Crowdkeeping in Last-mile Delivery *

Xin Wang, Okan Arslan, and Erick Delage

GERAD & Department of Decision Sciences, HEC Montréal, Montréal, QC H3T 2A7, Canada
{xin.wang@hec.ca, okan.arslan@hec.ca, erick.delage@hec.ca}

Abstract

In order to improve the efficiency of the last-mile delivery system when customers are possibly absent for deliveries, we propose the idea of employing the crowd to work as keepers and to provide storage services for their neighbors. Crowd-keepers have more flexibility, larger availability, and lower costs than fixed-storages such as automated lockers, and this leads to a more efficient and a more profitable system for last-mile deliveries. We present a bilevel program that jointly determines the assignment, routing, and pricing decisions while considering customer preferences, keeper behaviors, and platform operations. We develop an equivalent single-level program, which takes the form of a quadratic mixed-integer program with subtour elimination constraints, and it can be solved to optimality using a row generation algorithm. To improve the efficiency of the solution procedure, we further derive the exact best response sets for both customers and keepers, and approximate the optimal travel time using linear regression. We present a numerical study involving a real-world dataset. Both the fixed-storage and the no-storage systems are set as benchmarks to evaluate the performances of crowd-keepers. The results show that the crowdkeeping delivery system has the potential to yield more profits and produce less pollution due to its higher capability of consolidating deliveries and eliminating failed deliveries.

Keywords: last-mile delivery, crowdkeeping, bilevel program, pricing and routing, row generation

1 Introduction

E-commerce is thriving. The number of sales has almost tripled from 2014 to 2019 (Deloison et al. 2020). This boom has led to an unprecedented volume of goods being shipped every day. Customers are more demanding than ever in terms of the quality of delivery services: they are expecting to receive orders at any time they want (Ulmer and Savelsbergh 2020), and their expectations for speed forces e-tailers to offer same-day delivery with small time windows (Savelsbergh and Van Woensel 2016, Koch and Klein 2020). Such services lead to costly last-mile deliveries, which constitute the final stage in the delivery process

*This manuscript is under review at a scientific journal.

when a product is transported and delivered to a customer. Indeed, last-mile services can comprise up to 41% of the total cost to move goods (Jacobs et al. 2019).

Several novel technologies and business models, including crowd-shipping, drones, autonomous robots, and parcel lockers, have emerged. Common goals in such innovations are cost reduction or improved service quality. Sharing the same goals, we propose an innovative business model in last-mile delivery, which we refer to as *crowdkeeping*. Crowdkeeping is broadly defined as employing the crowd to keep parcels locally until customers pick them up. In other words, the *'crowd-keepers'* first attend the delivery on behalf of customers, and transfer parcels to customers on behalf of the delivery company. Compared to pickup points and automated lockers that are used in the real world, the availability and capacity of crowd-keepers are higher, the cost of using crowd-keepers is lower, and crowd-keepers are more flexible to adapt to different customer groups in different time periods. We consider an online service platform that coordinates customers and keepers to reduce the delivery costs and to improve the overall profitability of the delivery company.

Crowdkeeping is a new form of crowdsourcing in the last-mile delivery. It has the potential to reduce delivery costs, eliminate delivery failures, and improve the service quality. Moreover, it can be implemented without much infrastructure requirement and with modest operational cost. We contribute to the current research on delivery logistics from three aspects:

- We propose the idea of crowdkeeping for last-mile delivery systems, and present its concept, viability, benefits, and operational framework.
- To model the behaviors of all participants in the delivery system, including customers, keepers, and the platform, we present a bilevel program that jointly considers the assignment, routing, and pricing decisions. We make use of the duality theory to obtain an equivalent quadratic mixed-integer programming formulation, and develop a row generation algorithm to find the exact optimal solutions. To improve the efficiency of the solution procedure without sacrificing from the effectiveness, we propose an approximation model by approximating the optimal travel time using linear regression and by deriving explicit expressions for the best responses of customers and keepers.
- We carry out extensive experiments on a real-world dataset to investigate the effectiveness of the delivery system, the efficiency of the solution procedure, and how these are influenced by the number of customers, the customer absence ratio, the keeper service range, and various related costs. We find that crowdkeeping has the potential to improve the customer service level, increase the platform profits, and make the whole delivery system more cost-efficient and environmentally friendly.

The paper is organized as follows. Section 2 reviews the related literatures in last-mile delivery. We define the problem setting in Section 3, present a bilevel program in Section 4, and develop the solution methodology in Section 5. We finally carry out an extensive numerical study in Section 6 and present our conclusions in Section 7. We also refer the reader to Appendix A for a list of alternative delivery systems and Appendix B for all proofs.

2 Literature Review

In this section, we first make an overview on last-mile deliveries, including the three most common types and their respective challenges. We then review the most-recent innovations for handling these challenges in detail, including lockers in Section 2.2, crowdsourcing in Section 2.3, and demand management in Section 2.4. Undoubtedly, the problem considered here is closely related to the traveling salesman problem and its variants, which have a vast body of knowledge. We lastly review the related literature in Section 2.5.

2.1 Last-mile Delivery Types, Challenges, and Innovations

There is a wide range of products being shipped and delivered every day. According to the necessity of the customer presence and the coordination of delivery time windows, deliveries are categorized into three types. In 'unattended home delivery' (UHD), customer presence is not needed since a parcel is left at the doorstep with no attendance requirement. In 'attended home delivery' (AHD), the customer is required to be present at the time of delivery, for instance, an important document or a high-tech computer that requires a signature, or groceries shipped from a local store. In AHD, the company and the customer can either agree or not on a delivery time window, a.k.a resepctively coordinated AHD (c-AHD) and uncoordinated AHD (u-AHD).

Each delivery type poses different challenges. In UHD, coordination and customer absence is not a concern. However, theft and weather conditions pose important risks. There are also risks associated with denial-of-receipt or burglary at the house (McKinnon and Tallam 2003). In u-AHD, not finding the customer at home causes inefficiencies since it requires a second trip to the same customer. In c-AHD, timing is important. Companies offer limited number of delivery time slots to customers and each time slot comes potentially with a different delivery price (Ulmer 2020, Koch and Klein 2020). Time windows can increase the delivery costs significantly, because consolidation may not be possible for parcels destined to the same region. When customers cannot find a suitable time slot that fits their needs, the demand (and therefore the revenue) is lost.

Last-mile delivery is a growing field to deal with these challenges, that is, to make the deliveries on-time and low-risk, to eliminate failed deliveries, and to reduce the delivery costs. The goals are achieved either by improving the operational procedures with self-service lockers and crowdsourcing, or by managing the demand.

2.2 Self-service Locker Systems

Self-service locker systems are proposed to alleviate the risk of theft, to protect from unfavorable weather conditions, and to provide consolidation of parcels. Motivated by an example of Singaporean companies experimenting with a set of shared parcel lockers, Lin et al. (2020) propose a quantitative approach to determine the optimal locker locations with the objective to maximize the overall quality of services. Schwerdfeger and Boysen (2020) further consider the dynamic relocation of parcel lockers during the day.

Rohmer and Gendron (2020) extensively investigate different delivery concepts that exploit parcel locker stations and their associated decision problems.

There are unfortunately major disadvantages of employing lockers. First, setting up a network of lockers requires large initial investment costs, which can eventually lead to small returns. Currently, there is lack of a dense locker network. The ownership of lockers is also a major problem. Hasija et al. (2020) argue that, due to the proprietary nature of such systems, the utilization of lockers tends to be low. The lockers can also be shared among multiple firms, in which case the assignment of capacities becomes a concern. Shared or not, the use of lockers may result in rental costs, which could eventually be imposed on customers. Joerss et al. (2016) also report that *"Somewhat surprisingly, unattended delivery to parcel lockers does not really appeal to consumers despite the possibility of picking up their parcel 24/7"*. The authors report that customers put large value on home delivery instead of going to the lockers and conclude that their wide utilization is unlikely.

In our understanding, the locker system is a viable option for delivery, especially when the confidentiality of items is a concern. Therefore, lockers represent an important option of the last-mile delivery problem and add value to the overall system. However, due to the high initial cost, the benefits of using automated lockers are limited. The crowdkeeping system, on the other hand, provides a comparable service with no significant infrastructure requirement other than setting up an online platform. Even though the compensation is necessary to offer keepers proper incentives, it is not a major concern because the compensation can be adapted to capacity needs. In this case, the cost for unused capacities is avoided.

2.3 Crowdsourcing

Carbone et al. (2017) conceptualize the applications of crowdsourcing in logistics by reviewing the websites of 57 initiatives. The authors argue that most of these initiatives mainly offer two types of logistics services: crowdshipping and crowdstorage. Crowdshipping is the transportation of parcels by the crowd in return for a compensation and is offered as an option in the last-mile delivery. There is a high level of interest for crowdshipping both in practical applications and in the scientific literature. Arslan et al. (2019) report that several companies use crowdshipping partially or completely in their delivery operations. The authors investigate the benefits of crowdshipping by considering a platform that matches parcel delivery tasks and ad hoc drivers in real time. All requests are essentially served. A related problem is the online vehicle routing problem with occasional drivers (Archetti et al. 2021), in which a penalty is incurred for not serving a customer or for violating the time window constraints. Dayarian and Savelsbergh (2020) considers crowdshipping by employing in-store customers to deliver online orders. Ulmer and Savelsbergh (2020) study the problem of keeping a scheduled delivery workforce along with crowdsourcing to hedge against the uncertainty in crowdsourced delivery capacity. Qi et al. (2018) study shared mobility in last-mile delivery by optimally sizing the service zones. They argue that crowdshipping is not a scalable alternative of the conventional truck-only system in terms of operating costs, but that a combined operational mode can provide flexibilities and benefits. For a recent review on multiple dimensions of crowdshipping, we refer to Le et al. (2019) and Alnaggar et al. (2021). Crowdstorage, on the other hand, is considered as a logistics operation in rental of storage areas such as cellars, spare rooms, garages, or yards. It is considered as a local service that is particularly suitable in urban areas who need to store furniture or similar items for long terms. To the best of our knowledge, crowdstorage is still not considered for last-mile deliveries.

We define crowdkeeping by introducing the idea of crowdstorage into the last-mile delivery. That is, keepers can provide storage services for their neighbors, or neighbors can temporarily store parcels for absent customers. In fact, delivering a parcel to a neighbor is not an entirely new idea. Jacobs et al. (2019) reports that *"55% of consumers can accept the service of delivering products to neighbors in their vicinity"*. There are also empirical evidences that neighbors can cooperatively undertake delivery tasks with little or no compensation, and that 70% of customers in a survey reported that they can make deliveries for less than $5 (Devari et al. 2017). Nevertheless, there is no formal way of delivering to a neighbor. In the current operations, the courier needs to search for an available neighbor to deposit the parcel when the customer is absent from the delivery (McKinnon and Tallam 2003). In our study of crowdkeeping, neighbors are incentivized by a monetary compensation to participate in the delivery process as crowd-keepers. Then keepers are selected by the platform to serve multiple customers before deliveries. In this case, deliveries are consolidated, and the additional task of searching for an available neighbor is eliminated. Customers then pick up their parcels possibly by walking (Figure 1). At this point, it is noteworthy to mention that walking is also reported as a mode of transportation in crowd logistics (Carbone et al. 2017): *"Transport resources can be vans, cars, scooters, bicycles, public transport, or even walking"*. We identified single study in the literature considering walking as a form of transportation in crowdshipping. Martinez-Sykora et al. (2020) consider drivers making deliveries in dense urban areas by walking at the end of their vehicle trip in crowdshipping to avoid heavy traffic.

2.4 Demand Management

There is extensive research in the area of demand management for last-mile deliveries. Our work is most related to the service time slot, the service price, and customer incentives. E-tailers can offer different delivery time windows and associated prices to manage the demand. Several static and dynamic demand management strategies have been investigated including differentiated time slot allocation and differentiated time slot pricing.

In the time slot allocation dimension, Agatz et al. (2011) study assigning time slots to zip codes in a service region to minimize the delivery costs. Spliet and Gabor (2015) introduce the time window assignment vehicle routing problem, in which time windows have to be assigned before demand is realized. Bruck et al. (2018) study the problem of creating time slot tables and routing technicians in a cost-effective way. A decision support system for an Italian company is also developed for a related problem (Bruck et al. 2020).

In the time slot pricing dimension, Yang and Strauss (2017) present a delivery cost approximation scheme by decomposing the delivery problem into a collection of smaller problems. The customers' delivery time slot choices are estimated using a multinomial logit model. Klein et al. (2019) study differentiating the time-slot pricing by considering the routing phase. The customers' choice behavior is modeled as a general nonparametric rank-based choice model. These authors study two policies for incorporating the routing costs, by explicitly incorporating the routing constraints to their model or by using a model-based approximation and find that the latter can be used in real-world applications. In a similar line of research, Klein et al. (2018) present a cost approximation approach for dynamic time slot pricing decisions by forecasting the potential future customers. Koch and Klein (2020) additionally combines dynamic pricing with dynamic vehicle routing.

