is available to deliver the freight to the customers. Each secondary vehicle has a capacity $Q_2$, and from each satellite $s \in S$ at most $\nu_2 s$ secondary vehicles can be routed. Each customer $c \in C$ demands $d_c$ units of freight. A handling cost $h_s$ for loading/unloading operations is paid for each unit of freight handled at satellite $s \in S$.

The route-based optimization model (1)-(11) can be adapted to the 2E-CVRP as follows. The two sets of binary variables $y_d$ and $y_s$ are removed, while the flow variables $q_{rs} \geq 0$, with $r \in R_1, s \in S$, and the binary variables $x_r \in \{0, 1\}$, with $r \in R_1 \cup R_2$, are still used. We recall that each route $r \in R_2$ is feasible by construction. Then, the 2E-CVRP can be cast as the following MILP model:

$$\text{minimize} \quad \sum_{r \in R_1 \cup R_2} p_r x_r + \sum_{s \in S} h_s \sum_{r \in R_1} q_{rs}$$

subject to

$$\sum_{r \in R_1} x_r \leq \nu_1$$

$$\sum_{r \in R_2} x_r \leq \nu_2$$
\[ \sum_{r \in R^2, s \in r} x_r \leq \nu_s^2 \quad s \in S. \quad (15) \]

The main differences of the above formulation for the 2E-CVRP compared to model (1)-(11) are that in objective function (12) the costs for opening the facilities and the fixed costs of the vehicles are removed, whereas the total cost of handling freight at satellites is added. Furthermore, the capacity constraint for the depot and the satellites are removed, as well as the constraints imposing that each satellite must be visited by exactly one primary vehicle. Additionally, constraints (13) and (14) are added to impose that the number of routes used at each echelon does not exceed the number of vehicles available. Finally, constraints (15) limit the number of secondary vehicles routed at each satellite.

The 2E-CVRP is proved to be \( \mathcal{NP} \)-hard via a reduction from the CVRP, which is a special case of the 2E-CVRP arising when just one satellite is considered (see [45]). Note that if an assignment of the customers to each satellite is given, the 2E-CVRP reduces to \((1 + |S|)\) CVRPs, i.e. one for the first echelon, and \(|S|\) for the second echelon.

![Figure 3: An example of a feasible solution for a 2E-CVRP.](image)

Figure 3 shows an example of a feasible solution for a 2E-CVRP. The square represents the depot, the pentagons are the satellites, and the circles are the customers. The routes belonging to the first echelon are represented as dashed lines, whereas the routes belonging to the second echelon are depicted as solid lines.

### 3.3 Literature Review

Almost the entire research on 2E-VRPs concentrates on its basic version, the 2E-CVRP. We are aware of only one article, i.e., Crainic et al. [18], that studies a variant of the 2E-CVRP including several additional characteristics. As for 2E-LRPs, mathematical
formulations and solution algorithms are often inspired by the literature on related problems, namely CVRP, MDVRP and Split Delivery Vehicle Routing Problem (SDVRP, see Archetti and Speranza [2] for a survey). Heuristic and exact methods are equally considered in the literature. To the best of our knowledge, the best performing heuristic for the 2E-CVRP is the ALNS introduced in Hemmelmayr et al. [26], while the best exact algorithm is proposed in Baldacci et al. [3].

It is worth highlighting that some researchers compare the effectiveness of a delivery strategy implementing a 2E-CVRP with a single-echelon strategy adopting a CVRP. Crainic et al. [13] investigate the impact of instance modifications (e.g., the number and location of the satellites as well as the customers) on the total distribution cost. The same authors study in [15] how the inclusion in the travel costs of other components than the distance (namely, fixed costs for using a link, operational and environmental costs) affects the optimal solution. Both studies indicate that the overall performance of a delivery strategy adopting a 2E-CVRP is better than the one implementing a CVRP.

A general overview of the main characteristics of the papers on the 2E-CVRP is reported in Figure 4.

The application that motivated the introduction of the 2E-CVRP is city logistics. Several studies on the 2E-CVRP cite the paper by Crainic et al. [18] as the one that introduced the first formal definition of a 2E-CVRP, even if the term 2E-CVRP appeared later in the literature. This may also explain why in the 2E-CVRP the capacity restriction for each satellite is expressed in terms of a maximum number of secondary vehicles that can be routed from it. Indeed, in city logistics (see [18] for further details), satellites can usually accommodate simultaneously a maximum number of urban (i.e., primary)
vehicles, as well as a maximum number of city freighters (i.e., secondary vehicles). The problem tackled in Crainic et al. [18] is a 2E-CVRP with time-dependent, synchronized, multi-depot, multi-product, heterogeneous fleets (on each echelon), and time windows. The authors provide an ILP formulation of the problem and design solution methods, but do not report any computational experiment.

