# Integrating Public Transport in Sustainable Last-Mile Delivery: Column Generation Approaches

Diego Delle Donne delledonne@essec.edu[1], Alberto Santini alberto.santini@upf.edu[1,2,3,4], and Claudia Archetti archetti@essec.edu[1]

[1] Department of Information Systems, Decision Sciences and Statistics, ESSEC Business School, 3 Avenue Bernard Hirsch, 95021, Cergy, France
[2] Institute of Advanced Studies, CY Cergy Paris Université, 1 Rue Descartes, 95000, Neuville-sur-Oise, France
[3] Department of Economics and Business, Universitat Pompeu Fabra, Carrer Ramon Trias-Fargas, 25–27, 08005, Barcelona, Spain
[4] Data Science Centre, Barcelona School of Economics, Carrer Ramon Trias-Fargas, 25–27, 08005, Barcelona, Spain

11th January 2024

**Abstract**

We tackle the problem of coordinating a three-echelon last-mile delivery system. In the first echelon, trucks transport parcels from distribution centres outside the city to public transport stops. In the second echelon, the parcels move on public transport and reach the city centre. In the third echelon, zero-emission vehicles pick up the parcels at public transport stops and deliver them to customers. We introduce two extended formulations for this problem. The first has two exponential sets of variables, while the second has one. We propose column generation algorithms and compare several methods to solve the pricing problems on specially constructed graphs. We also devise dual bounds, which we can compute even when the graphs are so large that not a single round of pricing completes within the time limit. Compared to previous formulations, our models find 17 new best known solutions out of an existing dataset of 24 from the literature.

**Keywords:**last-mile delivery, logistics, column generation, routing problems, resource-constrained shortest path problems

## 1 Introduction

The e-commerce boom and its growth projections (Alfonso et al. 2021) raise new challenges as retailers and couriers innovate their supply chains to keep up with demand. Last-mile delivery (LMD), the segment of the supply chain which starts at the last distribution centre and ends at the customer's doorstep, is particularly affected. Its nature changed when retailers stopped delivering to stores and started delivering directly to consumers: couriers now handle many small parcels instead of fewer, larger shipments; they deliver during tight time windows when

Figure 1: Two pilot projects launched in 2021 in Germany and England to use spare capacity on public transport vehicles. On the left: operators loading parcels on a tram operated by AVG in Karlsruhe, Germany. In 2021, the Mobility Cluster of Karlsruhe started a pilot project named "RegioKargo" to transport light cargo on the region's tram and commuter train network. Photo by Michael Krauth, AVG, reproduced with permission. On the right: in 2021, Orion Railways launched a pilot project to deliver parcels on commuter trains to London Euston station, from where they are delivered using cargo bikes. Photo by Rails Ops Group reproduced with permission.

customers are at home; they deal in real-time with newly incoming orders while their fleet is already busy shipping other parcels.

The volume growth in LMD has also raised concerns, especially in dense urban environments where the externalities—traffic, emissions, noise, congestion—have become noticeable (see, e.g., the review by Viu-Roig and Alvarez-Palau (2020) and the recent papers by Wang, Rabinovich and Guda (2023), Majoral, Gasparín and Saurí (2021) and Caspersen (2021)). Several authors in Operational Research (OR), Environmental Engineering, Urban Planning, and Economics have proposed alternative LMD implementations to reduce externalities while guaranteeing a timely and cost-effective service. Solutions range from deliveries using autonomous drones, bicycles, or porters (Silva, Amaral and Fontes 2023) to congestion taxes and economic incentives. For an OR perspective on LMD services, we refer the reader to the two recent surveys by Archetti and Bertazzi (2021) and Boysen, Fedtke and Schwerdfeger (2021).

In the rest of this paper, we focus on a promising operational practice: integrating public transport within LMD to leverage the unused capacity of public transit vehicles and reduce the number of delivery vans in the city. This concept has emerged during the last years and is gaining traction, both academically (see Section 2) and in practice, as demonstrated by numerous pilot projects (Baron 2019; Sustainable Bus 2021; Der Spiegel 2020; Saito and Shimbun 2021; Clinnick 2020; Longhorn 2021; Antkowiak 2018), technical reports (Edrington et al. 2017; Deloison et al. 2020; Segura et al. 2020) and patents (Bhatt 2019).

In our study, we consider the three-echelon system introduced by Mandal and Archetti (2023). In the first echelon, transport trucks move parcels from distribution centres to public transport stops. In the second echelon, public transport vehicles transport the parcels towards the city centre and leave them at some of their scheduled stops. In the third echelon, zero-emission vehicles, such as cargo bikes, deliver the parcels to the customers' preferred locations.

This system reduces the number of vehicles on urban roads as parcels enter the city on public transport rather than on trucks. As mentioned above, pilot projects are exploring the feasibility of this approach using almost all types of transit vehicles: buses (Saito and Shimbun 2021), trams (Sustainable Bus 2021; Antkowiak 2018), commuter trains (Clinnick 2020; Longhorn 2021) and metro (Der Spiegel 2020). Most of these pilot projects started in 2020 or 2021 as a response to the increase in e-commerce which has followed the COVID-19 pandemic (see, e.g., Guthrie, Fosso-Wamba and Arnaud (2021), Villa and Monzón (2021b) and Beckers et al. (2021)). Figure 1 shows two examples of pilot projects launched in 2021, which use spare capacity on public transport vehicles to transport parcels. Mobility company AVG of Karlsruhe, Germany, uses trams, while Orion uses commuter trains in London, UK.

The main contributions of this paper are the following:

- We propose new formulations for the *Three-Tier Delivery Problem using Public Transportation* (3T-DPPT) introduced by Mandal and Archetti (2023). Our main formulation uses two exponential sets of variables and is able to refine the original model, forbidding first-echelon transport trucks to park and wait indefinitely at public transport stops—a practice that would be unacceptable in real implementations. Furthermore, we show that our formulation can be generalised to consider parcel capacity at stops. The formulation proposed by Mandal and Archetti (2023), on the other hand, cannot be adapted to this case without introducing additional variables (as we prove in Section 7 of the supplemental material).
- We propose a column generation algorithm, and we solve separate pricing subproblems for each of the two exponential sets of variables. In one of the two subproblems, the resulting shortest path problem features time-dependent label costs. We propose two novel labelling algorithms to solve this subproblem: one explicitly exploits the time dependence, while the other operates on a transformed graph where costs no longer depend on time. We embed the column generation approaches in a restricted master heuristic, thus obtaining a heuristic algorithm for the 3T-DPPT.
- We perform a computational campaign showing that our algorithm outperforms, especially on larger instances, the compact formulation and also a semi-compact formulation where one of the two sets of exponential variables mentioned above is replaced by a polynomial set of variables (with a different set of constraints). In addition, our heuristic outperforms the heuristics proposed by Mandal and Archetti (2023) by producing primal solutions with better objective values.

The paper is organised as follows. In Section 2, we survey the relevant literature. We define the problem in Section 3 and describe the extended formulations in Section 4. The latter section also introduces a semi-compact formulation containing only one set of exponentially many variables. The main characteristics of the column generation approach are illustrated in Section 5, while Section 6 describes additional speed-up techniques. We report computational results in Section 7 and conclusions in Section 8.

# 2 Related literature

Mandal and Archetti (2023) introduced the above-mentioned 3T-DPPT. In the problem name, the word "tier" is a synonym of "echelon", which is more popular in the logistics literature. The authors proposed a compact formulation, which they showed to be too large to be used in practice, and decomposition approaches. They split the problem into three subproblems, corresponding to the three echelons of the delivery network and solve each part in sequence, using the solution obtained in a subproblem to constrain the decision space of the next one. They obtain three heuristic algorithms, differing in the order in which the subproblems are solved.

The authors test their approach on instances with 10–80 customers, which they identify as the largest instances for which they can obtain primal feasible solutions using the decomposition heuristic.

In Section 3, we describe the 3T-DPPT in detail, while in the rest of this section, we position our work in the growing literature on integrating public transport with LMD. In this literature review, we identify four themes characterising the contributions. Many works present case studies, but not all employ optimisation techniques to make the integrated system more efficient.

**Integration with the metro** Kikuta et al. (2012) were among the first to study an LMD system using public transport. They carried out a 2-week pilot project in the city of Sapporo, Japan, where heavy snowfall poses significant challenges to parcel distribution in winter. The authors partnered with Sapporo's transport authority and a logistic operator to use the metro to carry hand-pushed carts with parcels during three off-peak times daily. The parcels moved from a suburban logistic centre to the inner city, where they went out for delivery. The authors concluded that this system reduced both congestion in surface roads and $CO_2$ emissions. Zhou and Zhang (2020) analysed the possibility of using a metro line for parcel delivery in three configurations: using trolley carts on a regular car, dedicating a car to freight cargo but within an otherwise passenger train, and using a freight-only train. They evaluate carbon emissions reductions of up to 50% for the scenario in which high parcel demand and commuters' off-hours justify using a dedicated train. Recently, Villa and Monzón (2021a) proposed to use the metro system of Madrid for parcel delivery. The authors devise a system in which, after travelling on the metro, packets are stored at station lockers, where customers can pick them up at their convenience. They compare two modes of operation: shared trains, used by both parcels and commuters and dedicated freight-only trains. They quantify the economic, environmental and social costs required to implement such a system and find that, after an initial investment, operators can reduce logistic costs by 11–14%.

**Integration with buses, water-buses and trams** Like our work, Masson et al. (2017) consider a three-echelon model with trucks moving parcels from distribution centres to bus stops. Buses bring the parcels—packed in roll containers—to designated stops, where tricycles perform deliveries to end customers. Unlike our setting, the authors assume that each tricycle can carry exactly one roll container. Therefore, no consolidation of parcels travelling on different buses but delivered by the same rider is possible; analogously, parcels carried on the same bus cannot be taken to different stops to be picked up by different riders. The authors minimise the number of tricycle riders needed to satisfy demand and, secondarily, their total travel time. They develop an Adaptive Large Neighbourhood Search (ALNS) heuristic, which they use to solve a case study in the city of La Rochelle, France, involving one bus line with eight stops and up to 303 customers. He and Yang (2018) studied the possibility of delivering parcels using urban buses in the city of Dalian, China. The authors minimise a combination of fixed and variable costs, plus compensation costs for excess carbon emissions and late-delivery penalties. Their results show savings of up to 10% and a reduction of $CO_2$ emissions of up to 13%.

Bruzzone, Cavallaro and Nocera (2021) present two case studies of integrating LMD and public transport, both focused on sparsely populated areas. The first case concerns the farthest islands of the Venice lagoon in Italy, which are served by two water-bus lines; the second case includes the town of Velenje, Slovenia and its neighbouring villages, collectively served by five bus lines. The authors assess the economic, environmental and social advantages of carrying parcels on the (water-)buses and conclude that these advantages are more pronounced when delivery and pickup locations are limited, travel demand is inelastic, and parcel recipients are willing to pick their parcels up at public transport stops (thus eliminating the need for the third echelon moving parcels from stops to customers' houses). Through an analysis of both technical and law literature, they conclude that it is the regulatory aspects, more than technical challenges, which hinder the wider adoption of such integrated systems.

In (Hörsting and Cleophas 2023), the authors study a system where a single tram line is used to move parcels, sharing capacity with passengers. Given a known passenger demand during the day, they aim at scheduling parcel transport with a bi-objective model minimising passenger inconvenience and delivery delays.

**Integration with generic scheduled lines** Like our work, Ghilas, Demir and Van Woensel (2016b) and Ghilas, Demir and Van Woensel (2016a) consider integrating generic scheduled public transport lines with LMD operations. The authors propose two variants of the Pickup and Delivery Problem with Time Windows (PDPTW), in which part of each request's journey can happen on a scheduled line. The first variant is the PDPTW with Scheduled Lines (PDPTWSL); the second is the PDPTWSL with Stochastic Demands (PDPTWSLSD). In both variants and different from our setting, the same trucks which bring the parcels to public transport stations can also perform the final delivery to customers. The authors also assume infinite capacity and no maximum waiting time at public transport stations and that parcel transhipments can only happen at end-of-line terminal stations. In the PDPTWSLSD, the amount of capacity that each parcel takes on the vehicles (i.e., the demand) is unknown—although distributed according to a known distribution—until the first truck picks it up, making the problem stochastic. If the realised demand for a parcel turns out to be too large for the capacity of the vehicle scheduled to pick it up, the authors assume the payment of a penalty due to outsourcing the shipment. In (Ghilas, Demir and Van Woensel 2016b), the authors propose an ALNS heuristic, and in (Ghilas, Demir and Van Woensel 2016a), they extend it, combining it with a Sample Average Approximation method to adapt it to the stochastic case. In both cases, they observe savings in operational costs compared to the truck-only scenario, but they do not analyse the potential environmental benefits.

**Strategic and tactical decisions** Another relevant problem is making long-term decisions, such as building infrastructure, which will affect the operational level of an integrated LMD. Delle Donne, Alfandari et al. (2023) study strategic and tactical decisions arising when planning a 3T-DPPT system. The decisions concern which public transport lines and stops to use in the 3T-DPPT, considering that the corresponding vehicles and stops will be equipped appropriately, at a cost. The objective is to maximise the number of parcels served through the system, subject to a maximum number of lines and stops to include. They propose a column generation approach and perform an extended computational campaign and a sensitivity analysis. They conclude that, given a limited budget, a planner should equip a limited number of vehicles and stops with high capacity, rather than expanding the 3T-DPPT network to many lines and many stops.

# 3 Problem definition

In this section, we formally define the 3T-DPPT, as introduced by Mandal and Archetti (2023) but with the modification mentioned in Section 1, i.e., that we forbid indefinite wait time of first-echelon trucks at public transport stops. The objective of the problem is to deliver parcels to a set $\mathcal{C}$ of customers. Each customer $c \in \mathcal{C}$ must receive one parcel of size $q_c \geq 0$ (size here refers generically to the amount of capacity the parcel would use on a vehicle) during a delivery time window $[\underline{T}_c, \bar{T}_c]$ at a location of choice. We assume that all delivery requests are known in advance and that all parcels start their journey at a Consolidation and Distribution Centre (CDC), denoted with $o$.