Another interesting idea in demand management is incentivizing customers. One of the first papers in customer incentives is Campbell and Savelsbergh (2006), who investigate the use of incentives for demand management to reduce the delivery costs. Ulmer (2020) considers anticipatory pricing and routing policy method for the same-day delivery, in which customers are incentivized to select delivery deadline options efficiently to align with routing considerations. Yildiz and Savelsbergh (2020) consider offering a discount to customers on their delivery fee in return for flexibility to adjust a previously agreed upon delivery window. The authors report that the cost savings of offering discounts can exceed 30%.

These studies on demand management reveal the importance of the time slot management, the pricing for services and incentives, and the integration of pricing and routing. However, some customers may be unavailable in any of the offered time slots, and this implies lost revenues. Our approach can provide on-time deliveries but does not have to enforce the time slot management, and it can jointly consider the service pricing, the incentive pricing, and the routing decisions.

2.5 Traveling Salesman Problem and Variants

The Traveling salesman problem (TSP) is one of the most touted problems in logistics and its extensions attract significant attention due to their economic importance, theoretical challenge, and applicability in many real-world contexts (Vidal et al. 2020). It consists of finding one route from a depot such that all customers are visited and the total cost is minimized. We refer the reader to Applegate et al. (2007) for reviews in TSP and to Toth and Vigo (2014) for related problems, methods, and applications. When customers are not necessarily visited by a vehicle and it is sufficient to visit another close-by node in the network, the problem is then called covering-tour problem (CTP) (Gendreau et al. 1997). The problem we consider in this paper is closely related to CTP, because keepers in the same vicinity of customers cover the customer nodes. Exact solutions of CTP and its variants are notoriously difficult to obtain. A branch-and-price algorithm is introduced by Jozefowiez (2014) for solving CTP. The pricing subproblem is a ring-star problem, which is solved using a branch-and-cut algorithm. Kartal et al. (2017) introduce the single allocation p-hub median location and routing problem with simultaneous pick-up and delivery. Other closely related problems are location-or-routing problem (LoRP) by Arslan (2021), and location-and-routing problem (LRP) by Kartal et al. (2017). In LoRP, the location decision is related to depots, the vehicles are dispatched from the selected depots, the depots are not connected to each other, and a customer is served either by being covered by a located depot or by being visited by a vehicle routing. In LRP, the depots are connected, the vehicles are dispatched from the selected depots, and all customers are visited by a vehicle routing. In our crowdkeeping delivery problem, the location decision is related to keepers and customers, a vehicle is assumed to visit a subset of nodes (i.e., the active nodes), and the inactive nodes are covered by the active nodes. Compared to them, pricing decisions are incorporated into our problem.

3 Problem Description

This section introduces the crowdkeeping framework by comparing it with the standard operational framework for deliveries. It describes the behaviors and decisions of keepers, customers, and the platform and lists the potential benefits and real-world applications of crowdkeeping.

3.1 Operational Framework

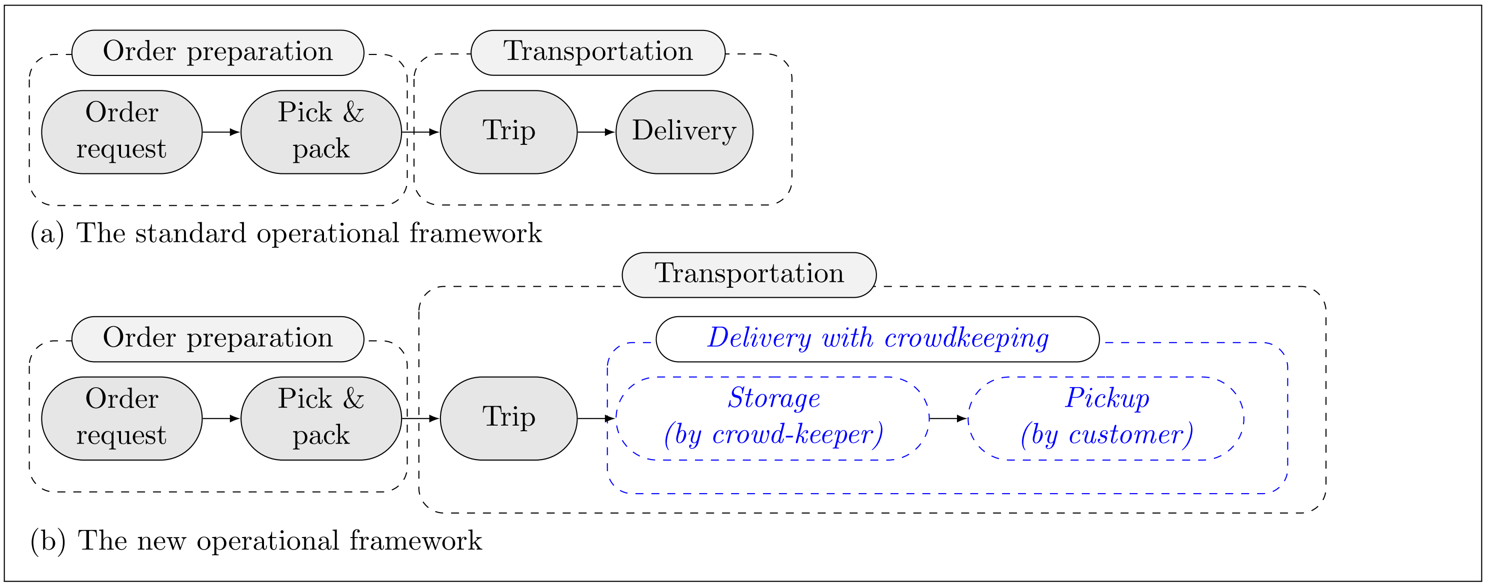


Figure 1: Comparison of the standard and the new operational frameworks for delivery systems

In the standard operational framework for delivery systems (Figure 1(a)), the delivery phase is a combination of order preparation and parcel transportation. In the order preparation phase, companies receive order requests, pick up items, and pack them as parcels for deliveries. In the transportation phase, parcels are in transit from their origin depots to their destination depots, and delivered to customers in last mile deliveries.

In our new business model (Figure 1(b)), we decompose the last-mile delivery process into two steps. First, parcels are delivered to keepers who store them for customers. Second, customers pick up their parcels from their selected keepers to finish the delivery process. Decomposition of tasks allows deliveries to be coordinated with absent customers and also with crowd-keepers who have more flexibility to consolidate orders in their neighborhood.

3.2 Participants and Their Behaviors

In the crowdkeeping delivery system, there are three groups of players: customers, keepers and the platform. The ***customers*** are people who purchase products online and expect their parcels to be delivered. The ***keepers*** are individuals, such as homemakers, stay-at-home parents, home-office workers, and unemployed persons, that can receive parcels and temporarily store them. This term makes a clear reference to the duty that such an individual performs and emphasizes the functional difference from the "*courier*" generally used in crowdshipping. Similar to other supply sides on crowdsourcing platforms, keepers work in reputation-based systems and normally receive a compensation for every customer they serve. Different from other crowdsourcing platforms, the entry to the crowdkeeping market is simpler, because it only requires a smart phone and no investment or special equipment is necessary. Coupled with the mobile application, a smart phone is capable of updating the tracking information, specifying the pickup location and duration, and collecting the customer signature to ensure the safety and the convenience of the delivery process. The third participant, the ***platform*** coordinates the delivery between customers and keepers. The delivery company delivers the product to the keeper, in that, it has more flexibility to consolidate orders in the same neighborhood. The online service platform we consider here acts as an intermediary between customers and delivery companies to ensure on-time delivery by employing keepers. From the platform's perspective, the customers represent the *demand*, and keepers represent the *supply*, and the objective is to match the demand and supply in this market.

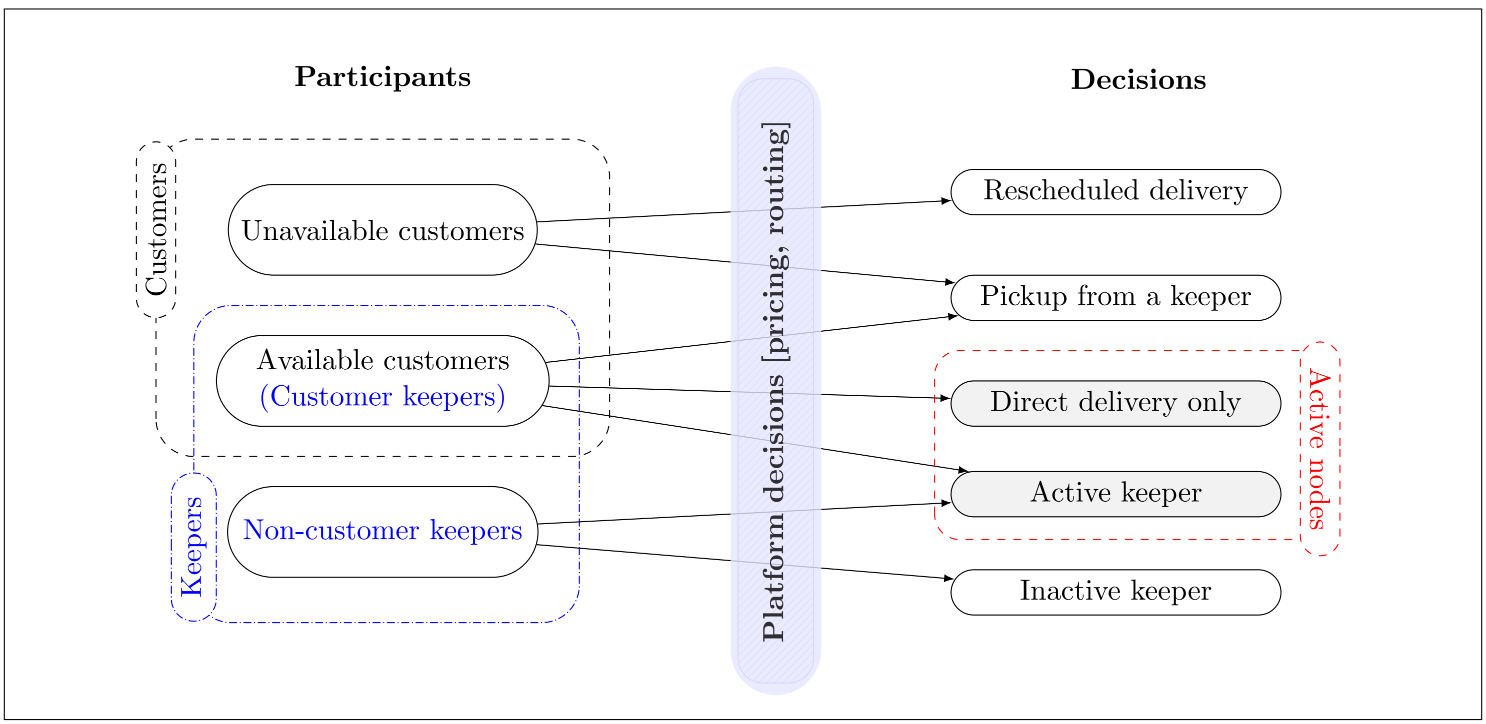


Figure 2: Problem setting of the crowdkeeping delivery system

Observing that customers may be absent during the deliveries, we categorize customers as *available customers* who are available for attended home deliveries, and *unavailable customers* who are absent for deliveries. The available customers are also referred to as *customer keepers* because they can additionally provide storage services in their neighborhood. In our problem setting, we also consider keepers who are not necessarily a customer receiving a parcel, but who may declare their availability to provide storage services. This group is called *non-customer keepers*. These participants are displayed in the left side of Figure 2. Each arrow in this figure represents a potential option for a participant. Each participant makes its own decisions, which are guided (or filtered) by the platform's decisions on the pricing and routing.

We now discuss each participant's choices. Unavailable customer can reschedule the delivery for another day or they can select one available keeper and then pick up their parcels from the specified keeper at their convenience. Available customers who are also customer keepers have three choices. They can choose to pick up their parcels from a selected keeper subject to a pickup fee, or they can choose the direct delivery to doorstep subject to a delivery fee. They can also choose to provide storage services as *available keepers*. If another customer chooses to pick up from them, they become *active keepers*. Finally, non-customer keepers can either choose to show their availability and offer storage services if the compensation is attractive, or they will be unavailable. Note that, if a customer or a keeper is directly visited by the delivery company, we refer to those participants as *active nodes*, which are the gray nodes on the right side of Figure 2. The active keepers are those available keepers who are selected by customers for keeping parcels, and the active customers are those who choose 'direct delivery only' or who choose direct delivery and also being active keepers.

3.3 Benefits and Applications

The idea of using pickup locations to eliminate failed deliveries and improve the efficiency of the last-mile delivery is not a new idea. For instance, Amazon and IKEA have several pickup points in cities to provide self-pickup services for customers. These points are generally managed by full-time employees, have fixed locations, intend to provide long-term services, and therefore are inflexible to changes. Similar to pickup points, automated lockers also provide self-pickup services, but without in-person supervision. We define such pickup points and lockers as *fixed-storages* since they both have fixed locations. For more details on the delivery system with fixed-storage, please refer to Appendix A.2.