The term 2E-CVRP was introduced in Perboli et al. [45] where a formal definition of the problem is provided. The authors propose a MILP formulation along with two families of valid inequalities and two matheuristics for the 2E-CVRP. The optimization model is derived from the literature on multi-commodity network design problems and uses the flow of freight on each arc as main decision variable. However, Jepsen et al. [30] point out that this mathematical formulation may not provide correct upper bounds when more than two satellites are selected in the solution. In [45], two families of valid inequalities are introduced. The first family of valid inequalities is derived from the subtour elimination constraints for the TSP, while the second family is based on the flow decision variables. Matheuristics, sometimes also called math-based heuristics, are optimization algorithms that combine elements of mathematical programming with elements of metaheuristics. The matheuristics proposed in [45] are based on information retrieved from the optimal solution of the linear relaxation of the proposed model. The authors also design a B&C algorithm to solve the 2E-CVRP where the two matheuristics are applied at the root node of the B&C tree only, and the best integer solution found is the initial solution of the algorithm. Then, the proposed valid inequalities are introduced when violated. Computational experiments are performed for four sets of instances: three sets are generated from benchmark instances for the CVRP, while the fourth is composed of a subset of the instances generated by Crainic et al. [13]. Crainic et al. [12] design two heuristics for the 2E-CVRP. Both algorithms are based on a two-phase approach where the first and the second echelon routing problems are separated and solved sequentially. In the first phase, a feasible solution for the second echelon routing problem is computed giving an assignment of customers to satellites. The two heuristics mainly differ in the approach used to tackle this phase: one decomposes the second echelon routing problem into a set of independent CVRPs, while the other treats it as a MDVRP. Given the customer-to-satellite assignment configuration, a feasible solution for the first echelon routing problem is computed solving a CVRP in which each satellite is associated with a demand equal to the sum of the demands of the customers assigned. The solution resulting from the first phase is possibly improved in the second phase by three improvement procedures focusing directly on the routes. The performance of the heuristics is assessed using the solutions found by a preliminary version of the exact method introduced in [45] for small-scale instances. The authors also generated three sets of instances comprising up to 5 satellites and 150 customers to analyze the impact of several instance modifications on algorithmic efficiency and solution quality. The detailed computational results for the latter sets of instances are not reported. Perboli et al. [44] propose several valid inequalities derived from the CVRP literature to strengthen the flow-based formulation introduced in [45]. Additionally, the presence of the network flows in the mathematical formulation allows the authors to define some valid inequalities based on the interaction between routing and arc activation variables. Other classes of valid inequalities are derived from considering connectivity and feasibility property of any feasible solution of routing problems. To asses
the proposed inequalities the authors implement a B&C. Computational results are given for a subset of the instances tested in a preliminary version of [45]. It is worth mentioning that Perboli et al. [45] claim that, after performing some preliminary experiments, they verified that the improvement after the introduction of these cuts was quite marginal compared to the additional computational effort. The idea of separating the first and the second level routing problems is also used in Crainic et al. [14], where a family of multi-start heuristics in which the two sub-problems are sequentially solved is proposed. The algorithms begin by assigning each customer to a satellite according to a distance-based greedy rule. Then, an initial solution is computed by solving the resulting first and second level CVRPs. Subsequently, a local search algorithm based on changing one customer-to-satellite assignment at a time is applied attempting to improve the initial solution. Finally, a multi-start procedure is run until a maximum number of iterations has been performed in order to avoid being trapped in local optima. Computational experiments are performed for a subset of the instances tested in [45], and those solved in [13]. The performance of the proposed heuristics solving these instances is compared with that of a preliminary version of the matheuristics introduced in [45] and validated using the lower bounds reported in [44].

The current state-of-the-art heuristic for the 2E-CVRP is the ALNS algorithm introduced by Hemmelmayr et al. [26]. The main idea of the ALNS algorithm is to remove, at each iteration, a subset of customers from the current solution by means of a destroy operator and, then, to re-insert the customers in other positions using a repair operator. Each operator is associated with a score and is selected randomly from a probability distribution function built on its past success. In other words, an operator that has found several improving solutions has a higher score than other operators, and thus a higher probability to be chosen. The destroy operators proposed in the paper are divided into two classes: those that change the configuration of the current solution by closing or opening a satellite, and those that affect a more restricted area of the search space, for instance removing a small number of customers and keeping the current satellite configuration unchanged. The destroy operators of the first class are used whenever a given number of iterations have been performed without any improvement. Every time that one of these operators is used, a local search is performed on the new solution. It is worth pointing out that the search is not restricted to the feasible solutions only. Indeed, violations of the constraints on the vehicle capacity and the number of vehicles available are allowed. A weighted penalty term is included in the objective function to consider those violations. Computational results are given for three sets of benchmark instances for the 2E-CVRP and a new set of large-scale instances (up to 10 satellites and 200 customers) obtained adapting LRP instances to the 2E-CVRP. In particular, the benchmark instances for the 2E-CVRP are the two sets of largest-scale test problems generated from benchmark instances for the CVRP by Perboli et al. [45], and the whole set of instances tested in [13]. Computational results for these sets of benchmark instances show that the proposed ALNS heuristic improves several best-known solutions previously published in [14] and [45].

