Proceedings of the 20th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 625–634 DOI: 10.15439/2025F1656 ISSN 2300-5963 ACSIS, Vol. 43

The price of customer presence in Attended Home Delivery with Customer Availability Profiles

Roberto Zanotti
0000-0002-3073-4895
Dept. of Clinical and Experimental Sciences,
Università degli Studi di Brescia,
viale Europa 11, 25123 Brescia, Italy
E-mail: roberto.zanotti@unibs.it

Daniele Manerba, Renata Mansini
0000-0002-3502-5289, 0000-0002-2194-0339
Dept. of Information Engineering,
Università degli Studi di Brescia,
via Branze 38, 25123 Brescia, Italy
E-mails: {daniele.manerba, renata.mansini}@unibs.it

***Abstract*—Attended Home Delivery is a last-mile distribution paradigm in which the customer must be present at home to receive the goods in person. In this context, we study a vehicle routing and scheduling problem in which the customers' availability is given as a time-dependent probability profile, and the company incurs a penalty cost proportional to the probability of not finding the customer at home during the selected timeslot for delivery. Using an efficient Mixed-Integer Linear Programming formulation for the problem as a black-box tool and lexicographic optimization procedures, we develop an economic analysis to support the company in exploiting the trade-off between basic optimization and the possibility of increasing the customer presence probabilities by paying additional costs. Managerial insights are derived from the change in the value and structure of optimal solutions under different budget levels allocated to improving customer availability profiles.**

I. Introduction and literature review

SINCE the advent of e-commerce, the delivery of goods purchased online directly to customers' homes has rapidly become the leading business model for retail companies and, in turn, one of the most important logistics operations within urban areas, and beyond. The largest share of deliveries performed in e-commerce consists of parcels, small packages, food, and groceries. The volume of this business is impressive and continues to grow rapidly. Over the past 10-12 years, the global e-commerce market has grown tenfold, reaching a value of 3.3 trillion dollars in 2022, approximately 22% of total retail sales. The COVID-19 pandemic significantly boosted this trend: the share of online retail rose from 15% in 2019 to 21% in 2021, exceeding even the most optimistic forecasts (see, e.g., [11]). Experts now agree that the sector still holds considerable long-term growth potential, with projections estimating it could reach 5.4 trillion dollars within the next 3–4 years [21].

Despite being very convenient (and sometimes necessary) for customers, home delivery services pose significant logistical challenges for companies and municipalities, in particular concerning last-mile deliveries. Last-mile delivery refers to the final stage of the distribution process, in which goods are transported from a local hub or depot to the end customer's address. From an economic point of view, last-mile operations represent by far the most expensive part of the fulfillment chain, accounting for more than 50% of the total shipping costs. On average, companies charge the consumers up to 80% of the shipping cost per package to cover them, taking the rest from the profit margins of sold products. Given the costs and the high impact in terms of emissions [8][15] and social implications [19], there is a considerable amount of literature devoted to the evaluation of alternative ways of handling the last mile, from parcel lockers [5] to drones [13], from autonomous robots [7] to crowd-shipping [2][3][12][22]. However, home delivery is one of the predominant delivery paradigms in the industry, whether because it is necessary given the types of goods being purchased (grocery, high-value items) or because customers perceive value when goods are delivered directly to their home. In this paper, we focus on a specific implementation of home delivery operations known as Attended Home Delivery (AHD). In the AHD setting, unlike standard home deliveries, the presence of the customer at home is required when the courier arrives. This requirement prominently brings the time dimension into the delivery operations, since couriers and customers must be synchronized with each other.

From a practical perspective, delivery synchronization is typically managed at the time of ordering by collecting not only the customer's location, but also a *time window* (*timeslot*) on a specific day for the expected delivery. While, in principle, the width of a customer's service time window can be arbitrarily chosen, it is intuitive that narrower time windows result in higher delivery costs for companies [18], which are often passed on to customers through additional delivery fees. Moreover, the delivery company needs to maintain a certain level of control over the time windows schedule available for its services, i.e., it is not realistic to leave an arbitrary choice to the online customer. For this reason, AHD services are often based on a discrete set of available time windows from which customers can select their favorite one. This way, the company still provides customers with a certain degree of flexibility while allowing itself the possibility of finding, at a tactical level, the best schedule to propose. An active stream

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).

in the literature has addressed the optimization of this AHD tactical aspect, using approximated or explicit evaluation of the corresponding routing costs (see, e.g., [1][9][23]).

A recent alternative to use multiple time windows (thus avoiding the need to solve tactical timeslot assignment and pricing problems) is the introduction of Customer Availability Profiles (CAPs), presented in [10]. After partitioning a working day into timeslots, a CAP assigns to each timeslot the probability that the customer will be at home. This framework grants the company full flexibility in planning delivery routes and schedules, at the cost of potentially incurring a penalty when a delivery is attempted in a timeslot during which the customer is actually not at home. For the company, such a penalty may represent an estimation of the costs to sustain for the recovery of the service missed (e.g., re-delivery at a different time or day) and/or a compensation to the customer for having lowered the quality of their service. Since the customer presence in a timeslot is not deterministically known a-priori but just estimated through a probability, then it makes sense to calculate the penalty as proportional to the probability of not finding the customer at home in the timeslot selected by the company.

AHD optimization problems that explicitly include CAPs are more tailored to support a company in the decision-making process at a day-ahead tactical level (see, e.g., the multi-attribute approach proposed in [4]). They can better assess the trade-off that derives from simultaneously minimizing the traveling cost and maximizing the so-called service *hit rate* (equivalent to minimizing the corresponding penalty cost). The total hit-rate cost reflects the quality level of the delivery service and should be considered by the company together with the operating costs. However, in practice, the company may adopt additional strategies, such as offering discounted fees or other forms of compensation to encourage customers to be available during timeslots that are operationally desirable but associated with low presence probabilities. To the best of our knowledge, this integrated approach, where CAPs are treated with a certain degree of flexibility, has not yet been explored in the context of AHD problems. Hence, in this work, by developing mathematical tools including Mixed-Integer Linear Programming (MILP) models and tailored lexicographic solution methodologies, we provide a detailed economic analysis on the trade-offs related to investing in customers' behavior modification along with classical routing and scheduling decisions.

