Int. J. Production Economics 279 (2025) 109463

Contents lists available at ScienceDirect

International Journal of Production Economics

journal homepage: www.elsevier.com/locate/ijpe

Optimizing Attended Home Delivery: Multiple recovery options and customer availability profiles to face synchronization failures

Valentina Bonomi [a], Daniele Manerba [b], Renata Mansini [b], Roberto Zanotti [c,*]

[a] CEGIST, Instituto Superior Tecnico, Universidade de Lisboa, av. Rovisco Pais, 1049-001, Lisbon, Portugal
[b] Department of Information Engineering, Università degli Studi di Brescia, via Branze 38, 25123, Brescia, Italy
[c] Department of Clinical and Experimental Sciences, Università degli Studi di Brescia, viale Europa 11, 25123, Brescia, Italy

ARTICLE INFO

Keywords:
Last-mile delivery
Attended Home Delivery
Customer Availability Profiles
Recovery options
Multiple visits
Parcel lockers

ABSTRACT

In the growing sector of Attended Home Delivery, unsynchronized deliveries between couriers and recipients affect both customers' satisfaction and companies' costs. Hence, reducing such failures improves companies' service quality and logistics efficiency. To address this issue, we study an Attended Home Delivery Problem with Recovery Options (AHDPRO) in which customers specify their probability of being at home during different timeslots of the day and their preferred recovery option in case of a synchronization failure. The options include leaving the package in a predefined safe location, bringing it to a generic collection point, or scheduling a second delivery attempt. Each alternative involves different costs and, in most cases, additional operational decisions. The AHDPRO aims to complete all customer deliveries while minimizing overall routing times as well as the overall penalty due to the recovery actions implemented and weighted by the probability of a synchronization failure to occur. We propose a branch-and-cut algorithm, including valid inequalities and heuristic procedures, to solve a Mixed-Integer Linear Programming model based on an expanded graph. Using the developed method as a tool for evaluating costs and operations, we conduct an experimental campaign on scenarios adapted from the literature involving lexicographic-based optimization procedures able to address the multiple attributes of the solutions. The results obtained allow us to assess the impact of the different recovery options on the optimal solutions and their values. Additionally, the results yield several managerial insights for companies operating in the Attended Home Delivery sector, such as the timeslot length, perceived service quality, and other key operational factors contributing to efficient planning and improved customer satisfaction.

1. Introduction

Over the recent years, the restrictions due to the COVID-19 pandemic have fostered further growth in the already booming online retail market. According to the 2023 European E-Commerce Report (Lone et al., 2023), e-commerce in Europe has experienced a 47% growth in B2C sales compared to 2019 and, in the same period, a growth of around 8% of internet users that purchased at least a product or service online. This pandemic-driven shift from in-person to online shopping has allowed (sometimes forced) more companies to enter the market by offering home delivery services. Moreover, given the growing competition in this sector, e-business companies must focus on developing highly efficient, customer-satisfaction-oriented services. The leading business model in the field is the classical *unattended* service, in which the customer is not required to be at home for the delivery. However, especially over the past 15 years, there has been growing interest in the Attended Home Delivery (AHD) business model where customer presence is required to complete the service (see, for example, Agatz et al., 2008 or Ehmke and Mattfeld, 2012). Our work will be focused on AHD services.

The presence of a customer could be necessary for many reasons linked to the type of items delivered. For example, the AHD paradigm has become increasingly popular for delivering items like e-groceries, high-end electronics, and large pieces of furniture (see Côté et al., 2024). In addition, many companies have started to provide AHD services, even for non-special items, to guarantee higher customer satisfaction. Anyway, AHD requirements prominently bring the time dimension into the delivery setting since the couriers and the customers must be synchronized for the service. In the most common case, the customer selects (at the time of the online purchase) a specific timeslot during which the service will take place based on his availability to attend the delivery. Then, at the operational level, the company tries

* Corresponding author.
E-mail address: roberto.zanotti@unibs.it (R. Zanotti).

https://doi.org/10.1016/j.ijpe.2024.109463
Received 30 April 2024; Received in revised form 1 November 2024; Accepted 3 November 2024
Available online 9 November 2024

to provide a routing and scheduling plan that satisfies the delivery demand of the customers in the desired timeslot while minimizing the service costs. Hence, given all the delivery requests to be satisfied during the current working day, this decisional setting boils down to solving a so-called *Vehicle Routing Problem with Time Windows* (VRPTW) or related variants.

However, in real environments, ensuring successful (i.e., synchronized) deliveries is by no means trivial for many reasons. First, customers tend to move around during the day (e.g., to run errands), so they might not be at home when the courier arrives despite their timeslot choice. In this regard, since the actual stop duration of the courier at the customer's home could be very short to stick with its schedule, synchronization is particularly critical, even if the timeslot length is reasonable. Second, especially in urban areas, uncertain travel times (due to traffic jams, congestion, or unexpected events) or some necessary schedule flexibility might prevent the courier from being on time for the selected timeslot of a customer. Hence, in practice, the courier and the customer are frequently not synchronized and the delivery fails, thus involving some recovery strategy for the company to accommodate the service anyway. Failed deliveries are, in general, very costly for the company and time-consuming for couriers (Song et al., 2009). Moreover, they negatively impact customers' satisfaction since the delivery does not occur when and as expected (e.g., in Amorim et al., 2024, the authors showed that the precision and timing of the delivery process are often considered more important by the customers than the expected time between the order placement and its delivery). So, to effectively handle a problem involving service to customers, it is important to include both the service provider and the customers' perspectives in the optimization model (see, e.g., Bonomi et al., 2023). AHD providers cannot indeed disregard the consideration of synchronization failures or successes when they want to control their costs but also when they care about the perceived quality of their service. In particular, it is fundamental to evaluate the impact on the costs of different operational recovery options to implement in case of failures.

To address all these aspects, we introduce a VRPTW variant called the *Attended Home Delivery Problem with Recovery Options* (AHDPRO). Unlike the standard VRPTW, where exactly one delivery timeslot is associated with each customer, we adopt a more general approach by considering the so-called *Customer Availability Profiles* (CAPs). These profiles represent the probability of a customer being present at home during the different timeslots that compose the working day (Florio et al., 2018). Moreover, each customer chooses a *recovery option*, i.e., the operative actions the carrier must perform in case of a failed delivery. In particular, we concurrently consider three recovery options:

- delivering the uncollected package to a predefined *safe place* near the customer's home (e.g., a trustful neighbor or a personal in-box);
- scheduling a *second attempt* of the delivery during an alternative timeslot, to increase the chance of finding the customer at home;
- relocating the package to a predefined *collection point*, such as an automated parcel locker or an in-store shared pick-up point, not necessarily close to the customer's home.

Given the described options, our AHDPRO seeks to minimize delivery costs while explicitly accounting for the additional operations required by recovery actions and penalties for failed deliveries. More precisely, we assume that i) recovery actions are always performed, somehow embracing a worst-case perspective, and (ii) the penalty incurred is proportional to the probability of not finding the customer at home during the timeslot in which the delivery is scheduled and depends on the selected recovery option. The lower the probability of a successful delivery according to the customer's preference, the higher the penalty. Such penalties, representing customers' dissatisfaction with a missed delivery, will be modeled as additional costs for the customers to retrieve their package from the place where it has been eventually relocated (a safe place, the depot, or a collection point). In fact, the company should compensate for such costs.

The complexity of the resulting problem led us to propose a tailored branch-and-cut approach based on a reformulation of the Mixed-Integer Linear Programming (MILP) problem and incorporating several valid inequalities, and primal heuristic procedures to easily provide upper bounds. In particular, the model exploits a smart problem representation over an ad-hoc expanded graph that allows us to obtain a compact two-index vehicle-flow formulation including additional scheduling variables to eliminate subtours in the solutions. The framework efficiently provides optimal solutions for realistic instances in a reasonable amount of time. This allows us to analyze the cost and the structure of such solutions and, in turn, to derive useful managerial insights for the underlying operations.

The contribution of this paper is three-fold. First, we introduce a novel VRPTW variant named AHDPRO that explicitly addresses, together with the classical cost minimization, the quality of the service by minimizing the penalty costs of the recovery actions implemented to accommodate synchronization failures. To the best of our knowledge, this is the first time multiple recovery options are integrated into an AHD environment and assessed against the presence of CAPs. Second, we propose a tailored branch-and-cut approach designed to serve as an evaluation tool to assess the features of the AHDPRO optimal solutions under different operating scenarios. Third, useful managerial insights are derived through an experimental campaign and the analysis of many interesting Key Performance Indicators (KPIs) under different perspectives.

Some important assumptions on our optimization setting must be clarified before going further. First, we assume that each customer is associated with a recovery option, chosen at the time of the online purchase or possibly set as a user account preference of an online shopping platform. Such a selection may lead to varying prices for the customer due to differences in service operation costs. For instance, leaving a package in a safe place incurs lower costs for the courier compared to scheduling a second delivery attempt. In this operational-level study, we do not explicitly consider the pricing aspect of the options as this is typically the focus of tactical optimization. Yet, the results of our cost analysis hold significant value for determining and accurately estimating the pricing strategy to implement. The second assumption regards the use of the CAPs probabilities to evaluate the penalty costs of the recovery actions, without resorting to any specific modeling paradigm to tackle decisions under uncertainty. We adopt instead a purely deterministic perspective and conceive the AHDPRO as a cost-assessment modeling tool that considers scenarios in which synchronization failures always happen, thus triggering recovery costs according to the favorite customer's option. Such costs always proportionally depend on the probability of not finding the customer at home. Clearly, such scenarios are inherently pessimistic since the recovery actions are not always implemented in practice, but only in response to actual failures. However, the solutions obtained by our model and the relative features (cost proportion, success rate, customers' satisfaction, etc.) could be used by the company to introduce tiered pricing or define the length of the most convenient timeslots. Finally, we assume that CAPs probabilities are strictly between 0 and 1 for any timeslot to always maintain a certain degree of uncertainty in the synchronization mechanism, regardless of whether customers are confident about being at home or not. A null presence probability of a customer during a specific timeslot would trivially eliminate it from the problem consideration, whereas a presence probability equal to 1 could lead to infallible visits, which are not meaningful in the context of our analysis.

The rest of the paper is structured as follows. In Section 2, we propose a brief analysis of the main contributions existing in the literature concerning AHD and related problems. In Section 3, we formally describe our AHDPRO detailing the logistic settings involved, whereas, in Section 4, we provide a solution framework based on the MILP model and algorithmic techniques. Section 5 describes the experimental setting adopted to obtain the results needed for our economic analysis, while in Section 6, we present and discuss the obtained results to derive managerial insights for the decisional process. Finally, in Section 7, we conclude the work by summarizing the main contributions and discussing directions for future research.

2. Literature review

Home delivery functionality is crucial for the economic success of online shopping business models. Hence, it is fundamental to provide a cost-efficient service while, at the same time, maintaining good quality of service to meet customers' expectations. As online shopping has increasingly become part of people's daily lives, customers demand flexibility and a service tailored to their needs. Thus, in addition to traditional routing issues such as the minimization of delivery costs, new logistic challenges arise from the so-called *logsumers* introduced by Norell and Gammelgaard (2020), i.e., customers who are allowed to individualize their orders deciding on aspects such as time, quality, and environmentally friendly options, directly affecting the supply chain. Therefore, the several strategies that companies adopt in responding to customers' needs pave the way for an equally varied literature (Baldi et al., 2019). The AHD paradigm, which incorporates all these aspects, has received a growing interest in the scientific literature since the seminal works by Agatz et al. (2008), which exploited its challenges and opportunities. The *Scopus* database indicates that in the last 12 years, more than 40 scientific papers have focused specifically on this topic, with 85% published in the last 5 years.

The very recent survey by Waßmuth et al. (2023) introduces a framework to categorize demand management decisions for AHD across distinct planning levels and demand management strategies. The authors systematically categorize prescriptive analytics approaches from existing literature for each planning level while pinpointing areas needing further research. Following this vision, we provide a comprehensive overview of the relevant literature concerning AHD, divided into tactical organization concerning timeslot design and pricing, operational routing aspects, and management of synchronization failures.