Crainic et al. [16] propose a hybrid heuristic to solve the 2E-CVRP that combines a GRASP algorithm with a path relinking procedure. Similar to other approaches previously proposed by the same authors (see [12] and [14]), the 2E-CVRP is tackled by decomposing
Table 3: A summary of solution algorithms and maximum size instances in each paper on 2E-VRPs.

| Reference          | Solution Algorithm            | |S|_{max} | |S'|_{max} |
|--------------------|--------------------------------|--------|--------|
| Perboli et al. [45]| B&C, Matheuristics             | 5      | 50     |
| Crainic et al. [12]| Two-Phase Heuristics           | 5      | 150    |
| Perboli et al. [44]| B&C                           | 5      | 50     |
| Crainic et al. [14]| Multi-Start Heuristics         | 5      | 50     |
| Hemmelmayr et al. [26]| ALNS                            | 10     | 200    |
| Crainic et al. [16]| GRASP with Path Relinking      | 5      | 50     |
| Jepsen et al. [30]| B&C                           | 5      | 50     |
| Baldacci et al. [3]  | Problem Specific Exact Algorithm | 6      | 100    |

The best exact method currently available for the 2E-CVRP is introduced in Baldacci et al. [3]. The authors propose a new ILP formulation that is used to derive both continuous and integer relaxations. They present a bounding procedure based on dynamic programming, a dual ascent method, and an exact algorithm that decomposes the 2E-CVRP into a limited set of MDVRPs with side constraints. Then, the optimal solution for the 2E-CVRP is obtained by solving the generated set of MDVRPs. The exact algorithm is tested on 207 instances, taken both from the literature (i.e., the set of benchmark instances solved in [30] and a subset of the large-scale instances generated by Hemmelmayr et al. [26] from LRP test problems) and newly generated, with up to 6 satellites
and 100 customers. The exact method is compared, whenever possible, with the B&C algorithms introduced in [45] and in [30]. The computational results show that the latter exact algorithms are outperformed by the exact method designed by Baldacci et al. [3] in terms of both size of the instances solved, number of problems solved to optimality and computing times.

A summary of the solution methods proposed in each paper on a 2E-VRP along with the maximum size instance solved is reported in Table 3.

4 Truck and Trailer Routing Problems

In this section, we first provide a general description of the class of TTRPs, then describe the Capacitated TTRP, which is the simplest and most studied member of the class, and, finally, review the papers.

4.1 Problem Class Definition

In the TTRP trucks or trucks pulling a trailer have to visit a set of customers with the following restrictions. The set of customers is partitioned into two subsets: one subset comprises the customers that must be served by a truck alone (without any trailer), referred to as truck customers, while the other subset includes the customers that can be served either by a truck pulling a trailer or by a truck alone, called vehicle customers. The rationale for partitioning the customer set in two subsets is related to the presence of real-life logistic constraints. Indeed, some customers may be located in inaccessible areas for a truck pulling a trailer but can be reached by the truck alone. Some examples are customers located in the countryside, in mountain areas or in city centers where it may be forbidden to drive large vehicles. Three types of routes can be identified in a feasible solution of a TTRP: pure truck routes, pure vehicle routes, and complete vehicle routes. A pure truck route visits truck and/or vehicle customers by a truck alone. A pure vehicle route visits vehicle customers by a complete vehicle (i.e., a truck pulling a trailer) and without any sub-tour. Finally, a complete vehicle route consists of a main tour, starting and ending at the depot and traveled by a complete vehicle, and one or more sub-tours, traveled by a truck alone. Each sub-tour starts and ends at a given location visited in the main tour, called the root of the sub-tour, where the trailer is temporarily parked. Specifically, at a location in the main tour the trailer is detached from the truck. Then, the truck alone serves some truck customers and returns to the location where the trailer is parked. The trailer is attached to the truck that can continue its main tour. TTRPs aim at determining the optimal set of vehicles routes such that each customer is visited by a compatible vehicle, while the total cost of the system is minimized.