Compared to fixed-storages, crowd-keepers has the advantages of offering more flexibility, larger availability and lower costs. To be specific, there are concrete benefits of crowdkeeping for distribution companies, customers, and keepers. The distribution companies can improve the operational efficiency and reduce costs because deliveries are consolidated in terms of both time and space. They can also expand the delivery capacity and eliminate failed deliveries because of the large flexibility and high availability provided by the crowd, who can be any available individuals. Since customers are potential crowd-keepers, keeper locations can change as customer locations change in different time periods. Additionally, crowd-keepers only need to be active if they are chosen by customers, leading to a higher utilization and higher flexibility than fixed-storages. Customers will receive better services because they can get their parcels in person whenever they are available. The safety of parcels is improved with the supervision of keepers. Moreover, except for paying the standard delivery fee to receive their parcels, customers have another choice to use the pickup service with a lower fee. Crowd-keepers can earn the compensation or rewards by participating in this system without much investment or setup cost for being keepers.

Even though the legal issues are beyond the scope of this study, there is a wealth of experience gathered in crowdshipping applications, which are directly transferable to crowdkeeping. Furthermore, local neighborhood is less prone to legal concerns due to the neighborhood relationship between the parties, and there are real-world applications such as Pickme[1], Voisinsrelais[2], and Cainiao Network[3].

4 Bilevel Program for Crowdkeeping Delivery Problem

We now define the Crowdkeeping Delivery Problem (CDP) and present the customer, the keeper, and the platform models. We then develop a bilevel program for the CDP by considering the platform as the leader, and customers and keepers making decision simultaneously as followers.

Definition 1. *The* ***Crowdkeeping Delivery Problem*** *is defined as pricing the compensation and the pickup fee that maximize the platform profit by respecting independent decision making mechanisms of customers and keepers, which involve minimization of the delivery service cost for each customer and maximization of the profit for providing storage services for each keeper.*

The decisions involved in the CDP by different participants are as presented by the arrows in Figure 2. Descriptions of the notation are given in Table 1. The timeline of decisions made by the platform, customers, and keepers are shown in Figure 3. We assume that delivery operations are carried out under

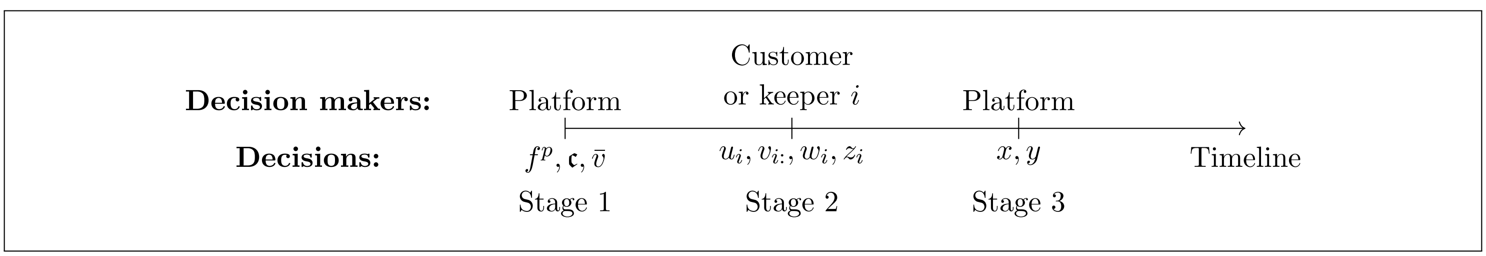


Figure 3: Timeline of decisions made by the platform, customers, and keepers

the condition of full information. The platform, as the leader, prices the pickup fee and compensation first; then customers, as followers, will choose the direct delivery, pickup, rescheduled delivery, and being a keeper according to the revealed decisions of the platform and that of other customers or keepers. Finally, the platform will visit all active nodes according to the decisions of all customers and keepers. That is, the service fee and the compensation are revealed before customers and keepers make their decisions, and customer demands and keeper availabilities are revealed before the routing planning takes place. These assumptions lead to a form of Stackelberg game.

[1]https://mypickme.zendesk.com/hc/fr, accessed February 12, 2022.
[2]http://www.voisinsrelais.org, accessed February 12, 2022.
[3]https://www.cainiao.com, accessed February 10, 2022.

Table 1: Notations

Sets	**Description**
$\mathcal{N}$	the set of all customers, including unavailable customers and customer keepers
$\mathcal{M}$	the set of non-customer keepers
Customer decision variables	
u_i	1 if customer i chooses the direct delivery, 0 otherwise
v_{ij}	1 if customer i chooses to pick up from keeper j, 0 otherwise
w_i	1 if participant i makes itself available as a keeper, 0 otherwise
z_i	1 if customer i chooses to reschedule the delivery, 0 otherwise
Platform decision variables	
f^p	the pickup fee offered to customers
$\mathfrak{c}$	the compensation offered to keepers
$\bar{v}_{ij}$	1 if platform accepts to show keeper j as being available to customer i in terms of limited capacities, 0 otherwise
y_j	1 if customer or keeper j is active and need to be visited, 0 otherwise
x_{ij}	1 if arc (i, j) appears on tour, 0 otherwise
Parameters	
f^d	the standard delivery fee offered to customers, i.e., the reference fee
t_{ij}	the travel time between i and j
a_i	1 if customer i is absent for deliveries, 0 otherwise
b_j	the capacity of keeper j
e_i	1 if node i is a customer, 0 if node i is a non-customer keeper
r_{ij}	1 if customer i is willing to pick up from keeper j in terms of restrictions on the service zone and walk time, 0 otherwise (with $r_{ii} = 0$ to represent impossibility of keeping for oneself)
c^p	the inconvenience cost per minute of walk time for picking up, i.e., pickup cost
c^d	the truck delivery cost per minute of travel time
c^r	the additional cost of rescheduling a delivery incurred by the platform
c^k	the inconvenience cost for providing keeping services, i.e., keeping cost
$\hat{v}_{ki}$	1 (in customer problem) if customer $k \neq i$ chooses to pick up from customer i, 0 otherwise
$\hat{w}_j$	1 (in customer problem) if customer $j \neq i$ accepts to work as a keeper, 0 otherwise

4.1 Customer and Keeper Models

There are two groups of customers and they have different choices. Unavailable customers cannot receive the parcel in person for the direct delivery, but they can pick up their parcel from a keeper or reschedule the delivery. Available customers have three choices. The first one is to pay the standard delivery fee and have their parcels delivered to their doorstep. The second one is to pay the pickup fee, select a keeper, and pick up their parcels from the keeper. They also have a third option, which is to work as a crowd-keeper and provide keeping services for other customers in addition to receiving their own parcels. The model for customer or keeper i is as follows:

$$
\begin{aligned}
H_i(f^p, \mathfrak{c}, \bar{v}, \hat{v}_{:i}, \hat{w}) \triangleq \min_{u_i, v_{i:}, w_i, z_i} \quad & f^d(u_i + z_i) + \left(c^k - \mathfrak{c} \sum_{k \in \mathcal{N}} \hat{v}_{ki}\right) w_i + \sum_{j \in \mathcal{M} \cup \mathcal{N}} (f^p + c^p t_{ij}) v_{ij} && \text{(1a)} \\
\text{s.t.} \quad & u_i + z_i + \sum_{j \in \mathcal{M} \cup \mathcal{N}} v_{ij} = e_i && \text{(1b)} \\
& v_{ij} \leq \bar{v}_{ij}, \forall j \in \mathcal{M} \cup \mathcal{N} && \text{(1c)} \\
& v_{ij} \leq r_{ij} \hat{w}_j, \forall j \in \mathcal{M} \cup \mathcal{N} && \text{(1d)} \\
& u_i \leq 1 - a_i && \text{(1e)} \\
& z_i \leq a_i && \text{(1f)} \\
& e_i w_i \leq u_i && \text{(1g)} \\
& u_i, v_{ij}, w_i, z_i \in \{0, 1\}, \forall j \in \mathcal{M} \cup \mathcal{N}, && \text{(1h)}
\end{aligned}
$$

where $\hat{v}_{:i}$ and $v_{i:}$ denote a column and and row vector respectively. Given the platform's decision on the pickup fee f^p, keeper compensation $\mathfrak{c}$, and available keepers $\bar{v}$ (with $\bar{v}_{ii} = 0$) and other customers and keepers' decisions $\hat{w}$ and $\hat{v}_{:i} := [\hat{v}_{1i}, \ldots, \hat{v}_{|\mathcal{N}|i}]$ (with $\hat{v}_{ii} = 0$), $H_i(f^p, \mathfrak{c}, \bar{v}, \hat{v}_{:i}, \hat{w})$ is the optimization model of customer i. The latter decides whether they prefer a direct delivery u_i, picking up from a keeper $v_{i:}$, acting as a keeper w_i, or rescheduling the delivery z_i.

The objective function (1a) states that each customer minimizes the total amount they pay to receive their parcels. If customers choose the delivery option, whether it is the direct same-day delivery or rescheduled next-day delivery, they need to pay the delivery fee f^d. If they work as crowd-keepers, they earn compensation $\mathfrak{c}$ for each customer they serve. Providing the service costs each keeper a fixed inconvenience cost c^k. If customers choose the pickup option, they need to pay the pickup fee f^p and the inconvenience cost c^p for each minute of pickup travel time.

The parameter e_i indicates whether participant i is a customer, i.e. $e_i = 1$ if $i \in \mathcal{N}$. In this case, constraint (1b) ensures that to receive the parcel, each customer must choose the direct delivery (if available), the rescheduled delivery (if absent), or the option of crowdkeeping by being served by one selected keeper. Constraints (1c) and (1d) specify that customers can only retrieve parcels from an available and approved (by the platform) keeper that is in the same zone and within the acceptable walk time. We note that $v_{ii} \leq r_{ii}\hat{w}_i$, is equivalent to $v_{ii} \leq 0$ given that $r_{ii} = 0$, hence $H_i(f^p, \mathfrak{c}, \bar{v}, \hat{v}_{:i}, \hat{w})$ is insensitive to $\hat{v}_{ii}$. This property represents that customers can receive their parcels by choosing direct delivery and working as keepers, but they cannot receive any compensation for keeping their own parcels.

Constraints (1e) and (1f) state that customers cannot choose the direct delivery and they may reschedule their deliveries if absent ($a_i = 1$), and that they have no reason to delay the delivery if available ($a_i = 0$). Constraint (1g) states that customers who can work as crowd-keepers and provide services for others only when they are available for direct deliveries. Constraints (1h) are domain restrictions.

When $e_i = 0$, the formulation (1) models non-customer keeper $i \in \mathcal{M}$ and, since u_i, v_{ij}, z_i all equal 0, reduces to:

$$H_i(f^p, \mathfrak{c}, \bar{v}, \hat{v}_{:i}, \hat{w}) \triangleq \min_{w} \quad \left(c^k - \mathfrak{c} \sum_{k \in \mathcal{N}} \hat{v}_{ki}\right) w_i \tag{2a}$$

$$\text{s.t.} \quad w_i \in \{0, 1\}. \tag{2b}$$

In this case, keepers are only willing to act as keepers if the total to-be-earned compensation is higher than the inconvenience cost.

Proposition 1. *Model* (1) *(and* (2)*) with relaxed integrality requirement always has an optimal solution for which all variables assume binary values.*

The proof of Proposition 1 is presented in Appendix B.1.

Remark 1. *Each participant has at most* $|\mathcal{M}| + |\mathcal{N}| + 2$ *choices.*

Remark 2. *Both the formulation (1) and (2) are always feasible.*

4.2 Platform Model

On the supply side, the platform attracts the crowd who provide storage services by offering them compensation, and on the demand side, it encourages customers to use storage services by offering them convenience and a lower fee. The platform matches supply and demand by pricing the compensation $\mathfrak{c}$ and the pickup fee f_p, and by showing keeper availabilities to customers using $\bar{v}$ variables. The platform model is:

$$\max_{f^p, \mathfrak{c}, \bar{v}} \quad \sum_{i \in \mathcal{N}} \left[f^d(u_i + z_i) + \sum_{j \in \mathcal{M} \cup \mathcal{N}} f^p v_{ij} \right] - \mathfrak{c} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M} \cup \mathcal{N}} v_{ij} - c^d h(u, v) - c^r \sum_{i \in \mathcal{N}} z_i \tag{3a}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{N}} \bar{v}_{ij} \le b_j, \forall j \in \mathcal{M} \cup \mathcal{N} \tag{3b}$$

$$\bar{v}_{ij} \in \{0, 1\}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \cup \mathcal{N} \tag{3c}$$

$$f^p \in [0, \bar{f}^p], \mathfrak{c} \in [0, \bar{\mathfrak{c}}], \tag{3d}$$

Objective function (3a) states that the platform maximizes its profit, which depends on how customers and keepers react to its decisions. The revenue is generated from the delivery and pickup fees paid by customers. The cost is due to the compensation, due to visiting active keepers and customers, and due to the rescheduled deliveries. The term $h(u, v)$ is the minimal travel time for visiting active nodes, which will be elaborated below. Constraints (3b) ensure that the platform does not offer more deliveries to each available keeper j than what they can handle in terms of capacity b_j. Constraints (3c) and (3d) are domain restrictions.