In the first tier, trucks carry parcels from the CDC to public transport stops. We denote with $\mathcal{D}$ the set of trucks. Following Mandal and Archetti (2023), we consider homogeneous vehicles with capacity $Q^{\mathrm{D}}$, although our formulation can be naturally extended to a heterogeneous fleet. Let $\mathcal{S}^{\mathrm{in}}$ be the set of public transport stops that the trucks can use to unload a parcel so that a bus will later pick it up (in the following, we use the term "bus" for simplicity, although any form of public transport could be equally used). We refer to stops of $\mathcal{S}^{\mathrm{in}}$ as "in-stops". Unloading parcels at stop $s \in \mathcal{S}^{\mathrm{in}}$ incurs a service time $T_s$. Parcels can wait at the stop for a maximum time $W^{\max}$. The service time models the handling of the parcel: a parcel unloaded by the truck at time $t$ will be ready for pick-up by a bus at time $t + T_s$. The maximum wait time ensures that bus stops are not used for long-term parcel storage.

The driving time between two locations $u, v \in \{o\} \cup \mathcal{S}^{\mathrm{in}}$ is denoted with $l_{uv}$. Consistent with the approach of Mandal and Archetti (2023), we assume that each truck route is elementary, i.e., it visits each in-stop at most once. This requirement, however, would be easy to drop given our method to handle the first tier (see Section 4) because it would correspond to dropping the elementarity requirement from a shortest-path problem. Finally, we impose that trucks depart from the in-stop as soon as they have unloaded the corresponding parcels. This is the only respect in which our definition differs from that of Mandal and Archetti (2023). We introduce this requirement because a truck parked at a stop would unduly occupy the public road and hinder the operations of public transport vehicles. If, for some particular application, this requirement does not apply, it can be straightforwardly removed from our model.

In the second tier, each parcel travels between two bus stops aboard a bus. We denote the bus fleet with $\mathcal{P}$. Each bus $p \in \mathcal{P}$ has capacity $Q_p$ and serves a route $\mathcal{S}_p = (s_p^1, \ldots, s_p^{|\mathcal{S}_p|})$ represented by the ordered list of stops the bus visits. To denote that a stop $s$ is part of the route of bus $p$, we write simply $s \in \mathcal{S}_p$. The scheduled arrival time of $p$ at stop $s \in \mathcal{S}_p$ is $t_p^s$.

In the third tier, couriers move parcels between bus stops and customer locations. The set of bus stops enabled for courier pick up is denoted as $\mathcal{S}^{\mathrm{out}}$; we refer to these as "out-stops". Because trucks do not enter the city centre and couriers do not go outside the city, each stop is either used for truck-to-bus transhipment (if outside the centre) or for bus-to-courier transhipment (if in the city centre), and thus $\mathcal{S}^{\mathrm{in}} \cap \mathcal{S}^{\mathrm{out}} = \emptyset$. For the same reason, in-stops precede out-stops in all bus routes. Each stop $s \in \mathcal{S}^{\mathrm{out}}$ has a service time $T_s$ and a maximum wait time $W^{\max}$, analogous to those of $\mathcal{S}^{\mathrm{in}}$. We denote with $\mathcal{K}$ the set of couriers who can perform the deliveries to the customers. A courier $k \in \mathcal{K}$ has capacity $Q^{\mathrm{F}}$ and is assigned to a stop $s_k \in \mathcal{S}^{\mathrm{out}}$. Each courier starts the route at $s_k$, visits a number of customers, and finally comes back to the same stop. Courier travel time between locations $u, v \in \mathcal{S}^{\mathrm{out}} \cup \mathcal{C}$ is denoted as $l_{uv}$ (because the set of locations which trucks and couriers can visit is disjoint, the notation is not ambiguous). Each courier route has a maximum duration of $L^{\max}$, which includes both the travel time and service times $T_c$ incurred when delivering parcels to customers $c \in \mathcal{C}$. Again, to be consistent with Mandal and Archetti (2023), we assume a homogeneous fleet of couriers, and we denote with $n_s$ the number of couriers which can start from stop $s \in \mathcal{S}^{\mathrm{out}}$. To facilitate delivery operations, each customer $c$ has an associated subset of stops $\mathcal{S}_c^{\mathrm{out}} \subseteq \mathcal{S}^{\mathrm{out}}$ from which they can be served. For example, stops $s \in \mathcal{S}^{\mathrm{out}}$ such that $l_{sc} + T_c + l_{cs} > L^{\max}$ are excluded from $\mathcal{S}_c^{\mathrm{out}}$ because they cannot be used to serve $c$ while respecting the courier's maximum route duration time. Other real-life operational considerations can lead to excluding further stops, e.g., if a stop is too far from the customer or if it is not possible to reach the customer using a safe bike path.

The goal of the 3T-DPPT is to determine a distribution plan to deliver all parcels at minimum cost. The cost is the sum of the route costs of the first- and third-tier vehicles.

# 4 Formulation

To introduce our formulation for the 3T-DPPT, we must consider some additional notation. We denote the set of in-stops which can be used to deliver a parcel to customer $c$ as $\mathcal{S}_c^{\text{in}}$. We build this set, as well as the other sets introduced in this section, in A. We denote with $\mathcal{R}_c^{\text{D}}$ the set of feasible truck routes which carry the parcel of customer $c$ and with $\mathcal{R}^{\text{D}}$ the set of all feasible truck routes. A feasible route must visit each in-stop at most once, deliver each parcel on board the truck to exactly one in-stop, start and end at the CDC, respect the capacity limit, and have an assigned start time from the CDC.

With an analogous notation, $\mathcal{R}_c^{\text{F}}$ denotes the set of feasible courier routes which deliver customer $c$'s parcel, while $\mathcal{R}^{\text{F}}$ is the set of all feasible courier routes. A feasible route must start and end at the same out-stop, visit each customer whose parcel is on-board exactly once and during the corresponding time window, and respect both the courier's capacity and the maximum route duration limit. We can also partition $\mathcal{R}^{\text{F}}$ based on the starting out-stop of the route; in this case, set $\mathcal{R}_s^{\text{F}}$ denotes all feasible courier routes starting and ending at $s \in \mathcal{S}^{\text{out}}$.

Finally, we introduce two more subsets of routes. The first, $\mathcal{R}_{spc}^{\text{D}} \subseteq \mathcal{R}_c^{\text{D}}$, contains all feasible routes that a truck can use to take customer $c$'s parcel to in-stop $s \in \mathcal{S}_c^{\text{in}}$, at a time at which bus $p \in \mathcal{P}$ may pick it up. The second, $\mathcal{R}_{spc}^{\text{F}} \subseteq \mathcal{R}_c^{\text{F}} \cap \mathcal{R}_s^{\text{F}}$, contains all feasible routes which a courier can use to deliver customer $c$'s parcel after bus $p \in \mathcal{P}$ delivers it at out-stop $s \in \mathcal{S}_c^{\text{out}}$. These two sets are characterised more precisely in A.

Both truck and courier routes have an associated cost $c_r > 0$ (for $r \in \mathcal{R}^{\text{D}} \cup \mathcal{R}^{\text{F}}$). Our formulation allows the cost to depend on the route characteristics in many ways to adapt to real-life circumstances, and the only assumption we make on route costs is that they are strictly positive. For example, in the case of trucks, it is common that the cost is proportional to the distance travelled (fuel cost) or to the travel time (driver hourly compensation). In the case of couriers, the cost can be proportional to the travel time or a step function of the travelled distance (as is common in crowdsourcing platforms). For consistency with Mandal and Archetti (2023), we consider route costs proportional to the travelled distance, with different multipliers for drivers and couriers.

Our formulation uses the following sets of variables (sets $\mathcal{P}_{sc}^{\text{in}}$, $\mathcal{P}_{sc}^{\text{out}}$ and $\mathcal{S}_{pc}^{\text{in}}$ are defined in A): $x_r \in \{0,1\}$ taking value 1 iff truck route $r \in \mathcal{R}^{\text{D}}$ is used; $y_r \in \{0,1\}$ taking value 1 iff courier route $r \in \mathcal{R}^{\text{F}}$ is used; $z_{spc}^{\text{in}} \in \{0,1\}$ with value 1 iff bus $p \in \mathcal{P}_{sc}^{\text{in}}$ picks up the parcel of customer $c \in \mathcal{C}$ at in-stop $s \in \mathcal{S}_c^{\text{in}}$; $z_{spc}^{\text{out}} \in \{0,1\}$ with value 1 iff bus $p \in \mathcal{P}_{sc}^{\text{out}}$ unloads the parcel of customer $c \in \mathcal{C}$ at out-stop $s \in \mathcal{S}_c^{\text{out}}$. The extended formulation of the 3T-DPPT then reads as follows:

$$\min \quad \sum_{r \in \mathcal{R}^{\text{D}}} c_r x_r + \sum_{r \in \mathcal{R}^{\text{F}}} c_r y_r \tag{1a}$$

$$\text{s.t.} \quad \sum_{r \in \mathcal{R}^{\text{D}}} x_r \leq |\mathcal{D}| \tag{1b}$$

$$\sum_{r \in \mathcal{R}_s^{\text{F}}} y_r \leq n_s \qquad \forall s \in \mathcal{S}^{\text{out}} \tag{1c}$$

$$\sum_{s \in \mathcal{S}_c^{\text{in}}} \sum_{p \in \mathcal{P}_{sc}^{\text{in}}} z_{spc}^{\text{in}} = 1 \qquad \forall c \in \mathcal{C} \tag{1d}$$

$$\sum_{s \in \mathcal{S}_{pc}^{\text{in}}} z_{spc}^{\text{in}} = \sum_{s \in \mathcal{S}_{pc}^{\text{out}}} z_{spc}^{\text{out}} \qquad \forall c \in \mathcal{C},\ \forall p \in \mathcal{P}_c \tag{1e}$$

$$\sum_{c \in \mathcal{C}} q_c \sum_{s \in \mathcal{S}^{\text{in}}_{pc}} z^{\text{in}}_{spc} \leq Q_p \qquad \forall p \in \mathcal{P} \tag{1f}$$

$$z^{\text{in}}_{spc} \leq \sum_{r \in \mathcal{R}^{\text{D}}_{spc}} x_r \qquad \forall c \in \mathcal{C},\ \forall p \in \mathcal{P}_c,\ \forall s \in \mathcal{S}^{\text{in}}_{pc} \tag{1g}$$

$$z^{\text{out}}_{spc} \leq \sum_{r \in \mathcal{R}^{\text{F}}_{spc}} y_r \qquad \forall c \in C,\ \forall p \in \mathcal{P}_c,\ \forall s \in \mathcal{S}^{\text{out}}_{pc} \tag{1h}$$

$$\sum_{r \in \mathcal{R}^{\text{D}}_c} x_r \geq 1 \qquad \forall c \in \mathcal{C} \tag{1i}$$

$$\sum_{r \in \mathcal{R}^{\text{F}}_c} y_r \geq 1 \qquad \forall c \in \mathcal{C} \tag{1j}$$

$$x_r \in \{0, 1\} \qquad \forall r \in \mathcal{R}^{\text{D}} \tag{1k}$$

$$y_r \in \{0, 1\} \qquad \forall r \in \mathcal{R}^{\text{F}} \tag{1l}$$

$$z^{\text{in}}_{spc} \in \{0, 1\} \qquad \forall c \in C,\ \forall s \in \mathcal{S}^{\text{in}}_c,\ \forall p \in \mathcal{P}^{\text{in}}_{sc} \tag{1m}$$

$$z^{\text{out}}_{spc} \in \{0, 1\} \qquad \forall c \in C,\ \forall s \in \mathcal{S}^{\text{out}}_c,\ \forall p \in \mathcal{P}^{\text{out}}_{sc}. \tag{1n}$$

The objective function (1a) minimises the combined costs of routing trucks and couriers. Constraints (1b) and (1c) ensure, respectively, that the maximum number of available trucks and couriers (the latter at each out-stop) is not exceeded. Constraint (1d) asserts that each parcel is picked up by one bus, while constraint (1e) ensures that the same bus picks up and delivers each parcel. Constraint (1f) ensures that the bus capacity is respected. Constraints (1g) and (1h) link, respectively, variables $z^{\text{in}}$ with $x$ and variables $z^{\text{out}}$ with $y$. Finally, constraints (1i) and (1j) state that the parcel of each customer must be included in at least one truck route and at least one courier route, respectively. These constraints are not required for a correct formulation of the problem. The following theorem, proved in the supplemental material's Section 1, shows that, despite their simplicity, (1i) and (1j) are not implied by the other inequalities in fractional solutions, and can considerably improve the quality of the linear relaxation of (1a)–(1n).

**Theorem 1.** *Denote with* ($1^{-}$) *formulation* (1) *without constraints* (1i) *and* (1j) *and with* *LP($\star$) the linear relaxation of a generic formulation $\star$. Then* $\text{LP}(1^{-})$ *can be arbitrarily bad, i.e.,* $\frac{\text{LP}(1)}{\text{LP}(1^{-})} \to \infty$ *in the worst case.*

As we will discuss in Section 5.3, using variables $x$ can be problematic when implementing a column generation algorithm to solve (1a)–(1n). To tackle this issue, we also propose an alternative model in which we replace the truck-route variables $x$ with three polynomial-sized sets of variables. We present the resulting *semi-compact formulation* in C.

# 5 Column generation

Formulation (1a)–(1n), which we denote as the Mater Problem (MP), uses two exponential sets of routes: $\mathcal{R}^{\text{D}}$ and $\mathcal{R}^{\text{F}}$. We call the corresponding variables $x$ and $y$, taken together, the *columns* of MP. Because sets $\mathcal{R}^{\text{D}}$ and $\mathcal{R}^{\text{F}}$ are too large to enumerate in practice, we use a column generation approach initially considering reduced sets of variables. We initialise these sets with a small number of columns, as explained in Section 4 of the supplemental material. The resulting formulation is the Reduced Master Problem (RMP). We then consider the continuous relaxation of RMP, known as the Reduced Relaxed Master Problem (RRMP), obtained by replacing integrality constraints (1k)–(1n) with non-negativity constraints.