The nature of the TTRP is that of a two-echelon routing problem since in a complete vehicle route the first level routing is represented by the main tour traveled by the complete vehicle among the vehicle customers, whereas the second level routing is represented by the sub-tours traveled by the truck alone starting and ending at a given location visited in the main tour.
4.2 The Capacitated TTRP

This section is devoted to the Capacitated TTRP, from now on referred to as the CTTRP, which is the basic and most studied version of the TTRPs. In the CTTRP one (un capacitated) depot is available at a given location, where the freight originates. A limited fleet of trucks and trailers is located at the depot. All trucks, as well as all trailers, are capacitated and homogeneous. Each customer has a given demand that has to be served by an appropriate vehicle. Trailers can be parked only at vehicle customers locations, where loads can be transferred between the truck and the trailer. The CTTRP aims at finding an optimal set of vehicles routes serving each customer by a compatible vehicle, while minimizing the total routing cost, respecting the capacity of the allocated vehicles and using a number of trucks and trailers not larger than the number of vehicles available.

Note that in the literature the CTTRP defined above is simply called TTRP. However, we prefer to adopt a slightly different terminology in order to clearly distinguish the class of problems (TTRPs), from the name of a member of the class (CTTRP).

More formally, the CTTRP may be described as the following problem. We consider a weighted undirected graph $G = (N,E)$, where $N = \{0\} \cup C$ is the set of vertices and $E$ is the set of edges $\{i,j\}$, for $i,j \in N$, $i < j$. The depot corresponds to vertex 0, whereas $C = \{1,2,\ldots,|C|\}$ is the set of customers. Moreover, set $C$ is partitioned as follows: set $C^K$ comprises the truck customers, while set $C^L$ contains the vehicle customers. Each customer $c \in C$ has a known and deterministic demand $d_c$ that must be served by exactly one compatible vehicle. A fleet of $\nu^K$ trucks and $\nu^L$ trailers, with $\nu^L \leq \nu^K$, is available at the depot to serve the customers. Each truck alone has a capacity $Q^K$, whereas the capacity of each trailer is $Q^L$. Hence, a complete vehicle has a capacity equal to $Q^K + Q^L$, i.e., the trailer can be interpreted as a “mobile depot” that increases the capacity of a truck.

Let $R^1$ be the set of pure truck routes, $R^2$ the set of pure vehicle routes, and $R^3$ the set of complete vehicle routes. Each pure truck route $r \in R^1$ is traveled by a truck alone, starts from the depot, visits one or several customers in $C$, and returns to the depot. Each pure vehicle route $r \in R^2$ is traveled by a complete vehicle, starts from the depot, visits one or several customers in $C^L$, and returns to the depot. Finally, each complete vehicle route $r \in R^3$ involves a first-echelon route, i.e., the main tour, and at least one second-echelon route, i.e., a sub-tour. The main tour is traveled by a complete vehicle, starts from the depot, visits one or several customers in $C^L$, and returns to the depot. Each sub-tour is traveled by the truck alone, starts from one of the customers in $C^L$ visited in the main tour, say $i \in C^L$, visits one or several customers in $C^K$ and possibly some customers in $C^L$, and returns to customer $i$. Furthermore, given a pure truck route $r \in R^1$ and a customer $c \in C$, let $\alpha_{rc} \in \{0,1\}$ be a binary parameter that takes value 1 if customer $c$ is visited in route $r$, and 0 otherwise. Likewise, given a pure vehicle route $r \in R^2$ and a customer $c \in C^L$, let $\beta_{rc} \in \{0,1\}$ be a binary parameter equal to 1 if customer $c$ is visited in route $r$, and 0 otherwise. Finally, given a complete vehicle route $r \in R^3$ and a customer $c \in C$, let $\gamma_{rc} \in \{0,1\}$ be a binary parameter that takes value 1 if customer $c$ is visited in route $r$, and 0 otherwise. The cost of each route $r \in R^1 \cup R^2 \cup R^3$ is $p_r$. Finally, we define the binary variable $x_r$, with $r \in R^1 \cup R^2 \cup R^3$, that takes value 1 if and only if route $r$ is selected in the solution. The CTTRP can be cast as the following route-based ILP formulation:
minimize $\sum_{r \in R^1 \cup R^2 \cup R^3} p_r x_r$ \hspace{1cm} (16)

subject to

$\sum_{r \in R^1} \alpha_{rc} x_r + \sum_{r \in R^3} \gamma_{rc} x_r = 1 \quad c \in C^K$ \hspace{1cm} (17)

$\sum_{r \in R^1} \alpha_{rc} x_r + \sum_{r \in R^2} \beta_{rc} x_r + \sum_{r \in R^3} \gamma_{rc} x_r = 1 \quad c \in C^L$ \hspace{1cm} (18)

$\sum_{r \in R^1 \cup R^2 \cup R^3} x_r \leq \nu^K$ \hspace{1cm} (19)

$\sum_{r \in R^2 \cup R^3} x_r \leq \nu^L$ \hspace{1cm} (20)

$x_r \in \{0, 1\} \quad r \in R^1 \cup R^2 \cup R^3$. \hspace{1cm} (21)

Objective function (16) minimizes the total routing cost. Constraints (17) ensure that each truck customer is visited exactly once, either in a truck route or in a complete vehicle route, whereas constraints (18) assure that each vehicle customer is visited exactly once. Constraints (19) and (20) impose an upper bound on the number of trucks and trailers used, respectively. Constraints (21) define the decision variables.