The term $h(u, v)$ in (3a) is the optimal travel time obtained by solving the following model:

$$h(u,v) := \min_{x,y} \quad \sum_{i\in\mathcal{M}\cup\mathcal{N}} \sum_{j\in\mathcal{M}\cup\mathcal{N}} t_{ij}x_{ij} \tag{4a}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{M}\cup\mathcal{N}} x_{ij} = y_j, \forall j \in \mathcal{M}\cup\mathcal{N} \tag{4b}$$

$$\sum_{i\in\mathcal{M}\cup\mathcal{N}} x_{ji} = y_j, \forall j \in \mathcal{M}\cup\mathcal{N} \tag{4c}$$

$$\sum_{i,j\in\mathcal{S}} x_{ij} \leq |\mathcal{S}| - 1, \forall \mathcal{S} \subset \mathcal{M}\cup\mathcal{N}, 2 \leq |\mathcal{S}| \leq |\mathcal{M}\cup\mathcal{N}| - 2 \tag{4d}$$

$$\sum_{i\in\mathcal{N}} v_{ij} \leq b_j y_j, \forall j \in \mathcal{M}\cup\mathcal{N} \tag{4e}$$

$$y_j \geq u_j, \forall j \in \mathcal{N} \tag{4f}$$

$$x_{ij} \in \{0,1\}, y_j \in \{0,1\}, \forall i \in \mathcal{M}\cup\mathcal{N}, \forall j \in \mathcal{M}\cup\mathcal{N}, \tag{4g}$$

where x defines the tour to the active nodes identified by y. Objective function (4a) is the travel time for visiting all active nodes. Constraints (4b) and (4c) are degree constraints, and (4d) are subtour elimination constraints. Constraints (4e) and (4f) require that keepers serving others and customers with direct delivery are active nodes that need to be visited. Finally, (4g) are domain restrictions.

4.3 Bilevel Program with Multiple Followers

Considering the platform as the leader and customers and keepers as followers, the bilevel program (BP) for the CDP is presented as follows:

$$(\text{BP}) \max_{\substack{f^p,\mathfrak{c},x,y,\bar{v}\\u,v,w,z}} \quad \left(\begin{array}{l} \sum_{i\in\mathcal{N}} \left[f^d(u_i + z_i) + \sum_{j\in\mathcal{M}\cup\mathcal{N}} f^p v_{ij} \right] \\ -\mathfrak{c} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} - c^d \sum_{i\in\mathcal{M}\cup\mathcal{N}} \sum_{j\in\mathcal{M}\cup\mathcal{N}} t_{ij}x_{ij} - c^r \sum_{i\in\mathcal{N}} z_i \end{array} \right) \tag{5a}$$

$$\text{s.t.} \quad (3b) - (3d), (4b) - (4g)$$

$$(u_i, v_{i:}, w_i, z_i) \in \arg\min H_i(f^p, \mathfrak{c}, \bar{v}, v_{:i}, w), \forall i \in \mathcal{M}\cup\mathcal{N}, \tag{5b}$$

where constraints (5b) indicate that $(u_i, v_{i:}, w_i, z_i)$ must be the optimal responses of each customer or keeper i.

The CDP is a generalization of the Covering Tour Problem (CTP) and therefore is NP-hard. Specifically, customers who choose delivery and keepers who serve customers are active points to be visited in the optimal tour, and customers who choose to pick up are inactive points to be covered by active nodes. Compared to CTP, the main difference here is that the decisions are decentralized in CDP and that the platform, keepers and customers optimize their respective objective functions. In addition to the covering and routing decisions, we also consider the pricing decisions. Note that, in this study, customers are assumed to pick up their parcels from the keeper, which is in line with the current practice. Nevertheless, the keeper could also do the deliveries to customers. A model that accommodates such a feature is presented in Appendix A.3.

5 Solution Procedure

In order to solve the BP, we first reformulate it into an equivalent singe-level model using the strong duality theorem and then solve the model exactly using a row generation algorithm. We then develop an approximation of the optimal travel time and derive the exact best response sets of followers to improve the efficiency of the solution procedure.

5.1 Reformulation as a Single-level Program

Due to Proposition 1, the integrality requirement of the variables u_i, v_{ij}, w_i, z_i can be relaxed into $\{u_i, v_{ij}, w_i, z_i \geq 0\}$. The upper bounds $\{u_i, v_{ij}, w_i, z_i \leq 1\}$ can be omitted since they are implied by constraints (1b) and (1g). The BP is then given as:

$$\begin{aligned} \max_{\substack{f^p,\mathfrak{c},x,y,\bar{v} \\ u,v,w,z}} \quad & \sum_{i\in\mathcal{N}} \left[f^d(u_i+z_i) + \sum_{j\in\mathcal{M}\cup\mathcal{N}} f^p v_{ij} \right] - \mathfrak{c} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} - c^d \sum_{i\in\mathcal{M}\cup\mathcal{N}} \sum_{j\in\mathcal{M}\cup\mathcal{N}} t_{ij}x_{ij} - c^r \sum_{i\in\mathcal{N}} z_i \\ \text{s.t.} \quad & (3b)-(3d), (4b)-(4g) \\ & (u_i, v_{i:}, w_i, z_i) \in \arg\min \bar{H}_i(f^p, \mathfrak{c}, \bar{v}, v_{:i}, w), \forall i \in \mathcal{M}\cup\mathcal{N}, \end{aligned}$$

with

$$\bar{H}_i(f^p, \mathfrak{c}, \bar{v}, \hat{v}_{:i}, \hat{w}) \triangleq \min_{u_i, v_{i:}, w_i, z_i} \quad f^d(u_i+z_i) + \left(c^k - \mathfrak{c} \sum_{k\in\mathcal{N}} \hat{v}_{ki} \right) w_i + \sum_{j\in\mathcal{M}\cup\mathcal{N}} (f^p + c^p t_{ij}) v_{ij} \tag{6a}$$

$$\text{s.t.} \quad (1b)-(1g)$$

$$u_i, v_{ij}, w_i, z_i \geq 0, \forall j \in \mathcal{M}\cup\mathcal{N} \tag{6b}$$

Due to Remark 2, the strong duality holds for each follower model. Let $\gamma, \lambda, \mu, \nu, \phi, \psi$ be the dual variables corresponding to constraints (1b)−(1g) respectively. Due to the strong duality theorem, the optimal objective function value of the customer model and its dual objective function value are always equal:

$$\begin{pmatrix} f^d(u_i+z_i) + \left(c^k - \mathfrak{c} \sum_{k\in\mathcal{N}} \hat{v}_{ki} \right) w_i \\ + \sum_{j\in\mathcal{M}\cup\mathcal{N}} (f^p + c^p t_{ij}) v_{ij} \end{pmatrix} = \begin{pmatrix} e_i\gamma_i - \sum_{j\in\mathcal{M}\cup\mathcal{N}} (\bar{v}_{ij}\lambda_{ij} + r_{ij}\hat{w}_j\mu_{ij}) \\ -(1-a_i)\nu_i - a_i\phi_i \end{pmatrix}, \forall i \in \mathcal{M}\cup\mathcal{N} \tag{7}$$

We then reformulate the BP into an equivalent single-level program (SP_1):

$$(\text{SP}_1) \quad \max_{\substack{f^p,\mathfrak{c},x,y,\bar{v},u,v,w,z,\\ \gamma,\lambda,\mu,\nu,\phi,\psi}} \quad \left(\begin{array}{l} \sum_{i\in\mathcal{N}} \left[f^d\left(u_i+z_i\right) + \sum_{j\in\mathcal{M}\cup\mathcal{N}} f^p v_{ij} \right] \\ -\mathfrak{c} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} - c^d \sum_{i\in\mathcal{M}\cup\mathcal{N}} \sum_{j\in\mathcal{M}\cup\mathcal{N}} t_{ij}x_{ij} - c^r \sum_{i\in\mathcal{N}} z_i \end{array} \right) \tag{8a}$$

$$\text{s.t.} \quad (1b)-(1g), (3b)-(3d), (4b)-(4g), (6b), (7)$$

$$\gamma_i - \nu_i + \psi_i \le f^d, \forall i \in \mathcal{M}\cup\mathcal{N} \tag{8b}$$

$$\gamma_i - \phi_i \le f^d, \forall i \in \mathcal{M}\cup\mathcal{N} \tag{8c}$$

$$\gamma_i - \lambda_{ij} - \mu_{ij} \le f^p + c^p t_{ij}, \forall i \in \mathcal{M}\cup\mathcal{N}, \forall j \in \mathcal{M}\cup\mathcal{N} \tag{8d}$$

$$-e_i\psi_i \le c^k - \mathfrak{c}\sum_{k\in\mathcal{N}} v_{ki}, \forall i \in \mathcal{M}\cup\mathcal{N} \tag{8e}$$

$$\lambda_{ij}, \mu_{ij}, \nu_i, \phi_i, \psi_i \ge 0, \forall i \in \mathcal{M}\cup\mathcal{N}, \forall j \in \mathcal{M}\cup\mathcal{N} \tag{8f}$$

$$\gamma_i \text{ is free }, \forall i \in \mathcal{M}\cup\mathcal{N}, \tag{8g}$$

where constraints (3b)−(3d) and (4b)−(4g) are constraints of the platform model, constraints (1b)−(1g) and (6b) ensure the primal feasibility and (8b)−(8g) ensure the dual feasibility of the customer model, and constraint (7) ensure the optimality of the customer model. Note that the $\hat{v}$ ($\hat{w}$) and v (w) refer to the same decision variable, and the former ones are used in one specific customer or keeper model to distinguish the decisions of themselves and that of other customers and keepers. Therefore, $\hat{v}$ ($\hat{w}$) in (1d) and (7) can be equivalently replaced with v (w).

5.2 Row Generation Method

We use the row generation method in order to solve the single-level model (8) with exponentially many subtour elimination constraints (SEC) (4d). We first solve the relaxed SP_1 model (SP_1^R), which is obtained by removing the SECs. Let h^* be an optimal solution of the SP_1^R. Given h^*, we then identify a violated SEC. We search for a subset $\mathcal{S} \subset \mathcal{M}\cup\mathcal{N}$ with $\sum_{i,j\in\mathcal{S}} x_{ij} > |\mathcal{S}|-1$ and $2 \le |\mathcal{S}| \le |\mathcal{M}\cup\mathcal{N}|-2$. If such subset $\mathcal{S}$ exists, then constraint $\sum_{i,j\in\mathcal{S}} x_{ij} \le |\mathcal{S}|-1$ is violated and is added to the SP_1^R, and the SP_1^R is resolved. Otherwise, there is no violated constraint and the current h^* is optimal. The pseudocode of the algorithm is given in Algorithm 1.

5.3 Approximation Model with Estimated Travel Time

Solving SP_1 model exactly for large instances may be computationally inefficient. When nodes are identically and independently distributed according to a probability density function f on a two-dimensional region R, Beardwood et al. (1959) show in their seminal work that:

$$\lim_{n\to\infty} \frac{\text{TSP}_n^*}{\sqrt{n}} \approx \beta \iint_R \sqrt{f(x,y)}\, dxdy,$$

Algorithm 1 Row generation algorithm

Initialization: Solve the relaxed SP_1^R without subtour elimination constraints (SEC), and obtain the optimal value h^*, the number of visited nodes N, the nodes set $U \leftarrow \{1, 2, .., n+m\}$, and the optimal routes set $E \leftarrow \{(i,j) : i \in U, j \in U, x_{ij}^* = 1\}$.
Repeat: Given U and E, identify a subtour $\mathcal{S}$.
  If $|\mathcal{S}| < N$, which means there exists a subtour $\mathcal{S}$,
    then add the lazy cut $\sum_{i,j \in \mathcal{S}} x_{ij} \leq |\mathcal{S}| - 1$ to the SP_1^R.
  Break
  Resolve SP_1^R with lazy cuts.
Until: The current route has no subtours.