In the remainder of this section, we explain how to solve the continuous relaxation of MP, denoted $\text{MP}_{\text{CONT}}$, by iteratively solving RRMP and generating new variables $x$ and $y$ which improve the objective value. At each iteration of the column generation algorithm, we perform three tasks:

1. We solve the RRMP and collect the dual values associated with the constraints involving variables $x$ and $y$. We denote with $\lambda^{(n)}$ the dual variable associated with constraint $(n)$ of the model. For example, $\lambda^{(1g)}_{spc}$ will refer to the dual variable associated with the inequality (1g) indexed by $s$, $p$ and $c$. We consider RRMP in its canonical form and therefore obtain non-negative dual variables.

2. We solve a column-generation subproblem $\text{SP}_x$ to find $x$ variables with negative reduced cost. Denoting with $\mathcal{C}_r$ the set of customers whose parcels are carried on a truck in route $r \in \mathcal{R}^{\text{D}}$, the reduced cost of a variable $x_r$ is

$$c_r + \lambda^{(1b)} - \sum_{c \in \mathcal{C}_r} \lambda^{(1i)}_c - \sum_{c \in \mathcal{C}_r} \sum_{p \in \mathcal{P}_c} \sum_{\substack{s \in \mathcal{S}^{\text{out}}_{pc} \\ \text{s.t. } r \in \mathcal{R}^{\text{D}}_{spc}}} \lambda^{(1g)}_{spc}. \tag{2}$$

   The reduced cost contains a fixed term (dual variable $\lambda^{(1b)}$) and dual prizes for each customer whose parcel is transported (dual variables $\lambda^{(1i)}_c$) and for each bus which can pick up the parcel at the in-stop (dual variables $\lambda^{(1g)}_{spc}$).

3. We solve a second column-generation subproblem $\text{SP}_y$ to find $y$ variables with negative reduced cost. Using again notation $\mathcal{C}_r$ to represent the set of customers whose parcels are carried by a courier in route $r \in \mathcal{R}^{\text{F}}$, the reduced cost of a variable $y_r$ is

$$c_r + \lambda^{(1c)}_{s_r} - \sum_{c \in \mathcal{C}_r} \lambda^{(1j)}_c - \sum_{c \in \mathcal{C}_r} \sum_{\substack{p \in \mathcal{P}_c \\ \text{s.t. } r \in \mathcal{R}^{\text{F}}_{s_r pc}}} \lambda^{(1h)}_{s_r pc}, \tag{3}$$

   where $s_r$ indicates the starting out-stop of route $r$. The reduced cost contains a fixed term (dual variable $\lambda^{(1c)}$) and dual prizes for each customer whose parcel is transported (dual variables $\lambda^{(1j)}_c$) and for each bus which can deliver the parcel to the out-stop (dual variables $\lambda^{(1h)}_{s_r pc}$).

The algorithm terminates when there are no more negative reduced cost columns to generate. We remark that subproblems $\text{SP}_x$ and $\text{SP}_y$ need not be solved in the given order; indeed, in our implementation, we solve $\text{SP}_x$ only when $\text{SP}_y$ does not produce any negative reduced cost column.

Although the proposed method can be used to find the optimal solution of $\text{MP}_{\text{CONT}}$, this task is computationally demanding. Therefore, in Section 6.1, we describe algorithms to get high-quality columns heuristically. Once we generate a sufficient number of columns, or when a time limit hits, we solve RMP with a black-box Mixed-Integer Programming (MIP) solver to obtain a primal solution for MP. This approach is known as "price-and-branch" or "restricted master heuristic" in the literature (Sadykov et al. 2019). To prove the quality of such a solution, in Section 5.3, we devise dual bounds for MP.

## 5.1 Column generation for truck routes

In this subsection, we explain how to solve the pricing subproblem $\text{SP}_x$ to generate truck routes with negative reduced costs or prove that no such route exists. A truck route $r \in \mathcal{R}^{\text{D}}$ is defined by three elements: (i) a sequence of in-stops, without repetitions; (ii) a set of parcels to deliver

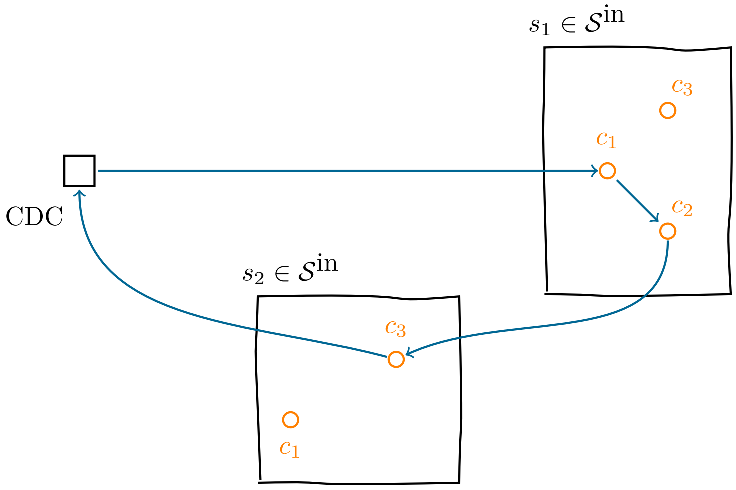


Figure 2: Small example of graph $G$ with two in-stops and three customers. Blue arrows denote a possible truck route.

to each visited in-stop such that each parcel is delivered once and the total size of the parcels does not exceed the truck capacity; (iii) the route start time from the CDC.

The route start time is particularly important: because trucks are not allowed to wait at stops, the start time determines the visit time of all in-stops. Moreover, the time at which the truck deposits parcels at a stop is crucial: a truck should not arrive too early or too late. If the truck arrives too early, the parcels will not be able to use some buses, which might be associated with large dual prizes (recall that parcels can stay a maximum of $W^{\max}$ units of time at an in-stop). Analogously, if the truck arrives too late, the parcels will miss some buses with a possibly high dual prize. Contrast this with, e.g., vehicle routing problems with time windows in which arriving late is forbidden or penalised, but vehicles can arrive early and wait at customer locations.

In what follows, we describe two different approaches to solve $\mathrm{SP}_x$. They use an underlying graph and solve special types of shortest-path problems with a dynamic programming labelling algorithm. The main difference between them is in how they model time. A third, MIP-based approach is discussed in Section 2 of the supplemental material.

#### 5.1.1 Labelling algorithm with a time-dependent cost function

We start by describing the directed graph $G = (V, A)$ used by the labelling algorithm. The vertex set is $V = \{o\} \cup \{(s, c) \ : \ c \in \mathcal{C}, \ s \in \mathcal{S}^{\text{in}}_c\}$ and it contains the CDC (o) and the set of all pairs of in-stops and customers $(s, c)$, taking care to add only those pairs which are compatible (see A for a formal definition of $\mathcal{S}^{\text{in}}_c$). The arc set contains three types of arcs. First, arcs from $o$ to all other vertices and from the other vertices to $o$. Second, arcs from vertices $(s, c)$ to $(s', c')$ such that $s \neq s'$ and $c \neq c'$. These arcs correspond to the truck moving from in-stop $s$, where it unloads parcel $c$, to in-stop $s'$, where it unloads parcel $c'$. Third, within-stop arcs from vertices $(s, c)$ to vertices $(s, c')$. These arcs indicate that the truck unloads both parcels $c$ and $c'$ at in-stop $s$. Because there is no inherent order in which parcels are unloaded, we can drastically reduce the number of arcs by defining an arbitrary ordering of the parcels (say $c_1, \dots, c_{|\mathcal{C}|}$) and only adding arcs from $(s, c_i)$ to $(s, c_j)$ if $j > i$. Figure 2 shows a small example in which rectangles correspond to in-stops and orange circles correspond to vertices of type $(s, c)$. The square vertex is the CDC, while blue arrows represent a possible truck route which delivers parcels for $c_1$ and $c_2$ at stop $s_1$, and for $c_3$ at stop $s_2$.

A path in $G$ can only define two of the three elements required to identify a truck route: the sequence of in-stops and the sets of parcels delivered to each in-stop. Therefore, a route $r \in \mathcal{R}^{\mathrm{D}}$ is determined by both a suitable path in $G$ and a start time from the CDC. To include the time dimension, we first associate each arc $a = \big((s,c),(s',c')\big) \in A$ $(s \neq s')$ with a traversing time $t_a$ equal to the travel time from the origin in-stop to the destination in-stop, $l_{ss'}$, plus the service time $T_{s'}$. If origin and destination in-stops coincide $(s = s')$, then the traversing time is $t_a = 0$. The traversing time of arcs arriving at $o$ only includes the travel time $(t_a = l_{so})$. We then define the *benefit* of arriving at a vertex $(s,c) \in V$ at time $t$ as

$$\nu_{sc}(t) = \lambda_c^{(1i)} + \sum_{\substack{p \in \mathcal{P}_{sc}^{\mathrm{in}} \text{ s.t.} \\ t_p^s - W^{\max} \leq t \leq t_p^s}} \lambda_{cps}^{(1g)}.$$

Interval $I_{ps} := \left[t_p^s - W^{\max}, t_p^s\right]$ defines all time instants when a parcel ready for pick-up at stop $s$ can be loaded onto bus $p$. Finally, we associate with a path $P = (o,(s_1,c_1),\ldots,(s_k,c_k),o)$ of $G$ its reduced cost function

$$C_P(t) = c_r + \lambda^{(1b)} - \sum_{i=1}^{k} \nu_{s_i c_i}(\hat{t}_i(t)), \tag{4}$$

where $t$ is the start time from the CDC, $c_r$ is the cost of the truck route associated with path $P$ and $\hat{t}_i$ is the truck arrival time at in-stop $s_i$, defined as

$$\hat{t}_i(t) = t + t_{(o,(s_1,c_1))} + \sum_{j=1}^{i-1} t_{(s_j,c_j),(s_{j+1},c_{j+1})}.$$

$C_P(t)$ is a step-wise non-convex function and, as such, needs tailored dominance rules, as explained in the following.

Finding a truck route of negative reduced cost corresponds to finding a path $P$ in $G$ such that its corresponding route is feasible and that $C_P(t) < 0$ for at least one start time $t$. To accomplish this task, we use a labelling algorithm which associates a label $L_P$ with each partial path $P$ from the CDC to a vertex $(s,c) \in V$. The label has the following components: $v_P = (s,c)$ is the end vertex of the path; $\mathcal{S}_P \subseteq \mathcal{S}^{\mathrm{in}}$ is the set of in-stops which can still be visited when departing from $v_P$; $\mathcal{C}_P \subseteq \mathcal{C}$ is the set of customers whose parcels can still be delivered when departing from $v_P$; $\tau_P \geq 0$ is the sum of traversing times of the arcs used in the path; $Q_P \geq 0$ is the spare capacity on the truck when departing from $v_P$; $C_P : \mathbb{R}_0^+ \to \mathbb{R}$ is the reduced cost function associated with $P$, obtained by truncating the sum in (4) to the end vertex $v_P$.

As shown in Section 1 of the supplemental material, to get a valid dominance rule, the cost function associated with each partial path has to be calculated with respect to the arrival time at the end vertex $v_P$. To this end, we introduce function $\bar{C}_P(t)$ representing the reduced cost of path $P$ when the truck arrives at the end vertex $v_P$ at time $t$ and we define $\bar{C}_P(t) = \infty$ for all $t < \tau_P$. By contrast, $C_p(t)$ is a function of the start time from the CDC. Because the arrival time at $v_P$ is completely determined by the start time of the route, one can always recover the value of $C_P(t)$ from that of $\bar{C}_P(t)$ and vice-versa.

When extending a path $P$ to a new vertex $(s',c') \in V$ along arc $a \in A$, the label associated with the new path $P'$ has the following components:

$$\begin{aligned} v_{P'} &= (s',c'), & \mathcal{S}_{P'} &= \mathcal{S}_P \setminus \{s'\}, & \mathcal{C}_{P'} &= \mathcal{C}_P \setminus \{c'\}, \\ \tau_{P'} &= \tau_P + t_a, & Q_{P'} &= Q_P - q_{c'}, & \bar{C}_{P'}(t) &= \nu_{s'c'}(t) + \bar{C}_P(t - t_a) + c_a, \end{aligned}$$

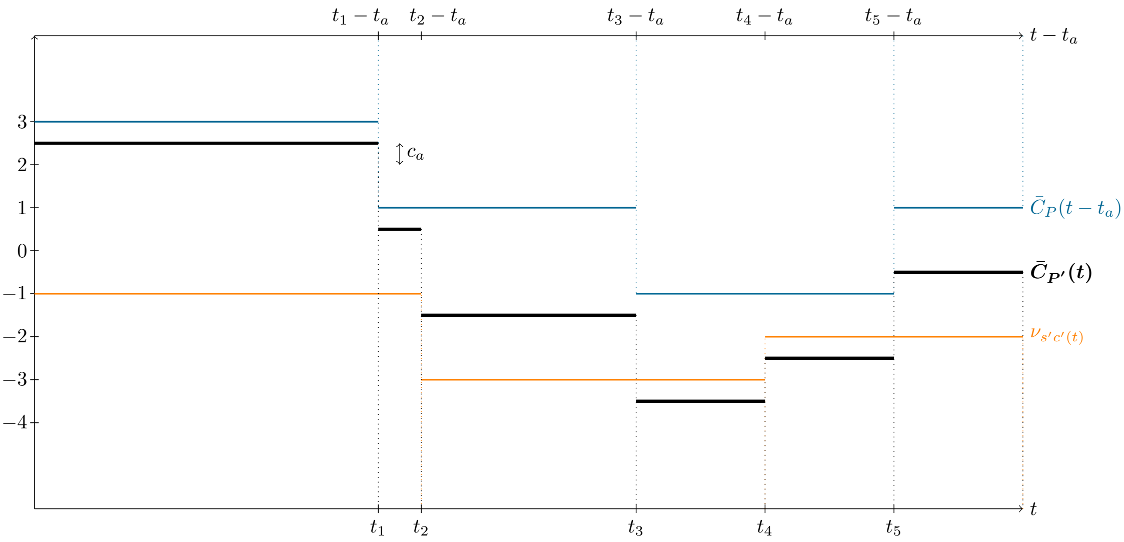


Figure 3: Example of updating cost function $\bar{C}_P$ to $\bar{C}_{P'}$ using the benefit function $\nu_{s'c'}$, during label extension. Time $t$ at the end-point of $P'$ is on the bottom horizontal axis. Time $t - t_a$ at the end-point of $P$ is on the top horizontal axis.

where $c_a$ is the routing cost associated with arc $a$. Figure 3 shows an example of how cost function $\bar{C}_P$ gets updated to $\bar{C}_{P'}$ using benefit function $\nu_{s'c'}$. In the figure, $\bar{C}_{P'}$ and $\nu_{s'c'}$ are represented as functions of $t$, i.e., the arrival time at the new end vertex $v_{P'}$; $\bar{C}_P$ is represented as a function of $t - t_a$, i.e., the arrival time at the old end vertex $v_P$.