2.1. Tactical aspects in AHD

The core aspect of AHD is the customer's presence requirement during delivery. Companies usually schedule deliveries in narrow timeslots to provide a good quality of service and avoid synchronization failures. Managing such timeslots can involve tactical aspects, such as the construction of the slots themselves (number and length), or more operational ones, such as their allocation to customers and pricing. From a logistic point of view, timeslot length represents a great trade-off between costs and quality of service. Customers expect a narrow and reliable time window, which may lead to high delivery costs. Some of the first authors to assess the impact of slot length on transportation costs are Punakivi and Saranen (2001). Their results indicate that a completely unattended service can reduce costs by up to a third compared to an attended delivery with two-hour timeslots. The same conclusion is reached in Boyer et al. (2009), in which the authors report that offering a three-hour delivery window is 45% more expensive than an unattended delivery. Furthermore, Gevaers et al. (2011) show that the impact of time windows on costs has proved to be proportionally larger in delivering low-value customer goods to the point of suppressing the benefits of online shopping. Finally, Manerba et al. (2018) assess the impact of time window length also from a sustainability perspective. The authors show how the environmental impact in kilometers and CO_2 emissions massively increase when moving from all-day delivery to two-hour timeslots.

Regarding timeslot allocation, Campbell and Savelsbergh (2005) are the first to introduce solutions in which the service provider decides which deliveries to accept or reject and the timeslot to assign to each delivery. Ehmke and Campbell (2014) further deepens the analysis on the acceptance of deliveries in Campbell and Savelsbergh (2005) incorporating time dependent and stochastic travel time information. In Côté et al. (2024), the authors propose a Stochastic Multi-period Time Window Assignment Problem. Given the zones into which the geographical area is divided, the problem consists of assigning a predefined number of time windows to each zone while minimizing the expected traveling costs required to visit the customers within the allocated time windows, plus expected penalty costs associated with unserved customers. Time windows are allocated to zones before determining customers' requests, locations, and service times. The impact on delivery costs of strict delivery time windows and demand unpredictability can be reduced, influencing customers' behavior through timeslot pricing. Campbell and Savelsbergh (2005) propose an optimization model involving two types of incentives, such as dynamic pricing, to influence customers' timeslot choices. Yang et al. (2016) extend (Campbell and Savelsbergh, 2006) approach incorporating potential future demand when anticipating routing costs. Finally, Klein et al. (2019) seek to manage demand through a differentiated pricing policy, in which timeslot-dependent prices allow the company to construct more cost-effective delivery schedules at an operational level.

In this paper, the tactical aspects are addressed only indirectly. In particular, using our operational model, we derive a cost analysis that could guide the development of pricing, design, and allocation policies for the timeslots.

2.2. Routing optimization in AHD

At an operational level, the AHD process corresponds to a basic VRP where orders must be dispatched to a set of predefined locations with the available fleet of vehicles to minimize the overall costs. More specifically, since customers must be served in defined time windows, the problem falls under the VRPTW category. The VRPTW is $\mathcal{NP}$-Hard and deriving optimal solutions can be computationally intractable. So, researchers have mainly developed several heuristic algorithms to approximate delivery costs, while the existing exact algorithms and model formulations are reviewed in Baldacci et al. (2012).

However, since AHD services may require flexible delivery options to increase efficiency and/or customer satisfaction, many authors extended the classical VRPTW setting in this perspective. Moccia et al. (2012) propose the first formulation of a VRPTW with alternative delivery locations, also developing an incremental tabu search, while a VRP with multiple time windows, related to some furniture and electronic retail operations, is addressed in Belhaiza et al. (2014) by using a hybrid variable neighborhood tabu search heuristic. Spliet and Desaulniers (2015) present a discrete time window assignment VRP that considers for each customer a set of candidate time windows from which a single one must be selected. The authors also develop a branch-and-price algorithm and column generation heuristics to solve large instances.

Another important generalization of a VRPTW is the Vehicle Routing Problem with Delivery Options (VRP-DO), in which customers can prioritize different delivery options, and a certain customer satisfaction level has to be achieved. Tilk et al. (2021) propose a VRP-DO in which requests can be shipped to alternative capacitated locations, and the carriers need to select the most effective option for each customer. The authors also present a new branch-price-and-cut to solve the problem exactly. Finally, in the most recent application of VRPTW with delivery options, Escudero-Santana et al. (2022) offer the customers the option to suggest different timeslot combinations and delivery locations. Moreover, the authors give an exhaustive overview of the performance of various ad-hoc heuristics and metaheuristics.

2.3. Failed deliveries management

Committing to time windows could cause a significant impact on delivery costs, however, they are an inevitable occurrence in many cases, thus justifying the interest in AHD services. Interesting works have studied how to avoid or mitigate failed deliveries, or recovery from them. A non-negligible part of the literature suggests overcoming the issue of failed deliveries by exploring alternative delivery methods. Moving the package to collection points or automated lockers is the most widely adopted option. Recently, Grabenschweiger et al. (2021) promote using locker stations through customer discounts, whereas Bonomi et al. (2022) show the effectiveness of collection points in reducing the environmental impact of last-mile deliveries when customers are willing to use zero-emissions transportation modes to collect their packages. Boysen et al. (2020) make an extensive survey on delivery options, including using drones. Finally, Reyes et al. (2017) and Ozbaygin et al. (2017) propose a VRP with roaming delivery locations allowing packages to be delivered to the trunks of the customers' cars.

However, if we stick with AHD, only recently researchers have started to incorporate the probability of finding customers at home during different timeslots into their analyses to evaluate more accurately the cost of failed deliveries. The few related papers are described below, highlighting the differences from our work. Florio et al. (2018) introduce the concept of CAP to analytically handle the customer presence probability and model their problem through a set-partitioning formulation to maximize the successful deliveries. Their route creation is made a priori and evaluated using an expected loss function. Also, they allow multiple visits to the same customer during the day as a recovery option in case of delivery failure. However, differently from our work, they only consider multiple attempt options and do not optimize over the routing/scheduling costs. Using the CAPs just described, Voigt et al. (2023) proposes a *VRP with Availability Profiles* where the novel aspect is considering the trade-off between transportation and failed-delivery cost. The cost of a failed delivery is customer-dependent and proportional to the probability of finding the customer at home. Differently from our work, they do not explicitly consider recovery options, while they use penalties as proxies for the possible operational costs induced by the failures. Under this perspective, their problem results in a special case of our AHDPRO where the only option available for all the customers does not imply any additional operation and simply triggers cost penalties (e.g., our proposed option to leave packages in a *safe place*). Again, Özarık et al. (2021) study an AHD problem with uncertain customer presences. The authors compute the probability of finding the customer at home during a timeslot starting from historical data about the success of past deliveries to the same customer. However, no recovery options are taken into consideration in case of failed delivery. Finally, the work by Özarık et al. (2023) represents the only one addressing a similar setting through stochastic optimization (despite the explicit use of presence probabilities, in fact, all the previous articles present deterministic programming). The authors propose a two-stage Stochastic Programming approach in which vehicle routes and schedules are decided at the first stage, whereas recovery actions are taken in the second stage after the realization of the actual presence of the customer. However, in their problem, a penalty is paid only after two failed deliveries on the same day and it is not quantified in proportion to operational costs, as in our model.

To the best of our knowledge, this is the first work that delves into an AHD setting tackled by considering CAPs, the trade-off optimization between delivery and recovery costs, as well as the concurrent presence of multiple recovery options in case of a synchronization failure, thus making the overall solution cost and structure policy-dependent.

3. Problem description

This section outlines the Attended Home Delivery Problem with Recovery Options (AHDPRO). The logistic setting and its main features are explained in Section 3.1 while the problem statement is described in Section 3.2.

3.1. Logistic setting and assumptions

Differently from any operational AHD setting addressed as a VRPTW, our AHDPRO assumes that the synchronization between the courier and the customer fails with a certain probability and, to face such failures, introduces multiple and concurrent recovery options. In the following, these two main features are described in detail from a logistic operations perspective.

In our problem, the planning horizon (which corresponds to the working day of the couriers) is partitioned into a discrete number of consecutive and non-overlapping timeslots of equal length.[1] For each timeslot, the likelihood of the customer's presence at home must be available and it is treated as a probability. To this aim, we adopt the use of the so-called *Customer Availability Profiles* (CAPs), i.e., tailored functions that map the likelihood of a customer being present at home during various timeslots to a probability measure (Florio et al., 2018). In realistic settings, information to derive the CAP of each customer can be obtained directly from the users when they place their orders (e.g., on an online platform), estimated from historical data (van Duin et al., 2016), or even learned by using IoT technologies (Fadda et al., 2019). It is important to clarify that, originally, CAPs were presented as continuous piece-wise linear functions taking values in the interval $[0,1]$ to simulate, in a versatile way, diverse practical scenarios encountered during the delivery process and representing daily behaviors of the customers. For instance, a CAP with higher probabilities during the first, the last, and the central part of the working day mainly simulates a customer going to work in the morning and the afternoon, while returning home for lunchtime. However, to address our problem consistently, we discretize the most interesting CAPs from Florio et al. (2018) and generate profiles in which the presence probability is constant for the entire duration of a timeslot. Moreover, as already discussed in Section 1, we exclude probabilities exactly equal to 0 or 1, since scenarios not involving even a little chance of finding or missing the customers are not in line with the aim of our analysis.

The AHDPRO also includes the definition, for each customer, of a specific recovery option in case the delivery attempt fails, i.e., when the customer is not found at home. In this work, we consider three different concurrent options[2], each with a different impact on the overall costs:

- *Safe Place* (SP) option: the carrier is allowed to leave the package in a predefined safe spot around the house (the garden, a garage, a private locker, a neighbor's home, etc.) specified by the customer at the time of the order. This option does not imply additional operations impacting the vehicle schedule, apart from a slightly longer service time. However, we still need to consider the customers' dissatisfaction with receiving a package in their absence.
- *Second Attempt* (SA) option: a second visit is scheduled to the same customer on the same day, in a timeslot successive to the first attempt one. Consequently, the courier must plan more complex routes, possibly involving subtours. We remark that always performing a second visit is an assumption underlying our worst-case perspective and the proposed cost-assessment model. Finally, we also assume that packages are eventually brought back to the depot where customers not found both times can retrieve them.

[1] Although delivery patterns with overlapping or varying timeslot durations may exist, such an assumption is realistic for most existing AHD services.

[2] The three options considered may be non-exhaustive but are at least representative of the current and most common last-mile business models. However, the proposed problem and the corresponding mathematical formulation are flexible enough to accommodate new recovery options with their respective operations and costs.

- *Collection Point* (CP) option: packages are eventually moved to a shared collection point, such as a self-service parcel locker or an attended retail location. The customer must be informed of the relocation and provided with all necessary details to self-retire the package (e.g., a code for opening the locker). Such a collection point is unique for each customer and is preselected as a preference, together with the recovery option choice. From an operational point of view, we assume that the couriers store in their vehicles all the packages that must be relocated and visit the collection points only at the end of their delivery tour, just before returning to the depot. Again, note that always performing a visit to the collection point is coherent with the worst-case perspective of our setting. Finally, we assume that the collection points have enough capacity to always accommodate all failed deliveries. This is not unrealistic, since delivery companies could book in advance space inside dedicated stores or lockers, and the price for the customer of selecting this option can compensate for the booking cost.

Any time a synchronization failure occurs, the problem penalizes the decision-maker's satisfaction with a cost proportional to the probability of absence in the decided timeslot. In the SA case, where two different timeslots are visited per customer, the overall penalty is calculated as the average of the penalties associated with both visits. Such penalties represent additional costs for the customers to retrieve their packages and may depend both on the option and on the customer itself (see Section 5.1 for the details).

3.2. Optimization problem statement

We consider a set P of customers requiring deliveries throughout the day. Each customer $p \in P$ is characterized by a location and a specific delivery demand d_p. Deliveries are handled by a fleet $K = \{1, \dots, \lambda_K\}$ of homogeneous vehicles, each with a fixed capacity Q. All vehicles start from and must return to the same depot within a designated working time t_{max} corresponding to the daily planning time horizon of the service. The delivery time horizon is partitioned into a set $T = \{1, \dots, \lambda_T\}$ of consecutive and non-overlapping timeslots. Each timeslot $t \in T$ is defined by its start and end times, a_t and b_t, respectively. Let ρ_{pt} denote the probability of customer $p \in P$ being at home during timeslot $t \in T$. We assume that $0 < \rho_{pt} < 1$ for all $p \in P$ and $t \in T$, meaning that no customer is expected to be either always present or absent during any single timeslot with absolute certainty. The set of customers is partitioned into three subsets according to the recovery option chosen (SP, SA, and CP), i.e. $P = P_{SP} \cup P_{SA} \cup P_{CP}$ with $P_{o'} \cap P_{o''} = \emptyset \ \forall o', o'' \in \{SP, SA, CP\}, o' \neq o''$. Moreover, let $C = \{1, \dots, \lambda_C\}$ be the set of collection points associated with option CP. Each customer $p \in P_{CP}$ selects a specific collection point $c_p \in C$ to which their packages must be delivered at the end of the route. Each recovery option incurs a distinct penalty: the penalty α^{SP} for option SP remains constant for all customers, the SA and CP penalties, α_p^{SA} and α_p^{CP}, are customer-specific as they depend on the geographical position of the customer and the selected collection point. Penalties are paid proportionately to the likelihood of the customer not being home at the delivery time. The service time for each customer is assumed to be constant and equal to s.