where TSP_n^* is the optimal travel time, n is the number of nodes, $\iint_R \sqrt{f(x,y)}\, dxdy$ is the integral density of the region R, and β is a constant. When nodes are uniformly and independently scattered, the integral density is then equal to the area of the region, A. In this case, $\beta\sqrt{nA}$ is asymptotically a good approximation for the optimal travel time as $n \to \infty$. Considering that our model is meant to serve real-world cases, where the node dispersion is unknown, the integral density cannot be computed. Therefore, we consider $\iint_R \sqrt{f(x,y)}\, dxdy$ as part of the approximation, similar to (Cavdar and Sokol 2015). We then use regression to approximate the term $\beta \iint_R \sqrt{f(x,y)}\, dxdy$ as $\hat{\beta}(n)$ for each instance region, since β also depends on the number of nodes (Franceschetti et al. 2017) and $\iint_R \sqrt{f(x,y)}\, dxdy$ depends on the node distribution. That is,

$$\text{TSP}(n) \approx \hat{\beta}(n)\sqrt{n}, \tag{9}$$

where $\text{TSP}(n)$ is the approximated optimal travel time of visiting n number of nodes. Both $\text{TSP}(n)$ and $\hat{\beta}(n)$ are functions of n, which is a auxiliary decision variable in our model with $n = \sum\limits_{j \in \mathcal{M} \cup \mathcal{N}} y_j$. That is, the number of nodes to be visited depends on the decisions to be made and therefore $\hat{\beta}$ is not known before the problem is optimized. In order to estimate $\hat{\beta}$ as a function of n for each region in our dataset, we first solve the SP_1 with the exact algorithm to get the optimal cost $\text{TSP}^*(n)$ under different customer densities, and then use the pair $(n, \text{TSP}^*(n))$ as input for model fitting. This approximation (9) yields a more efficient formulation. We present the approximation model (AM_1) below:

$$(\text{AM}_1) \quad \max_{\substack{f^p, \mathfrak{c}, y, \bar{v}, u, v, w, z,\\ n, \gamma, \lambda, \mu, \nu, \phi, \psi}} \quad \left(\begin{array}{l} \sum\limits_{i \in \mathcal{N}} \left[f^d (u_i + z_i) + \sum\limits_{j \in \mathcal{M} \cup \mathcal{N}} f^p v_{ij} \right] \\ -\mathfrak{c} \sum\limits_{i \in \mathcal{N}} \sum\limits_{j \in \mathcal{M} \cup \mathcal{N}} v_{ij} - c^d \hat{\beta}(n)\sqrt{n} - c^r \sum\limits_{i \in \mathcal{N}} z_i \end{array} \right)$$

$$\begin{aligned} \text{s.t.} \quad & (1b) - (1g), (3b) - (3d), (4\text{e}), (4\text{f}), (6b), (7), (8b) - (8g) \\ & n = \sum_{j \in \mathcal{M} \cup \mathcal{N}} y_j \\ & y_j \in \{0, 1\}, \forall j \in \mathcal{M} \cup \mathcal{N}, \end{aligned}$$

where n represents the number of active nodes to be visited. We elaborate more on the shape of the function $\hat{\beta}(n)$ in Section 6.2.

5.4 Customer Best Response Set

We now derive a linear programming representation of each customer and keeper's best response set in order to further improve the solution efficiency of BP. According to Remark 1, each customer has at most $|\mathcal{M}|+|\mathcal{N}|+2$ choices. Therefore, it is possible to enumerate the objective values achieved by all possible choices to confirm that a response is indeed best.

Proposition 2. *Given any $i \in \mathcal{M} \cup \mathcal{N}$, f_p, $\mathfrak{c}$, $\bar{v}$, $\hat{v}_{:i}$, and $\hat{w}$, a candidate solution $(u_i, v_{i:}, w_i, z_i)$ is optimal for model (1) if and only if (1) it satisfies constraints* (1b)-(1h) *and (2) there exists a $\eta_i \in \Re$ such that:*

$$f^d(u_i + z_i) + \left(c^k - \mathfrak{c}\sum_{k\in\mathcal{N}} \hat{v}_{ki}\right) w_i + \sum_{j\in\mathcal{M}\cup\mathcal{N}} (f^p + c^p t_{ij}) v_{ij} \le \eta_i \tag{10a}$$

$$\eta_i \le e_i f^d + (1 - e_i a_i)\left(c^k - \mathfrak{c}\sum_{k\in\mathcal{N}} \hat{v}_{ki}\right) \tag{10b}$$

$$\eta_i \le e_i f^d \tag{10c}$$

$$\eta_i \le \begin{cases} f^p + c^p t_{ij} + M_i(2 - \bar{v}_{ij} - r_{ij}\hat{w}_j), \forall j \in \mathcal{M}\cup\mathcal{N}\setminus i & \text{if } i \in \mathcal{N} \\ 0 & \text{otherwise} \end{cases}, \tag{10d}$$

for some $M_i \ge \min\left\{f^d, c^p \max_{j:r_{ij}=1} t_{ij}\right\} - c^p \min_{j:r_{ij}=1} t_{ij}$ when $i \in \mathcal{N}$.

The proof of Proposition 2 is presented in Appendix B.2.

Using Proposition 2, we can reformulate the BP into an equivalent single-level program (SP_2):

$$\begin{aligned}
(\mathrm{SP}_2) \max_{\substack{f^p,\mathfrak{c},x,y,\bar{v},\\ u,v,w,z,\eta}} \quad & \sum_{i\in\mathcal{N}} \left[f^d(u_i + z_i) + \sum_{j\in\mathcal{M}\cup\mathcal{N}} f^p v_{ij}\right] - \mathfrak{c}\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} - c^d \sum_{i\in\mathcal{M}\cup\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} t_{ij}x_{ij} - c^r \sum_{i\in\mathcal{N}} z_i \\
\text{s.t.} \quad & (1b)-(1h), (3b)-(3d), (4b)-(4g), (10a)-(10d) \\
& \eta_i \in \Re, \forall i \in \mathcal{M}\cup\mathcal{N},
\end{aligned}$$

and its approximation model (AM_2) with the estimated delivery time is:

$$\begin{aligned}
(\mathrm{AM}_2) \max_{\substack{f^p,\mathfrak{c},y,\bar{v},\\ n,u,v,w,z,\eta}} \quad & \sum_{i\in\mathcal{N}} \left[f^d(u_i + z_i) + \sum_{j\in\mathcal{M}\cup\mathcal{N}} f^p v_{ij}\right] - \mathfrak{c}\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} - c^d\hat{\beta}(n)\sqrt{n} - c^r \sum_{i\in\mathcal{N}} z_i \\
\text{s.t.} \quad & (1b)-(1h), (3b)-(3d), (4e), (4f), (10a)-(10d) \\
& n = \sum_{j\in\mathcal{M}\cup\mathcal{N}} y_j \\
& \eta_i \in \Re, y_j \in \{0,1\}, \forall i \in \mathcal{M}\cup\mathcal{N}, \forall j \in \mathcal{M}\cup\mathcal{N}.
\end{aligned}$$

Similarly, $\hat{v}$ ($\hat{w}$) in (10a), (10b), and (10d) can be equivalently replaced with v (w).

6 Numerical Study

We now present the implementation details, the experimental settings, the computational performances, and the results.

6.1 Dataset and Implementation Details

We use a real-world dataset of vehicle routes that were executed by Amazon delivery trucks between July 19, 2018 and August 26, 2018 (Merchan et al. 2021). These routes are located in densely populated urban areas across the United States. The number of customers ranges between 33 and 238 with an average value of 146. The dataset contains information on customer locations including their latitudes, longitudes, zone ids, and the travel time between customers. A sample customer set is shown in Figure 4.

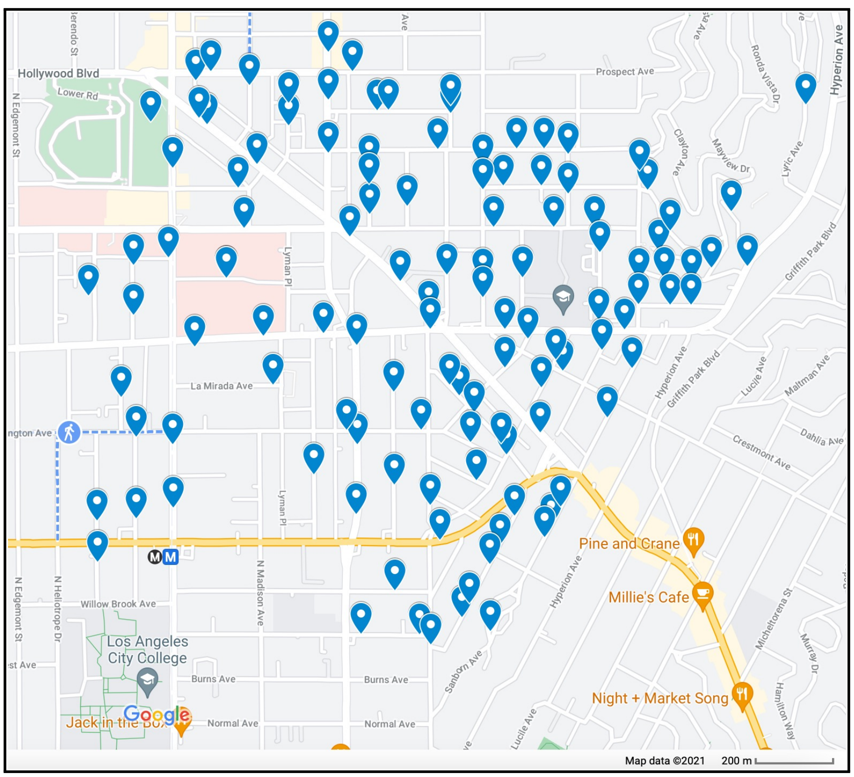


Figure 4: The region where a sample set of customers are assigned to a single vehicle in Los Angeles. The exact coordinates are perturbed by the data providers to anonymize the data. Google Map data ©2021.

We now describe parameter settings in the benchmark instance and in the experimental design. We use 50 randomly selected instances for numerical studies. In each instance, we randomly draw 80% of points as customer locations who have parcels to be delivered and can be potential crowd-keepers.

We change this ratio between 20% and 100% to obtain samples with different number of customers. Customer absence ratio is 5% for the benchmark instance and changes between 0 and 100% in the sensitivity analysis. The remaining 20% of the locations in the benchmark instance are taken as non-customer keepers who do not have parcels to be delivered but can store parcels for their neighbors. For each area, the customer set dynamically changes in different time periods (e.g. days). Thus, each instance will be draw for 20 times to obtain the samples over 20 periods, which are used to evaluate the average performances of crowd-keepers and fixed-storages. The standard delivery fee is set to an integer value and must be high enough to cover the cost of delivering all parcels to customers' doorstep. For example, if 80 dollars must be spent for visiting 100 customer locations by truck, the delivery fee is then 1 dollar. The rescheduling cost is the same value as the delivery fee. The capacity of each keeper is 10 parcels. We take the truck speed to be 4 times the walking speed. Considering the oil prices and the driver wages, the truck delivery cost per minute is set as 1 dollar in the benchmark instance (implying truck travel costs equal to travel time) and changes between 0 and 2 in the sensitivity analysis. We take the customer inconvenience cost per minute of walking as 0.1 dollar and change it between 0 and 2 in the sensitivity analysis. The keeper inconvenience cost is taken as 0.1 in the benchmark instance and changes between 0 and 2 in the sensitivity analysis. The crowd-keepers can only serve those customers located in the same zone and within a limited walk time. Zone id is given in the dataset, and the maximum walk time is set as 4 minutes in the benchmark instance and changes between 0 and 6 in the sensitivity analysis.

To evaluate the performances of different formulations and different systems, we compare the platform profit (i.e., the optimal value of the platform model), the customer costs (i.e., the optimal value of the follower models), the truck delivery time implying pollution (i.e., $h(u, v)$), and the customer average walk time for picking up. We also report the standard delivery fee, and the optimal value of the pickup fee and compensation, to show how pricing decisions adjust to different scenarios. Additionally, the pickup proportion defined as the percentage of all customers who choose the pickup option and the pickup proportion of absent customers are reported, to investigate whether keepers can consolidate deliveries and eliminate failed deliveries.

We implement our algorithms using Python 3.7 on a computer with one 2 GHz Quad-Core Intel Core i5 processor and 16GB of RAM. We use Gurobi 9.0.2 as the solver. The time limit is set as two hours.

6.2 Selection and Calibration of Optimal Travel Time Estimator

According to an overview on TSP continuous approximation by Franceschetti et al. (2017), there are several approximation functions for $\hat{\beta}(n)$, such as a constant, $\hat{\beta}(n) = \beta_1 + \beta_2 \frac{1}{n}$, $\hat{\beta}(n) = \beta_1 \sqrt{n} + \beta_2 \frac{1}{\sqrt{n}}$, or $\hat{\beta}(n) = \beta_1 + \beta_2 \frac{1}{\sqrt{n}}$. In our study, we consider all of them and find the best fit. In each region, we repeatedly draw different samples with different number of nodes n, find the optimal travel time $TSP^*(n)$, and use the pair data $(n, TSP^*(n))$ to estimate the continuous approximation formulation by linear regression. For example, Figure 5 is the estimation model of the optimal travel time under the sample region in Figure 4, where the total number of nodes is 118. For this instance, the function $TSP(n) = 8.58\sqrt{n} - 8.66$ for $n \in [23, 118]$ can accurately estimate the travel time of visiting n nodes with the out-of-sample $R^2 = 0.96$. The case that 118 nodes are visited occurs when all customers (i.e., all nodes) are visited and there is no keeper in the system. The case that 23 nodes are visited rather occurs when only keepers (i.e., 20% nodes) are visited and customers are all assigned to keepers. For every setting with the same number of customers and keepers, 20 samples are randomly drawn as the training dataset for model fitting, and 10 samples are randomly drawn as the testing dataset for obtaining the out-of-sample performance.