The following proposition, proven in supplemental material's Section 1, establishes a dominance relation allowing to discard labels which cannot possibly be extended to an optimal complete path, i.e., to a path with the lowest possible reduced cost.

**Proposition 1.** *Given two paths $P_1$ and $P_2$, and their corresponding labels $L_{P_1}$ and $L_{P_2}$, ending at the same vertex $v_{P_1} = v_{P_2}$, $L_{P_1}$ dominates $L_{P_2}$ if: $\mathcal{S}_{P_2} \subseteq \mathcal{S}_{P_1}$, $\mathcal{C}_{P_2} \subseteq \mathcal{C}_{P_1}$, $Q_{P_2} \leq Q_{P_1}$, $\bar{C}_{P_1}(t) \leq \bar{C}_{P_2}(t)$ for all $t \geq \tau_{P_2}$, and at least one condition holds strictly. We remark that the condition on $\bar{C}$ implies that $\tau_{P_1} \leq \tau_{P_2}$.*

The dominance rule presented in Proposition 1 implies a point-to-point pairwise comparison of the two labels. Indeed, contrary to standard scalar costs, the dominance should be checked for any value of the time-dependent cost function (4). A similar dominance rule was proposed by Tagmouti, Gendreau and Potvin (2007) for a time-dependent arc routing problem. However, the authors do not specify how they implement dominance checking and calculate the cost of a path with respect to the starting time from the depot, which leads to a non-valid dominance rule in our case, as discussed in Section 1 of the supplemental material. More recently, Baum et al. (2020) and Klein and Schiffer (2022) also proposed a labelling algorithm for a subproblem with a time-dependent piece-wise linear convex function. Their pairwise dominance rule is similar to the one proposed in Proposition 1. Moreover, Klein and Schiffer (2022) use a set-based dominance rule, where a label can be jointly dominated by a set of labels. Preliminary tests showed that set-based dominance did not improve the performance of our algorithm.

#### 5.1.2 Labelling algorithm with a scalar cost

The dual prizes collected at the vertices of $G$ only change, as a function of the arrival time at the in-stop, at the intersection of the time intervals $I_{sp}$ defined in Section 5.1.1. In other words, the dual prizes change at time instants when a bus becomes available or is no longer available

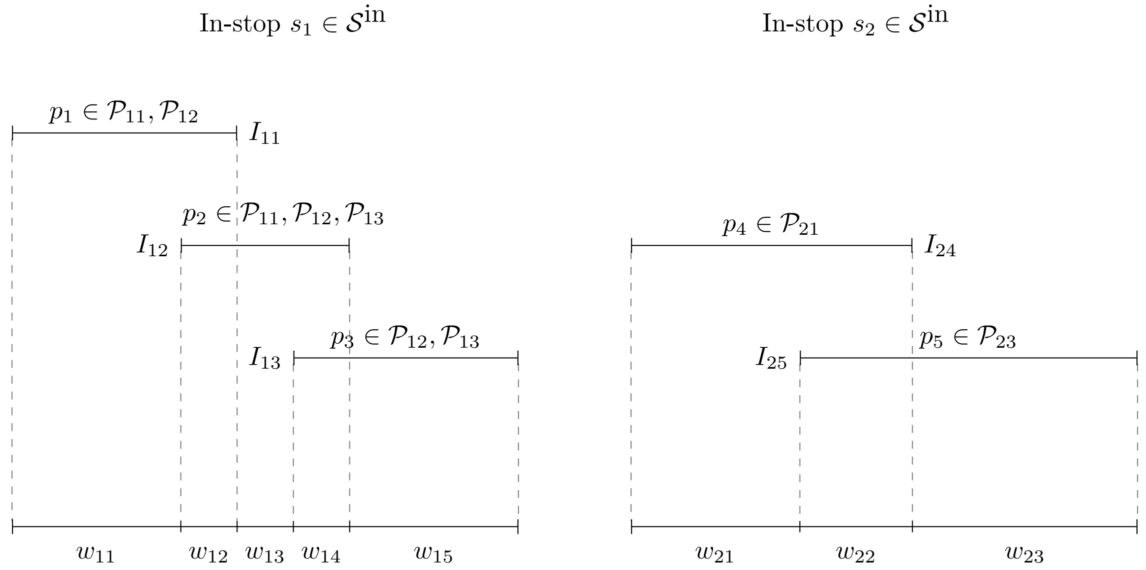


Figure 4: Time intervals $w_{s\ell}$ for the instance presented in Figure 2. Buses $p_1, p_2, p_4$ are compatible with customer $c_1$; $p_1, p_2, p_3$ with customer $c_2$; and $p_2, p_3, p_5$ with customer $c_3$.

to pick up parcels. Therefore, for each in-stop, it is possible to partition the time horizon into intervals where the dual prizes are constant.

Consider, for example, the instance presented in Figure 2. Assume that buses $p_1, p_2, p_3$ serve in-stop $s_1$, and buses $p_4, p_5$ serve in-stop $s_2$. A possible time horizon partitioning for both in-stops is depicted in Figure 4; the left part of the figure refers to $s_1$ and the right one to $s_2$. The horizontal lines in the top part of the picture represent time intervals $I_{sp}$. The line at the bottom is the time horizon, partitioned into intervals $w_{s\ell}$, where $s \in \mathcal{S}^{\text{in}}$ is the in-stop and $\ell$ is the index of the interval ($\ell \in \{1, \dots, \Lambda_s\}$).

As a consequence of the above observation, we propose an alternative approach to solving $\text{SP}_x$ as a shortest path problem where, contrary to the problem described in Section 5.1.1, the cost is no longer a time-dependent function but a scalar. We create a directed graph $H = (W, B)$ with vertex set:

$$W = \{o\} \cup \left\{ (s, c, \ell) \ : \ c \in \mathcal{C}, \ s \in \mathcal{S}^{\text{in}}_c, \ \ell \in \{1, \dots, \Lambda_s\}, \ \exists p \in \mathcal{P}_{cs} \text{ s.t. } I_{sp} \cap w_{s\ell} \neq \emptyset \right\}.$$

Each vertex $(s, c)$ of $V$ is copied multiple times in $W$: it gets one copy for each time interval $w_{s\ell}$ during which at least one bus can pick up the parcel of $c$ at $s$. The arc set $B$ contains three types of arcs. First, arcs from $o$ to all other vertices, and from the other vertices to $o$. Second, arcs from vertices $(s, c, \ell)$ to vertices $(s', c', \ell')$ such that $s \neq s'$ and $c \neq c'$. Third, within-interval arcs from vertices $(s, c, \ell)$ to vertices $(s, c', \ell)$, indicating that the truck delivers both parcels for $c$ and $c'$ at $s$ during interval $\ell$. Similar to what was noted in Section 5.1.1, the number of such arcs can be drastically reduced through an arbitrary ordering of the customers. Figure 5 shows the same instance and route of Figure 2, but on graph $H$ with time intervals defined as in Figure 4. The route delivers $c_1$'s and $c_2$'s parcels at in-stop $s_1$ during time interval $w_{11}$, and $c_3$'s parcel at in-stop $s_2$ during time interval $w_{23}$. Therefore, the only possible continuation of the parcels' journeys is on bus $p_1$ for $c_1$ and $c_2$ and on bus $p_5$ for $c_3$.

The labelling algorithm used to find truck routes of negative reduced cost (or prove that none exists) on graph $H$ is similar to classical algorithms to solve the resource-constrained shortest path problem. We give a detailed description of it in B.1. Here we only note that using this

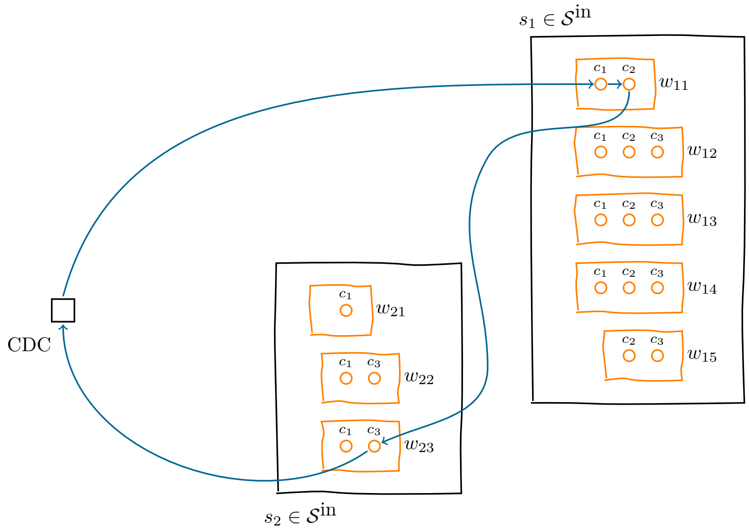


Figure 5: Small example of graph $H$ with two in-stops and three customers. Buses $p_1, p_2, p_3$ serve in-stop $s_1$; buses $p_4, p_5$ serve in-stop $s_2$. Time intervals $w$ are as in Figure 4. Blue arrows denote a possible truck route.

approach, we no longer need to associate a cost function with each label. Because the vertices already encode the information about buses compatible with a given route, we can use scalar costs. The trade-off is that graph $H$ is much larger than graph $G$. We will introduce speed-up techniques to overcome the challenges of solving $\text{SP}_x$ both on $G$ and $H$ in Section 6.

## 5.2 Column generation for courier routes

We now explain how to solve the pricing subproblem $\text{SP}_y$. We first remark that $\text{SP}_y$ is decomposable by out-stop. The lowest reduced cost of any courier route can be written as

$$\min_{s \in \mathcal{S}^{\text{out}}} \min_{r \in \mathcal{R}^{\text{F}}_s} \left\{ c_r + \lambda^{(1\text{c})}_s - \sum_{c \in \mathcal{C}_r} \lambda^{(1\text{j})}_c - \sum_{c \in \mathcal{C}_r} \sum_{\substack{p \in \mathcal{P}_c \\ \text{s.t. } r \in \mathcal{R}^{\text{F}}_{spc}}} \lambda^{(1\text{h})}_{spc} \right\},$$

and all inner minimisation problems are independent of each other.

For a given out-stop $s \in \mathcal{S}^{\text{out}}$ and customer $c \in \mathcal{C}_s$, the dual prizes $\lambda^{(1\text{h})}_{spc}$ which a courier route can collect only depend on the route start time from $s$ because this time determines which buses can deliver the customer's parcel. Using the same key ideas from Section 5.1.2, we can partition the time horizon into intervals in which these dual prizes stay constant. Keeping the same notation, we denote these intervals with $w_{s\ell}$ ($w \in \{1, \dots, \Lambda_s\}$). We then decompose $\text{SP}_y$ both by out-stop and by time interval, and we let $\text{SP}_y(s, \ell)$ be the subproblem associated with $s \in \mathcal{S}^{\text{out}}$ and $\ell \in \{1, \dots, \Lambda_s\}$, in which we use $w_{s\ell}$ as the out-stop time window limiting the possible start times of the courier routes. Considering customer time windows, the capacity of the courier, and that each customer must be visited at most once, we model $\text{SP}_y(s, \ell)$ as an elementary shortest-path problem with time windows. B.2 presents a labelling algorithm to solve $\text{SP}_y(s, \ell)$.

## 5.3 Dual bounds

A straightforward way to obtain a dual bound for MP is to solve $\text{MP}_{\text{CONT}}$ to optimality, i.e., until there are no more negative reduced cost columns. In practice, this approach only works for small instances. For medium-size instances, the column generation procedures do not usually terminate within the time limit of one hour that we impose in computational tests. For large instances, even solving one iteration of the column generation algorithm may take more than one hour; in particular, solving a single pricing problem $\mathrm{SP}_x$ is computationally prohibitive. Pricing problem $\mathrm{SP}_y$, on the other hand, is usually solved in less than a second. This is mainly due to the fact that we solve an independent subproblem $\mathrm{SP}_y(s, \ell)$ for each out-stop $s$ and each time interval $\ell$. Hence, the size of these subproblems remains tractable even for large instances. Additionally, courier capacity is usually limited, preventing sets $\mathcal{R}^{\mathrm{F}}$ from growing large.

We can speed up the solution of the pricing problems by approaching them heuristically (see Section 6.1). In this case, however, we can no longer guarantee that all remaining columns have a non-negative reduced cost and, thus, that we have solved $\mathrm{MP}_{\mathrm{CONT}}$. Therefore, in the following, we introduce an alternative dual bound for the 3T-DPPT, which can be computed off-line when it is intractable to solve $\mathrm{SP}_x$ or $\mathrm{SP}_y$. In Section 3 of the supplemental material, we also describe a way to obtain an on-line Lagrangian bound while running the restricted master heuristic.

We obtain the off-line bound considering the relaxation of the 3T-DPPT in which we ignore the second echelon and solve the first- and third-echelon problems separately. In particular, we no longer specify which buses transport parcels from in-stop to out-stop but rather devise truck and courier routes independently. The resulting solution is likely unfeasible: for example, it can involve a truck delivering a parcel at in-stop $s_1$ and a courier delivering the same parcel starting from out-stop $s_2$, even when there is no bus connecting $s_1$ and $s_2$. Furthermore, we relax the in-stop elementarity and no-wait constraints from the first-echelon problem: we allow trucks to visit the same stop multiple times and to wait at a stop before delivering parcels. In the following, we describe the proposed solution approach for the first- and third-echelon problems.