Figure 5: An example of a feasible solution for a CTTRP.
As the CTTRP can be seen as an extension of the CVRP, the CTTRP is an $\mathcal{NP}$-hard problem. Indeed, the CTTRP reduces to the CVRP if there are only truck customers, i.e., $C^L = \emptyset$ and $C = C^K$ (see Chao [8]). If only vehicle customers are present, i.e., $C^K = \emptyset$ and $C = C^L$, the problem may still be solved as a CVRP (with heterogeneous fleet of vehicles if $\nu^L < \nu^K$), as there is no need for detaching the trailers (see [51]).

In Figure 5 an example of a CTTRP feasible solution is depicted. The square represents the depot, the octagons are the truck customers, while the circles are the vehicle customers. The edges traveled by a complete vehicle are represented as dashed lines, whereas those traveled by a truck alone are represented as solid lines.

### 4.3 Literature Review

The CTTRP is the most studied member of the class of TTRPs. It is tackled in Chao [8], Scheurer [51], Caramia and Guerriero [6], Lin et al. [34], Villegas et al. [56], and Villegas et al. [57]. Limited research has been carried on variants of the CTTRP. The only variant that received some attention is the CTTRP with time windows studied in Lin et al. [36] and in Derigs et al. [21]. Some authors consider a generalization of the CTTRP consisting in introducing the option of detaching the trailer from a truck in locations that do not necessarily correspond to a customer site (see Villegas et al. [55] and Drexl [22]). Specific variants have been tackled by some authors motivated by real-life applications (e.g., see Semet and Taillard [50], Caramia and Guerriero [7] and Drexl [22]). As far as the solution method is considered, almost all of the papers propose a heuristic algorithm. To the best of our knowledge, the best performing heuristic for the CTTRP is the matheuristic introduced in [57]. On the other hand, we are aware of only the contribution by Drexl [22] where an exact method is proposed for a generalization of the CTTRP.

A general overview of the main characteristics of the papers on the TTRPs is reported in Figure 6. The CTTRP is the most studied problem, and some specific variants have been investigated. As far as the solution methods are considered, almost all the research has concentrated on designing heuristic methods. Hence, a promising open research direction is the definition of exact solution methods. To this aim, we believe that the ILP model (16)-(21) might act as a starting point for the design of exact methods based, for instance, on column generation.

To the best of our knowledge, the first paper where a problem belonging to the class of TTRPs has been tackled is due to Semet and Taillard [50]. Their study stems from a practical application faced by a Swiss company that must deliver goods to grocery stores located in Switzerland. To perform the deliveries, the company owns a heterogeneous fleet of vehicles: each vehicle is characterized by a capacity (defined both in terms of a maximum volume and weight that can be transported) and a cost per kilometer traveled. Each truck can pull a subset (possibly empty) of the available trailers. Additionally, each truck customer can be served only by a subset of the trucks available. Finally, each customer must be visited by a compatible vehicle within a given time window. A TS algorithm, along with some strategies to speed up computing times, is proposed to solve the problem. A simplified version of the TTRP introduced in [50] has been studied in Semet [49], where the available trucks are heterogeneous, and, in any feasible solution, all of them are used. An assignment first-route second heuristic is proposed to solve the problem. In the first phase, trailers are assigned to trucks and customers are assigned to
Figure 6: A summary of the papers on TTRPs.
each truck and complete vehicle. Given these assignments, routes are constructed in the second phase by solving either a TSP (if a pure truck or a pure vehicle route has to be determined) or a TSP under accessibility constraints (if a complete vehicle route has to be defined) by means of appropriate algorithms available in the literature.