The AHDPRO aims to establish a set of vehicle routes and their corresponding delivery schedules to minimize the total cost including the travel costs and penalties resulting from recovery actions with weights determined by the probability of synchronization failures.

4. Modeling and solution approach

In the following, we outline the approach devised to solve the AHDPRO, namely, a branch-and-cut algorithm based on a tailored MILP formulation, valid inequalities, and heuristic ideas.

4.1. A compact two-index vehicle-flow MILP formulation

As a variant of the VRPTW, it is possible to model our AHDPRO using MILP formulations from the literature. However, to capture specific problem characteristics and obtain tighter bounds, we reformulate the problem using an extended graph and a MILP model including additional scheduling variables to eliminate subtours. This results in a compact two-index vehicle-flow model that does not need a vehicle index on arc variables.

4.1.1. Graph definition

The graph is constructed as follows. Let us define $N_P = \{1, \dots, \lambda_N\}$ as the ordered set of nodes corresponding to the delivery requests made by the customers. For each $p \in P_{SP} \cup P_{CP}$, a single node n_p is associated with the customer, while for customers requiring a second delivery attempt ($p \in P_{SA}$), two nodes sharing the same location are created, namely n_p^1 and n_p^2. So, in general, $\lambda_N \geq |P|$. We introduce the ordered set $N_S = \{\lambda_N + 1, \dots, \lambda_N + \lambda_K\}$ of λ_K dummy nodes representing replicas of the depot as a *starting* point for each one of the λ_K vehicles and the ordered set $N_C = \{\lambda_N + \lambda_K + 1, \dots, \lambda_N + \lambda_K + \lambda_C \cdot \lambda_K\}$ of $\lambda_C \cdot \lambda_K$ dummy nodes representing the collection points duplicated for each vehicle. In the following, we will identify an element of set N_C as the node n_c^k corresponding to the kth replica of collection point $c \in C$. For example, n_2^3 corresponds to the $(\lambda_K + 3)$-th element of N_C, i.e., the node associated with the third duplicate of the second collection point. The *ending* point for all vehicles is represented by the node $e = \lambda_N + \lambda_K + \lambda_C \cdot \lambda_K + 1$. The problem can then be defined over a directed graph $G = (V, A)$ where $V = N_P \cup N_S \cup N_C \cup \{e\}$ and $A = \{(i,j) : i \in N_P, j \in N_P \cup N_C\} \cup \{(i,j) : i \in N_S, j \in N_P\} \cup \{(i,e) : i \in N_P \cup N_C\} \cup \{(n_{\bar{c}}^k, n_{\tilde{c}}^k) : \bar{c} \neq \tilde{c} \in C, k \in K\}$. For each set $\bar{V} \subset V$, let $\delta^+(\bar{V}) = \{(i,j) \in A : i \in \bar{V}, j \notin \bar{V}\}$ and $\delta^-(\bar{V}) = \{(i,j) \in A : i \notin \bar{V}, j \in \bar{V}\}$ be the set of arcs leaving from and entering set $\bar{V}$, respectively. For each arc $(i,j) \in A$, we also define $t_{ij} > 0$ as the time required to travel from node i to node j.

In the above-described expanded graph, the route of each vehicle $k \in K$ corresponds to an oriented path starting at node $\lambda_N + k$ and ending at node e and visiting (if necessary) one or more collection points represented by nodes $n_c^k, c \in C$. This allows us to use two-indexed binary variables representing the selection of an arc without losing the information on which vehicle is visiting which customer (the technique of duplicating the depots was originally proposed in Luo et al., 2015 for a multi-period VRP).

4.1.2. MILP formulation

Let us define a binary variable x_{ij} taking value 1 if arc $(i,j) \in A$ is traversed, and 0 otherwise, and a continuous variable $z_{ij} \geq 0$ representing the arrival time at node j when coming from node i on arc $(i,j) \in A$. Moreover, let us define, for each node $i \in V$, each timeslot $t \in T$, and each vehicle $k \in K$, a binary variable y_{it}^k taking value 1 if node i is visited in timeslot t by vehicle k, and 0 otherwise. Finally, for each $k \in K$, the visit at a collection point $c \in C$ is regulated by the binary variable w_c^k, taking value 1 if collection point node n_c^k is visited by vehicle k, and 0 otherwise. Then, the AHDPRO can be formulated as follows:

$$\min \quad f_{WT} + f_{SP} + f_{SA} + f_{CP} \tag{1}$$

where

$$f_{WT} := \sum_{(i,j) \in \delta^-(\{e\})} z_{ij} \tag{2}$$

$$f_{SP} := \sum_{p \in P_{SP}} \sum_{t \in T} \sum_{k \in K} \alpha^{SP} \, (1 - \rho_{pt}) \, y_{n_p,t}^k \tag{3}$$

$$f_{SA} := \sum_{p \in P_{SA}} \frac{1}{2} \sum_{i \in \{n_p^1, n_p^2\}} \sum_{t \in T} \sum_{k \in K} \alpha_p^{SA} \, (1 - \rho_{pt}) \, y_{it}^k \tag{4}$$

$$f_{CP} := \sum_{p \in P_{CP}} \sum_{t \in T} \sum_{k \in K} \alpha_p^{CP} \, (1 - \rho_{pt}) \, y_{n_p,t}^k \tag{5}$$

subject to

$$\sum_{(i,j)\in\delta^+(\{h\})} x_{ij} \le 1 \quad h \in N_S \tag{6}$$

$$\sum_{(i,j)\in\delta^-(\{e\})} x_{ij} = \sum_{k\in K}\sum_{(i,j)\in\delta^+(\{\lambda_N+k\})} x_{ij} \tag{7}$$

$$\sum_{(i,j)\in\delta^+(\{h\})} x_{ij} = \sum_{(i,j)\in\delta^-(\{h\})} x_{ij} = \sum_{t\in T}\sum_{k\in K} y^k_{ht} \quad h \in N_P \tag{8}$$

$$\sum_{k\in K}\sum_{t\in T} y^k_{ht} = 1 \quad h \in N_P \tag{9}$$

$$y^k_{\lambda_N+k,1} = y^k_{e,\lambda_T} = \sum_{(i,j)\in\delta^+(\{\lambda_N+k\})} x_{ij} \quad k \in K \tag{10}$$

$$\sum_{t\in T} y^k_{n^1_p,t} = \sum_{t\in T} y^k_{n^2_p,t} \quad p \in P_{SA}, k \in K \tag{11}$$

$$y^k_{n^2_p,t} \le \sum_{\bar{t}\in\{1,\dots,t-1\}} y^k_{n^1_p,\bar{t}} \quad p \in P_{SA}, k \in K, t \in \{2,\dots,\lambda_T\} \tag{12}$$

$$\sum_{(i,j)\in\delta^+(\{n^k_c\})} x_{ij} = \sum_{(i,j)\in\delta^-(\{n^k_c\})} x_{ij} = w^k_c \quad c \in C, k \in K \tag{13}$$

$$\sum_{t\in T}\sum_{p\in P_{SP}\cup P_{CP}} d_p y^k_{n_p,t} + \sum_{t\in T}\sum_{p\in P_{SA}} \frac{d_p}{2}\left(y^k_{n^1_p,t} + y^k_{n^2_p,t}\right) \le Q \quad k \in K \tag{14}$$

$$\sum_{t\in T} y^k_{n^k_c,t} = w^k_c \quad c \in C, k \in K \tag{15}$$

$$\sum_{t\in T} y^k_{n_p,t} \le w^k_{c_p} \quad p \in P_{CP}, k \in K \tag{16}$$

$$\sum_{t\in T}\sum_{k\in K} a_t y^k_{ht} \le \sum_{(i,j)\in\delta^-(\{h\})} z_{ij} \le \sum_{t\in T}\sum_{k\in K} b_t y^k_{ht} \quad h \in N_P \tag{17}$$

$$\sum_{(i,j)\in\delta^+(\{h\})} z_{ij} - \sum_{(i,j)\in\delta^-(\{h\})} z_{ij} = \sum_{(i,j)\in\delta^+(\{h\})} (s+t_{ij})x_{ij} \quad h \in N_P \cup N_C \tag{18}$$

$$z_{ij} = t_{ij}x_{ij} \quad (i,j) \in \delta^+(\{h\}),\ h \in N_S \tag{19}$$

$$(t_{\lambda_N+1,i} + t_{ij} + s)x_{ij} \le z_{ij} \le (t_{max} - t_{je} - s)x_{ij} \quad (i,j) \in A \mid i,j \notin N_S \cup \{e\} \tag{20}$$

$$\sum_{t\in T} y^k_{jt} \ge \sum_{t\in T} y^k_{it} + x_{ij} - 1 \quad (i,j) \in A, k \in K \tag{21}$$

$$z_{ij} \ge 0 \quad (i,j) \in A \tag{22}$$

$$x_{ij} \in \{0,1\} \quad (i,j) \in A \tag{23}$$

$$y^k_{it} \in \{0,1\} \quad i \in V, t \in T, k \in K \tag{24}$$

$$w^k_c \in \{0,1\} \quad k \in K, c \in C. \tag{25}$$

Constraints (6) and (7) ensure that each vehicle departs from its designated starting node and returns to the common final depot e. Constraints (8) require each node $h \in N_P$ to be visited exactly once, activating one incoming and one outgoing arc. Constraints (9) impose that the visit is made by only one vehicle $k \in K$ in a single timeslot $t \in T$. Constraints (10) specify that if vehicle $k \in K$ is used, the initial timeslot for its starting node is activated (i.e., $y^k_{\lambda_N+k,1} = 1$), and the same applies to the last timeslot at the final depot (i.e., $y^k_{e,\lambda_T} = 1$).[3] For customers in set P_{SA}, the constraints (11) ensure that both visits n^1_p and n^2_p are made by the same vehicle $k \in K$, while constraints (12) ensure that n^2_p is visited in a timeslot after n^1_p. Constraints (13) regulate flow at collection points, stating that the collection point node n^k_c is visited if and only if it is activated ($w^k_c = 1$). Constraints (14) ensure the total demand of customers served by a vehicle does not exceed its capacity Q. If a customer requires a second visit, their demand is halved since two separate y variables will take value 1 in any feasible solution, one for each visit. Constraints (15) and (16) establish the link between variables y and w. Constraints (15) ensure a y variable corresponding to a vehicle's collection point replica can be 1 if and only if the associated w variable is also 1. Constraints (16) require activation of collection point c_p for vehicle $k \in K$ ($w^k_{c_p} = 1$) if customer $p \in P_{CP}$ is assigned to the vehicle ($y^k_{n_p,t} = 1$) in any given timeslot t. Similar to starting and ending nodes, there is no penalty for choosing a specific timeslot to visit a collection point. To keep the notation as concise as possible, we defined a y variable for all collection points and all timeslots, though only one y variable per collection point node is necessary. Constraints (17)–(20) regulate arrival times at customer nodes and, as a by-product, avoid the creation of subtours in the solution (see, e.g., Maffioli and Sciomachen, 1997 and Gobbi et al., 2019). In particular, constraints (17) state that if node $h \in N_P$ is visited by a vehicle in timeslot $t \in T$, the arrival time must occur within the corresponding time window $[a_t, b_t]$. Constraints (18) handle the time between consecutive nodes, ensuring that if a vehicle visits node j immediately after node i, the elapsed time accounts for the service time at node i and the travel time between nodes i and j. Constraints (19) initialize the arrival time at the first node after the starting depot for each vehicle, while constraints (20) establish upper and lower bounds on the arrival time at node j when coming directly from node i. This time has to be greater than the time to reach node i from the depot ($t_{\lambda_N+1,i}$) plus the time required to travel from i to j plus the time to service node i. The arrival at node j must also be lower than the maximum working time minus the time required to go from node j to the end depot minus the time to service node j. Constraints (21) ensure that y variables identify which vehicle visited which node, requiring consecutive nodes to be visited by the same vehicle. Exploiting the fact that each vehicle starts from a unique replica of the same starting node, node-vehicle assignments can be determined without requiring an additional vehicle index for variables x. Finally, constraints (22)–(25) impose binary and non-negative conditions on the variables.