With the above relaxations, the first-echelon problem becomes a variant of the Generalised Vehicle Routing Problem with Time Windows (GVRPTW). The GVRP is an extension of the classical Vehicle Routing Problem in which the customers are partitioned into clusters. Vehicles must only visit one customer in each cluster, and, in doing so, they deliver the entire cluster demand. In the GVRPTW, additionally, each customer can be visited only within a given time window.

Consider the graph $G$ introduced in Section 5.1.1, in which each vertex other than the CDC corresponds to a pair of in-stop and customer. We partition these vertices grouping together those that correspond to the same customer. Because a customer's parcel must be delivered at most once, it is sufficient to visit only one vertex in each cluster; this vertex determines the in-stop used to deliver the parcel. We obtain time windows for each vertex $(s, c)$ considering the delivery time window for customer $c$, the possible out-stops which can handle $c$'s parcel, and the possible buses which can carry the parcel from $s$ to these out-stops. The corresponding time window $\Theta_{sc}$ is defined in A. In a GVRPTW solution, the same vehicle can visit two vertices referring to the same in-stop at any point in its route. Thus, we cannot guarantee the elementarity of the in-stops. Section 3 of the supplemental material details the solution approach used to solve the GVRPTW.

The third-echelon problem, on the other hand, is a Multi-Depot Vehicle Routing Problem with Time Windows (MDVRPTW). The capacity constraints of the couriers considerably limit the number of customers served by the same route. Therefore, we solve the MDVRPTW using a straightforward extended formulation enumerating all feasible routes.

# 6 Other elements of the column generation algorithm

In this section, we describe the auxiliary components of the column generation algorithm. We will focus on subproblem $\mathrm{SP}_x$ because it is, in practice, much harder than $\mathrm{SP}_y$. Section 6.1 describes how to speed up the solution of $\mathrm{SP}_x$ using graph sparsification heuristics. We obtain a further speed-up by considering the order in which we extend and dominate labels, as explained in Section 6.2. Finally, in Section 6.3, we explain how to post-process truck routes to exploit eventual residual capacity. Section 4 of the supplemental material describes the greedy algorithm we use to generate an initial set of columns.

## 6.1 Heuristic column generation

We speed up the generation of columns in subproblem $\mathrm{SP}_x$, reducing the number of arcs in graphs $G$ and $H$, respectively introduced in Sections 5.1.1 and 5.1.2. We start by describing the techniques used on graph $G$. Recall that, for a given in-stop $s$, we only generate the within-stop arc from $(s, c_i)$ to $(s, c_j)$ if $i < j$ in an arbitrary ordering of the customers $(c_1, \dots, c_{|\mathcal{C}|})$. In our heuristic arc reduction, we sort the customers within the same in-stop by increasing (negative) maximum reduced cost $\mathrm{mrc}_s(c)$, given by

$$\mathrm{mrc}_s(c) = \max_t \left\{ \nu_{sc}(t) \right\} = \max_t \left\{ \lambda_c^{(1i)} + \sum_{\substack{p \in \mathcal{P}_{sc}^{\mathrm{in}} \text{ s.t.} \\ t_p^s - W^{\max} \leq t \leq t_p^s}} \lambda_{cps}^{(1g)} \right\}.$$

Then, for each node $(s, c)$, we only keep the within-stop arc to $(s, c')$, where $c'$ is the customer following $c$ in the given order, over those customers not yet visited by the route being constructed. In practice, this technique forces a truck at an in-stop to deliver the parcel of the $k$-th best customer by reduced cost only if it also delivers the parcels of the best, second-best, …, $(k-1)^{\text{th}}$-best customers for the stop, either at the current or at previous stops. We denote this technique as PATH heuristic.

We use an analogous procedure to reduce the number of arcs in the graph $H$. Given an in-stop $s$ and a time interval $\ell$, we sort customers by increasing (negative) reduced cost, expressed in this case as

$$\lambda_{c'}^{(1i)} + \sum_{\substack{p \in \mathcal{P}_{s'c'} \text{ s.t.} \\ I_{s'p} \cap w_{s'\ell'} \neq \emptyset}} \lambda_{s'pc'}^{(1g)}.$$

We only generate arcs from $(s, c, \ell)$ to the $(s, c', \ell)$, with $c'$ immediately following $c$ in the given order.

In addition to the PATH heuristic, we propose another arc reduction technique, which we call $\mathrm{BEST}_k$, for a fixed value of $k \in \mathbb{N}$. Again, we sort the customers at every stop $s$ by the value of $\mathrm{mrc}_s(c)$. For each stop, we only keep the best $k$ customers (and delete the other vertices) and keep all the corresponding internal arcs.

## 6.2 Using checkpoints in truck columns labelling

Graph $G$ of Section 5.1.1 contains $\mathcal{O}(|\mathcal{S}^{\mathrm{in}}|^2 |\mathcal{C}|^2)$ "travelling" arcs of type $\big((s, c), (s', c')\big)$ because each pair $(s, c)$ of in-stop and customer is connected to each other pair $(s', c')$ (when $s \neq s'$ and $c < c'$). In this section, we explain how to reduce these arcs to $\mathcal{O}(|\mathcal{S}^{\mathrm{in}}|^2 + |\mathcal{S}^{\mathrm{in}}| \cdot |\mathcal{C}|)$, at the expense of adding $\mathcal{O}(|\mathcal{S}^{\mathrm{in}}|)$ vertices. Because in practical applications $|\mathcal{S}^{\mathrm{in}}| \ll |\mathcal{C}|$, this considerably reduces the number of arcs in $G$ and speeds up the labelling algorithm to solve $\mathrm{SP}_x$. The main idea is to add to each in-stop $s$ two "check-point" vertices denoted $v_s^-$ (entry

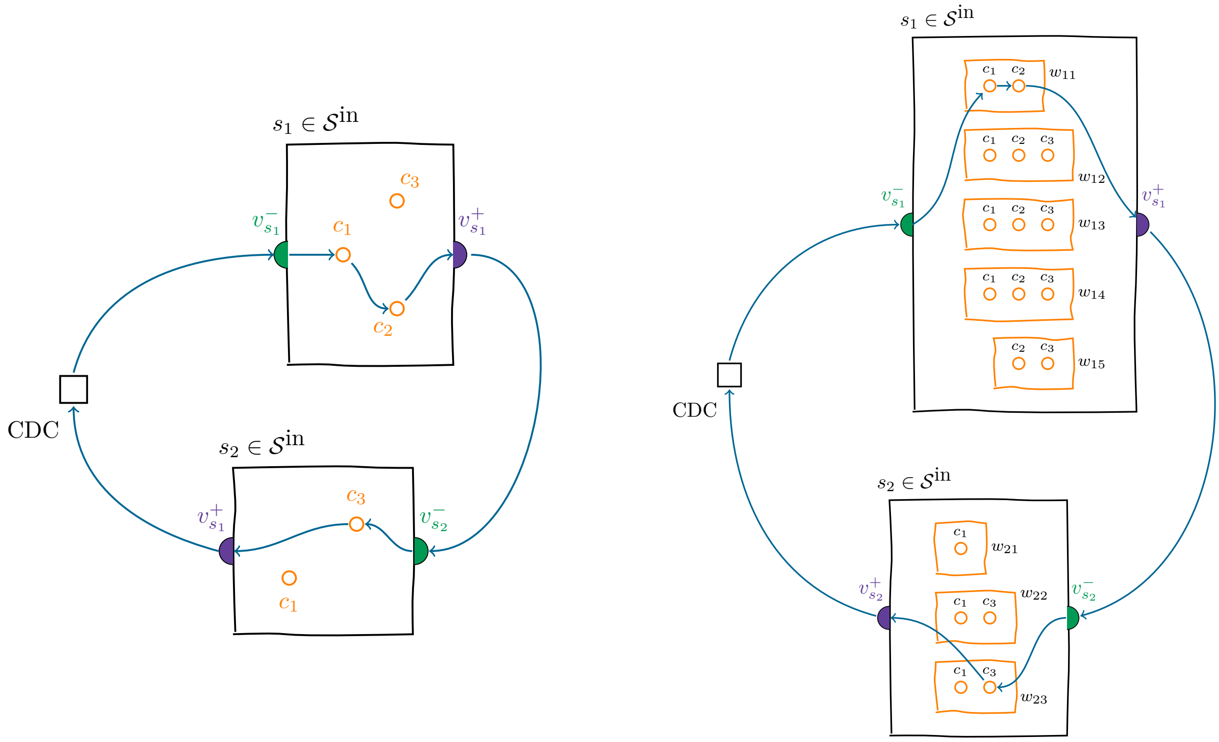


Figure 6: Introduction of checkpoints in graph $G$ of Section 5.1.1 (left) and in graph $H$ of Section 5.1.2 (right).

point) and $v_s^+$ (exit point). We remove from the graph the travelling arcs described above, and we replace them with: (i) $\mathcal{O}(|\mathcal{S}^{\text{in}}|^2)$ new travelling arcs of type $(v_s^+, v_{s'}^-)$ for any pair of in-stops $s \neq s'$; (ii) $\mathcal{O}(|\mathcal{S}^{\text{in}}| \cdot |\mathcal{C}|)$ entry-to-customer arcs from $v_s^-$ to $(s,c)$ for all $c \in \mathcal{C}_s$; (iii) $\mathcal{O}(|\mathcal{S}^{\text{in}}| \cdot |\mathcal{C}|)$ customer-to-exit arcs from $(s,c)$ to $v_s^+$ for all $c \in \mathcal{C}_s$. Figure 6 (left) shows the same instance and route as Figure 2 after the introduction of checkpoints. The same idea can be applied to graph $H$ introduced in Section 5.1.2, as depicted in Figure 6 (right).

### 6.3 Post-processing of columns

We post-process the truck routes generated when solving subproblem $\text{SP}_x$ to exploit eventual spare capacity. In the following, we give a brief description of the post-processing algorithm for a route $r \in \mathcal{R}^{\text{D}}$ with its associated start time. The algorithm consists of two phases.

First, we order the set of customers not covered by $r$ in ascending order of ratio $\lambda_c^{(1\text{i})}/q_c$. Then, we go through the in-stops visited along route $r$, and we deliver at each of them as many additional parcels as possible in the given order. The truck can deliver a parcel at a stop if there is at least one bus compatible with that parcel, if there is enough capacity on the truck, and if the parcel is not already included in $r$. Performing this post-processing step does not increase the route's cost but makes it cover more customers.

In the second phase, we use a procedure which might increase the route's cost. We first compute, for each in-stop $s$ not visited by $r$, the best dual prize we could collect if we visit $s$ and deliver parcels using the same criterion as in the first phase. Because the best dual prize depends on the position in which we insert $s$ in the route, we consider all possible positions. For each stop $s$, giving dual prizes larger (in absolute value) than the cost of the detour, we create a copy of $r$ in which we insert stop $s$. We repeat this procedure until no stop can be inserted.

# 7 Computational experiments

In this section, we present the results of a wide range of computational experiments. These experiments have two objectives. First, we want to assess the trade-offs associated with the different solution approaches we presented:

- Using the extended formulation with two exponential sets of variables, the semi-compact formulation with one exponential set of variables, or the compact formulation of Mandal and Archetti (2023) amended to forbid trucks waiting at stops indefinitely.
- Using the cost-function (Section 5.1.1) or the scalar-cost (Section 5.1.2) labelling algorithms to solve the truck routes pricing problem.
- Using the PATH or the BEST$_k$ column generation heuristics.

Results concerning other algorithmic parameters were so overwhelmingly biased towards one of the possible choices that we fixed them during preliminary experiments. In particular, we always use the checkpoints introduced in Section 6.2, activate a column generation heuristic, apply postprocessing (Section 6.3), generate initial columns (using the procedure described in Section 4 of the supplemental material) and prematurely stop a column generation iteration and re-solve RRMP when we generate five columns with negative reduced cost.

The second objective is to compare our approach with the current state-of-the-art, i.e., the decomposition heuristics of Mandal and Archetti (2023).

We ran our experiments on a cluster with Intel Xeon CPUs running at 2.4GHz. Each run was limited to using one core and 32GB of memory. The RRMP solver was Cplex version 20.1, used through its `C++` API with default parameters. We generated the instance set following the same procedure as Mandal and Archetti (2023), which uses the instance generator developed by Delle Donne, Alfandari et al. (2023). The dataset contains ten instances for each number of customers in $\{20, 25, 30, 40, 50\}$, for a total of 50 instances. Each instance contains between 21 and 32 stops, and between 45 and 60 buses. All instances and results are available in a GitHub repository (Delle Donne, Santini and Archetti 2024).

## 7.1 Impact of algorithmic components

The main components of our proposed algorithm revolve around pricing the $x$ variables, which is by far the most challenging part of solving the 3T-DPPT. We explored four main possibilities: pricing using the labelling algorithms with the time-dependent cost function or the scalar costs or avoiding pricing altogether using the semi-compact or compact formulations. For the first two options, we must decide what pricing heuristic to use. After a tuning phase whose results are reported in Section 5 of the supplemental material, we selected the BEST$_k$ heuristic with $k = 3$ for the algorithm using time-dependent cost function and the PATH heuristic for the algorithm using the scalar cost. We denote these two algorithms "**CG1**" and "**CG2**"; the semi-compact formulation is denoted with "**SCF**", and the compact formulation with "**CF**". A maximum running time of one hour has been set for all algorithms. Specifically, for CG1, CG2 and SCF, we reserve a minimum time of five minutes for the final reduced master heuristic. Therefore, if the column generation phase is not over after 55 minutes, we stop it and start the final MIP.

The compact formulation is exactly as presented in (Mandal and Archetti 2023, Section 3.1), but adding a constraint forbidding indefinite truck waits at stops, as detailed in Section 6 of the supplemental material. The semi-compact formulation also includes this additional constraint. In Section 1, we mentioned that we identified a second shortcoming of the compact formulation of Mandal and Archetti (2023), namely, that it does not model parcel capacity at public transit stops, allowing them to be used as long-term storage options. In our experiments, however, we do not introduce this extra constraint for several reasons. First, a manual inspection of the results shows that long truck waits occur more often, while it is rare that stops host more than a handful of parcels at a time. Second, it is hard to justify in practice that a delivery van can occupy public space next to a public transit stop, but it is not unfathomable to equip such stops with lockers that allow parcels to be stored for a few hours. Third, the compact formulation cannot be used to model parcel capacity at stops without adding new variables. We prove this result in Section 7 of the supplemental material. Adding new variables to the already large model of Mandal and Archetti (2023) would worsen its performance and does not allow a fair comparison between the compact formulation and our proposed approach. Finally, the results published in Mandal and Archetti 2023 refer to the version without capacities. While it would be straightforward to re-run the compact formulation with the new variables on a black-box solver, we would need to reimplement the decomposition heuristics—a task that is out of scope for the present paper.