The term truck and trailer routing problem appeared the first time in Chao [8] where the CTTRP is tackled and the following heuristic proposed. The algorithm consists of a procedure computing an initial feasible solution for the CTTRP followed by an improvement phase based on a TS heuristic. The procedure computing the initial solution is based on the solution of a relaxed generalized assignment problem to assign one route type (i.e., either pure truck, or pure vehicle, or complete vehicle routes) to each customer. The resulting three types of routes are treated as TSP routes that are constructed by means of a cheapest insertion heuristic. Computational results are given for 21 instances adapted from the CVRP literature comprising up to 17 trucks, 9 trailers and a total of 199 customers, partitioned in different ways among truck customers and vehicle customers. This set of instances has become the test bed for several of the following papers on the CTTRP and, to the sake of brevity, it is hereafter referred to as the Chao instances. Scheuerer [51] designs two constructive heuristics and a TS algorithm to solve the CTTRP. The first constructive heuristic, called T-Cluster, is a cluster-based sequential insertion procedure where routes are constructed one-by-one up to full vehicle utilization. The second constructive heuristic, called T-Sweep, is based on the classical sweep algorithm, i.e., feasible routes are constructed by rotating a ray centered at the depot and customers are gradually added to the current route. The TS algorithm starts its search from the best solution found by the T-Cluster heuristic, is allowed to visit intermediate infeasible solutions, and uses random sampling to reduce the number of moves to evaluate at each iteration. Computational results for the Chao instances show that the TS proposed in [51] outperforms the algorithm designed by Chao [8]. A Simulated Annealing (SA) algorithm to solve the CTTRP is introduced in Lin et al. [34]. The authors propose a standard SA procedure with a random neighborhood structure that features three types of moves, namely insertion, swap, and change of vehicle type used for the vehicle customers (e.g., change from the truck alone to the complete vehicle). Computational experiments are given for the Chao instances and a comparison of the solutions obtained by the SA algorithm with those reported in [8] and [51] is provided. The SA algorithm clearly outperforms the TS heuristic introduced in [8], and finds slightly better solutions than the TS method designed in [51]. Caramia and Guerriero [6] propose an approach based on the solution of two mathematical formulations and a local search procedure. The two formulations, solved sequentially, model two different sub-problems: the first one aims at minimizing the number of vehicles used to serve the customers, while the second one aims at minimizing the total route length given the set of customers assigned to each vehicle. A local search procedure is used to recover feasibility if the solution found solving the latter two sub-problems violates any constraint. The authors provide a comparison of the performance of the proposed heuristic with the algorithms introduced in [8] and [51]. Computational results on the Chao instances show that the algorithm designed in [6] produces solutions of comparable quality than those reported in [51]. Villegas et al. [56] propose a route-first, cluster-second procedure embedded within a hybrid metaheuristic based on GRASP, VNS and path relinking (GRASP/VNS, hereafter) to solve the CTTRP. The VNS is used as
an improvement mechanism for feasible solutions, and as a repair operator for infeasible solutions. Path relinking is tested as a post-optimization procedure, as an intensification mechanism, and within evolutionary path relinking. Computational experiments are carried out on the Chao instances and a comparison with the solutions reported in [8], [51], [34], and [6] is provided. The comparison shows that the GRASP/VNS with evolutionary path relinking performs on average better than the other algorithms considered. The results also show that the GRASP/VNS with path relinking used as a post-optimization mechanism is, on average, significantly faster than the GRASP/VNS with evolutionary path relinking achieving slightly worse solutions.

The current state-of-the-art heuristic for the CTTRP is the matheuristic proposed in Villegas et al. [57]. The authors combine a set partitioning formulation for the CTTRP with a hybrid metaheuristic. The resulting matheuristic follows an iterative two-phase approach. In the first phase, a GRASP algorithm is first used to populate a pool of routes with a subset of all possible routes, and an ILS procedure is then used to improve the quality of the routes found. Then, in the second phase, the set partitioning formulation is solved on the subset of routes previously identified to obtain a feasible solution. Two variants are proposed: the first variant considers a large pool of routes, while the second one contemplates a small pool of routes. The matheuristic is tested on the Chao instances and those originally proposed for the relaxation of the CTTRP studied in Lin et al. [35] (see below for further details) after including the data about the fleet size available. The first variant finds considerably better solutions than the other methods previously available in the literature, in comparable or shorter computing times. The second variant spends one third of the running time taken by the first variant and finds slightly worse solutions.

Other members of the class of TTRPs are studied by the following authors. The CTTRP with Time Windows (CTTRPTW) is introduced in Lin et al. [36]. Drawing on the SA algorithm proposed in [34] for the CTTRP, the authors design a SA heuristic to solve the CTTRPTW that can also be used to tackle the CVRP with Time Windows (CVRPTW). Two sets of computational experiments are performed. In the first set, some benchmark instances for the CVRPTW are solved with the proposed SA heuristic to assess its effectiveness in comparison with the best-known solutions available in the literature. In the second set of experiments, some benchmark instances for the CVRPTW are used to generate 54 instances for the CTTRPTW that are solved with the proposed SA approach.