Finally, over the described feasibility set, the objective function (1) consists of four components. The first, represented by expression (2), considers the sum of all vehicles' arrival times at the final depot. The remaining three components, Eqs. (3)–(5), represent customers' dissatisfaction penalties for recovery options SP, SA, and CP, respectively. These penalties are proportional to the probability of not finding customer $p \in P$ at home $(1 - \rho_{pt})$ in the selected timeslot. Note that customers $p \in P_{SA}$ will be visited twice in two different timeslots. Therefore, the objective function includes the average of the penalties associated with these two visits. The separation of the objective function components has been made to support the methodology implemented for the cost analysis (see Section 5.2), which is tailored for multi-attribute optimization problems.

4.2. Valid inequalities

The proposed MILP formulation can be improved by adding different valid inequalities that strengthen its linear relaxation.

First, we can impose lower bounds on the arrival time at the collection points and at the second visit for a customer with the SA option. In particular, the constraints

$$\sum_{(i,j)\in\delta^-(\{n^k_{c_p}\})} z_{ij} \ge \sum_{t\in T}(a_t + s + t_{n_p,n^k_{c_p}})y^k_{n_p,t} \quad p \in P_{CP}, k \in K \tag{26}$$

state that, given a certain vehicle and a certain customer that has chosen the CP option, the arrival time of such vehicle to the designated collection point must be at least greater than the starting time of the timeslot in which the visit to the customer happened plus the service time plus the time needed to move from that customer to the collection point itself. Moreover, the constraints

$$\sum_{(i,j)\in\delta^-(\{n^2_p\})} z_{ij} \ge \sum_{k\in K}\sum_{t\in T} b_t y^k_{n^1_p,t} + 1 \quad p \in P_{SA} \tag{27}$$

[3] Notice that there is no penalty in the objective function for the timeslot selected for the starting and ending nodes. This fictitious and always feasible assignment does not restrict the set of meaningful and feasible solutions but simplifies the problem notation.

state that, given a certain customer that has chosen the SA option, the arrival time of any vehicle to the node representing the second visit must be strictly greater than the ending time of the timeslot in which the first visit to the customer happened.

Second, we can impose several lower bounds to the number of arcs exiting the k-th replica of the depot. More precisely, for each vehicle $k \in K$, this quantity must be at least one if the vehicle visits a collection point, as defined in constraints

$$\sum_{(i,j)\in\delta^+(\lambda_N+k)} x_{ij} \geq w_c^k \quad c \in C, k \in K, \tag{28}$$

it must be at least as the number of arcs going from all the collection points replicas of that vehicle to the ending node e, as defined in constraints

$$\sum_{(i,j)\in\delta^+(\lambda_N+k)} x_{ij} \geq \sum_{p\in P_{CP}} x_{n^k_{c_p},e} \quad k \in K, \tag{29}$$

and at least as the number of timeslots in which the vehicle visits any customer, as defined in constraints

$$\sum_{(i,j)\in\delta^+(\lambda_N+k)} x_{ij} \geq \sum_{t\in T} y^k_{ht} \quad h \in N_P, k \in K. \tag{30}$$

Third, some arcs can be made impossible to traverse, given the special problem's properties. More precisely, inequalities

$$x_{n_p,e} \leq 0 \quad p \in P_{CP}, \tag{31}$$

$$x_{n_p^1,j} \leq 0 \quad p \in P_{SA}, j \in N_C \cup \{e\}, \tag{32}$$

$$x_{i,n_p^2} \leq 0 \quad p \in P_{SA}, i \in N_S \tag{33}$$

state that it is impossible to directly move from a customer that has chosen the CP option to the depot, from the first-visit node of a customer that has chosen the SA option to the depot or any collection point, and from any depot to the second-visit node of a customer that has chosen the SA option, respectively.

Finally, we impose the following general connectivity constraints

$$\sum_{(i,j)\in\delta^+(S)} x_{ij} \geq \sum_{k\in K}\sum_{t\in T} y^k_{ht} \quad S \subset V, |S| \geq 2, h \in S \tag{34}$$

ensuring that, for each subset of nodes $S \subset V$, there are at least as many arcs exiting S as the number of nodes visited in S by any vehicle and during any timeslot.

The cuts (26)–(30) are added directly to the model while conditions (31)–(33) are managed at a preprocessing stage by simply eliminating the arcs involved in the inequalities. Instead, cuts (34), which are exponential in the number of nodes, are separated dynamically throughout the branch-and-bound search tree through the MIP solver callbacks. The separation problem, which resorts to a max-flow/min-cut, is solved through the algorithm proposed in Boykov and Kolmogorov (2004). We refer to Hanafi et al. (2020) or Gobbi et al. (2023) for more details.

4.3. Heuristic initial solution

A simple and fast (greedy) heuristic procedure has been developed to identify a starting feasible solution, obtaining an upper bound for the problem and a warm start for the branch-and-cut. We begin by placing all the vehicles at the depot at time 0, thus creating dummy routes. Then, we select a random customer among those not yet visited and insert it into one of the existing routes. The insertion procedure, which includes scheduling the visit to the chosen customer and selecting a timeslot, follows a q-th *cheapest insertion* rule. This means we identify the q-th best way to insert the customer into a route between any two existing customers, considering both the working times and the customer dissatisfaction. The integer parameter q is randomly selected in the interval $[1, \lambda_K]$. The procedure ends when all the customers are visited. We conclude the routes by providing a second visit in a different timeslot for the SA customers and by adding the visit to the designated collection points just before returning to the depot for those vehicles that have been assigned to CP customers.

Since the above procedure cannot guarantee to always obtain a feasible solution, anytime we face a situation in which no feasible insertion is allowed, we simply restart the procedure looking for a different solution's construction.

5. Experimental setting and methodology

This section outlines the experimental setup and methodologies used to evaluate the AHDPRO under various scenarios. In Section 5.1, we discuss how benchmark instances are generated, detailing customer CAPs and recovery options. Section 5.2 introduces the solution methodology, describing a two-step optimization approach to understand how different components of the objective function influence the results. Finally, in Section 5.3, we report some details on the implementation and the resolution of our model.

5.1. Instances generation

To evaluate the problem, benchmark instances with 10 or 20 customers ($\lambda_P = \{10, 20\}$) and 4 vehicles ($\lambda_K = 4$) are considered. Customers are randomly placed within a 20000×20000 m^2 area and Euclidean distances between them are calculated. Then, traveling times are derived by assuming a constant vehicle speed of 11.11 m/s (i.e., 40 km/h) and truncating the results to integer values.[4] Each customer has a demand d_p ranging from 5 to 30 units and a service time s of 7 min. The collection point c_p of each customer $p \in P_{CP}$ is randomly selected among those within 15 min of the customer location. To proportion the total fleet capacity to the overall demand, we set the vehicles' capacity Q equal to $3 \cdot \frac{\sum_{p\in N_P} d_p}{\lambda_K}$. Moreover, we assign a maximum work shift duration of $t_{max} = 16000$ seconds (approximately 4.5 h, covering a morning shift) for 10-customer instances, and $t_{max} = 28800$ seconds (8 h, covering both morning and afternoon shifts) for 20-customer instances. This means that we are considering an average time of about 25 min to reach and serve a single customer. Finally, to simulate different timeslot lengths, the working shift is either divided into 4 or 8 timeslots of equal duration ($\lambda_T \in \{4, 8\}$).

Each customer is randomly assigned to one out of seven CAPs, each representing a specific customer's daily behavior. Six CAPs are inspired by Florio et al. (2018), whereas the remaining one is a random profile. Table 1 shows the probabilities of each CAP considered for 4- and 8-slot configurations. Each row corresponds to a CAP while each column indicates the probability of finding a customer at home in the respective timeslot. The *V-Shape* profile, for example, represents the behavior of customers who are most likely to be home during the initial and final timeslots.

Moreover, Figs. 1 and 2 provide a visual representation of the non-random CAPs features for instances 4- and 8-timeslots, respectively. It is important to note that, despite differences in granularity, a consistent *shape* of the profiles is maintained between $\lambda_T = 4$ and $\lambda_T = 8$, thus ensuring that customers with similar profiles exhibit comparable probabilities of being at home at the same times of day, regardless of the number of timeslots.

Finally, instances can be divided into 4 classes, depending on the recovery options adopted by the customers. The first three classes include customers all sharing the same recovery option and are called S_{SP}, S_{SA}, and S_{CP} to indicate option SP, SA, and CP, respectively. The fourth class, called $\mathcal{M}$, comprises instances with multiple options (approximately, one-third of the customers is assigned to each option). The relative penalties $\alpha^{SP}, \alpha_p^{SA}$, and α_p^{CP} depend on the required recovery option and, in some cases, on the customer location. They are defined as follows:

[4] Given the applied truncation, the travel time matrix is eventually triangularized to guarantee the triangle inequality to hold.

Table 1
CAPs features for $\lambda_T = 4$ and $\lambda_T = 8$.

CAP	$\lambda_T = 4$				$\lambda_T = 8$							
V-Shape	0.9	0.1	0.1	0.9	0.9	0.6	0.3	0.1	0.1	0.3	0.6	0.9
A-Shape	0.1	0.9	0.9	0.1	0.1	0.4	0.7	0.9	0.9	0.7	0.4	0.1
Linear-Dec	0.9	0.7	0.4	0.1	0.9	0.9	0.8	0.7	0.6	0.5	0.3	0.1
Linear-Inc	0.1	0.4	0.7	0.9	0.1	0.3	0.5	0.6	0.7	0.8	0.9	0.9
M-Shape	0.1	0.9	0.1	0.9	0.1	0.5	0.9	0.5	0.1	0.7	0.9	0.7
W-Shape	0.9	0.1	0.9	0.1	0.9	0.5	0.1	0.5	0.9	0.7	0.1	0.9
Random	$\mathcal{U}[0.1, 0.9]$				$\mathcal{U}[0.1, 0.9]$							

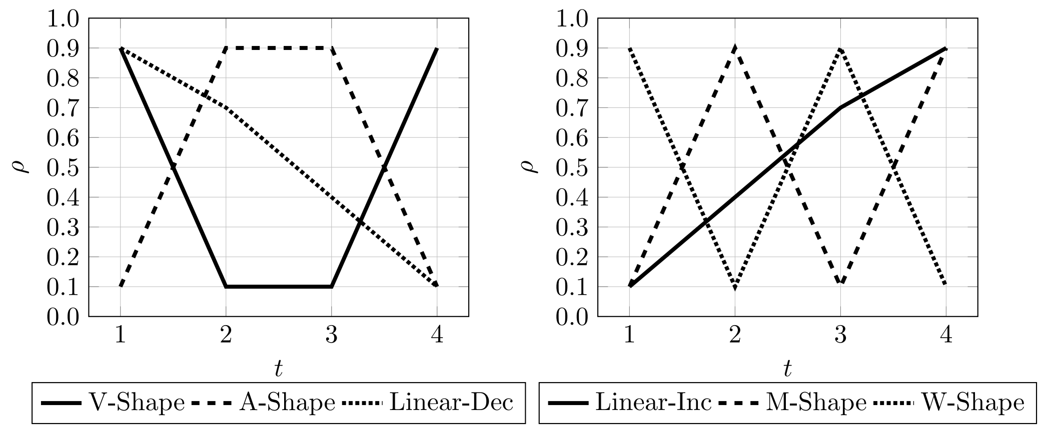


Fig. 1. Probability of finding the customer at home according to the considered CAPs ($\lambda_T = 4$).

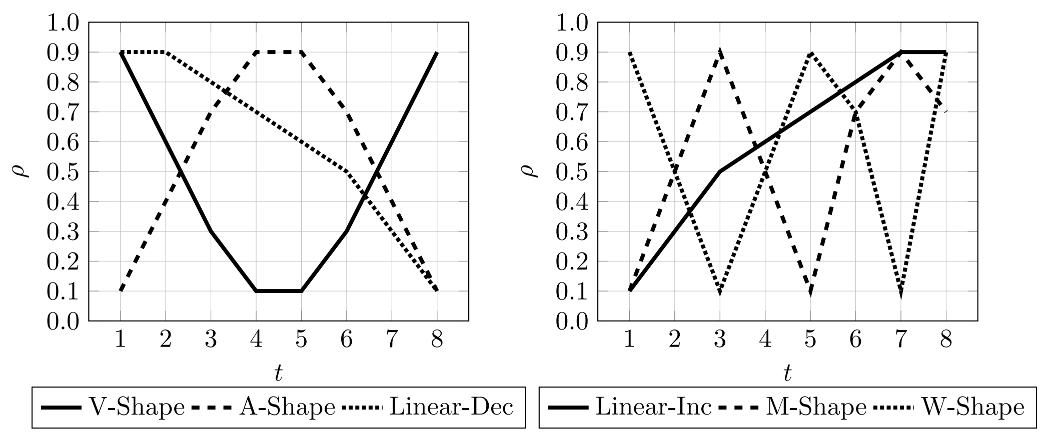


Fig. 2. Probability of finding the customer at home according to the considered CAPs ($\lambda_T = 8$).