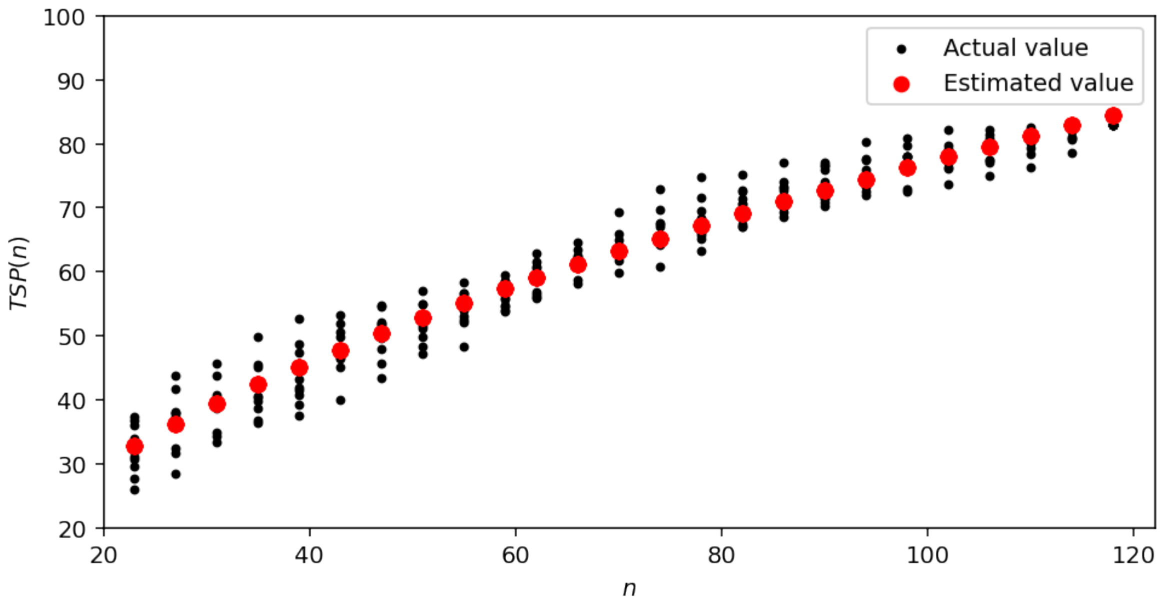


Figure 5: Estimation of the optimal travel time

With this approximation for the delivery time, we do not have to run the exact algorithm for each customer group everyday in this instance region, but instead use the approximation model and obtain the approximated solution with high efficiency and accuracy.

6.3 Effectiveness and Efficiency of Solution Procedures

We compare the efficiency and effectiveness of four different formulations: SP_1, SP_2, AM_1, and AM_2. The *effectiveness* represents the quality of the solutions in terms of realized costs, while the *efficiency* represents the computing time for obtaining the solutions.

Regarding efficiency, Figure 6 shows the runtime of SP_1, SP_2, AM_1, and AM_2 models for instances with different number of customers. The approximated reformulation with the best response set AM_2 yields the best performance with the highest efficiency.

Regarding effectiveness, we compare three different values. The *exact solution* is the output of exact models SP_1 and incidently SP_2, which are solved to optimality using the row generation algorithm. The *approximation* is the output of approximation models AM_1 and incidently AM_2, which are directly solved given that they employ the approximated delivery time. Finally, the *approximated solution* is the realized output by applying the solution of AM_1 (and AM_2) and computing its exact optimal travel time (as measured using $h(u, v)$). Figure 7 presents our results. We find that the crowdkeeping delivery

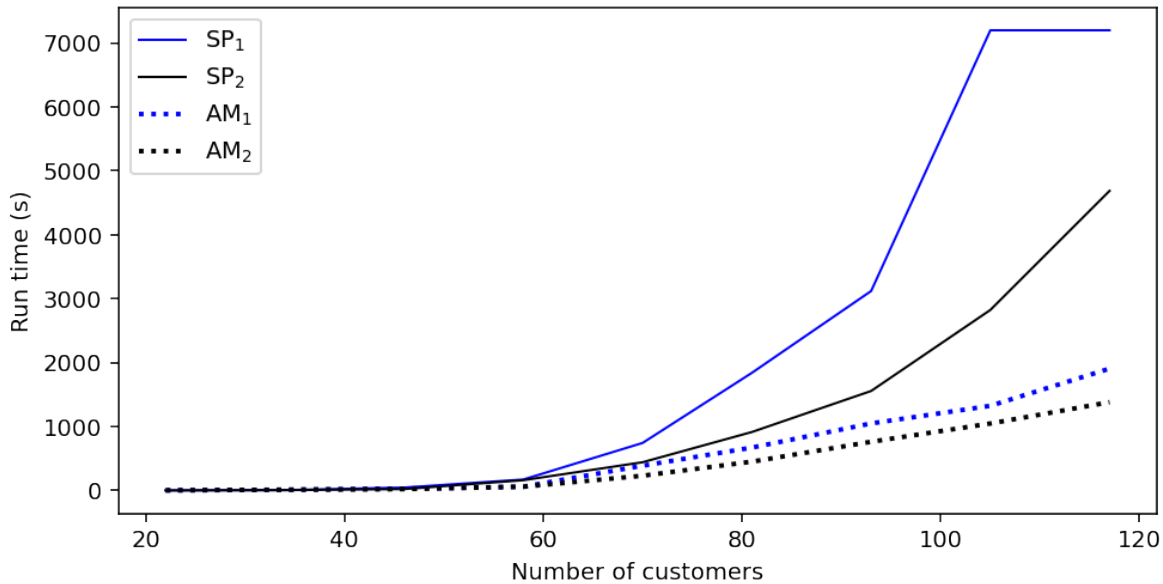


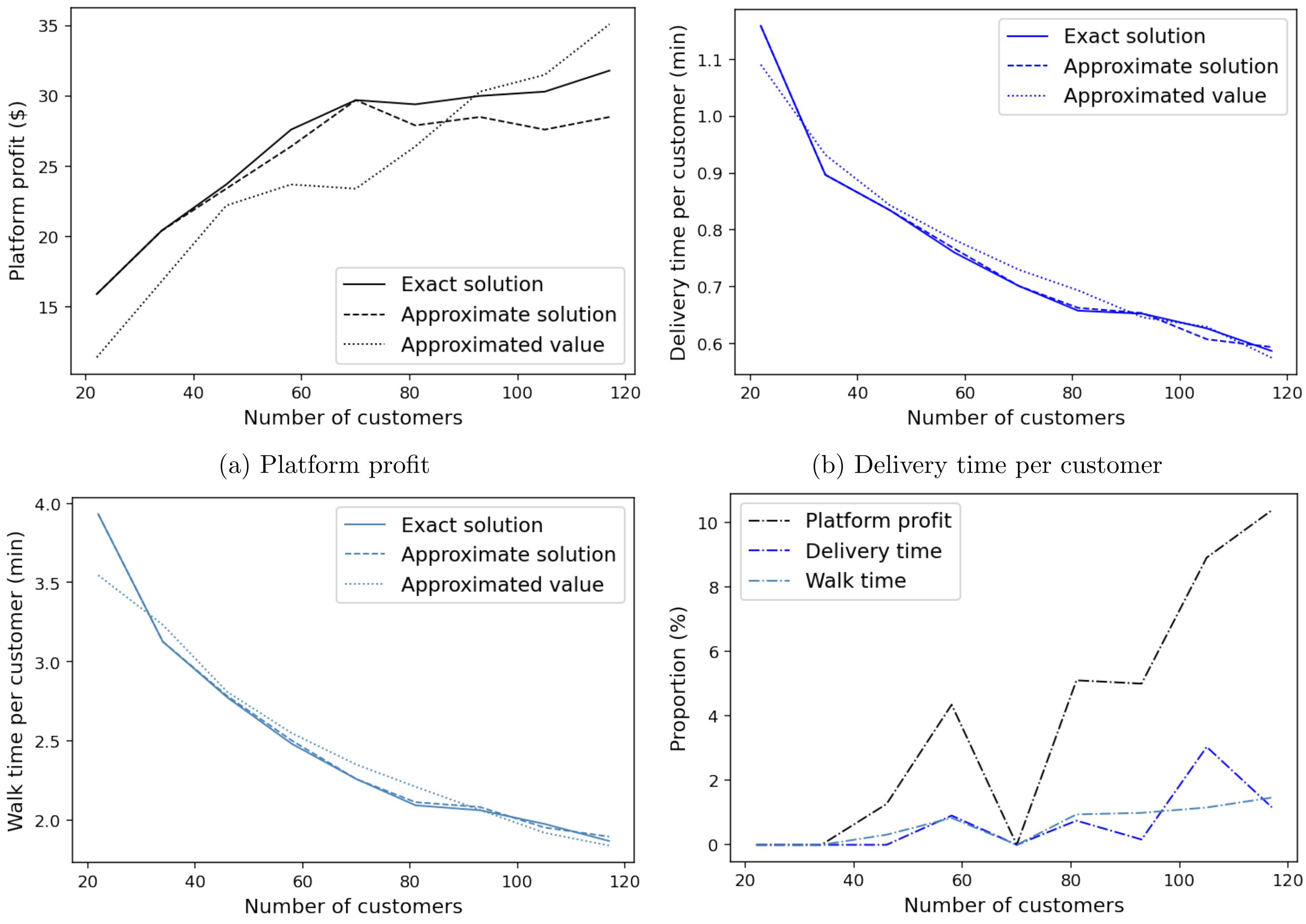
Figure 6: Runtime of four different solution procedures for the benchmark instance

system benefits from economies of scale owing to the observation in Figures 7(a)−(c) that the platform obtains more profits by serving a larger group of customers and that both the marginal walk time and delivery time of serving one more customer decrease. Additionally, Figure 7(d) shows that the gap between the exact and the approximate profits is less than 10%, and Figure 6 shows that the runtime of the approximation model decreases more than 90% compared to the exact model. Therefore, both AM_1 and AM_2 have good performances on efficiency and effectiveness, but that AM_2 is better overall because it offers the same accuracy with a higher efficiency.

6.4 Sensitivity Analysis

We now investigate the factors that may affect the decisions of participants in the delivery system and lead to different results. These sensitivity analyses will help us identify the conditions under which crowd-keeping model is profitable. We compare the performances of three systems: the "crowd-keeper", the "fixed-storage", and the "no-storage" systems. The main difference between the crowd-keeper and fixed-storage systems is that fixed-storage locations are always fixed in different periods, while crowd-keeper selection decisions can adapt to the changing customer sets in different periods. The no-storage system represents that there is no storage in the system, only delivering to doorsteps is allowed, but rescheduling deliveries is possible. Different special cases of our model can solve these systems and the details are presented in Appendix A.

(1) The impact of the service range: The keeper service range is the customer walk range. The longer the maximum walk time that customers can tolerate for picking up, the larger the service range. When the service range is zero, keepers are not able to serve any customers, leading to the case that pickup proportion is zero. Otherwise, if keepers can serve customers, the pickup proportion may be higher. Note



(a) Platform profit

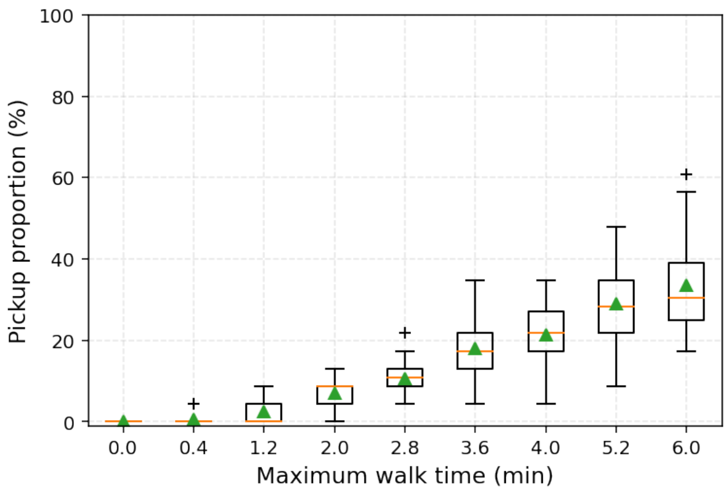
(b) Delivery time per customer

(c) Walk time per customer

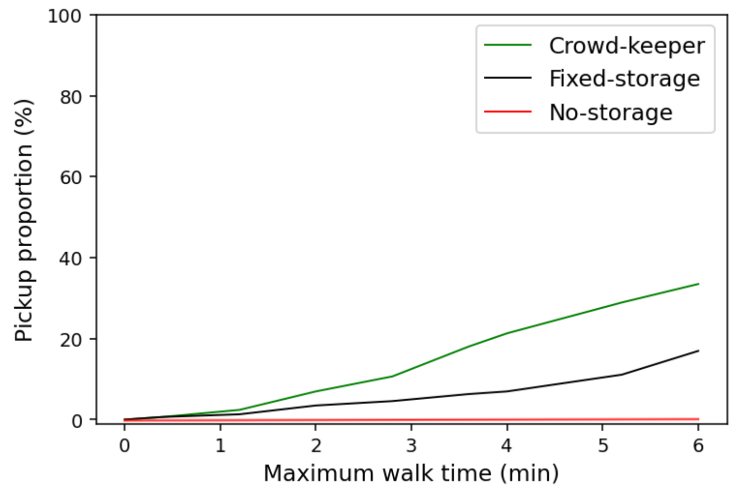
(d) Gap between the exact and approximate solution