Figure 7 presents the performance profiles of the four algorithms CG1, CG2, SCF, and CF. The x-axis reports the gap with respect to the best dual bound, i.e., the tightest among all the dual bounds produced during the experimental campaign. These bounds are obtained by the following techniques: (i) running the column generation algorithm solving the pricing subproblem for the $x$ variables using the MIP model described in Section 2 of the supplemental material (and the subproblem for the $y$ variables to optimality) and applying the Lagrangian bound described in Section 3 of the supplemental material; (ii) computing the decomposition bound described in Section 5.3; (iii) solving the semi-compact or the compact formulation and taking the best dual bound reported when the solver terminates. For any given gap on the x-axis, the y-axis reports the number of instances solved within the gap. Each curve corresponds to one of the four algorithms (dark green for CG1, light green for CG2, orange for SCF, and red for CF). The curves for SCF and CF do not reach the top-right corner of the figure because these algorithms find feasible solutions for 27 and 22 instances only. On the other hand, CG1 and CG2 produce a feasible solution for all instances but one. On the bottom-left part of the chart, the curves relative to SCF and CF dominate those relative to CG1 and CG2. Indeed, for small instances, SCF and CF find optimal or almost optimal solutions. For larger instances, however, CG1 and CG2 either give better gaps than SCF and CF or are the only two methods yielding primal solutions at all. We finally note that, although no column generation method dominates the other completely, CG2 is often slightly better than CG1.

Figure 8 provides disaggregated results. The number of customers is reported on the x-axis, with the instances sorted by increasing gap provided by the best method; the y-axis shows the percentage gap. The numbers above the chart count how many methods produced a feasible solution. This same information is also conveyed visually: a darker background corresponds to an instance for which fewer algorithms produce a feasible solution. Finally, Table 1 summarises the number of instances solved by each method. We do not report solution times because, except for a few exceptions (when the gap is zero in Figure 7), all algorithms reach the 1-hour time limit.

Overall, we highlight two main results. First, despite the difficulty of the problem, column generation methods that use two exponential sets of variables can find feasible solutions for most instances. By contrast, the CF and SCF fail to do so for most instances with 30 or more customers. Second, for instances solved by all methods, our SCF usually achieves the lowest gaps. There are only a few instances for which the CF obtains the lowest gap uncontested.

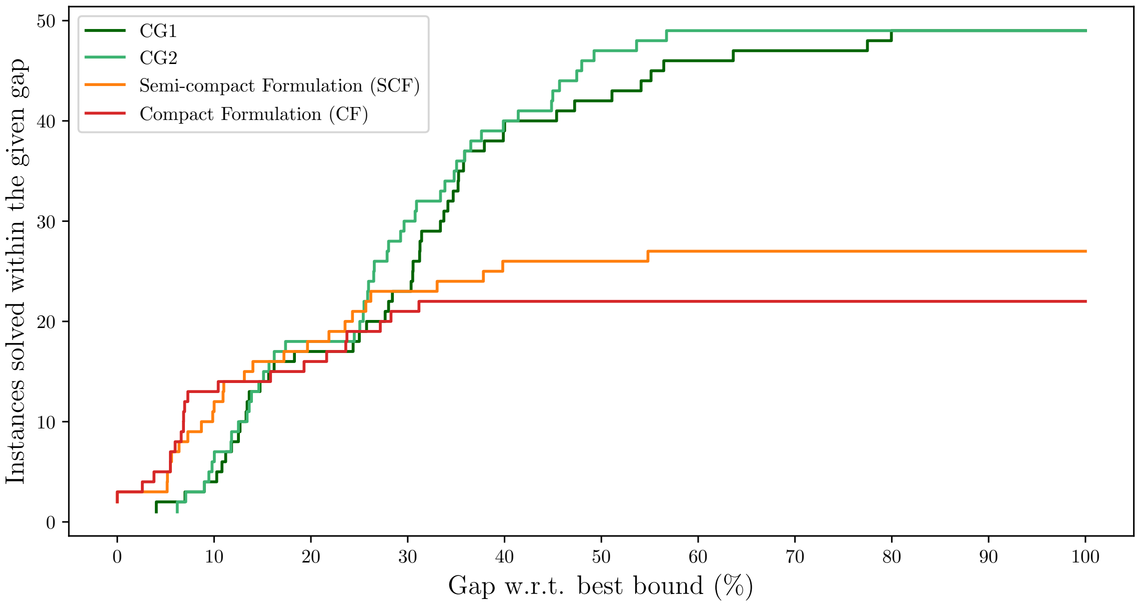


Figure 7: Performance profiles of the four tested algorithms.

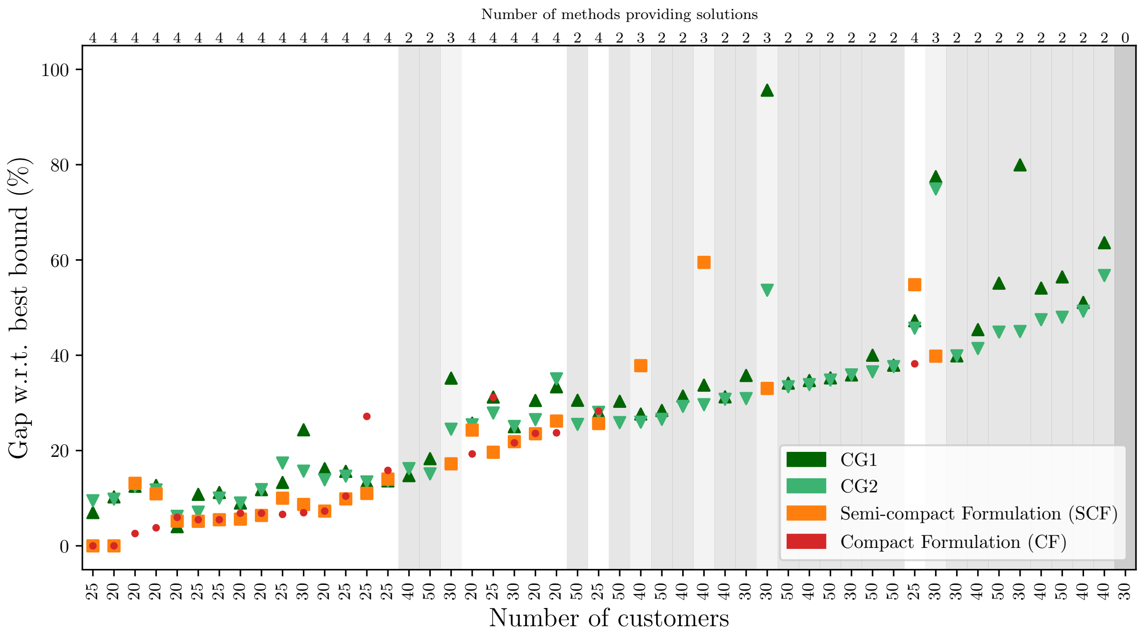


Figure 8: Disaggregated results for the four tested algorithms.

| | # of customers | | | | | |
|---|---|---|---|---|---|---|
| Method | 20 | 25 | 30 | 40 | 50 | Total |
| CF | 10 | 10 | 2 | 0 | 0 | 22 |
| SCF | 10 | 10 | 5 | 2 | 0 | 27 |
| CG1 | 10 | 10 | 9 | 10 | 10 | 49 |
| CG2 | 10 | 10 | 9 | 10 | 10 | 49 |

Table 1: Number of instances in which each method finds a feasible solution. The instances are grouped by number of customers.

| | % Improvement vs. | | | | % Improvement vs. | | |
|---|---|---|---|---|---|---|---|
| Ins | CF | DH | Best | Inst | CF | DH | Best |
| 1 | 0.00 | 0.00 | **0.00** | 13 | — | -6.22 | -6.22 |
| 2 | 0.00 | 5.43 | **0.00** | 14 | — | 3.06 | **3.06** |
| 3 | 0.00 | 8.37 | **0.00** | 15 | — | 4.45 | **4.45** |
| 4 | 0.00 | 5.13 | **0.00** | 16 | — | 10.17 | **10.17** |
| 5 | 3.94 | 8.68 | **3.94** | 17 | — | 9.85 | **9.85** |
| 6 | 0.00 | 12.36 | **0.00** | 18 | — | 3.86 | **3.86** |
| 7 | 7.27 | 10.86 | **7.27** | 19 | — | 6.99 | **6.99** |
| 8 | -0.02 | 8.52 | -0.02 | 20 | — | 10.10 | **10.10** |
| 9 | 6.00 | 10.03 | **6.00** | 21 | — | 11.50 | **11.50** |
| 10 | — | 2.35 | **2.35** | 22 | — | 14.78 | **14.78** |
| 11 | 4.77 | 10.41 | **4.77** | 23 | — | 5.92 | **5.92** |
| 12 | 11.16 | 5.06 | **5.06** | 24 | — | 5.35 | **5.35** |

| | Comparison vs. | | |
|---|---|---|---|
| | CF | DH | Best |
| # Better cost: | 18 | 22 | 17 |
| # Worse cost: | 1 | 1 | 2 |
| # Same cost: | 5 | 1 | 5 |

Table 2: Left: relative improvements of the solutions obtained by our methods on the 24 instances from Mandal and Archetti 2023, compared with the CF, DH, and the best between CF and DH. Right: summary of the number of instances in which our methods find a solution with a better, worse, or the same cost with respect to CF, to DH and to the best between these.

## 7.2 Comparison with the state of the art

We compare the performance of our methods against the decomposition heuristic methods proposed in Mandal and Archetti 2023, denoted with DH. In the aforementioned paper, the authors devise several DHs, and none dominates over the others. We use the best solution reported by Mandal and Archetti (2023) and, for a fair comparison, we also use the best solution obtained by our proposed methods, as detailed in Section 5 of the supplemental material. Because Mandal and Archetti (2023) use a 3-hour time limit, we apply the same timeout to our algorithms. In particular, for the SCF and CG methods, we set a maximum time of two hours for the column generation phase and, therefore, a minimum time of one hour for the final reduced master heuristic. We recall that Mandal and Archetti (2023) allow trucks to wait at in-stops and we do not. Therefore, the problems being solved are slightly different, and the solution space of Mandal and Archetti's formulation is larger than ours. This implies that the problem solved by the DHs may admit optimal solutions with lower objective values than in our formulation.

Table 2 (left) shows the results obtained on the 24 instances tested in Mandal and Archetti (2023). Column *Ins* is the instance number. Column *CF* reports the percentage objective cost improvement of the best of our methods compared with the CF. Analogously, column *DH* is the percentage improvement compared with the best of Mandal and Archetti's decomposition heuristics. Finally, column Best reports the improvement of the best of our methods compared with the best between CF and DH. An empty cell indicates that the method could not find any feasible solution for the instance. The summary table in the right of Table 2 reports the number of instances in which the best solution found by our method has better, worse or the same cost as the best solution from Mandal and Archetti 2023, either CF or DH. Absolute solution values can be found in the supplemental material of this paper.

Compared with DH, our algorithm found better solutions in 22 out of the 24 instances. The average improvement in these 22 instances is 7.87%, reaching almost 15% in the best cases. We tie in one instance (number 1), for which our method finds an optimal solution (optimality is proven because CF terminates within the time limit and finds the same solution). We find a worse solution in only one instance (number 13). However, by inspecting this instance, we verified that the best solution obtained by DH includes a route in which the truck waits at an in-stop for a large part of the planning horizon. Therefore, DH's solution is infeasible for our formulation. With respect to CF, there is just one instance (number 8) for which our solution is worse, but the cost difference is smaller than 0.1%. For the other instances, we get the same result on five, and we improve the solution for the other 18 (including the 13 instances for which CF fails to find a feasible solution).

To summarise, when we compare our results against the best result between CF and DH for each instance (last column of the right table in Table 2), we improve in 17 out of 24 instances, tie in five and find worse solutions in two cases. In one of these (instance 13), DH's solution is considered infeasible to our methods and in the other (instance 8) the cost difference is extremely small.

# 8 Conclusions

In this paper, we develop a decision support tool for a last-mile delivery system that uses spare capacity on public transport to bring parcels from an outside distribution centre into the city. In particular, we proposed extended formulations for the Three-Tier Delivery Problem using Public Transportation (3T-DPPT) introduced by Mandal and Archetti (2023). Two characteristics set the main extended formulation apart from most other models from the literature: first, the presence of two exponential sets of variables giving rise to two pricing subproblems; second, that one of these subproblems involves finding the shortest path on a graph with time-dependent costs. The complete price-and-branch algorithm involves several components. Apart from the algorithms to solve the pricing problems to optimality, we implemented an initial column generation heuristic, a column postprocessing algorithm, sparsification techniques to accelerate pricing, and dual bounding approaches based on problem decomposition and on the Lagrangian bound. Computational tests on benchmark instances show that column generation applied to the extended formulations outperforms other approaches from the literature, including a compact formulation and decomposition heuristics.

While our method produces high-quality primal solutions, more work is needed to devise better dual bounds. In small instances, we observed that the continuous relaxation of the extended formulation provides tight bounds. In larger instances, however, solving the continuous relaxation to optimality is prohibitive because of the high computation time required by exact pricing. The other bounding techniques we propose (based on decomposition and the Lagrangian bound) do not provide equally good dual bounds.

# Acknowledgements

We are extremely grateful to the authors of (Pessoa et al. 2023) for sharing with us their VrpSolver implementation of the GVRP. This work was partially funded by the CY Initiative of Excellence (grant "Investissements d'Avenir" ANR-16-IDEX-0008). The work of Alberto Santini has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement number 945380.