Derigs et al. [21] study the following variants of the TTRP: the CTTRPTW, and the CTTRP with and without the option of load transfer between the truck and the trailer. Regarding the option of load transfer, the authors point out that if it is forbidden, the demands of all customers served on a sub-tour must be loaded onto the truck at the depot and, on the other hand, only the demands of vehicle customers visited in the main tour can be loaded onto the trailer. The authors present a hybrid algorithm which combines local search and large neighborhood search moves guided by two simple metaheuristic control strategies. Computational experiments are given for the standard CTTRP on the Chao instances. The results show that the performance of the proposed heuristic is comparable to that of the hybrid heuristic designed in [56]. Some remarks are also given concerning the solutions obtained if load transfers between the truck and the trailer are forbidden. Further computational experiments are provided on the benchmark instances for the CTTRPTW introduced in [36]. The computational results show that the method
proposed in [21] outperforms the algorithm defined in [36]. It is worth highlighting that the authors claim that the benchmark instances for the CTTRPTW available in the literature, which have been constructed from benchmark instances for the CVRPTW by clustering the customers into the two classes of truck and vehicle customers, are not appropriate since some instances do not offer the opportunity to perform sub-tours, therefore being almost equivalent to CVRPTW instances.

Some authors studied TTRPs where it is possible to detach the trailer and transfer loads between the truck and the trailer in locations that do not necessarily correspond to a customer site. These locations are referred to as trailer points, and are sometimes called transshipment locations or satellite depots. Villegas et al. [55] study a TTRP with trailer points where only one complete vehicle is available to serve the customers. Additionally, the customer set comprises only truck customers, i.e., \( C = C^K \) and \( C^L = \emptyset \), and, as a consequence, the trailer points do not correspond to customer locations. The authors call the problem the Single TTRP with Satellite Depot (STTRPSD) and mention that milk collection and postal delivery with park-and-loop routes can be modeled as a STTRPSD. Indeed, in milk collection customers (i.e., farms) are usually visited by a single tanker with a removable tank trailer. Trailer points are in general parking locations located on main roads, whereas farms are often located in areas that can be reached only by driving on narrow streets that are inaccessible for the vehicle with the trailer. In postal delivery with park-and-loop routes the postman drives a vehicle from the postal facility to a parking location, loads a sack, delivers mails by walking the streets and then returns to the vehicle. The authors propose an ILP formulation and two metaheuristics for the STTRPSD. The two metaheuristics are a GRASP algorithm hybridized with a VND procedure (hybrid GRASP/VND, hereafter) and a multi-start evolutionary local search, respectively. Both methods are tested on 32 randomly generated instances comprising up to 200 customers and 20 trailer points. The computational experiments show that the multi-start evolutionary local search is more accurate, faster, and scales better as the number of customers increases than the hybrid GRASP/VND algorithm. Drexl [22] introduces the Generalized TTRP (GTTRP) as a generalization of the CTTRP. Trailers can be parked and load can be transferred between truck and trailer at vehicle customers locations, as in the CTTRP, but also at trailer points, as in the STTRPSD. Additionally, in the GTTRP fixed costs for using the vehicles are considered (they are added to the cost of the links emanating from the depot), all customers and trailer points have hard time windows associated with them, and the fleet is composed of heterogeneous vehicles. The author proposes a MILP formulation for the GTTRP based on binary arc-flow variables, and design a Branch-and-Price (B&P) algorithm, based on a path-flow reformulation of the MILP model, as well as some heuristic variants of the exact algorithm. Computational experiments are run for randomly generated instances structured to resemble real-world situations (the motivating application is also here milk collection) and on the Chao instances for the CTTRP. The computational results show that the instances of realistic structure and size can be solved in short computing times with high solution quality with a heuristic column generation approach. Conversely, the performance on the Chao instances is not competitive with other approaches previously designed to solve the problem the CTTRP.

Lin et al. [35] study a Relaxation of the CTTRP (henceforth referred to as the RTTRP) where the limited fleet constraints are removed (i.e., \( \nu^K = \nu^L = +\infty \)) in order to evaluate
their impact on the solution cost. The RTTRP is solved by means of a SA algorithm, and the heuristic solutions it finds for the RTTRP are compared with the solutions for the CTTRP available in [51] and [34] for the Chao instances. To gain some further insights into the differences between RTTRP solutions and TTRP solutions, the authors generated 36 new instances adapting benchmark instances taken from the VRP literature. For this second set of instances, the RTTRP solutions obtained by the proposed SA heuristic are compared with the CTTRP solutions obtained running the SA heuristic described in Lin et al. [34]. The computational results show that significant reduction of the total routing cost, especially on the second set of instances, can be achieved increasing the number of trailers available.