A. Contribution

Building on the current state of the art, this paper provides three main contributions:

- We propose a novel integrated problem in the AHD context, along with its mathematical formulation, that extends existing settings by allowing the company to actively invest a budget to improve customer presence probabilities. In this framework, availability profiles are no longer treated as fixed inputs but as strategic decision variables within the optimization process.
- We develop a multi-step analytical methodology based on lexicographic optimization that enables a structured economic analysis of the trade-offs involved, first by determining a maximum budget for influencing customer availability and then by systematically evaluating the impact of investing partial amounts of that budget.
- We conduct an extensive computational analysis on a large set of instances to derive actionable managerial insights. In particular, we quantify the strategic trade-off in which the company accepts higher penalties on some deliveries to obtain significant savings in travel costs, making the *price of customer presence* a clearly identifiable and decision-relevant concept.

B. Structure of the paper

The remainder of the paper is organized as follows. Section II introduces the optimization problem and provides its mathematical formulation. Section III describes the improved decisional setting in which the company has additional flexibility regarding the CAPs. Section IV details our experimental campaign, describing both the adopted methodology and the instances used, while Section V shows the results obtained and comments on them, deriving managerial insights on the process. Conclusions are drawn in Section VI.

II. Problem statement and mathematical formulation

In this section, we state the optimization problem addressed in this work by building upon an analogous framework presented in the literature [6]. We then formulate the problem as a Mixed-Integer Linear Programming (MILP) model.

A. Problem definition

We consider the problem of addressing daily delivery operations for a set P of geographically dispersed customers. The deliveries are performed by a uniform fleet of vehicles, $K = \{1, \ldots, \lambda_K\}$, which originates from and returns to a central depot $\mathcal{D}$. Each vehicle has a uniform capacity, Q, and all routes must be completed within a maximum operational period, t_{max}. For each customer $p \in P$, there is a specific demand d_p, and a constant service time s is required for every stop. The road network defines the feasible routes along which vehicles can travel, both from the depot to each customer and between any two customer locations. Let t_{ij} denote the travel time between any two locations i and j, where each location may be either the depot or a customer. To incorporate customer availability profiles, the planning horizon is discretized into a set of sequential, non-overlapping timeslots, $T = \{1, \ldots, \lambda_T\}$. Each timeslot $t \in T$ is defined by a start time a_t and an end time b_t. Let ρ_{pt} represent the likelihood that customer p is available at home during timeslot t. A key assumption is that $0 < \rho_{pt} < 1$ for all customers and timeslots, ruling out complete certainty of presence or absence. If a delivery attempt fails, a penalty is incurred. This penalty is calculated as the product of a failure cost parameter α_p, potentially different for each customer $p \in P$, multiplied by the probability $(1 - \rho_{pt})$

of the customer p not being home during the delivery during timeslot $t \in T$. The failure costs are assumed to be calculated in time units, thus representing the additional time spent by the courier to recover the service (e.g., by scheduling a second visit or by moving the goods to a collecting point) and/or the additional time spent by the customer to retrieve their parcel (e.g., roundtrip from the house to the collecting points where the goods have been left), which must be compensated by the company as well. Finally, while in many operational settings the failure costs can be seen as constant for all the customers, in this work we prefer to consider the general case in which they depend on the customer (e.g., on their distance to the depot, fidelity program, or delivery priority). The optimization objective is to determine a set of vehicle routes, along with their corresponding delivery schedules, that minimizes a composite cost function. This function aggregates the total vehicle travel costs with the overall penalty cost arising from potential synchronization failures (i.e., missed deliveries).

Unlike some prior studies that incorporate numerous operational details (see [6][10][24]), our model focuses solely on the core features of the optimization problem to keep the analysis clear and concise. For this reason, some simplifying assumptions are made concerning, e.g., the homogeneity of the vehicle fleet, the constant service time for all the customers, and the presence of a unique depot where all the vehicles depart and come back. Note, however, that these assumptions are realistic in many AHD services, as the e-grocery ones.

B. Problem formulation

The problem is modeled on an *extended* directed graph $G = (V, A)$. The ordered node set V is constructed from three distinct subsets:

- the ordered set $N_P = \{1, \ldots, \lambda_N\}$ of λ_N nodes, each representing a customer.
- the ordered set $N_S = \{\lambda_N + 1, \ldots, \lambda_N + \lambda_K\}$ of λ_K nodes, each being a replica of the unique depot $\mathcal{D}$. This duplication allows us to have one starting node per vehicle, thereby distinguishing each vehicle's route without the need to introduce additional vehicle-indexed arc variables.
- a single node $e = \lambda_N + \lambda_K + 1$, representing the common terminal node for all the routes of the vehicles. This node is a further replica of the unique depot $\mathcal{D}$, and its presence allows us to model vehicle routes as elementary paths.

The arc set A is defined as the set of all feasible connections, i.e., $A = \{(i,j) : i \in N_P, j \in N_P\} \cup \{(i,j) : i \in N_S, j \in N_P\} \cup \{(i,e) : i \in N_P\}$. A weight corresponding to the travel time t_{ij} is assigned to each arc $(i,j) \in A$. For those arcs starting or arriving in a replica of the depot, the corresponding travel time is replicated as well. Given the above graph, the problem asks for a set of λ_K directed elementary paths, one for each vehicle $k \in K$, which originate at the corresponding starting depot node $\lambda_N + k$ and terminate at e. Figure 1 shows an example of the creation of our extended graph starting from a standard network. The total travel time of a vehicle is the sum of the weights of the arcs along its path, and the total travel cost is the aggregate travel time of all vehicles.

The following decision variables are defined to model the problem:

- x_{ij}: a binary variable that is equal to 1 if the direct arc from node i to node j is part of a vehicle's route, and 0 otherwise;
- z_{ij}: a continuous variable denoting the time of arrival at node j, when it is reached directly from node i via arc (i,j);
- y_{it}^k: a binary variable that is equal to 1 if vehicle k services node i during timeslot t, and 0 otherwise.