- $\alpha^{SP} = s$, $\forall p \in P_{SP}$: it is assumed that a customer takes a time very similar to the service time to retrieve the package from the designated safe place.
- $\alpha^{SA}_p = 2t_{\lambda_N+1,n_p^1}$, $\forall p \in P_{SA}$: it is assumed that a customer takes twice the travel time from the depot to the customer's location to retrieve the package.
- $\alpha^{CP}_p = 2t_{n_p,n_c^1}$, $\forall p \in P_{CP}$: it is assumed that a customer takes twice the travel time from the customer's location to the designated collection point to retrieve the package.

5.2. *A lexicographic approach for multi-attribute analysis*

To evaluate the impact of the four components that make up the overall objective function $f = f_{WT} + f_{SP} + f_{SA} + f_{CP}$, we employ a two-step lexicographic optimization approach. In the first step, we optimize one or more components of f, leaving the remaining ones unconstrained. In the second step, we optimize f, but with a restriction on the value of the prioritized components. We define $f|f_{PRI}$ as the problem version where function f_{PRI} is prioritized in the first step, and we denote the optimal value obtained in this step as z^*_{PRI}. The second step involves solving the standard AHDPRO model with the complete objective function f, with the additional constraint that $f_{PRI} \leq z^*_{PRI}$, bounding the value of f_{PRI}.

Each one of the four components of the objective function (1) can be prioritized in our lexicographic approach. Moreover, we also want to test the prioritization of $f_{HR} = f_{SP} + f_{SA} + f_{CP}$, i.e., a function that comprises the overall quality of customer service, or the *hit-rate* (HR). Finally, prioritizing the entire objective function (i.e., $f|f$) boils down to solving our AHDPRO model. This leads to six lexicographic optimization versions to test, one of which ($f|f$) degenerates to a single step. This approach allows us to formulate and evaluate different business strategies for the delivery company. For example, $f|f_{WT}$ prioritizes minimizing travel time over the penalty costs from failed deliveries, while $f|f_{HR}$ explores a strategy where the delivery hit-rate is maximized before focusing on reducing travel costs.

5.3. Implementation and resolution details

All tests were conducted on a machine with an Ubuntu 22.04 operating system featuring an AMD Ryzen 9 3950x 16-core (32 threads) processor and 32 GB of RAM. The two-step approach was implemented in Java, and the MILP models were solved using Gurobi 10.0.1 with default parameter settings.

For $\lambda_P = 10$ instances, no time limit was set, ensuring all optimizations terminated only upon finding an optimal solution. The average time to complete a two-step optimization was 242 s, although this varied significantly among different optimizations. Solving optimization f took the most time, i.e., 667 s on average. In contrast, the $f|f_{HR}$, $f|f_{SP}$, and $f|f_{SA}$ optimizations took between 45 and 50 s. The $f|f_{WT}$ and $f|f_{CP}$ optimizations required more time, averaging 444 and 196 s, respectively. Generally, the first step in the two-step approach required less time than the second (47 vs 110 s). The number of timeslots also impacted the solution times, with differing trends. On average, instances with $\lambda_T = 4$ required more computing time than those with $\lambda_T = 8$, averaging 328 and 156 s, respectively. However, for $f|f_{HR}$, $f|f_{SP}$, $f|f_{SA}$, and $f|f_{CP}$, the solution time increased when going from 4 to 8 timeslots. For $\lambda_P = 20$ instances, we could not close the optimality gap in many cases within a time limit of 5 h. For 35 out of 90 instances, we have proven optimality while, on average, we have considered solutions with a 4.7% optimality gap.

6. Multi-attribute economic analysis and managerial insights

This section presents a comprehensive economic analysis of the impact of recovery options and timeslot granularity in AHD services. The aim is to quantify the advantages of accounting for potential synchronization failures in operational planning. All the presented values are obtained by averaging results over random repetitions of any combination of option class and number of timeslots, namely, 20 repetitions when $\lambda_P = 10$ and 3 when $\lambda_P = 20$.

Given the complexity of our analysis, we outline the goals and main features of each section:

- Section 6.1: We test all the lexicographic approaches on instances where customers select the same recovery option. For each class of instances and each approach, we analyze the proportion of the objective function factors (Section 6.1.1) and the delivery hit-rate by comparing the probability of selected and available timeslots (Section 6.1.2).
- Section 6.2: We first replicate the analysis from Section 6.1 for instances with multiple recovery options, presenting cost analysis in Section 6.2.1 and hit-rate analysis in Section 6.2.2. We then compare the cost proportion of the same instance when single or multiple options are available (Section 6.2.3).
- Section 6.3: We conduct a sensitivity analysis on the different costs by solving the same instances with increased penalty values.

Moreover, in Table 2, we summarize the features of the optimization versions in terms of prioritizing function f_{PRI} and instances solved, thus gathering the relevant notation used in the following. Note that, for the $\mathcal{M}$ instances, all six optimizations are performed. However, for instances $\mathcal{S}_o$ with only one recovery option $o \in \{SP, SA, CP\}$, only three optimizations are meaningful, namely f, $f|f_{WT}$, and $f|f_o$.

6.1. Results on instances with a single recovery option

In this section, we present the main results obtained from solving instances where all customers share the same recovery option, categorized as $\mathcal{S}_{SP}$, $\mathcal{S}_{SA}$, and $\mathcal{S}_{CP}$.

Table 2
Prioritized functions and instances tested in each optimization process.

Lexicographic version ($f|f_{PRI}$)	f_{PRI}	Instances
f (a.k.a. $f|f$)	$f_{WT} + f_{SP} + f_{SA} + f_{CP}$	$\mathcal{S}_{SP}$, $\mathcal{S}_{SA}$, $\mathcal{S}_{CP}$, $\mathcal{M}$
$f|f_{HR}$	$f_{SP} + f_{SA} + f_{CP}$	$\mathcal{M}$
$f|f_{WT}$	f_{WT}	$\mathcal{S}_{SP}$, $\mathcal{S}_{SA}$, $\mathcal{S}_{CP}$, $\mathcal{M}$
$f|f_{SP}$	f_{SP}	$\mathcal{S}_{SP}$, $\mathcal{M}$
$f|f_{SA}$	f_{SA}	$\mathcal{S}_{SA}$, $\mathcal{M}$
$f|f_{CP}$	f_{CP}	$\mathcal{S}_{CP}$, $\mathcal{M}$

6.1.1. Cost analysis

Fig. 3 shows, for each set of instances ($\mathcal{S}_{SP}, \mathcal{S}_{SA}$, and $\mathcal{S}_{CP}$), the average values of the objective function components for all the meaningful lexicographic optimization strategies. The analysis is divided per number of customers ($\lambda_P = 10$ in Fig. 3(a) and $\lambda_P = 20$ in Fig. 3(b)).

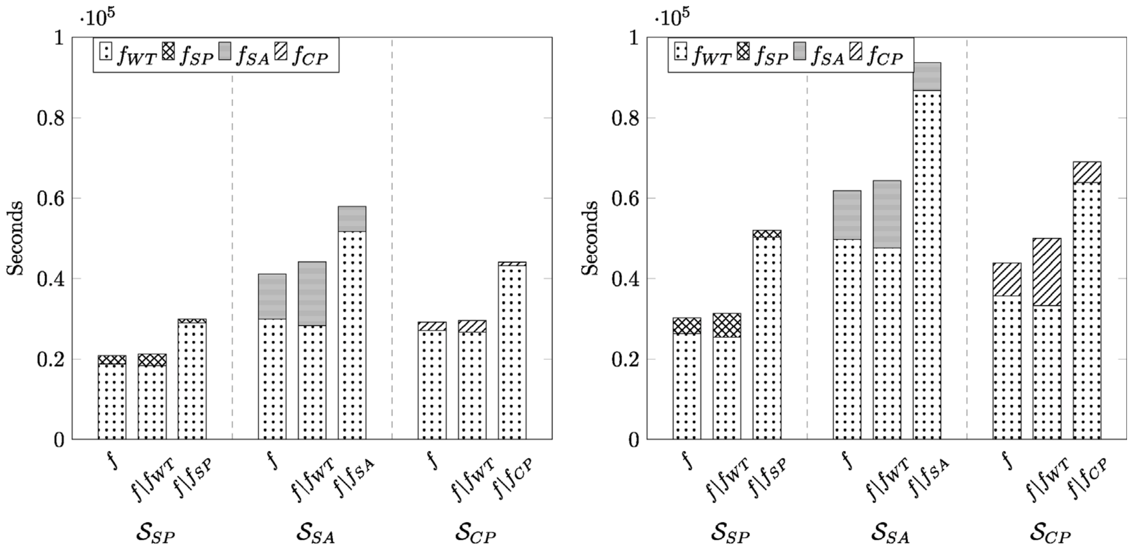
In both figures and across all optimizations, f_{WT} accounts for the largest portion of the objective function. Regardless of the number of customers, instances $\mathcal{S}_{SA}$ exhibit consistently higher values for all versions examined, while SP is the least expensive recovery option. This difference is justified by the nature of the actions required when synchronization fails. In fact, the SP option does not require additional operational adjustments, such as a second visit as in SA, or a visit to a collection point as in CP. Interestingly, while the cost contribution of CP remains moderate in the 10-customer case, its impact significantly increases in the 20-customer instances. This shift can be attributed to the larger operational costs required to manage deliveries to many different collection points, especially when handling more customers. In the comparative analysis of all the optimizations, f consistently yields the lowest values for the overall objective, as expected. Interestingly, switching to the $f|f_{WT}$ variant does not lead to significant changes in the outcomes for both values of λ_P. The value of the objective function increases slightly, with a minor trade-off between reduced routing costs and a marginal increase in penalties. This suggests that the standard model effectively balances working times and penalties, focusing on routing efficiency without substantially affecting hit-rates. This outcome is reasonable, given that routing costs are the primary driver for our setup and real-world service providers. Thus, even when optimizing f, routing costs continue to play a central role. From an operational perspective, this makes the f optimization preferable to $f|f_{WT}$, as it reduces penalties and avoids worst-case scenarios where a minor reduction in routing costs significantly increases penalties. When examining scenarios where recovery option costs are prioritized, there is a noticeable escalation in total cost due to increased f_{WT}. However, the strategy is beneficial for reducing failed synchronization penalties, cutting them by more than half compared to the f optimization.

From a managerial perspective, logistics companies should consider establishing policies that encourage customers to choose the most cost-effective recovery option, perhaps by offering discounts or loyalty rewards. This could lower operational expenses and improve customer satisfaction by simplifying the delivery process. Moreover, logistics providers should favor moving unsynchronized packages to shared collection points over multiple individual re-attempts.

6.1.2. Hit-rate analysis

We define the *hit-rate* of a solution as the probability, averaged over all the customers, of finding them at home during the selected timeslots. Tables 3 and 4 present, for single-option instances and for $\lambda_P = 10$ and $\lambda_P = 20$, respectively, the average hit-rates over all the instances associated with each option. The hit-rate achieved by each lexicographic optimization is reported in the right part of the table, while the left part reports the average presence probability (*nominal* probability) in the worst, mean, and best timeslot of the availability profiles.

The comparison highlights the effectiveness of the solutions in minimizing failed deliveries for different recovery options and optimization



(a) $\lambda_P = 10$ (b) $\lambda_P = 20$

Fig. 3. Optimal cost composition for single-option instances.