Figure 7: Effectiveness of the approximate solution under different number of customers

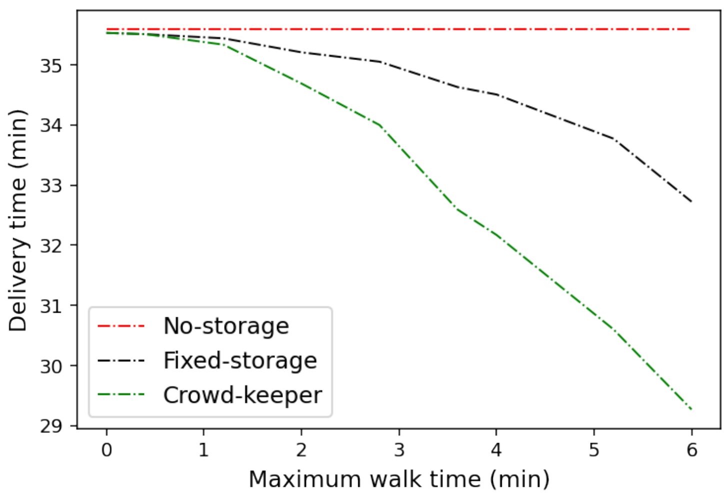
that this ratio is not necessarily 100% since, for some customers, the closest pickup option might already create too much of an inconvenience compared to a delivery. We show results for different sizes of the service range in Figure 8, where (a) and (b) present how the average pickup proportion changes as the maximum walk time changes, where (c), (d), and (e) show how the delivery time, platform profit, and customer costs change, and where (f) represents how those fees change, including the delivery fee, pickup fee, and compensation. We find that when the service range increases, more customers choose the pickup option (see Figure 8(a) and (b)), and the platform earns more profits (see Figure 8(d)). Moreover, the total delivery time for visiting all active nodes decreases (see Figure 8(c)), and this leads to less pollution for the environment (due to less truck utilization). In other words, the delivery system becomes more cost-efficient and environmentally friendly with larger service range. In Figure 8(f), the pickup fee first increases and then decreases, because a larger service range leads to a higher pickup proportion, and the platform can make more profits by increasing the pickup fee. However, when the service range is larger



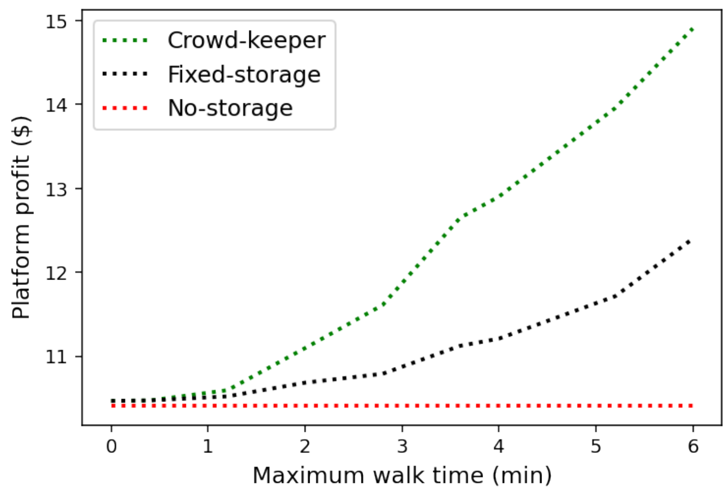
(a) Pickup proportion in crowdkeeping



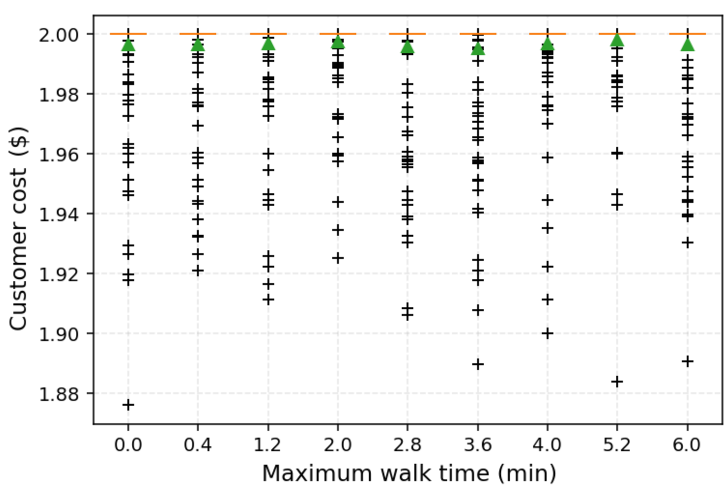
(b) Pickup proportion



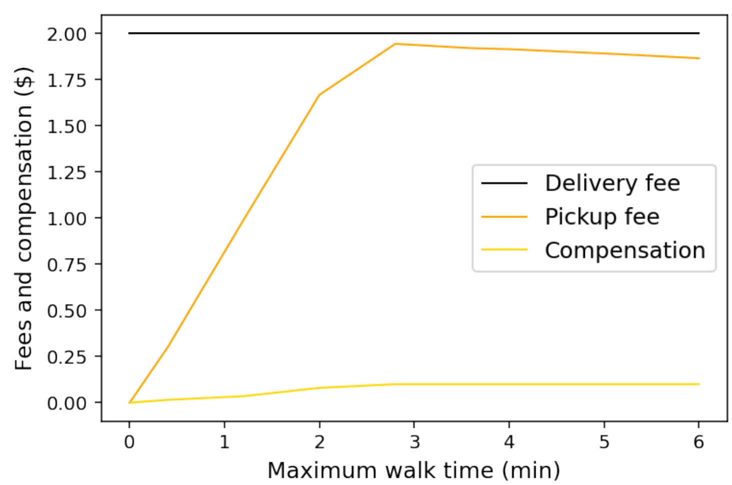
(c) Delivery time



(d) Platform profit



(e) Customer cost



(f) Fees and compensation

Figure 8: The impact of the service range

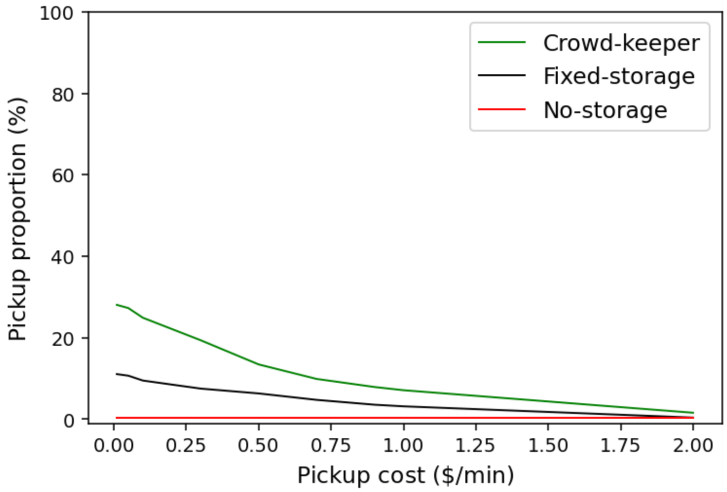
than 3 minutes walk time, more customers may find the pickup option less efficient than the delivery, and the platform has to decrease the pickup fee to make the pickup option more attractive. As shown in Figure 8(e), customer costs are always no higher than the delivery fee, since the direct or rescheduled delivery to doorstep is always an alternative for customers and they have the potential to pay less for receiving parcels by choosing the pickup option and to earn compensation by working as keepers. In addition, the crowd-keeper system always outperforms the no-storage and fixed-storage systems in terms of the delivery time and platform profit (see Figure 8(c) and (d)). Therefore, crowdkeeping is beneficial for the platform, the system, and customers, no matter what the keeper service range is. The larger the service range, the better the crowdkeeping performances.

(2) The impact of the pickup cost: Customers who choose the pickup option need to walk to their appointed keepers, and this creates inconvenience for them. Therefore, in addition to the maximum pickup walk time, the inconvenience cost per minute for picking up (i.e., pickup cost) may also affect customer decisions. We show results in Figure 9, where (a), (b), and (c) represent how the average pickup proportion, delivery time, and platform profit change, respectively, as the pickup cost changes, and where (d) shows how fees and compensation change.

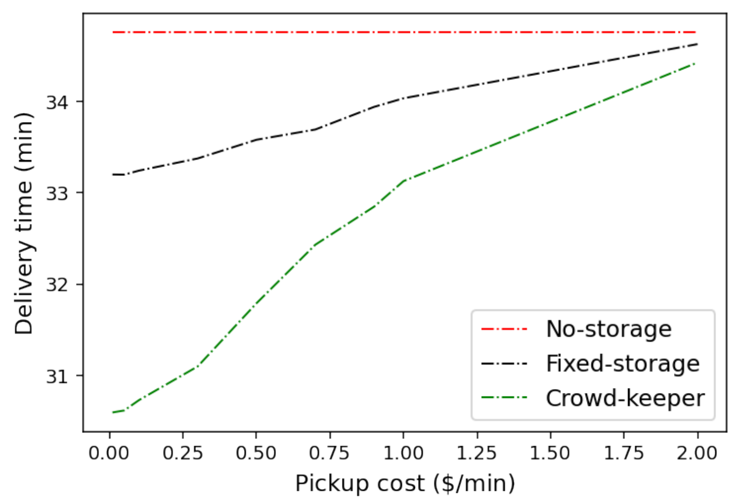
We find that the increasing pickup cost makes the delivery option more attractive for more customers and thus there is a tendency for customers to choose the pickup option less often (see Figure 9(a)). This tendency reduces the efficiency of the system (see Figure 9(b)) and cuts down the benefits of the platform (see Figure 9(c)) because both the platform and the delivery system benefit from the consolidation of deliveries. The lower pickup proportion leads to less consolidation, and performance deterioration. Therefore, to discourage more customers from changing their minds and relinquishing the pickup option, the platform must keep lowering the pickup fee to make up for the increasing pickup cost (see Figure 9(d)). The higher the pickup cost, the larger the gap between the standard delivery fee and the pickup fee, and the less are the platform's benefits. There might exist small fluctuations (e.g. see when pickup cost=1) in this trend because in some instances, some customer's decision to pick up might be insensitive to small changes in the pickup fee, hence encouraging the platform to increase it. Although the system with crowd-keepers continues to perform better than the one with fixed-storages, the increasing pickup cost narrows the gap between their performances. Therefore, to maintain the high efficiency of the delivery system and guarantee a decent profit for the platform, crowd-keepers with more accessible locations should be selected and used to decrease the pickup cost and reduce the inconvenience for customers.

(3) The impact of the delivery cost: Truck delivery costs account for a significant part of the total cost. Thus, the delivery cost per minute of travel time influences the efficiency of the delivery system. Figure 10(a), (b), (c), and (d) present how the average pickup proportion, truck delivery time, walk time per customer, and platform profit change, respectively, as the delivery cost changes.