Then, the problem under study can be efficiently formulated through the following MILP model, in which, given a subset of nodes $\bar{V} \subset V$, $\delta^-(\bar{V}) = \{(i,j) \in A : i \notin \bar{V}, j \in \bar{V}\}$ and $\delta^+(\bar{V}) = \{(i,j) \in A : i \in \bar{V}, j \notin \bar{V}\}$ represent the set of arcs entering and leaving the set of nodes $\bar{V}$, respectively:

$$\min \sum_{(i,j)\in\delta^-(\{e\})} z_{ij} + \sum_{p\in P}\sum_{t\in T}\sum_{k\in K} \alpha_p \left(1 - \rho_{pt}\right) y_{pt}^k \quad (1)$$

subject to

$$\sum_{(i,j)\in\delta^+(\{h\})} x_{ij} \le 1, \quad h \in N_S \quad (2)$$

$$\sum_{(i,j)\in\delta^-(\{e\})} x_{ij} = \sum_{k\in K} \sum_{(i,j)\in\delta^+(\{\lambda_N+k\})} x_{ij} \quad (3)$$

$$\sum_{(i,j)\in\delta^+(\{h\})} x_{ij} = \sum_{t\in T}\sum_{k\in K} y_{ht}^k, \quad h \in N_P \quad (4)$$

$$\sum_{(i,j)\in\delta^-(\{h\})} x_{ij} = \sum_{t\in T}\sum_{k\in K} y_{ht}^k, \quad h \in N_P \quad (5)$$

$$\sum_{k\in K}\sum_{t\in T} y_{ht}^k = 1, \quad h \in N_P \quad (6)$$

$$y_{\lambda_N+k,1}^k = \sum_{(i,j)\in\delta^+(\{\lambda_N+k\})} x_{ij}, \quad k \in K \quad (7)$$

$$y_{e,\lambda_T}^k = \sum_{(i,j)\in\delta^+(\{\lambda_N+k\})} x_{ij}, \quad k \in K \quad (8)$$

$$\sum_{t\in T}\sum_{p\in P} d_p y_{pt}^k \le Q, \quad k \in K \quad (9)$$

$$\sum_{t\in T}\sum_{k\in K} a_t y_{ht}^k \le \sum_{(i,j)\in\delta^-(\{h\})} z_{ij} \le \sum_{t\in T}\sum_{k\in K} b_t y_{ht}^k, \quad h \in N_P \quad (10)$$

$$\sum_{(i,j)\in\delta^+(\{h\})} z_{ij} - \sum_{(i,j)\in\delta^-(\{h\})} z_{ij} = \sum_{(i,j)\in\delta^+(\{h\})} (s + t_{ij}) x_{ij}, \quad h \in N_P \quad (11)$$

$$z_{ij} = t_{ij} x_{ij}, \quad h \in N_S, (i,j) \in \delta^+(\{h\}) \quad (12)$$

$$(t_{\lambda_N+1,i} + t_{ij} + s) x_{ij} \le z_{ij} \le (t_{max} - t_{je} - s) x_{ij}, \quad (i,j) \in A \mid i,j \notin N_S \cup \{e\} \quad (13)$$

$$\sum_{t\in T} y_{jt}^k \ge \sum_{t\in T} y_{it}^k + x_{ij} - 1, \quad (i,j) \in A, k \in K \quad (14)$$

$$x_{ij} \in \{0,1\}, \quad (i,j) \in A \quad (15)$$

$$y_{it}^k \in \{0,1\}, \quad i \in V, t \in T, k \in K \quad (16)$$

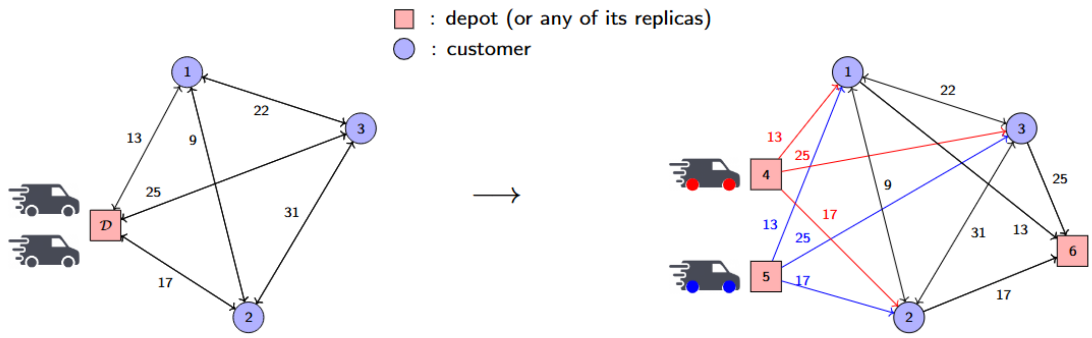

Fig. 1: Example of creation of our extended graph (on the right) starting from a standard network (on the left) involving a single depot, three customers, and two vehicles. In the extended graph, some arcs can be traversed only by a certain vehicle (red and blue arcs), while some others are in common (black arcs). Each vehicle starts from its own depot node, but they both arrive at the same final depot node e.

$$z_{ij} \geq 0, \quad (i,j) \in A. \tag{17}$$

The objective function (1) minimizes the total cost for the company, which is composed of the overall traveling time of the vehicles (first term) and the overall penalty cost paid for visiting customers in timeslots where there is a non-null probability of missing them (second term). Note that, for the sake of simplicity, this cost function is expressed in terms of time units as the scheduling variables and, therefore, also the penalties are measured in time.

Constraints (2) and (3) impose that all vehicle tours must originate from their specified start nodes and end at the common depot, e. The requirement for each customer in the set N_P to be visited exactly once is established through the flow conservation constraints (4) and (5). Constraints (6) guarantee that each customer visit is assigned to a single vehicle $k \in K$ and occurs within a single timeslot $t \in T$. If a vehicle $k \in K$ is used, constraints (7) and (8) ensure its corresponding start (i.e., $y^k_{\lambda_N+k,1} = 1$) and end (i.e., $y^k_{e,\lambda_T} = 1$) depot timeslots are activated. Furthermore, vehicle capacity limits are handled by constraints (9), which prevent the total demand of customers on any given route from exceeding the vehicle capacity, Q. The temporal feasibility of the routes is governed by the set of constraints from (10) to (13). Specifically, (10) enforces the arrival time within the time window $[a_t, b_t]$ for each service performed within timeslot $t \in T$, while (11) defines the progression of time between consecutive visits to customers i and j, accounting for service and travel times. The initial arrival time for each route's first customer is set by constraints (12). Constraints (13) introduce instead lower and upper bounds on the start of the service in a specific node j. This time has to be greater than the time required to reach the previous node from the depot, plus the time needed to reach j from the previous node, plus the service time in the previous node. The start of the service in node j also needs to make a return to the depot feasible, i.e., it needs to be lower than the maximum working time minus the time to serve the node minus the time to return to the depot. Together, these temporal constraints also serve the crucial function of eliminating subtours. Finally, the link between the routing variables (x) and the vehicle assignment variables (y) is established by constraints (14), to ensure that any two consecutively visited nodes are served by the same vehicle. These last constraints are needed since our formulation determines vehicle assignments without needing an explicit vehicle index on the x variables, a feature enabled by the unique starting depot for each vehicle. The domains of decision variables are defined in constraints (15) through (17).