# References


Alfonso, Viviana et al. (Jan. 2021). *E-commerce in the pandemic and beyond*. Bulletin 36. Bank for International Settlements. URL: https://web.archive.org/web/20210126163639/https://www.bis.org/publ/bisbull36.htm (visited on 26/01/2021).

Antkowiak, Tifany (Jan. 2018). 'Le projet de tram fret est arrêté à Saint-Étienne'. In: *France Bleu*. URL: https://web.archive.org/web/20180217163752/https://www.francebleu.fr/infos/transports/le-projet-de-tram-fret-est-arrete-a-saint-etienne-1514458535 (visited on 17/02/2018).

Archetti, Claudia and Luca Bertazzi (2021). 'Recent challenges in Routing and Inventory Routing: E-commerce and last-mile delivery'. In: *Networks* 77.2, pp. 255–268. DOI: 10.1002/net.21995.

Baron, Ethan (Feb. 2019). 'Amazon looks to turn public buses into mobile delivery stations'. In: *Star Tribune*. URL: https://web.archive.org/web/20190811055124/http://www.startribune.com/amazon-looks-to-turn-public-buses-into-mobile-delivery-stations/505136612/ (visited on 11/08/2019).

Baum, Moritz et al. (2020). 'Modeling and engineering constrained shortest path algorithms for battery electric vehicles'. In: *Transportation Science* 54.6, pp. 1571–1600.

Beasley, John Edward and Nicos Christofides (1989). 'An algorithm for the resource constrained shortest path problem'. In: *Networks* 19 (4), pp. 379–394. DOI: 10.1002/net.3230190402.

Beckers, Joris et al. (2021). 'COVID-19 and retail: The catalyst for e-commerce in Belgium?' In: *Journal of Retailing and Consumer Services* 62. DOI: 10.1016/j.jretconser.2021.102645.

Bhatt, Kushal Mukesh (2019). *Mobile Pickup Locations*. US Patent n. US10192189B2, holder: Amazon Technologies, Inc.

Boysen, Nils, Stefan Fedtke and Stefan Schwerdfeger (2021). 'Last-mile delivery concepts: a survey from an operational research perspective'. In: *OR Spectrum* 43, pp. 1–58. DOI: 10.1007/s00291-020-00607-8.

Bruzzone, Francesco, Federico Cavallaro and Silvio Nocera (2021). 'The integration of passenger and freight transport for first-last mile operations'. In: *Transport Policy* 100, pp. 31–48. DOI: 10.1016/j.tranpol.2020.10.009.

Caspersen, Elise (2021). 'Freight trip generation and consumer preferences for reducing externalities from last mile deliveries'. PhD Dissertation. Norwegian University of Life Sciences.

Clinnick, Richard (Apr. 2020). 'GB Railfreight uses electric commuter unit for London freight trial'. In: *Rail Magazine*. URL: https://web.archive.org/web/20201201151339/https://www.railmagazine.com/news/network/gbrf-runs-first-uk-freight-emu-trial (visited on 01/12/2020).

Delle Donne, Diego, Laurent Alfandari et al. (2023). 'Freight-on-Transit for urban last-mile deliveries: A strategic planning approach'. In: *Transportation Research Part B: Methodological* 169, pp. 53–81. DOI: 10.1016/j.trb.2023.01.004.

Delle Donne, Diego, Alberto Santini and Claudia Archetti (2024). *Instances and Results for the Paper "Integrating Public Transport in Sustainable Last-mile Delivery: Column Generation Approaches"*. DOI: 10.5281/zenodo.10493847. URL: https://github.com/alberto-santini/public-transport-lmd.

Deloison, Thomas et al. (Jan. 2020). *The future of the last-mile ecosystem*. Tech. rep. World Economic Forum. URL: https://web.archive.org/web/20210327151647/https://www.weforum.org/reports/the-future-of-the-last-mile-ecosystem (visited on 27/03/2021).

Der Spiegel (Feb. 2020). 'Scheuer will Paketauslieferung per U-Bahn testen'. In: *Der Spiegel*. URL: https://web.archive.org/web/20201130023547/https://www.spiegel.de/wirtschaft/unternehmen/scheuer-will-paketauslieferung-per-u-bahn-testen-a-ac3bd289-92d9-4c31-a969-4dbd8b504712 (visited on 30/11/2020).

Edrington, Suzie et al. (Mar. 2017). *Using public transportation to facilitate last mile delivery package delivery*. Guidebook TxDOT-0-6891-P2. Texas A&M Transportation Institute and Texas Department of Transportation. URL: https://web.archive.org/web/20170910010040/https://static.tti.tamu.edu/tti.tamu.edu/documents/0-6891-P2.pdf (visited on 10/09/2017).

Ghilas, Veaceslav, Emrah Demir and Tom Van Woensel (2016a). 'A scenario-based planning for the pickup and delivery problem with time windows, scheduled lines and stochastic demands'. In: *Transportation Research Part B: Methodological* 91, pp. 34–51. DOI: 10.1016/j.trb.2016.04.015.

Ghilas, Veaceslav, Emrah Demir and Tom Van Woensel (2016b). 'An adaptive large neighborhood search heuristic for the Pickup and Delivery Problem with Time Windows and Scheduled Lines'. In: *Computers & Operations Research* 72, pp. 12–30. DOI: 10.1016/j.cor.2016.01.018.

Guthrie, Cameron, Samuel Fosso-Wamba and Jean Brice Arnaud (2021). 'Online consumer resilience during a pandemic: An exploratory study of e-commerce behavior before, during and after a COVID-19 lockdown'. In: *Journal of Retailing and Consumer Services* 61. DOI: 10.1016/j.jretconser.2021.102570.

He, Yunzhu and Zhongzhen Yang (2018). 'Parcel delivery by collaborative use of truck fleets and bus-transit vehicles'. In: *Transportation Journal* 57.4, pp. 399–428. DOI: 10.5325/transportationj.57.4.0399.

Hörsting, Lena and Catherine Cleophas (2023). 'Scheduling shared passenger and freight transport on a fixed infrastructure'. In: *European Journal of Operational Research* 306 (3), pp. 1158–1569. DOI: 10.1016/j.ejor.2022.07.043.

Irnich, Stefan and Guy Desaulniers (2005). 'Shortest Path Problems with Resource Constraints'. In: *Column Generation*. Ed. by Guy Desaulniers, Jacques Desrosiers and Marius Solomon. Springer, pp. 33–65. DOI: 10.1007/0-387-25486-2_2.

Kikuta, Jun et al. (2012). 'New subway-integrated city logistics system'. In: *Procedia Social and Behavioral Sciences* 39, pp. 476–489. DOI: 10.1016/j.sbspro.2012.03.123.

Klein, Patrick Sean and Maximilian Schiffer (2022). 'Electric vehicle charge scheduling with flexible service operations'. In: *arXiv preprint arXiv:2201.03972*.

Longhorn, Danny (July 2021). 'Passenger trains converted to deliver parcels to city centres'. In: *Rail Business Daily*. URL: https://web.archive.org/web/20210707152625/https://news.railbusinessdaily.com/passenger-trains-converted-to-deliver-parcels-to-city-centres/ (visited on 07/07/2021).

Majoral, Genís, Francesc Gasparín and Sergi Saurí (2021). 'Reucing e-commerce delivery externalities with taxation. Application to Barcelona.' In: *Proceedings of the Transportation Research Board 100$^{th}$ Annual Meeting, 05–29 January 2021, Washington DC, United States*. Transportation Research Board 100$^{\text{th}}$ Annual Meeting (5th–29th Jan. 2021). Washington DC, United States, pp. 642–655. DOI: 10.1177/03611981211012412.

Mandal, Minakshi Punam and Claudia Archetti (2023). *A Decomposition Approach to Last Mile Delivery Using Public Transportation Systems*. arXiv: 2306.04219 [math.OC].

Masson, Renaud et al. (2017). 'Optimization of a city logistics transportation system with mixed passengers and goods'. In: *EURO Journal on Transportation and Logistics* 6 (1), pp. 81–109. DOI: 10.1007/s13676-015-0085-5.

Pessoa, Artur et al. (2023). 'A unified exact approach for Clustered and Generalized Vehicle Routing Problems'. In: *Computers & Operations Research* 149. DOI: 10.1016/j.cor.2022.106040.

Sadykov, Ruslan et al. (2019). 'Primal heuristics for branch and price: The assets of diving methods'. In: *INFORMS Journal on Computing* 31.2, pp. 251–267. DOI: 10.1287/ijoc.2018.0822.

Saito, Shigeo and Yomiuri Shimbun (Feb. 2021). 'Next stop, package delivery: bus and courier link up to serve rural area'. In: *The Japan News*. URL: `https://web.archive.org/web/20210219030006/https://the-japan-news.com/news/article/0007140779` (visited on 19/02/2021).
Segura, Vicente et al. (Feb. 2020). *Last mile logistics: challenges and solutions in Spain.* Tech. rep. Deloitte. URL: `https://web.archive.org/web/20210802053540/https://www2.deloitte.com/content/dam/Deloitte/es/Documents/operaciones/deloitte-es-operations-last-mile.pdf` (visited on 02/08/2021).
Silva, Vasco, António Amaral and Tânia Fontes (2023). 'Sustainable Urban Last-Mile Logistics: A Systematic Literature Review'. In: *Sustainability* 15.3. DOI: `10.3390/su15032285`.
Sustainable Bus (July 2021). 'What if trams were also to deliver goods? RegioKargo project launched in Karlsruhe'. In: *Sustainable Bus*. URL: `https://web.archive.org/web/20210708123229/https://www.sustainable-bus.com/news/trams-deliver-goods-regiokargo-karlsruhe-project/` (visited on 08/07/2021).
Tagmouti, Mariam, Michel Gendreau and Jean-Yves Potvin (2007). 'Arc routing problems with time-dependent service costs'. In: *European Journal of Operational Research* 181.1, pp. 30–39.
Villa, Rafael and Andrés Monzón (2021a). 'A Metro-Based System as Sustainable Alternative for Urban Logistics in the Era of E-Commerce'. In: *Sustainability* 13 (8). DOI: `10.3390/su13084479`.
Villa, Rafael and Andrés Monzón (2021b). 'Mobility Restrictions and E-Commerce: Holistic Balance in Madrid Centre during COVID-19 Lockdown'. In: *Economies* 9.2 (57). DOI: `10.3390/economies9020057`.
Viu-Roig, Marta and Eduard Alvarez-Palau (2020). 'The Impact of E-Commerce-Related Last-Mile Logistics on Cities: A Systematic Literature Review'. In: *Sustainability* 12.16. DOI: `10.3390/su12166492`.
Wang, Lina, Elliot Rabinovich and Harish Guda (2023). 'An analysis of operating efficiency and policy implications in last-mile transportation following Amazon's integration'. In: *Journal of Operations Management* 69 (1), pp. 9–35. DOI: `10.1002/joom.1172`.
Zhou, Fangting and Jin Zhang (2020). 'Freight Transport Mode Based on Public Transport: Taking Parcel Delivery by Subway as an Example'. In: *Proceedings of the 6th International Conference on Transportation Engineering, 20–22 September 2019, Chengdu, China.* 6th International Conference on Transportation Engineering (20th–22nd Sept. 2019). Chengdu, China, pp. 745–754.

# A Additional notation

We first describe how we build set $\mathcal{S}^{\text{in}}_c$ introduced in Section 4, i.e., the set of in-stops that a truck can use to unload a parcel for customer $c \in \mathcal{C}$. Given two stops $s_1 \in \mathcal{S}^{\text{in}}$ and $s_2 \in \mathcal{S}^{\text{out}}$, let the indicator parameter $\beta_{s_1 s_2} \in \{0, 1\}$ take value 1 iff a bus route links the two stops. Then, for each customer $c \in \mathcal{C}$, we define the following.

For each out-stop $s_2 \in \mathcal{S}^{\text{out}}_c$, feasible arrival times of $c$'s parcel at $s_2$ are constrained by the customer's time window, the maximum wait time at the stop and the maximum courier route time. We denote this set of feasible arrival times as $\Theta_{s_2 c}$:

$$\Theta_{s_2 c} = \big[\underline{T}_c - T_{s_2} - W^{\max} - (L^{\max} - l_{c s_2}), \bar{T}_c - T_{s_2} - l_{s_2 c}\big].$$

Feasible times $\Theta_{s_2 c}$ constrain the set of buses which can carry $c$'s parcel, when the parcel is delivered starting from $s_2$. We denote with $\mathcal{P}^{\text{out}}_{s_2 c}$ such set:

$$\mathcal{P}^{\text{out}}_{s_2 c} = \big\{p \in \mathcal{P}_{s_2} \ : \ t^{s_2}_p \in \Theta_{s_2 c}\big\},$$

where $\mathcal{P}_{s_2}$ is the set of buses which serve stop $s_2$. Moving backwards to the first tier, we can then consider the feasible arrival times of a parcel at an in-stop $s_1 \in \mathcal{S}^{\text{in}}$ if the parcel must then travel on bus $p$. These times are constrained by the maximum wait time at the in-stop and the minimum time it takes to travel from the CDC to $s_1$:

$$\Theta_{s_1 p} = \left[\max\{t_p^{s_1} - T_{s_1} - W^{\max}, l_{os_1}\}, t_p^{s_1} - T_{s_1}\right].$$

The set of feasible arrival times for a parcel $c \in \mathcal{C}$ at in-stop $s_1 \in \mathcal{S}_c^{\text{in}}$ is similarly defined as

$$\Theta_{s_1 c} = \bigcup_{p \in \mathcal{P}_c} \Theta_{s_1 p},$$

where $\mathcal{P}_c$ is defined below. We can then denote with $\mathcal{S}_{s_2 c}^{\text{in}}$ the set of in-stops which a bus can use to unload a parcel which will be delivered to customer $c$ starting from out-stop $s_2$:

$$\mathcal{S}_{s_2 c}^{\text{in}} = \left\{s_1 \in \mathcal{S}^{\text{in}} \ : \ \beta_{s_1 s_2} = 1 \text{ and } \exists p \in \mathcal{P}_{s_2 c}^{\text{out}} \text{ s.t. } \Theta_{s_1 p} \neq \emptyset\right\}.$$