Table 3
Nominal probability and obtained hit-rates for single-option instances and $\lambda_P = 10$.

	worst	*mean*	*best*	f	$f\|f_{WT}$	$f\|f_{SP}$	$f\|f_{SA}$	$f\|f_{CP}$
$\mathcal{S}_{SP}$	0.1008	0.5211	0.8768	0.7073	0.5965	0.8700	–	
$\mathcal{S}_{SA}$	0.1045	0.5227	0.8745	0.6690	0.5378	–	0.8245	–
$\mathcal{S}_{CP}$	0.1075	0.5262	0.8785	0.6068	0.4953	–	–	0.8298

Table 4
Nominal probability and obtained hit-rates for single-option instances and $\lambda_P = 20$.

	worst	*mean*	*best*	f	$f\|f_{WT}$	$f\|f_{SP}$	$f\|f_{SA}$	$f\|f_{CP}$
$\mathcal{S}_{SP}$	0.1083	0.5183	0.8750	0.7200	0.5775	0.8767	–	–
$\mathcal{S}_{SA}$	0.1083	0.5183	0.8850	0.7138	0.6029	–	0.8479	–
$\mathcal{S}_{CP}$	0.1100	0.5133	0.8667	0.7775	0.5508	–	–	0.8600

versions, demonstrating their impact on the hit-rate. In particular, $\mathcal{S}_{SP}$ instances consistently achieve the highest hit-rates across all optimizations for $\lambda_P = 10$, with values that range from 0.5965 in the $f|f_{WT}$ optimization to 0.8700 in $f|f_{SP}$, coming remarkably close to the best-case hit-rate of 0.8768. For 20 customer instances, however, the results become more homogeneous across the different recovery options, without any outperforming the others in all the optimizations. Even so, when recovery options are prioritized (hence in columns $f|f_{SP}$, $f|f_{SA}$, and $f|f_{CP}$), $\mathcal{S}_{SP}$ instances continue to maintain the highest hit-rate, while $\mathcal{S}_{SA}$ consistently yields the lowest. These findings underscore the operational challenges posed by the SA recovery option, which struggles to secure customer presence across two timeslots. Note that, the SA hit-rate is calculated as the average over the two visited timeslots, thus inherently penalizing the best possible customer satisfaction. Overall, hit-rates are higher when optimizing f in the larger instances, since, for example, it becomes easier from an operational point of view to fit two visits to the same customer in higher-probability timeslots without incurring high routing costs. Significantly, the CP option goes from being the one with the lowest hit-rate when optimizing f in the $\lambda_P = 10$ case to that with the highest hit-rate when $\lambda_P = 20$. This suggests that consolidation offers potential benefits, as customers are more likely to choose the same collection point, leading to better routing and a stronger balance between routing costs and hit-rate. It is noteworthy that, except for the resolution $f|f_{WT}$ for the class of instances $\mathcal{S}_{CP}$ and $\lambda_P = 10$, all other resolutions yield hit-rates above the mean-case benchmark. This indicates a performance that exceeds expectations. In particular, our proposed model f obtains a hit-rate up to 26% better than the mean-case, with an average difference of 18%.

The main takeaway from this hit-rate analysis is that prioritizing penalties over operational costs can lead to higher hit-rates, but this improvement may be limited when consolidation opportunities or operational flexibility are available. This typically occurs when serving a sufficiently large customer base, suggesting that a promising direction for future research is identifying the optimal workload that balances operational efficiency and penalty costs to obtain overall best performances.

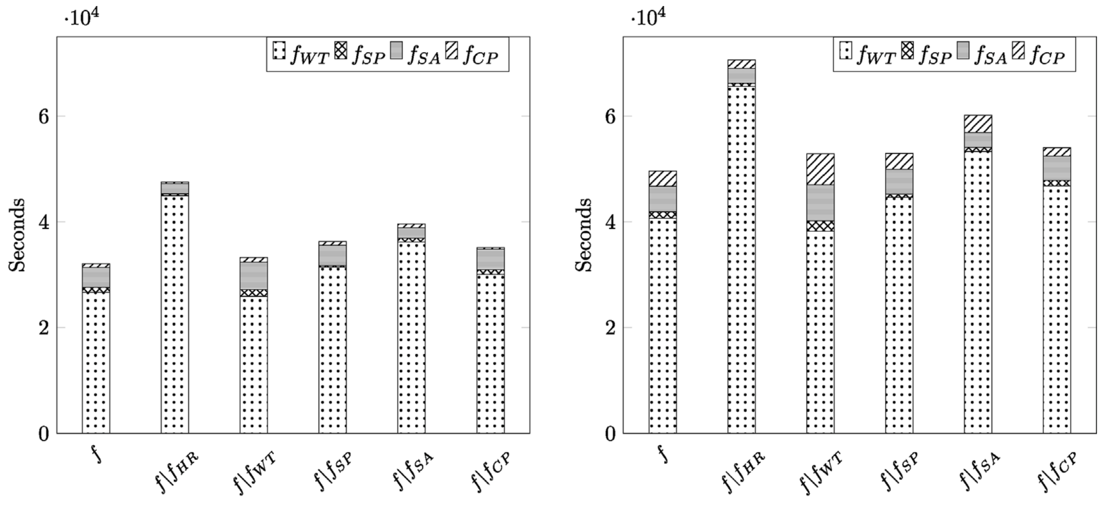
6.2. Results on instances with multiple recovery options

In the following, we expand our analysis by presenting the same indicators used in the previous section on the results obtained from solving instances with a heterogeneous distribution of recovery options ($\mathcal{M}$). We then compare the performance of the same instances in the case of single- and multiple-option environment.

6.2.1. Cost analysis

The cost proportion in the optimal solutions is analyzed in Fig. 4 where, for each of the six lexicographic optimizations, the contribution of all four components of the objective function is shown using different patterns. The results are again divided per number of customers (10 customers for the left chart and 20 customers for the right one).

As for the single-option instances, f_{WT} always accounts for the largest portion of the objective function, while f_{SA} remains the component with the highest impact among the recovery option penalties. Interestingly enough, the f_{CP} component is negligible for $\lambda_P = 10$



(a) $\lambda_P = 10$

(b) $\lambda_P = 20$

Fig. 4. Optimal cost composition for $\mathcal{M}$ instances.

Table 5
Nominal probability and obtained hit-rates for $\mathcal{M}$ instances and $\lambda_P = 10$.

	worst	*mean*	*best*	f	$f\|f_{HR}$	$f\|f_{WT}$	$f\|f_{SP}$	$f\|f_{SA}$	$f\|f_{CP}$
SP	0.1058	0.5187	0.8728	0.6023	0.8461	0.5048	0.8628	0.7145	0.6222
SA	0.1054	0.5298	0.8864	0.6691	0.8381	0.5570	0.6589	0.8392	0.6675
CP	0.1100	0.5222	0.8788	0.6298	0.8456	0.5429	0.6290	0.6067	0.8575
$\mathcal{M}$	0.1071	0.5236	0.8793	0.6337	0.8433	0.5349	0.7169	0.7201	0.7157

Table 6
Nominal probability and obtained hit-rates for $\mathcal{M}$ instances and $\lambda_P = 20$.

	worst	*mean*	*best*	f	$f\|f_{HR}$	$f\|f_{WT}$	$f\|f_{SP}$	$f\|f_{SA}$	$f\|f_{CP}$
SP	0.1000	0.5200	0.8850	0.7410	0.8875	0.5778	0.8875	0.8285	0.7951
SA	0.1033	0.5150	0.8667	0.6774	0.8316	0.5771	0.6882	0.8316	0.7035
CP	0.1067	0.5000	0.8583	0.7174	0.8625	0.5132	0.7208	0.6861	0.8625
$\mathcal{M}$	0.1033	0.5117	0.8700	0.7119	0.8605	0.5560	0.7655	0.7821	0.7870

but tends to be as costly as f_{SA} when considering 20 customers in many cases. Apart from this aspect, the trends in proportional and overall costs remain similar for different numbers of customers. The high costs associated with second-attempt deliveries and, sometimes, with collection point recoveries, further emphasize the need to consider and incentivize alternative recovery strategies.

Comparing optimization $f|f_{WT}$ with the standard version f reveals only a slight increase in the overall values of the objective function. The increase is due to the higher penalties in $f|f_{WT}$, which are not offset by a significant reduction in working times, thus indicating that our modeling approach to balancing costs and penalties is robust even when dealing with multiple recovery options. From a managerial perspective, this suggests that applying our model is generally more beneficial, as it provides a balanced trade-off between penalties and working time optimization. Instead, prioritizing f_{HR} leads to a substantial increase in working times (not compensated by the decrease in penalties), thus making the solutions less acceptable from a company's perspective. These insights suggest a strategic re-evaluation of cost–benefit dynamics associated with each recovery option and demonstrate that blindly optimizing service quality might be impractical and counterproductive.

6.2.2. Hit-rate analysis

In Tables 5 and 6, we report the hit-rate analysis for the multiple-option instances with 10 and 20 customers, respectively. The last row of each table shows the average hit-rate among all customers, while the first three rows specify those of customers requiring each available option.

The results indicate that the probability of failed synchronization is minimal when f_{HR} is optimized or when specific recovery options are prioritized. Moreover, it is rare to achieve the theoretical best hit-rate, consistently with insights from single-option instances, thus indicating that it is unlikely to visit all customers in their preferred timeslots even if we do not care about routing costs. In particular, when multiple recovery options are in play, it seems that the model attempts to balance the impact of expensive recovery options like SA and CP, possibly at the expense of performance for SP customers. On the other side, it is important to note that (for both $\lambda_P = 10$ and $\lambda_P = 20$) the total hit-rate gained by our model is higher by an amount between 0.1 and 0.2 than the mean values in all cases. Instead, prioritizing working times results in hit-rates very close to the mean, and sometimes slightly lower (see the SP option in Table 5). This implies that ignoring penalties when creating routing schedules leads to an increased probability of failed synchronization and very low customer satisfaction.

6.2.3. Cross-comparison of cost proportions between single- and multiple-option instances

In the following, we compare the results obtained on the same instance (topology, demand, etc.) when all the customers share the same recovery option and when, instead, the recovery options are considered concurrently. Fig. 5 presents the percentage variation in objective function components (the working times f_{WT} and the penalties f_{HR}) passing from single-option instances to the $\mathcal{M}$ ones.[5] A positive

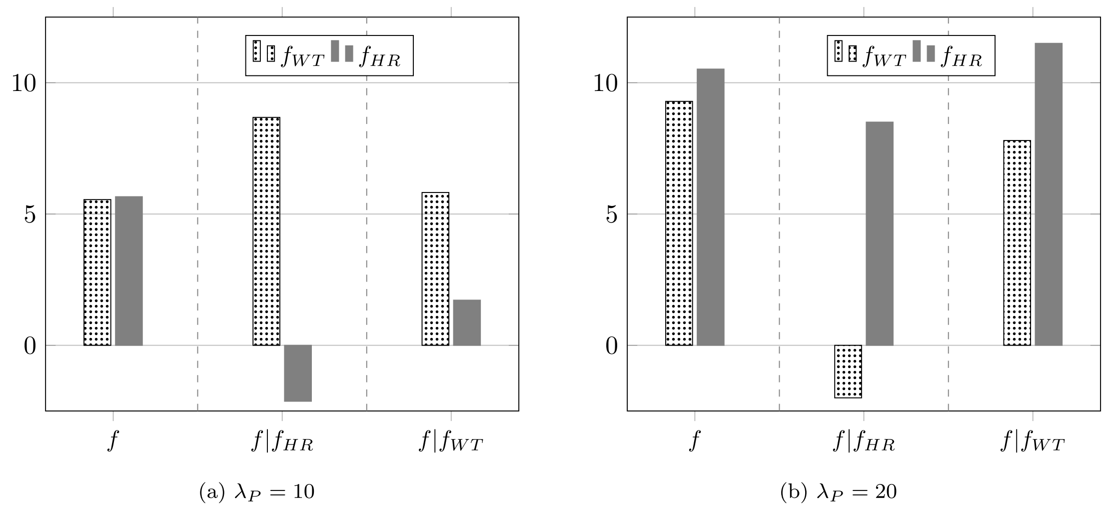


Fig. 5. Comparison of objective function components for $\mathcal{M}$ instances versus single-option instances.

percentage indicates that the cost obtained in the $\mathcal{M}$ instances is higher than that obtained in the single-option ones, on average.

Regarding the optimization versions f, we notice a 5% increase in both the costs when passing from single- to multiple-option for $\lambda_P = 10$, which almost doubles for $\lambda_P = 20$ instances. This indicates that the overall costs are generally negatively affected by the presence of multiple options. Intuitively, implementing recovery options that are expensive in terms of routing costs (SA and CP) also affects those options implying null routing costs (SP), thus worsening the average performance. Interestingly enough, the $f|f_{HR}$ optimization seems very sensitive to the number of customers. In fact, we observe a little improvement in f_{HR} and a consistent loss in f_{WT} for $\lambda_P = 10$, while the contrary occurs for $\lambda_P = 20$. This erratic behavior should further warn companies from adopting pure hit-rate maximization strategies.