The above formulation is similar to that presented in [4] and adopts a perspective that differs from that previously available in the literature [10][14][20][24]. In particular, the presented extended graph allowed us to obtain a more compact and more efficient MILP model through the introduction of:

- two-indexed binary variables for arc selection, which embed vehicle-specific route information without requiring a third index, similar to the method in [16];
- continuous scheduling variables to enforce time-based subtour elimination, drawing inspiration from modeling techniques for ordering problems [17].

The model efficiency obtained allows us to adopt state-of-the-art MILP solvers to find optimal solutions for real-world instance sizes.

III. An improved decisional setting including CAPs flexibility

As stated in the introduction, the above formulation assumes that customer availability profiles are fixed input data and cannot be modified. However, in practice, companies often have the opportunity to foster some specific customer behavior of interest by offering some form of incentives (e.g., a discount on delivery fees). Assuming that the company has a budget B to support such modifications and a communication channel to reach the customers (e.g., email), it becomes possible to evaluate when and to what extent it is convenient for the company to implement such an influence on the customers' behavior.

To this aim, let us introduce an additional continuous decision variable γ_{pt} representing the increment (induced by the company) with respect to ρ_{pt} of the presence probability of customer $p \in P$ during timeslot $t \in T$[1]. Then, it is possible to create a new optimization model by modifying the one already presented in Section II-B, namely model (1)–(17). More precisely, the objective function (1) is substituted with the following one

$$\min \; f_{TC} + f_{HRC} - f_{HRI} \tag{18}$$

where

$$f_{TC} = \sum_{(i,j)\in\delta^-(\{e\})} z_{ij} \tag{19}$$

$$f_{HRC} = \sum_{p\in P}\sum_{t\in T}\sum_{k\in K} \alpha_p \,(1-\rho_{pt})\, y^k_{p,t} \tag{20}$$

$$f_{HRI} = \sum_{p\in P}\sum_{t\in T} \alpha_p \gamma_{pt} \tag{21}$$

and the following inequalities

$$0 \leq \gamma_{pt} \leq (1-\rho_{pt}) \sum_{k\in K} y^k_{pt}, \quad p \in P, t \in T \tag{22}$$

$$\sum_{p\in P}\sum_{t\in T} \alpha_p \gamma_{pt} \leq B \tag{23}$$

are added to the system of constraints (2)–(17).

The new objective function (18), in which the overall traveling cost is named f_{TC} and the overall hit-rate cost based on the pure CAPs initially given by the customers is named f_{HRC}, now incorporates a third term representing the saving on the hit-rate cost when considering the CAPs modified by the company investment, which is named f_{HRI}. In constraints (22), the new decision variables are allowed to take non-negative values up to the residual presence probability of a customer in a certain timeslot. However, since it makes sense to pay for improving presence probabilities only for timeslots included in the visiting schedule, constraints (22) set to 0 the corresponding variables γ when it is not the case. Finally, constraint (23) limits the company's investment in the modification of the CAPs to a certain budget B.

With respect to the model just stated, it is important to clarify two points:

- despite the definition of the decision variables γ, in reality, the company cannot directly adjust the customers' probability profiles. The model instead quantifies the amount of additional presence probability the company is aiming for and returns the proportional compensations to be offered to the customers for being at home during the selected timeslot. While the customers still have the freedom to choose what to do, the offered compensation is meant to convince them to stay at home during the desired timeslot.
- the third term of objective function (18) represents a saving only for the problem as it is modeled, not on the actual overall cost for the company, which is of course made up by the total objective function value plus the amount invested to influence the customers, i.e., the same amount of the saving. Hence, the company wants to evaluate the possibility of investing more than the traveling and expected hit-rate costs to better balance them.

From now on, we will refer to the MILP model (18) subject to constraints (2)–(17) and (22)–(23) as the *Attended Home Delivery Problem with CAPs Flexibility* (AHDP-CF).

IV. Experimental setting

This section describes the experimental setup adopted to evaluate how the budget constraint affects the optimization problem considered. Section IV-A outlines the generation of benchmark instances, while Section IV-B and Section IV-C describe the analytical methodology developed and the corresponding implementation details, respectively.

A. Instances generation

To obtain statistically meaningful results, our experiments consider a very large set of benchmark instances, in which many of their characteristics are randomly varied. More precisely, we generated 500 instances involving 10 customers (i.e., $|P| = 10$) and 3 vehicles (i.e., $\lambda_K = 3$). Customers are randomly distributed within a 20000×20000 m^2 area. Euclidean distances are computed between each node[2], assuming a constant speed of 11.11 m/s (i.e., 40 km/h). They are then truncated to integer values. Customer demands d_p range from 5 to 30 units, with a fixed service time of 7 minutes. Vehicle capacity is set to $Q = 3 \cdot \frac{\sum_{p\in N_P} d_p}{\lambda_K}$ to align total fleet capacity with aggregate demand. Each vehicle's work shift is limited to $t_{max} = 14400$ seconds (i.e., 4 hours), while the time horizon is divided into 8 equal timeslots (i.e., $\lambda_T = 8$). This time horizon has been chosen to mimic the typical half-day shift of many couriers; however, the model can tackle wider time horizons and denser timeslot division of the shift. Each customer is assigned to one of seven CAPs, each representing different customer presence behaviors during the day. Six profiles are adapted from [10] and eventually discretized, namely, *V-shape*, *A-shape*, *M-shape*, *W-shape*, *Linear-Increase*, and *Linear-Decrease*, whereas the last profile is randomly generated. A detailed visualization of the six non-random profiles is reported in Figure 2. Clearly, each profile is intended to mimic a peculiar behavior of a customer during the time horizon. Failure costs α_p are drawn from a Normal distribution with mean equal to the service time s and standard deviation equal to $s/2$, truncated in $[2, 15]$ minutes.

The generated instances are publicly available at https://or-dii.unibs.it/instances/AHDPCF.zip.

[1]Note that this variable represents an increment and not a free variation with respect to ρ_{pt} since it never makes sense for the company to pay for diminishing the presence probability of a customer during a timeslot.

[2]While real distances between customers seldom correspond to Euclidean ones, the absence of a real street map and the randomness used to generate the customer locations justify the adoption of this simple metric.