Caramia and Guerriero [7] study a CTTRP with incompatible products and heterogeneous fleets. Their study stems from a problem faced by an Italian dairy company that collects milk from farmers with truck and trailers of different capacities. The company must collect different milk types. To this aim, the capacity of each truck and each trailer is organized in compartments, every one characterized by its capacity. When a compartment has been filled with milk of a particular type, that compartment cannot be used to store any other type of milk until it has been cleaned at the end of the day. Further specific characteristics of the problem are the presence of a constraint on route duration that cannot exceed the work shift, the loads that cannot be transferred between the truck and the trailer, and the fact that multiple vehicles can collect milk from a particular farmer. The solution method proposed is based on the heuristic algorithm designed by the same authors in [6] for the standard CTTRP.

A summary of the solution methods proposed in each paper on a TTRP along with the maximum size instance solved is reported in Table 4.

| Reference                    | Solution Algorithm                  | $|C^k|_{\text{max}}$ | $|C^L|_{\text{max}}$ |
|------------------------------|-------------------------------------|---------------------|---------------------|
| Semet and Taillard [50]      | TS                                  | 36                  | 9                   |
| Semet [49]                   | Two-Phase Heuristic                 | 70                  | 30                  |
| Chao [8]                     | TS                                  | 50                  | 50                  |
|                              |                                     | 99                  | 100                 |
|                              |                                     | 149                 | 50                  |
| Scheuerer [51]               | Constructive Heuristics, TS         |                     | Chao instances      |
| Lin et al. [34]              | SA                                  |                     | Chao instances      |
| Caramia and Guerriero [6]    | Matheuristic with Local Search      |                     | Chao instances      |
| Villegas et al. [56]         | Hybrid GRASP/VNS with (Evolutionary) Path Relinking | | Chao instances |
| Villegas et al. [57]         | Matheuristic with GRASP and ILS     |                     | Chao instances and inst. in [35] |
| Lin et al. [36]              | SA                                  | 50                  | 150                 |
|                              |                                     | 100                 | 100                 |
|                              |                                     | 150                 | 50                  |
| Derigs et al. [21]           | Hybrid Local Search and Large Neighborhood Search | | Chao instances and inst. in [36] |
| Villegas et al. [55]         | Hybrid GRASP/VND, Multi-Start Evolutionary Local Search | | 200                  |
| Drexl [22]                   | B&P, Heuristic Variants             |                     | Chao instances      |
| Lin et al. [35]              | SA                                  | 37                  | 113                 |
|                              |                                     | 75                  | 75                  |
|                              |                                     | 113                 | 37                  |
| Caramia and Guerriero [7]    | Matheuristic with Local Search      | 26                  | 14                  |

Table 4: A summary of solution algorithms and maximum size instances in each paper on TTRPs.
5 Conclusions and Open Research Areas

In this survey we have provided an extensive overview of the operations research literature on two-echelon routing problems, that is a class of problems that study the optimal routing of freight in two-echelon distribution systems. This research area is attracting an increasing attention both from practitioners and academics due to the relevant real-life applications (among others, city logistics and multi-modal transportation) and the intellectual challenges that their study poses. We have classified the literature on two-echelon routing problems into three classes: the two-echelon location routing problems, the two-echelon vehicle routing problems and the truck and trailer routing problems. For each class of problems we have provided a general description, formally introduced a basic problem of the class, identified the main variants studied in the literature and reviewed the foremost exact and heuristic solution approaches proposed.

Despite its importance in practical applications, this research area is still relatively unexplored. Most of the contributions cited in this survey focus on the basic problems and propose heuristic solution approaches. Hence, promising research directions are, on one side, the study of variants of the basic problem that include more realistic features and, on the other side, the design of exact methods. Particularly, only few papers deal with variants of the basic problems where time is taken into consideration. For instance, only few authors have studied variants of the problems with time windows, or variants where a limit on the total duration of each route is given. Other variants that it is worthwhile to investigate are the two-echelon location routing and the two-echelon vehicle routing problems with satellite synchronization constraints. In these variants, time constraints are introduced on the arrival of the vehicles at a satellite such that once a primary vehicle has unloaded its freight, it is immediately loaded onto a secondary vehicle. Such synchronization constraints model the delivery of perishable goods (e.g., frozen food) where the freight has to be moved, without interruption, from one vehicle to another. Another interesting variant of the two-echelon vehicle routing problem is its multi-depot version, i.e., more than one depot is available to serve the satellites in the first echelon. As far as the truck and trailer routing problem is concerned, on the one hand more realistic versions of the problem should consider a cost (and/or the time spent) for transferring load between a truck and its trailer, as well as to attach and detach the trailer. On the other hand, the development of exact methods is, at this moment, very limited and is also an interesting research area. Finally, the study of dynamic versions of the problems and stochastic models also seem to be promising research directions.

Acknowledgments. The authors would like to express their appreciation for the valuable comments made by the guest editors and two anonymous reviewers, which helped us to improve the quality of the paper.

References