Finally, the set of in-stops that a truck can use to unload a parcel bound to customer $c$ is

$$\mathcal{S}_c^{\text{in}} = \bigcup_{s_2 \in \mathcal{S}_c^{\text{out}}} \mathcal{S}_{s_2 c}^{\text{in}}.$$

Next, we characterise set $\mathcal{R}_{s_1 pc}^{\text{D}}$ introduced in Section 4. We can first determine a subset of buses $\mathcal{P}_c \subseteq \mathcal{P}$, which can carry the parcel to customer $c$. Each bus in this set must reach an out-stop from where the parcel can reach $c$'s location and must do so at a time which is compatible with $c$'s time windows once the maximum wait time at the out-stop, the courier travel time and the courier maximum route duration have been taken into account. Therefore, we define

$$\begin{aligned}\mathcal{P}_c = \big\{p \in \mathcal{P} \mid \exists s_2 \in \mathcal{S}_p \cap \mathcal{S}_c^{\text{out}} \ : \ & t_p^{s_2} + T_{s_2} + W^{\max} + L^{\max} \geq \underline{T}_c \text{ and} \\ & t_p^{s_2} + T_{s_2} + l_{cs_2} \leq \bar{T}_c\big\}.\end{aligned}$$

We also consider the set of buses which can carry customer $c$'s parcel when a truck unloads it at in-station $s_1 \in \mathcal{S}_c^{\text{in}}$. This set is simply $\mathcal{P}_{s_1 c}^{\text{in}} = \mathcal{P}_{s_1} \cap \mathcal{P}_c$. We define $\mathcal{R}_{s_1 pc}^{\text{D}}$ for each customer $c$, each in-stop $s_1 \in \mathcal{S}_c^{\text{in}}$ and each bus $p \in \mathcal{P}_{s_1 c}^{\text{in}}$ as:

$$\mathcal{R}_{s_1 pc}^{\text{D}} = \left\{r \in \mathcal{R}_c^{\text{D}} \mid t_p^{s_1} - W^{\max} \leq t_r^{s_1} \leq t_p^{s_1}\right\},$$

where $t_r^{s_1}$ indicates the arrival time of the truck arrives at in-stop $s_1$ in route $r$, plus the handling time $T_{s_1}$. The above definition ensures that the truck arrives at the stop before the bus (leaving enough time for parcel handling) but not so early that it violates the maximum parcel wait time.

We now characterise set $\mathcal{R}_{ps_2 c}^{\text{F}}$ for each customer $c \in \mathcal{C}$, each out-stop $s_2 \in \mathcal{S}_c^{\text{out}}$, and each bus $p \in \mathcal{P}_{s_2 c}^{\text{out}}$:

$$\mathcal{R}_{ps_2 c}^{\text{F}} = \left\{r \in \mathcal{R}_{s_2}^{\text{F}} \cap \mathcal{R}_c^{\text{F}} \mid t_p^{s_2} \leq t_r^{s_2} \leq t_p^{s_2} + W^{\max}\right\},$$

where $t_r^{s_2}$ denotes the start time of the courier from out-stop $s_2$ in route $r$, plus the handling time $T_{s_2}$. The above definition ensures that the bus arrives at the out-stop before the courier leaves (leaving enough time for parcel handling) but not so early that it violates the maximum parcel wait time. Finally, we introduce the following convenient notation for the set of stops at which a parcel can be picked up and, respectively, delivered by a given bus:

$$\mathcal{S}_{pc}^{\text{in}} = \mathcal{S}_p \cap \mathcal{S}_c^{\text{in}}, \quad \mathcal{S}_{pc}^{\text{out}} = \mathcal{S}_p \cap \mathcal{S}_c^{\text{out}}.$$

# B Details of the route generation subproblems

## B.1 Labelling algorithm introduced in Section 5.1.2

As for the labelling algorithm on graph $H$, we associate a label $L_P$ to each partial path $P$ from the CDC to a vertex $(s,c,\ell) \in W$. The label has the following components:

- $v_P = (s,c,\ell)$ is the end vertex of the path.
- $\mathcal{S}_P \subseteq \mathcal{S}^{\text{in}}$ is the set of in-stops which can still be visited when departing from $v_P$.
- $\mathcal{C}_P \subseteq \mathcal{C}$ is the set of customers whose parcels can still be delivered when departing from $v_P$.
- $\tau_P \geq 0$ is the traversing time of the path, i.e., the sum of traversing times of the arcs used in the path.
- $Q_P \geq 0$ is the spare capacity on the truck when departing from $v_P$.
- $C_P \in \mathbb{R}$ is the reduced cost associated with the path.

The relevant differences with the algorithm presented in Section 5.1.1 are the presence of time windows on the vertices $(s,c,\ell)$, implicitly defined by the intervals $w_{s\ell}$, and the scalar cost $C_P$. Time windows are easily accounted for using a time resource and the corresponding resource windows (Beasley and Christofides 1989). Regarding the cost, when path $P$ is extended to a new vertex $(s',c',\ell') \in W$ along arc $b \in B$, $C_P$ is updated as follows:

$$C_{P'} = C_P + c_b + \lambda^{(1i)}_{c'} + \sum_{\substack{p \in \mathcal{P}_{s'c'} \text{ s.t.} \\ I_{s'p} \cap w_{s'\ell'} \neq \emptyset}} \lambda^{(1g)}_{s'pc'}.$$

## B.2 Labelling algorithm introduced in Section 5.2

Given an out-stop $s \in \mathcal{S}^{\text{out}}$, and an interval $w_{s\ell}$, we first define the complete, simple, directed graph $G_{s\ell}$ used to solve the shortest-path problem mentioned in Section 5.2. The vertex set, denoted $V_{s\ell}$, consists of $s$ and all customers which can be served from $s$ starting at interval $\ell$:

$$V_{s\ell} = \left\{ c \in \mathcal{C} \ : \ s \in \mathcal{S}^{\text{out}}_c, \ \Theta_{sc} \cap w_{s\ell} \neq \emptyset \right\},$$

where $\mathcal{S}^{\text{out}}_c$ and $\Theta_{sc}$ are defined, respectively, in Section 3 and A. A feasible route corresponds to a path in $G_{s\ell}$ which: (i) starts from $s$ at a time contained in interval $w_{s\ell}$; (ii) ends in $s$; (iii) visits each other vertex of $V_{s\ell}$ at most once; (iv) respects the courier capacity $Q^{\text{F}}$; (v) has a maximum travel time, defined as the sum of the arc traversing times, of $L^{\max}$; (vi) visits each vertex $c$ not later than $\bar{T}_c$. As is standard in the vehicle routing literature, we allow a courier to visit customer $c$ before the beginning of $c$'s time window (time $\underline{T}_c$) but, in that case, the courier must wait until $\underline{T}_c$ before performing the delivery. The cost of a path in $G_{s\ell}$ is equal to the sum of the costs of the used arcs, minus dual prizes

$$\lambda^{(1j)}_c + \sum_{\substack{p \in \mathcal{P}^{\text{out}}_{sc} \\ t^s_p + T_s \in w_{s\ell}}} \lambda^{(1h)}_{spc}$$

collected at each visited customer $c$, plus the constant dual price $\lambda^{(1c)}_s$.

As mentioned in Section 5.2, the problem of finding the shortest path in $G_{s\ell}$ corresponding to a feasible courier route is a resource-constrained shortest-path problem. As is common in the literature, elementarity is ensured by associating a binary resource with each customer. To respect capacity, time windows, and maximum duration constraints, we further introduce their respective continuous resources. Finally, we solve $\mathrm{SP}_y(s,\ell)$ with a labelling algorithm with the usual dominance rules. We refer the reader to (Irnich and Desaulniers 2005) for a thorough introduction to labelling algorithms for resource-constrained shortest-path problems.

# C Semi-compact formulation

We present the complete semi-compact formulation for 3T-DPPT introduced in Section 4. Using notation $\bar{\mathcal{S}}^{\text{in}} = \mathcal{S}^{\text{in}} \cup \{o\}$ and $\bar{\mathcal{S}}^{\text{in}}_c = \mathcal{S}^{\text{in}}_c \cup \{o\}$, the three sets of variables we use to replace variables $x$ are as follows. First, $w_{dij} \in \{0,1\}$, taking value one if and only if truck $d \in \mathcal{D}$ travels from $i$ to $j$ ($i, j \in \bar{\mathcal{S}}^{\text{in}}$. Second, $\pi_{ds} \in \mathbb{R}$, indicating the time when truck $d \in \mathcal{D}$ completes parcel delivery at stop $s \in \mathcal{S}^{\text{in}}$. (We also use variable $\pi_{do}$ to indicate the return time at the CDC.) Finally, $\delta_{dcs} \in \{0,1\}$, taking value one if and only if truck $d \in \mathcal{D}$ delivers parcel $c \in \mathcal{C}$ at stop $s \in \mathcal{S}^{\text{in}}$. In the following, $c_{ij}$ is the travel cost of a truck driving from $i$ to $j$ and $M_s$ (for $s \in \mathcal{S}^{\text{in}}$) is a big constant, e.g., the last bus arrival time at stop $s$.

$$\min \sum_{d\in\mathcal{D}} \sum_{\substack{i,j\in\bar{\mathcal{S}}^{\text{in}}_c \\ i\neq j}} c_{ij} w_{dij} + \sum_{r\in\mathcal{R}^{\text{F}}} c_r y_r \tag{5a}$$

$$\sum_{s\in\bar{\mathcal{S}}^{\text{in}}} w_{dos} = 1 \qquad \forall d \in \mathcal{D} \tag{5b}$$

$$\sum_{i\in\bar{\mathcal{S}}^{\text{in}}\setminus\{s\}} w_{dis} = \sum_{i\in\bar{\mathcal{S}}^{\text{in}}\setminus\{s\}} w_{dsi} \qquad \forall d \in \mathcal{D},\ \forall s \in \mathcal{S}^{\text{in}} \tag{5c}$$

$$\sum_{s\in\bar{\mathcal{S}}^{\text{in}}} w_{dso} = 1 \qquad \forall d \in \mathcal{D} \tag{5d}$$

$$\pi_{di} + l_{ij} - M_j(1 - w_{dij}) \leq \pi_{dj} - T_j \qquad \forall d \in \mathcal{D},\ \forall i \in \bar{\mathcal{S}}^{\text{in}},\ \forall j \in \mathcal{S}^{\text{in}} \setminus \{i\} \tag{5e}$$

$$\pi_{dj} - T_j \leq \pi_{di} + l_{ij} + M_j(1 - w_{dij}) \qquad \forall d \in \mathcal{D},\ \forall i \in \bar{\mathcal{S}}^{\text{in}},\ \forall j \in \mathcal{S}^{\text{in}} \setminus \{i\} \tag{5f}$$

$$\sum_{p\in\mathcal{P}_c} t^s_p z^{\text{in}}_{pcs} - W^{\max} - M_s(1 - \delta_{dcs}) \leq \pi_{ds} \qquad \forall d \in \mathcal{D},\ \forall c \in \mathcal{C},\ \forall s \in \mathcal{S}^{\text{in}}_c \tag{5g}$$

$$\pi_{ds} \leq \sum_{p\in\mathcal{P}_c} t^s_p z^{\text{in}}_{pcs} + M_s(1 - \delta_{dcs}) \qquad \forall d \in \mathcal{D},\ \forall c \in \mathcal{C},\ \forall s \in \mathcal{S}^{\text{in}}_c \tag{5h}$$

$$\delta_{dcs} \leq \sum_{i\in\mathcal{S}^{\text{in}}\cup\{o\}, i\neq s} w_{dis} \qquad \forall d \in \mathcal{D},\ \forall c \in \mathcal{C},\ \forall s \in \mathcal{S}^{\text{in}}_c \tag{5i}$$

$$\sum_{s\in\mathcal{S}^{\text{in}}_c} \sum_{d\in\mathcal{D}} \delta_{dcs} = 1 \qquad \forall c \in \mathcal{C} \tag{5j}$$

$$\sum_{p\in\mathcal{P}_{cs}} z^{\text{in}}_{pcs} = \sum_{d\in\mathcal{D}} \delta_{dcs} \qquad \forall c \in \mathcal{C},\ \forall s \in \mathcal{S}^{\text{in}}_c \tag{5k}$$

$$\sum_{c\in\mathcal{C}} \sum_{s\in\mathcal{S}^{\text{in}}_c} q_c \delta_{dcs} \leq Q_d \qquad \forall d \in \mathcal{D} \tag{5l}$$

$$\text{Constraints (1c)-(1f), (1h), (1j)} \tag{5m}$$

$$w_{dij} \in \{0,1\} \qquad \forall d \in \mathcal{D},\ \forall i,j \in \bar{\mathcal{S}}^{\text{in}} \tag{5n}$$

$$\pi_{ds} \in \mathbb{R}^+ \qquad \forall d \in \mathcal{D},\ \forall s \in \bar{\mathcal{S}}^{\text{in}} \tag{5o}$$

$$\delta_{dcs} \in \{0,1\} \qquad \forall d \in \mathcal{D},\ \forall c \in C,\ \forall s \in \mathcal{S}^{\text{in}}_c \tag{5p}$$

$$y_r \in \{0,1\} \qquad \forall r \in \mathcal{R}^{\text{F}} \tag{5q}$$

$$z_{spc}^{\text{in}} \in \{0,1\} \qquad \forall c \in C,\ \forall s \in \mathcal{S}_c^{\text{in}},\ \forall p \in \mathcal{P}_{sc}^{\text{in}} \tag{5r}$$

$$z_{spc}^{\text{out}} \in \{0,1\} \qquad \forall c \in C,\ \forall s \in \mathcal{S}_c^{\text{out}},\ \forall p \in \mathcal{P}_{sc}^{\text{out}}. \tag{5s}$$

Constraints (5b)–(5d) are flow conservation constraints. Constraints (5e) and (5f) set the arrival times $\pi_{ds}$ of trucks to stops. Constraints (5g) and (5h) ensure that these times (and the corresponding delivered parcels) are compatible with the bus arrival times. Constraint (5i) forbids delivering a package unless the corresponding truck visits the stop. Constraint (5j) ensures that each package is delivered to one stop by one truck. Constraint (5k) ensures that each package is collected by a bus at the same stop where it was delivered. Finally, constraint (5l) asserts that the truck capacities are respected.