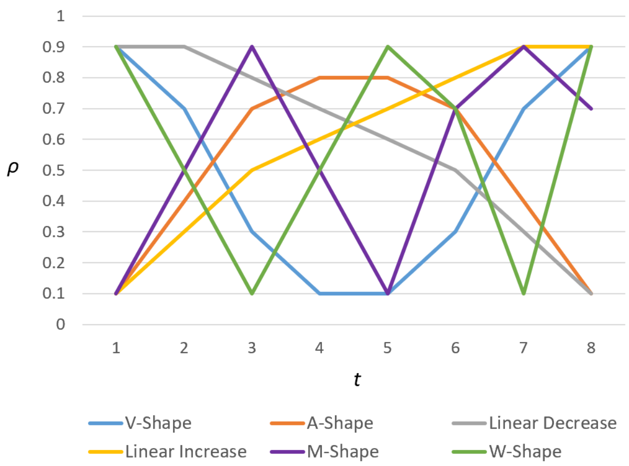

Fig. 2: A plot of the six non-random discrete Customer Availability Profiles adopted, across 8 timeslots.

B. Analytical methodology

The methodology adopted for our quantitative analysis is mainly based on the possibility to solve the MILP model AHDP-CF proposed in Section III and a two-step lexicographic approach that interprets the proposed multi-attribute optimization problem as an actual multi-objective one. The analysis includes a systematic change of the available budget B. To this aim, let B_{max} represent the maximum sensible budget that the company can consider and let β be a real value in $[0, 1]$. Then, let us define $B = \beta B_{max}$ as the available budget. In our analysis, β is used as a sensitivity parameter to evaluate the behavior of the optimal solution for the company against different budgets. The value of B_{max}, instead, which is instance-dependent, is calculated during the optimization process described below.

More precisely, for each of the generated instances, the following optimization steps are implemented:

- **Step 1:** A lexicographic optimization is performed on AHDP-CF where, first, the priority is given to the minimization of the overall objective function (18) and, then, the sole minimization of the hit-rate investment is pursued. This second optimization step is needed since there could be multiple optimal solutions, all with the same objective function value but different proportions of the three objective function terms f_{TC}, f_{HRC}, and f_{HRI}. Among these solutions, the one with the lowest possible value of f_{HRI} is of our interest. More in detail:
  - **Step 1.a**: The MILP model AHDP-CF is solved to optimality without considering the budget constraint (23). This way, the model returns a solution in which the presence probabilities of interest (i.e., those corresponding to the timeslots visited) are increased at least up to the most convenient value with respect to the traveling time minimization. Let us call $f^{Step1.a}$ the value of the optimal solution just obtained.
  - **Step 1.b**: The MILP model AHDP-CF is solved to optimality again without considering the budget constraint (23) but with the new objective function
  $$\min \ f_{HRI} \tag{24}$$
  and with the following additional constraint
  $$f_{TC} + f_{HRC} - f_{HRI} \leq f^{Step1.a}. \tag{25}$$
  Constraint (25) ensures that the value of the primary objective function obtained in the previous step is not deteriorated. Let us call $f_{HRI}^{Step1.b}$ the value of f_{HRI} calculated in the optimal solution.
- **Step 2**: The value of B_{max} is set to $f_{HRI}^{Step1.b}$, since it does not make sense to consider larger budgets for the current instance.
- **Step 3**: The parameter β is set to a value in $[0, 1]$ and the actual budget B is calculated.
- **Step 4**: The MILP model AHDP-CF is solved to optimality, now including also the budget constraint (23). Let us call sol_β the optimal solution obtained. This solution is the one analyzed in the later discussion of the results.

C. Implementation and resolution details

Experiments were performed on a machine equipped with an *Intel Xeon Gold* 6140x (18 physical cores, 36 logical threads) CPU, 64GB RAM, and running a *Windows 11* 64-bit operating system. The solution framework, including the lexicographic procedure and MILP models, was implemented in Java 16. Gurobi v12.0.0 was adopted for solving the MILPs; the default parameter setting is used, and no time limit is imposed on the resolution, thus ensuring termination only upon optimality.

V. Results and discussion

This section presents and discusses the results of the computational experiments conducted to evaluate the impact of investing a certain budget for improving the expected hit-rate (representing the quality of the delivery service) on our AHDP-CF. Our analysis focuses on the trade-offs between the investment in higher customer presence probabilities and the resulting operational costs and solution structure. To provide clear managerial insights, we always compare the solutions sol_β obtained by using an increasing budget (from $\beta = 0.05$ up to $\beta = 1$, with a 0.05 increment) against solution sol_0, which represents the baseline scenario where no budget is available (i.e., when $\beta = 0$).

Our analysis is designed to provide quantitative insights into how different budget levels β reshape the optimal solution. In each part of the discussion, results are reported as sequences of boxplots, one for each β value that has been tested, thus aggregating the results obtained over all the 500 instances generated. In a boxplot, the cross denotes the average value, while the thick line indicates the median. The upper part of the colored box corresponds to the 75th percentile, while the lower one is associated with the 25th percentile. Moreover, the whiskers span from the minimum to the maximum value,

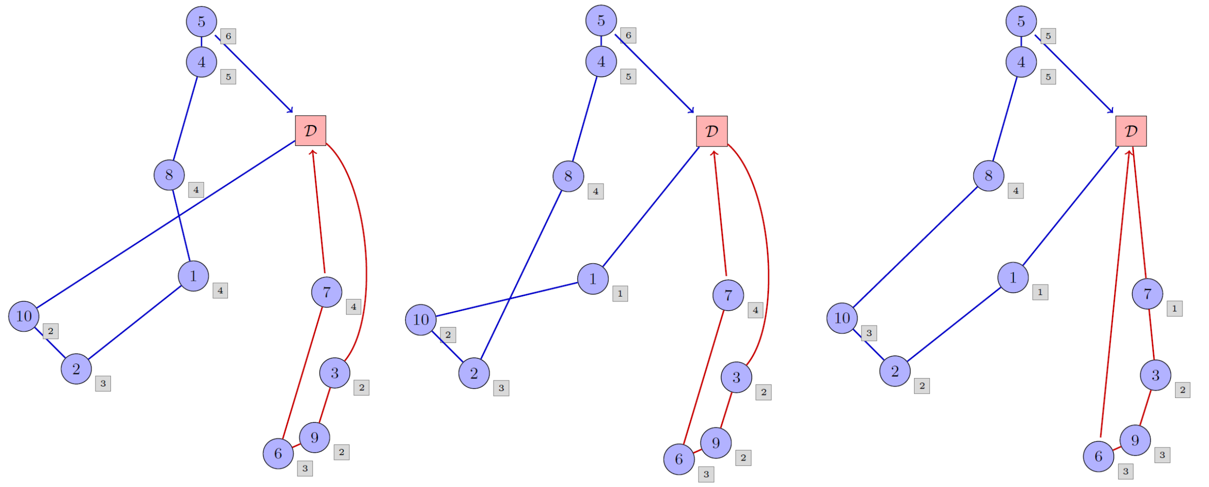

Fig. 3: Solution examples for different β values: $\beta = 0$ (left), $\beta = 0.50$ (center), $\beta = 1.00$, (right). The number in a gray square near each node indicates the timeslot selected for the corresponding visit.