In general, the obtained results suggest that it is important to pay great attention to the concurrent use of many recovery options, in particular when they affect costs that are different in nature since this practice could be quite costly in terms of routing costs and customer satisfaction. However, this quantitative analysis can help companies in setting correct pricing for the options, in order to compensate the expected losses.

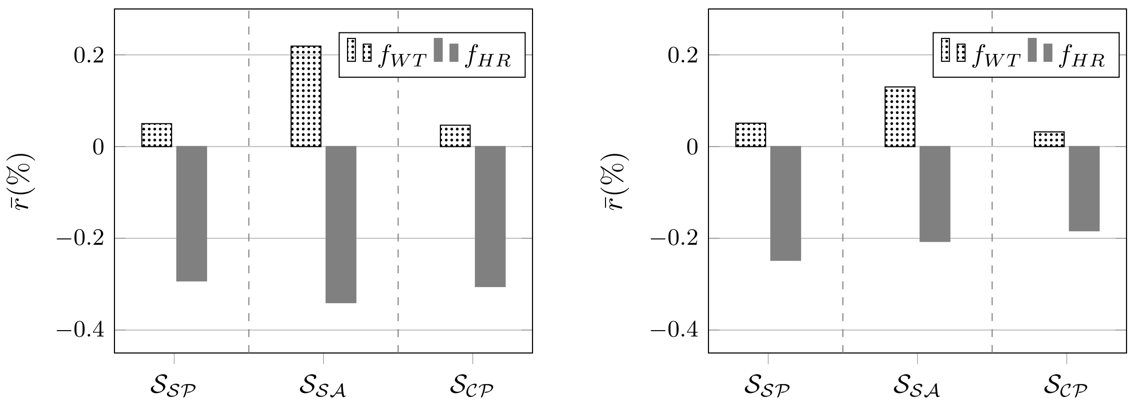
6.3. Evaluation of the impact of different penalty values

Determining the right value to attribute to penalties and their relative weight compared to the travel time cost is not straightforward. In this section, we analyze what happens if the weights of all penalties are changed, evaluating the interplay between penalties and total working time. In particular, we conducted all meaningful optimizations by using functions where the penalty coefficients for the recovery options are tripled, whereas the working time component is left unchanged. Then, we wanted to quantify how much tripling the penalties impacts the solution cost compared to a solution with standard penalties that maintains the same structure but triples the hit-rate cost. To this aim, we calculated the ratio $\bar{r}$ between the value of each component in the solutions obtained with tripled penalty functions ($z_o^{TP}, o \in \{SP, SA, CP\}$, where TP stands for *tripled-penalties*) and the standard ones (z_o), computed as $\bar{r} = (z_o^{TP} - 3z_o)/3z_o$. Instead, for f_{WT}, we computed the ratio without tripling the value of the standard optimization, i.e., $\bar{r} = (z_{WT}^{TP} - z_{WT})/z_{WT}$. For example, a ratio of 0 for f_o (i.e., $z_o^{TP} = 3z_o$) indicates that a solution for this component with standard penalties is still optimal for the optimization with tripled penalties, with a solution value that scales up due to the change in penalty magnitude. A positive $\bar{r}$ value means that f_o has more than tripled its value, implying that f_o has worsened (i.e., is less optimized) by tripling the penalties. The opposite is true when $\bar{r}$ is negative. Such percentage ratios concerning single-option instances are presented in Figs. 6 and 7 for $\lambda_p = 10$ and $\lambda_p = 20$, respectively. Similarly, Figs. 8 and 9 present the *tripled-penalties* analysis concerning $\mathcal{M}$ instances for both $\lambda_P = 10$ and $\lambda_P = 20$.

Concerning single-option instances, note that the only meaningful optimization to analyze is f, since tripling the penalties does not impact the outcomes of $f|f_{WT}$ or the three versions prioritizing the single options.[6] We can see that tripling the penalties affects the overall solution of the standard function f, leading to increased working times (f_{WT}) due to the significantly higher penalties. The $\bar{r}$ value for f_{WT} remains low, at most 0.05 for $\mathcal{S}_{SP}$ and $\mathcal{S}_{CP}$, but reaches a value of 0.21 for $\mathcal{S}_{SA}$. This result confirms earlier findings, suggesting that the SA option is more sensitive to the value of the penalties. The negative $\bar{r}$ values for the other components of the objective function across all instance sets indicate that tripling the penalties slightly worsens routing costs while substantially improving recovery options' effectiveness. A tripled penalty ratio does not correspond to a tripled objective function. Across all recovery options, the corresponding bars are always below −0.18, implying that the model performs at least 18% better than expected. Moreover, tripling the penalties seems to have a more significant impact when there are fewer timeslots. This trend suggests that with a limited selection of timeslots, changes in penalty estimations can significantly affect solution outcomes, leading to larger variations in routing and scheduling decisions. In such settings, therefore, tactical pricing policies assume a larger importance. Concerning the larger instances with $\lambda_P = 20$, the results show identical trends compared to the instances with $\lambda_P = 10$. In some cases, the values of $\bar{r}$ are almost identical. The most significant difference is for $\mathcal{S}_{CP}$ instances

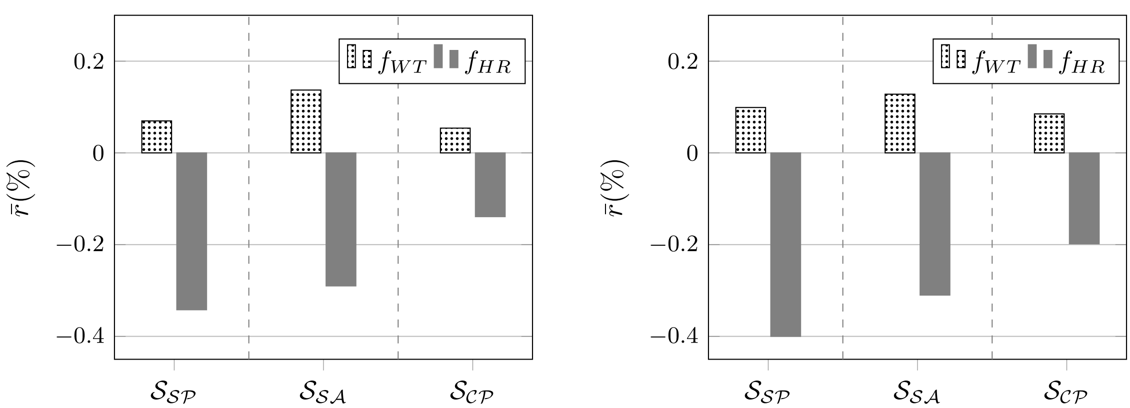
[5] For the single-option, the value is obtained by averaging costs over all the instances in $\mathcal{S}_{SP}$, $\mathcal{S}_{SA}$, and $\mathcal{S}_{CP}$.

[6] In the $f|f_{WT}$ case, the routing decisions at the first step are not influenced by tripled penalties. Thus, in the second step, the value of f_{WT} remains consistent with the standard optimization, and the timeslots chosen also remain the same. Instead, for any $f|f_o$, the selected timeslots in the tripled-penalty optimization will align with those in f despite the increase in costs. Thus, the second step will yield identical routing for both the standard and tripled-penalties optimizations, with the only difference being the penalties magnitude.



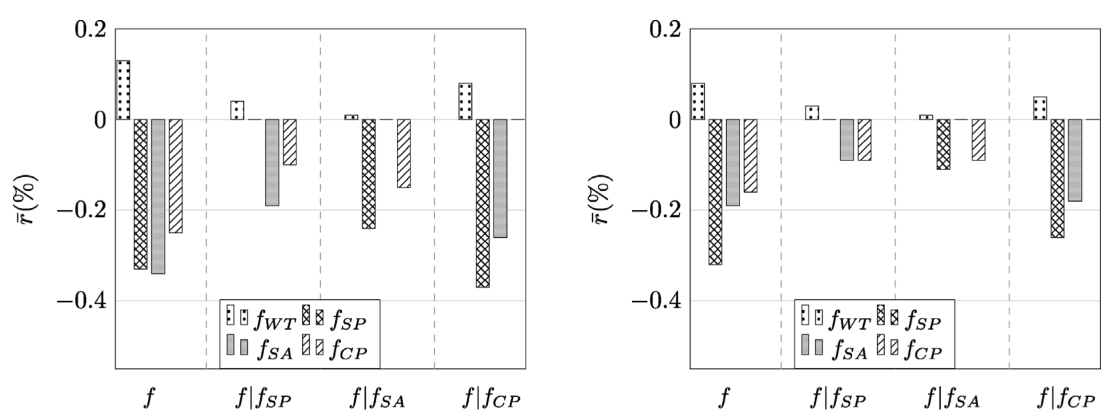
(a) $\lambda_T = 4$ (b) $\lambda_T = 8$

Fig. 6. Objective function components when penalties are tripled versus standard penalties for $\lambda_P = 10$.



(a) $\lambda_T = 4$ (b) $\lambda_T = 8$

Fig. 7. Objective function components when penalties are tripled versus standard penalties for $\lambda_P = 20$.



(a) $\lambda_T = 4$ (b) $\lambda_T = 8$

Fig. 8. Objective function components for $\mathcal{M}$ instances and $\lambda_P = 10$: tripled-versus standard-penalties.

and $\lambda_T = 4$, where the value of $\bar{r}$ for f_{CP} is noticeably higher (in absolute value) when $\lambda_P = 10$.

We now discuss the results for $\mathcal{M}$ instances (as before, the results for $f|f_{WT}$ and $f|f_{HR}$ are not shown as they are not affected by an increase in penalties). Across all versions, an increase in penalties leads to an increase in f_{WT}. This worsening is consistently higher in Figs. 8(a) and 9(a) compared to Figs. 8(b) and 9(b), reaffirming that having more timeslots available helps mitigate the effect of increased penalties, as there is more flexibility in timeslot selection. This is true also for the $\lambda_P = 20$ case, with one notable exception, that, once again is the $f|f_{CP}$ case, where the increase in f_{WT} is higher in the $\lambda_T = 8$ case. The negative values corresponding to f_{SP}, f_{SA}, and f_{CP} suggest that an increase in penalties is not directly related to a proportional increase in the objective function. For example, in the f optimization in Fig. 8(a),

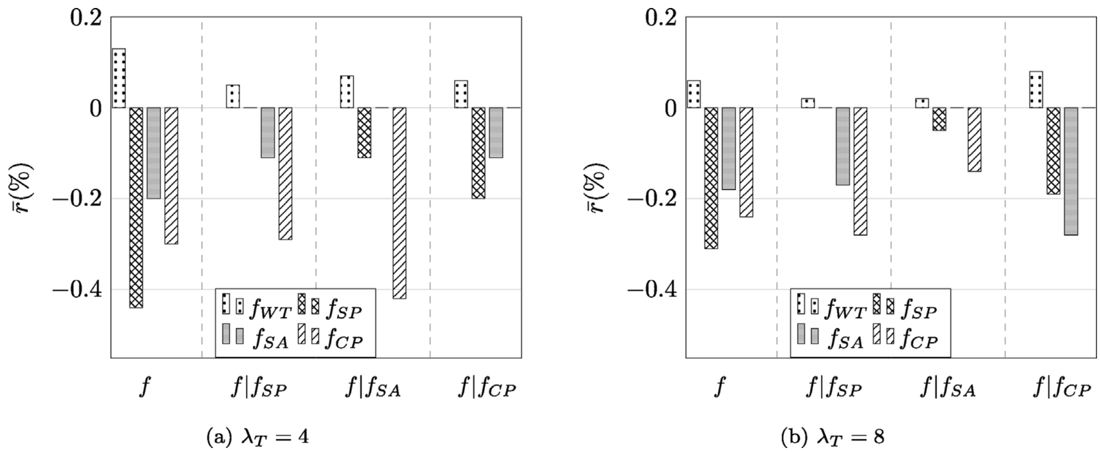


Fig. 9. Objective function components for $\mathcal{M}$ instances and $\lambda_P = 20$: tripled-versus standard-penalties.

the f_{SP} component performs 33% better than expected, showing that the increase in penalties does not lead to a tripled outcome across the board. Comparing the results obtained for the two values of λ_P, the objective function component that stands out the most is f_{CP}. In the larger instances, the increase in f_{CP} is much lower than expected when tripling the associated penalty, while the values for the other two penalties often get closer to zero compared to the $\lambda_P = 10$ case. For example, when $\lambda_T = 4$ and in the $f|f_{SA}$ optimization, the $\bar{r}$ value for f_{CP} goes from −0.15 for $\lambda_P = 10$ to −0.42 for $\lambda_P = 20$, which is the highest change across all objective function components.