excluding the outliers (which are reported as simple dots). Finally, we want to assess the impact of the available budget from several perspectives, each of which is related to a different Key Performance Indicator (KPI) that can be calculated on the solutions obtained by implementing the analytical method proposed in Section IV-B. In this discussion, we decided to focus on the following KPIs of interest:

- CSS (change in start times): represents, averaged over all the customers, the absolute value of the difference in seconds between the service start times scheduled in sol_β and those scheduled in sol_0.
- PCT (percentage of changed timeslots): represents the percentage of customers for which the timeslot scheduled for the delivery in sol_β is different from that scheduled in sol_0.
- PAC (percentage of affected customers): represents the percentage of customers in sol_β for which the company has invested at least something greater than 0 (clearly, in sol_0 where the budget is null, the baseline comparison value is always 0).
- PCTC (percentage change in traveling cost): represents the difference between the traveling cost of sol_β and that of sol_0, as a percentage of the traveling cost of sol_0.
- PCHC (percentage change in hit-rate cost): represents the difference between the hit-rate cost f_{HRC} of sol_β and that of sol_0, as a percentage of the available budget B.

To illustrate how investment drives structural changes in a single instance, Figure 3 compares the routes obtained with three different β values, namely, 0.0, 0.50, and 1.00. Although node-to-vehicle assignments remain fixed, increasing β reorders visits. The solution with $\beta = 1.00$ corresponds to the solution that could be obtained by solving a standard VRP with travel time minimization. The other solutions show more complicated routes, since the cost of visiting customers in less favorable timeslots cannot be fully compensated by the budget.

Figure 4 illustrates the distribution of the change in the service start times (CSS) for customers across various budget levels. A noticeable trend emerges, characterized by three

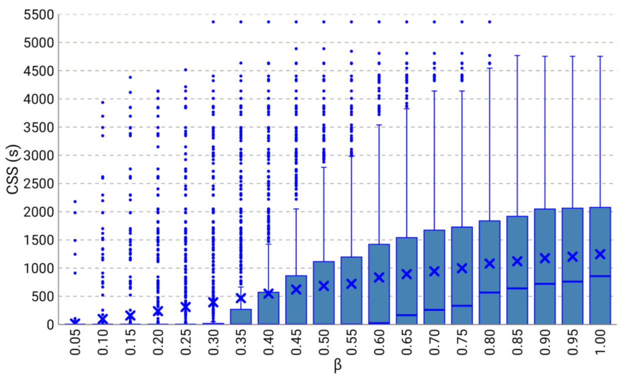

Fig. 4: Change of the start of the service (CSS), averaged over all customers.

distinct phases. Initially, for low budget levels ($\beta \leq 0.30$), CSSs are negligible, as the median, 25th, and 75th percentiles are all concentrated close to zero. As the budget increases into the intermediate range of $0.35 \leq \beta \leq 0.60$, the median CSS remains at zero, but the distribution skews significantly upward. This indicates that, while the start time for more than half of the customers is unaffected, a growing portion of them experiences a substantial shift, as seen in the 75th percentile progressively increasing to a value of nearly 1500 seconds. For the highest budget levels ($\beta \geq 0.65$), a different operational behavior emerges where the entire distribution shifts upwards. In particular, the median CSS itself starts to increase, reaching nearly 900 seconds for $\beta = 1$, signifying that the investment now alters the start times for the majority of the customers. This tripartite behavior is a pattern that can be noticed in most analyzed KPIs, indicating a fundamental shift in the solution's characteristics based on the amount of budget invested. In contrast to such a behavior of the median, the average CSS has an almost linear increase from 0 to around 1250 seconds.

Figure 5 presents the distribution for the percentage of customers (PCT) whose selected timeslot changes with respect to the zero-budget baseline. The analysis reveals the same clear

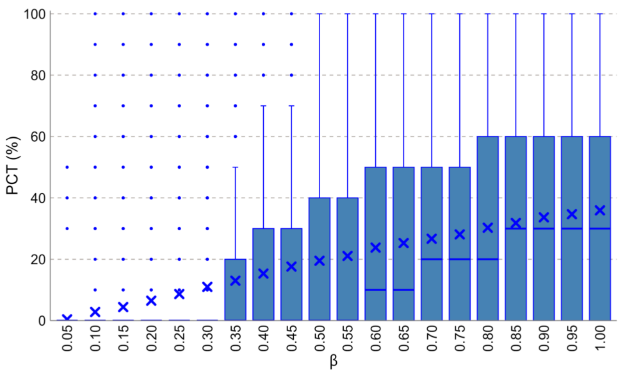

Fig. 5: Percentage of customers for which the timeslot selected by the company changes (PCT).

threshold-based behavior as in the previous KPI. The average PCT still increases linearly across β values, going from 0 to about 35%. The median's behavior, however, is more distinctly segmented. In the first segment ($\beta \leq 0.30$), the median PCT is null, thus indicating that most customers are unaffected. A sharp transition then occurs at $\beta = 0.35$ where the 75th percentile rises while the median stays at 0. For $\beta \geq 0.60$, the median increases in three steps and stabilizes at 30% for $\beta \geq 0.85$.

To understand the mechanism driving these outcomes, Figure 6 illustrates the investment strategy identified in the optimal solutions, by showing the percentage of customers (PAC) for which the company invests in changing their behavior. In

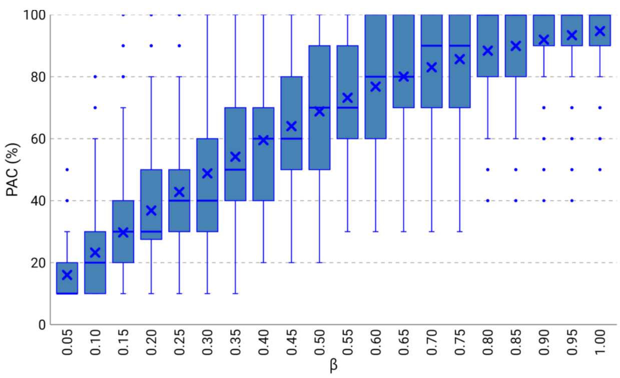

Fig. 6: Percentage of customers on which the company invests its budget (PAC).

contrast to the complex, threshold-based behavior seen in the previous KPIs, the application of the budget is remarkably direct and nearly linear. As β increases, the percentage of affected customers also rises monotonically and with relatively low variability. The median and average PAC values remain closely aligned throughout the entire range, starting near 15% and climbing steadily until they approach 100% for higher investment levels. A notable insight emerges when comparing this strategy to its impact on rescheduling (Figure 5), i.e., the percentage of customers receiving investment is considerably higher than the percentage whose timeslot is changed. For instance, if we consider median values at $\beta = 0.70$, the budget is invested in a 90% of the customers, while the percentage of customers with an altered timeslot is just 20%. This implies that, for a significant portion of customers (around 70%), the budget is invested not to change their assigned timeslot, but to increase the likelihood of delivery success for the scheduled timeslot in sol_0.

Figure 7 evaluates the impact of the budget on the routing decisions by showing the percentage change in traveling costs (PCTC). Note that, differently from the previous ones, this KPI

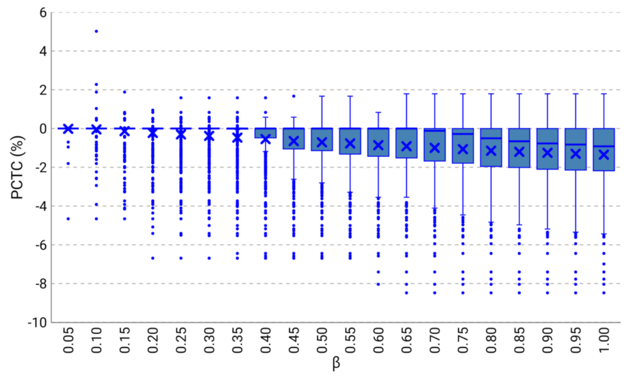

Fig. 7: Percentage change in traveling costs (PCTC).

may result in both negative and positive values. When they are negative, the f_{TC} derived from the solution including some investment is lower than the no-budget solution. This is the more frequent behavior. The resulting trend reveals a complex behavior that strongly mirrors (even if horizontally specular) the tripartite behavior of the first two KPIs analyzed rather than the linear investment strategy itself. Such a behavior appears, in this case, with thresholds at $\beta = 0.40$ and $\beta = 0.70$. The average PCTC decreases linearly from 0% to -1.4%, while the median PCTC value reaches a minimum of about -1%. Interestingly enough, there are some cases in which the budget investment increases the traveling cost, meaning that the model favors some decrease in hit-rate cost at the expense of a smaller increase in traveling cost. These situations correspond to noisy results, due to the possible existence of multiple optimal solutions for the problem. In fact, when a certain non-null budget is available, it is never optimal to invest in changing the solution structure without improving the traveling cost, since the hit-rate saving must always be compensated by the budget investment. This means, for every solution with an increase in traveling costs, there must be an equally optimal solution in which the traveling cost remains the same or improves compared to sol_0. Anyway, across all the β values, these cases fall beyond the 75th percentile, and, therefore, they do not disrupt the validity of our analysis.

The final analysis, presented in Figure 8, examines the portion of the budget dedicated to compensating for changes in the objective function component f_{HRC}, as a percentage of the available budget to invest (PCHC). As before, this KPI may

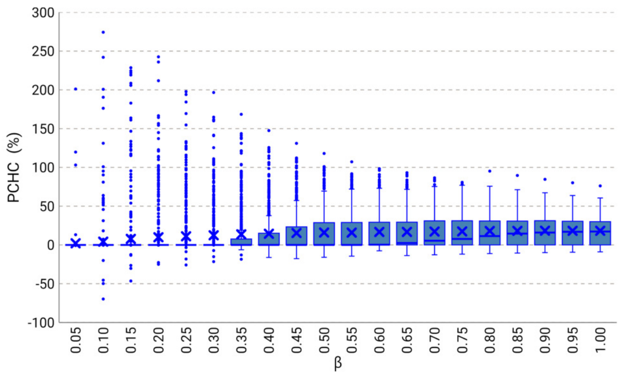

Fig. 8: Percentage change in hit-rate cost (PCHC).

result in both negative and positive values. Here, a positive value means that the value of f_{HRC} when a budget is available is greater than that obtained with zero budget. This is the more common situation. The trend follows a clear phased pattern, like in most previous KPIs. The median PCHC is initially zero, becomes positive and rises through the mid-range of β values, and finally stabilizes just under 20%. Notably, the high variability observed at low-to-moderate investment levels decreases significantly as the solution stabilizes. Interpreting these results provides a deeper insight into the solution's trade-offs. When the median PCHC value is close to zero, it suggests that the budget is primarily used to increase success probability within the timeslots scheduled for visits in sol_0. Instead, when the median PCHC stabilizes just under 20% for high β values, it indicates that roughly a fifth of the budget is now used to offset the corresponding increase in the hit-rate cost. This increase reveals a strategic selection of different timeslots for the visits that allows the traveling cost reduction seen previously for the same β values. The PCHC distribution also reveals a wide range of outcomes. Extreme cases show that the PCHC can reach almost 275%, where the hit-rate cost increase is 2.75 times the available budget. Conversely, very few times, negative values may also occur. These situations reveal the presence of multiple optimal solutions, as seen for the previous KPI. Again, the frequency of these cases does not affect the insights derived.

VI. Conclusions

In this paper, we studied the Attended Home Delivery problem with Customer Availability Profiles and introduced a novel and managerially relevant dimension, namely, the ability for a company to invest a budget to actively influence customer presence probabilities during certain timeslots of interest. We proposed the corresponding new optimization setting together with its MILP formulation, called the Attended Home Delivery problem with CAPs flexibility (AHDP-CF). Through a structured analytical methodology based on lexicographic optimization, we conducted an extensive computational study to quantify the impact of the possible investment. Our analysis revealed that the relationship between the investment in service quality and the resulting operational outcomes is characterized by two distinct thresholds that define three different ranges in which the invested budget has different impacts. We showed that small, incremental investments (within the first range) yield negligible changes to the solution, whereas substantial budgets (within the third range) enable a fundamental restructuring of the delivery routes and schedules. The primary managerial insight derived is the possible quantification of a strategic trade-off. Many times, it is optimal for the company to leverage the budget to accept higher penalties on some customer visits in exchange for significant savings in overall traveling costs. This work, therefore, provides a clear and quantitative understanding of the *price of customer presence* in AHD logistics. The framework and insights presented can serve as a valuable decision-support tool for logistics companies, allowing them to move beyond treating customer availabilities as fixed data and instead see them as a strategic lever that can be priced and optimized.