Overall, there are two main insights that can be obtained from this analysis. First, even a substantial increase in penalties does not lead to an equally substantial increase in routing costs in any of the optimizations. When recovery options are multiple, the impact on routing costs seems to be even lower. Second, our optimization model opts for a slight increase in routing costs to limit the growth in penalty costs when they are more impactful in proportion. In particular, we observe a major saving in the CP option costs when the number of customers increases for $\mathcal{M}$ instances. This might be due to the fact that, when more customers select the same collection point, more consolidation is possible. Moreover, a larger number of customers makes it easier to schedule visits to collection points that do not require large detours.

7. Conclusions

In this work, we addressed a variant of the Vehicle Routing Problem designed to optimize last-mile operations in the context of Attended Home Delivery. In our problem, customers have availability profiles indicating their probability of being at home during different timeslots throughout the day, and three recovery actions are available to mitigate synchronization failures. We developed a compact Mixed-Integer Linear Programming formulation relying on a smart graph-based representation of the problem and a branch-and-cut algorithm, including valid inequalities and heuristic procedures, that can provide optimal solutions for realistic instances within reasonable computational time. This tool is then used to analyze several performance indicators of interest through an extensive experimental campaign to collect managerial insights on the process and its associated costs.

To conclude, we want to highlight some promising research directions. The first involves developing tailored solution approaches to handle much larger instances. This might involve exploiting the problem's structure to derive other exact methods, heuristic algorithms, or hybrid approaches that combines the potential of both classes. Second, it could be interesting to explicitly consider stochastic parameters within the problem and to provide, following well-known optimization paradigms under uncertainty, models that mix the robustness of the planned routes and the flexibility of the recovery options in case of synchronization failures. For example, the two-stage Stochastic Programming framework, which is based on the implementation of recourse policies to hedge against uncertainty, seems particularly attractive in this operational setting (see, e.g., Fadda et al., 2020, 2021). Finally, we believe that our AHDPRO could be extended toward environmental and social sustainability (see Mangiaracina et al., 2015 for a comprehensive review of the literature on this aspect in B2C services). In fact, several factors are beyond the control of the delivery company. For example, recovery strategies may involve externalities caused by customers' behavior (e.g., driving by car to a collection point to retrieve the items) or municipalities may introduce policies limiting AHD services (Mansini et al., 2024). Hence, a potential future research direction could be to adopt the multi-attribute lexicographic approach proposed in this paper to study not only the operative costs and the customers' satisfaction but also the overall environmental footprint and the social impact of AHD services with recovery policies.

CRediT authorship contribution statement

Valentina Bonomi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Daniele Manerba:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Renata Mansini:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Conceptualization. **Roberto Zanotti:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been supported by "ULTRAOPTYMAL - Urban Logistics and sustainable TRAnsportation: OPtimization under uncertainTY and MAchine Learning", a PRIN2020 project funded by the Italian University and Research Ministry (grant number 20207C8T9M, website: https://ultraoptymal.unibg.it). The authors are grateful to two anonymous referees for having contributed to improving the quality of this paper.

Data availability

Data will be made available on request.

References

Agatz, N., Campbell, A.M., Fleischmann, M., Savelsbergh, M., 2008. Challenges and opportunities in attended home delivery. In: Golden, B., Raghavan, S., Wasil, E. (Eds.), The Vehicle Routing Problem: Latest Advances and New Challenges. Springer US, Boston, MA, pp. 379–396.

Amorim, P., DeHoratius, N., Eng-Larsson, F., Martins, S., 2024. Customer preferences for delivery service attributes in attended home delivery. Manage. Sci. 70 (11), 7559–7578.

Baldacci, R., Mingozzi, A., Roberti, R., 2012. Recent exact algorithms for solving the vehicle routing problem under capacity and time window constraints. European J. Oper. Res. 218 (1), 1–6.

Baldi, M.M., Manerba, D., Perboli, G., Tadei, R., 2019. A generalized bin packing problem for parcel delivery in last-mile logistics. European J. Oper. Res. 274 (3), 990–999.

Belhaiza, S., Hansen, P., Laporte, G., 2014. A hybrid variable neighborhood tabu search heuristic for the vehicle routing problem with multiple time windows. Comput. Oper. Res. 52, 269–281.

Bonomi, V., Mansini, R., Zanotti, R., 2022. Last Mile Delivery with Parcel Lockers: evaluating the environmental impact of eco-conscious consumer behavior. IFAC-PapersOnLine 55 (5), 72–77.

Bonomi, V., Mansini, R., Zanotti, R., 2023. Mediating governance goals with patients and nurses satisfaction: a multi-actor multi-objective problem including fairness. Int. J. Prod. Res. 62 (10), 3753–3780.

Boyer, K.K., Prud'homme, A.M., Chung, W., 2009. The last mile challenge: Evaluating the effects of customer density and delivery window patterns. J. Bus. Logist. 30 (1), 185–201.

Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Anal. Mach. Intell. 26 (9), 1124–1137.

Boysen, N., Fedtke, S., Schwerdfeger, S., 2020. Last-mile delivery concepts: a survey from an operational research perspective. OR Spectrum 43 (1), 1–58.

Campbell, A.M., Savelsbergh, M.W.P., 2005. Decision support for consumer direct grocery initiatives. Transp. Sci. 39 (3), 313–327.

Campbell, A.M., Savelsbergh, M., 2006. Incentive schemes for attended home delivery services. Transp. Sci. 40 (3), 327–341.

Côté, J.-F., Mansini, R., Raffaele, A., 2024. Multi-period time window assignment for attended home delivery. European J. Oper. Res. 316 (1), 295–309.

Ehmke, J.F., Campbell, A.M., 2014. Customer acceptance mechanisms for home deliveries in metropolitan areas. European J. Oper. Res. 233 (1), 193–207.

Ehmke, J.F., Mattfeld, D.C., 2012. Vehicle routing for attended home delivery in city logistics. Procedia - Soc. Behav. Sci. 39, 622–632.

Escudero-Santana, A., Muñuzuri, J., Lorenzo-Espejo, A., Muñoz-Díaz, M.-L., 2022. Improving E-commerce distribution through last-mile logistics with multiple possibilities of deliveries based on time and location. J. Theor. Appl. Electron. Commer. Res. 17 (2), 507–521.

Fadda, E., Fotio Tiotsop, L., Manerba, D., Tadei, R., 2021. Optimization problems under uncertainty in smart cities. In: Augusto, J.C. (Ed.), Handbook of Smart Cities. Springer International Publishing, Cham, pp. 1465–1492.

Fadda, E., Perboli, G., Tadei, R., 2019. A progressive hedging method for the optimization of social engagement and opportunistic IoT problems. European J. Oper. Res. 277 (2), 643–652.

Fadda, E., Tiotsop, L.F., Manerba, D., Tadei, R., 2020. The stochastic multipath traveling salesman problem with dependent random travel costs. Transp. Sci. 54 (5), 1372–1387.

Florio, A.M., Feillet, D., Hartl, R.F., 2018. The delivery problem: Optimizing hit rates in e-commerce deliveries. Transp. Res. B 117, 455–472.

Gevaers, R., de Voorde, E.V., Vanelslander, T., 2011. Characteristics and typology of last-mile logistics from an innovation perspective in an urban context. In: Macharis, C., Melo, S. (Eds.), City Distribution and Urban Freight Transport, Chapter 3. Edward Elgar Publishing.

Gobbi, A., Manerba, D., Mansini, R., Zanotti, R., 2019. A kernel search for a patient satisfaction-oriented nurse routing problem with time-windows. IFAC-PapersOnLine 52 (13), 1669–1674.

Gobbi, A., Manerba, D., Mansini, R., Zanotti, R., 2023. Hybridizing adaptive large neighborhood search with kernel search: a new solution approach for the nurse routing problem with incompatible services and minimum demand. Int. Trans. Oper. Res. 30 (1), 8–38.

Grabenschweiger, J., Doerner, K.F., Hartl, R.F., Savelsbergh, M.W.P., 2021. The vehicle routing problem with heterogeneous locker boxes. CEJOR Cent. Eur. J. Oper. Res. 29 (1), 113–142.

Hanafi, S., Mansini, R., Zanotti, R., 2020. The multi-visit team orienteering problem with precedence constraints. European J. Oper. Res. 282 (2), 515–529.

Klein, R., Neugebauer, M., Ratkovitch, D., Steinhardt, C., 2019. Differentiated time slot pricing under routing considerations in attended home delivery. Transp. Sci. 53 (1), 236–255.

Lone, S., Weltevreden, J., Luharuwala, A., 2023. European E-commerce Report. Amsterdam Univ. Appl. Sci. Ecommerce Eur. URL: https://www.eurocommerce.eu/european-e-commerce-report/.

Luo, Z., Qin, H., Che, C., Lim, A., 2015. On service consistency in multi-period vehicle routing. European J. Oper. Res. 243 (3), 731–744.

Maffioli, F., Sciomachen, A., 1997. A mixed-integer model for solving ordering problems with side constraints. Ann. Oper. Res. 69, 277–297.

Manerba, D., Mansini, R., Zanotti, R., 2018. Attended home delivery: reducing last-mile environmental impact by changing customer habits. IFAC-PapersOnLine 51 (5), 55–60.

Mangiaracina, R., Marchet, G., Perotti, S., Tumino, A., 2015. A review of the environmental implications of B2C e-commerce: a logistics perspective. Int. J. Phys. Distrib. Logist. Manage. 45 (6), 565–591.

Mansini, R., Ranza, F., Zanotti, R., 2024. Towards sustainable last-mile delivery: Introducing an off-peak urban policy to mitigate environmental and social impacts. IFAC-PapersOnLine 58 (2), 186–191.

Moccia, L., Cordeau, J.-F., Laporte, G., 2012. An incremental tabu search heuristic for the generalized vehicle routing problem with time windows. J. Oper. Res. Soc. 63 (2), 232–244.

Norell, V., Gammelgaard, B., 2020. E-commerce and the birth of the logsumer. Effektivitet 1, 8–10.

Özarık, S.S., Lurkin, V., Veelenturf, L.P., Woensel, T.V., Laporte, G., 2023. An adaptive large neighborhood search heuristic for last-mile deliveries under stochastic customer availability and multiple visits. Transp. Res. B 170, 194–220.

Özarık, S.S., Veelenturf, L.P., Woensel, T.V., Laporte, G., 2021. Optimizing e-commerce last-mile vehicle routing and scheduling under uncertain customer presence. Transp. Res. E: Logist. Transp. Rev. 148, 102263.

Ozbaygin, G., Karasan, O.E., Savelsbergh, M., Yaman, H., 2017. A branch-and-price algorithm for the vehicle routing problem with roaming delivery locations. Transp. Res. B 100, 115–137.

Punakivi, M., Saranen, J., 2001. Identifying the success factors in e-grocery home delivery. Int. J. Retail Distrib. Manage. 29, 156–163.

Reyes, D., Savelsbergh, M., Toriello, A., 2017. Vehicle routing with roaming delivery locations. Transp. Res. C 80, 71–91.

Song, L., Cherrett, T., McLeod, F., Guan, W., 2009. Addressing the last mile problem. Transp. Res. Rec. J. Transp. Res. Board 2097 (1), 9–18.

Spliet, R., Desaulniers, G., 2015. The discrete time window assignment vehicle routing problem. European J. Oper. Res. 244 (2), 379–391.

Tilk, C., Olkis, K., Irnich, S., 2021. The last-mile vehicle routing problem with delivery options. OR Spectrum 43 (4), 877–904.

van Duin, J., de Goffau, W., Wiegmans, B., Tavasszy, L., Saes, M., 2016. Improving home delivery efficiency by using principles of address intelligence for B2C deliveries. Transp. Res. Procedia 12, 14–25.

Voigt, S., Frank, M., Fontaine, P., Kuhn, H., 2023. The vehicle routing problem with availability profiles. Transp. Sci. 57 (2), 531–551.

Waßmuth, K., Köhler, C., Agatz, N., Fleischmann, M., 2023. Demand management for attended home delivery—A literature review. European J. Oper. Res. 311 (3), 801–815.

Yang, X., Strauss, A.K., Currie, C.S.M., Eglese, R., 2016. Choice-based demand management and vehicle routing in E-fulfillment. Transp. Sci. 50 (2), 473–488.