Future research could extend this work in several directions. Developing scalable heuristic or metaheuristic algorithms would allow the AHDP-CF to be applied to larger, real-world instances. Moreover, exploring dynamic settings where investment decisions can be adjusted during the operational day presents another promising avenue for future investigation. Finally, one may also want to incorporate into the problem the *willingness to accept* of a customer against the company's offers by adopting more sophisticated models specifically tailored for individual choices.

Acknowledgment

The authors gratefully acknowledge the financial support of "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).

References

[1] N. Agatz, A. Campbell, M. Fleischmann, and M. Savelsbergh. Time slot management in attended home delivery. *Transportation Science*, 45(3):435–449, aug 2011. doi: 10.2307/23018537.

[2] M. Bierlaire, E. Fadda, L. F. Tiotsop, and D. Manerba. A chance-constraint approach for optimizing social engagement-based services. page 301–304, 2022. doi: 10.15439/2022F235.

[3] M. Bierlaire, E. Fadda, L. F. Tiotsop, and D. Manerba. Workforce allocation for social engagement services via stochastic optimization. *Studies in Computational Intelligence*, 1158 SCI:71–87, 2024. doi: 10.1007/978-3-031-57320-0_5.

[4] V. Bonomi, D. Manerba, R. Mansini, and R. Zanotti. Optimizing attended home delivery: Multiple recovery options and customer availability profiles to face synchronization failures. *International Journal of Production Economics*, 279, 2025. doi: 10.1016/j.ijpe.2024.109463.

[5] V. Bonomi, R. Mansini, and R. Zanotti. Last mile delivery with parcel lockers: evaluating the environmental impact of eco-conscious consumer behavior. *IFAC-PapersOnLine*, 55(5):72–77, 2022. doi: 10.1016/j.ifacol.2022.07.642.

[6] V. Bonomi, R. Mansini, and R. Zanotti. Mediating governance goals with patients and nurses satisfaction: a multi-actor multi-objective problem including fairness. *International Journal of Production Research*, 62(10):3753–3780, 2023. doi: 10.1080/00207543.2023.2248522.

[7] N. Boysen, S. Schwerdfeger, and F. Weidinger. Scheduling last-mile deliveries with truck-based autonomous robots. *European Journal of Operational Research*, 271(3):1085–1099, 2018. doi: 10.1016/j.ejor.2018.05.058.

[8] A. Cassia, O. Jabali, F. Malucelli, and M. Pascoal. The electric vehicle shortest path problem with time windows and prize collection. pages 313–322, 2022. doi: 10.15439/2022F186.

[9] J.-F. Côté, R. Mansini, and A. Raffaele. Multi-period time window assignment for attended home delivery. *European Journal of Operational Research*, 316(1):295–309, 2024. doi: 10.1016/j.ejor.2024.01.021.

[10] A. M. Florio, D. Feillet, and R. F. Hartl. The delivery problem: Optimizing hit rates in e-commerce deliveries. *Transportation Research Part B: Methodological*, 117:455–472, nov 2018. doi: 10.1016/j.trb.2018.09.011.

[11] Forrester Research, Inc. Forrester Data: Online Retail Forecast, 2017 to 2022 (Western Europe), 2016.

[12] M. Gajda, O. Gallay, R. Mansini, and F. Ranza. Optimizing last-mile delivery through crowdshipping on public transportation networks. *Transportation Research Part C: Emerging Technologies*, 179:105250, 2025. doi: 10.1016/j.trc.2025.105250.

[13] V. Garg, S. Niranjan, V. Prybutok, T. Pohlen, and D. Gligor. Drones in last-mile delivery: A systematic review on efficiency, accessibility, and sustainability. *Transportation Research Part D: Transport and Environment*, 123:103831, 2023. doi: 10.1016/j.trd.2023.103831.

[14] A. Gobbi, D. Manerba, R. Mansini, and R. Zanotti. A kernel search for a patient satisfaction-oriented nurse routing problem with time-windows. volume 52, page 1669 – 1674, 2019. doi: 10.1016/j.ifacol.2019.11.440.

[15] A. Lorenz, N. Madeja, and C. Leyh. A framework for assessing the sustainability impact of intelligent transport systems in the smart city context. pages 161–169, 2023. doi: 10.15439/2023F6002.

[16] Z. Luo, H. Qin, C. Che, and A. Lim. On service consistency in multi-period vehicle routing. *European Journal of Operational Research*, 243(3):731–744, 2015. doi: 10.1016/j.ejor.2014.12.019.

[17] F. Maffioli and A. Schiomachen. A mixed-integer model for solving ordering problems with side constraints. *Annals of Operations Research*, 69:277–297, 1997. doi: 10.1023/A:1018989130169.

[18] D. Manerba, R. Mansini, and R. Zanotti. Attended home delivery: reducing last-mile environmental impact by changing customer habits. *IFAC-PapersOnLine*, 51(5):55–60, 2018. doi: 10.1016/j.ifacol.2018.06.199.

[19] R. Mansini, F. Ranza, and R. Zanotti. Towards sustainable last-mile delivery: Introducing an off-peak urban policy to mitigate environmental and social impacts. *IFAC-PapersOnLine*, 58(2):186–191, 2024. doi: 10.1016/j.ifacol.2024.07.112.

[20] E. Marzal and L. Sebastia. Team orienteering problem with time windows and variable profit. page 347–350, 2022. doi: 10.15439/2022F158.

[21] Morgan Stanley. Here's why e-commerce growth can stay stronger for longer. https://www.morganstanley.com/ideas/global-ecommerce-growth-forecast-2022, 2022. Accessed: 2025-07-18.

[22] K. Mousavi, M. Bodur, and M. J. Roorda. Stochastic last-mile delivery with crowd-shipping and mobile depots. *Transportation Science*, 56(3):612–630, 2022. doi: 10.1287/trsc.2021.1088.

[23] R. Spliet and G. Desaulniers. The discrete time window assignment vehicle routing problem. *European Journal of Operational Research*, 244(2):379–391, jul 2015. doi: 10.1016/j.ejor.2015.01.020.

[24] S. Voigt, M. Frank, P. Fontaine, and H. Kuhn. The vehicle routing problem with availability profiles. *Transportation Science*, 57(2):531–551, mar 2023. doi: 10.1287/trsc.2022.1182.