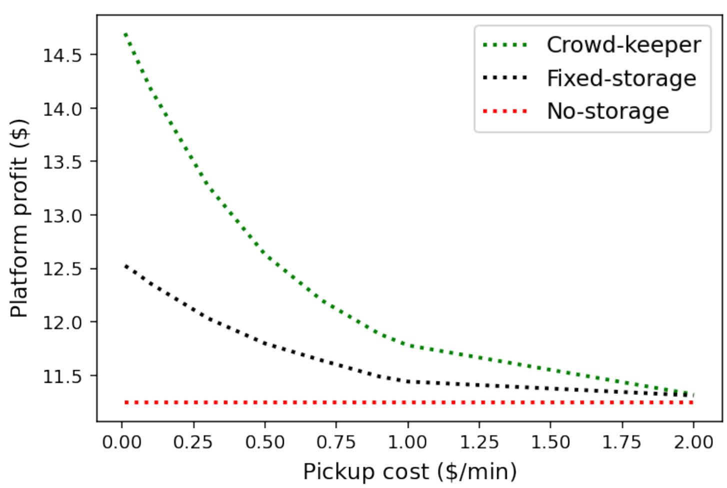
Figure 10(a) shows that the increasing delivery cost increases the pickup proportion both for the crowd-keeper system and for the fixed-storage system, leading to a decrease in truck delivery time (see Figure 10(b)) accompanied with a small increase in customer walk time (see Figure 10(c)). The pickup proportions and the customer walk time stabilize due to the limited keeper service range, and the crowd-keeper system always has a higher pickup proportion than the fixed-storage system due to the higher



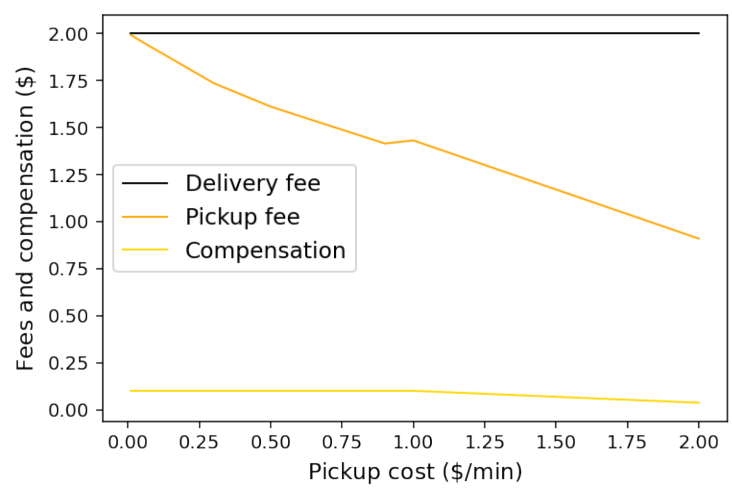
(a) Pickup proportion



(b) Delivery time

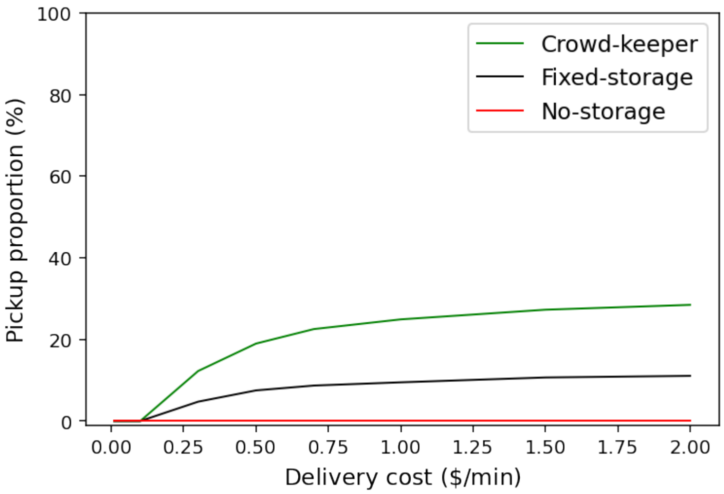


(c) Platform profit

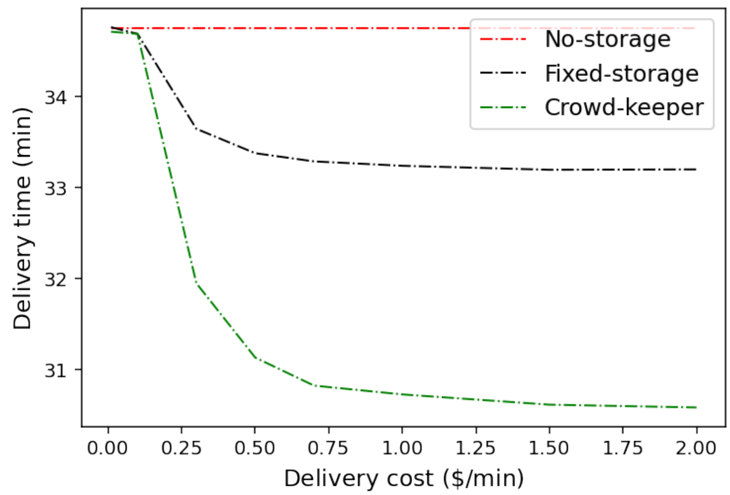


(d) Fees and compensation

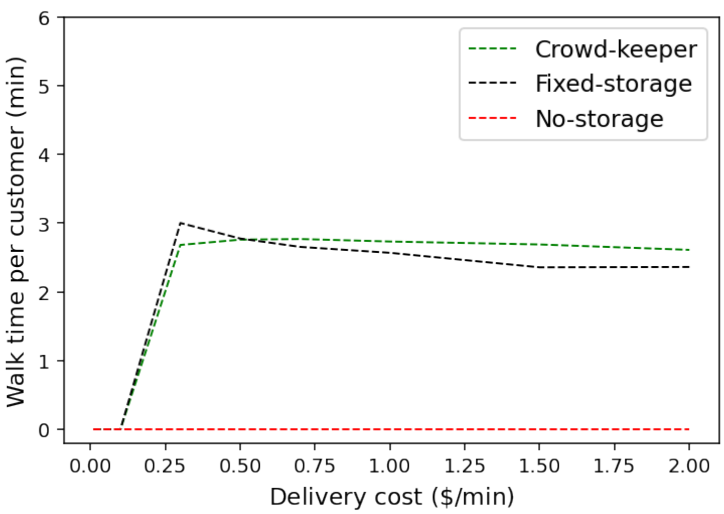
Figure 9: The impact of the pickup cost



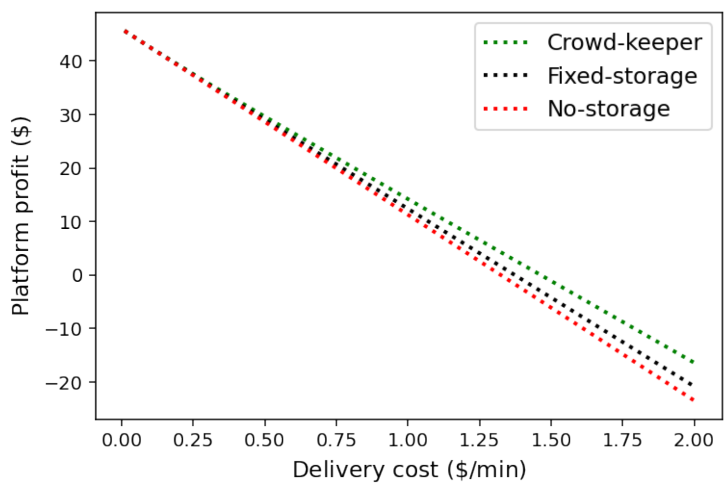
(a) Pickup proportion



(b) Delivery time



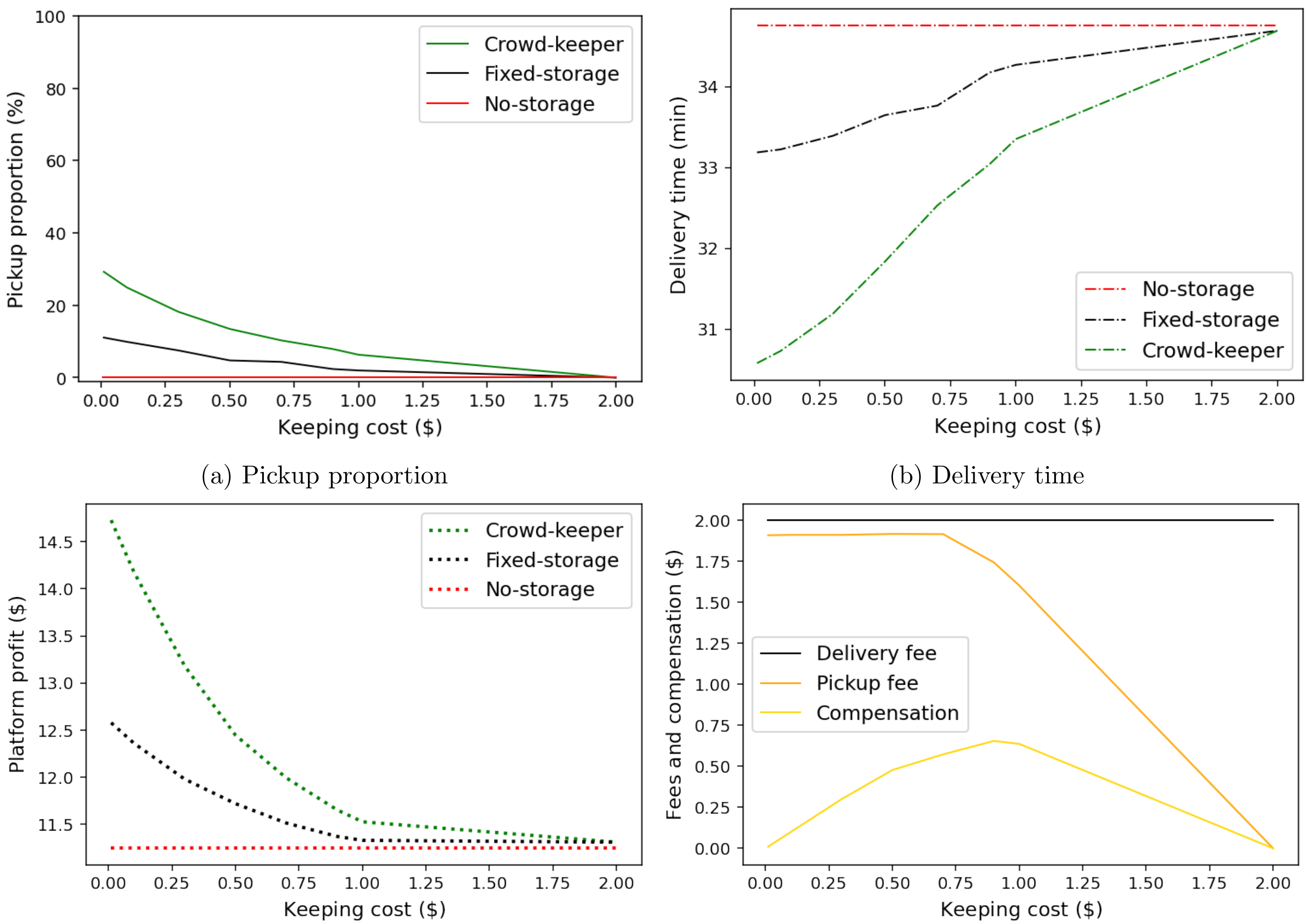
(c) Walk time per customer



(d) Platform profit

Figure 10: The impact of the delivery cost

availability and flexibility. For a fixed delivery fee, the platform will inevitably suffer some losses when the delivery cost becomes larger (see Figure 10(d)), but the crowd-keeper system always yields the best performance in terms of the delivery time and platform profit, compared to the no-storage and fixed-storage systems (see Figure 10(b) and (d)).



(a) Pickup proportion

(b) Delivery time

(c) Platform profit

(d) Fees and compensation

Figure 11: The impact of the keeping cost

(4) The impact of the keeping cost: There exists an inconvenience cost for keepers to keep parcels (i.e., keeping cost), and this inconvenience may be due to staying at home to guarantee their availability and using the smart phone to update the tracking information. This keeping cost is also viewed as the minimum earning for keepers being available. Therefore, keeper availability highly depends on the keeping cost, and keepers will declare their availability only when the total to-be-earned compensation is higher than the keeping cost. We show the impacts of keeping cost change in Figure 11, where (a), (b), and (c) present how the average pickup proportion, delivery time, and platform profit change, respectively, as the keeping cost changes, and where (d) shows how fees and compensation change.

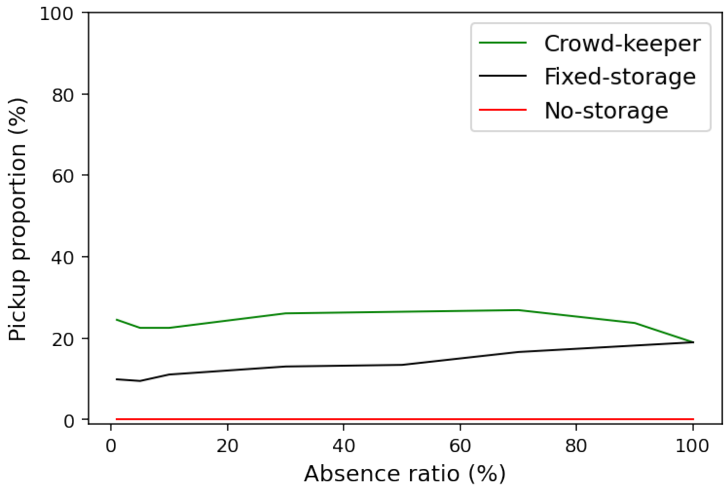
When the keeping cost increases, the pickup proportion decreases (see Figure 11(a)), the delivery time increases (see Figure 11(b)), and the platform profit decreases (see Figure 11(c)). In other words, the increasing keeping cost makes the system less efficient and reduces the platform benefits. When crowd-keepers suffer higher inconvenience, the platform has to increase the compensation offered to them to ensure their availability thus sacrificing part of its profits (see Figure 11(d)). Even in this case, many customers end up switching to the delivery option due to the reduced availability. When the keeping cost increases to a large value (e.g., 2), the pickup proportion, the pickup fee, and the compensation all decrease to zero. In this case, both the crowd-keeper and fixed-storage systems converge to the no-storage system. The crowd-keeper system has a decreased performance when the keeping cost increases, but it still dominates the fixed-storage and no-storage systems.

(5) The impact of the absence ratio: In the no-storage system, deliveries have to be rescheduled when customers are absent, leading to inefficiencies. We investigate if the fixed-storage and the crowd-keeper systems can eliminate this inefficiency in Figure 12, where (a), (b), (c), and (d) present how the pickup proportion of all customers, pickup proportion of absent customers, delivery time, and platform profit change, respectively, as the customer absence ratio changes.

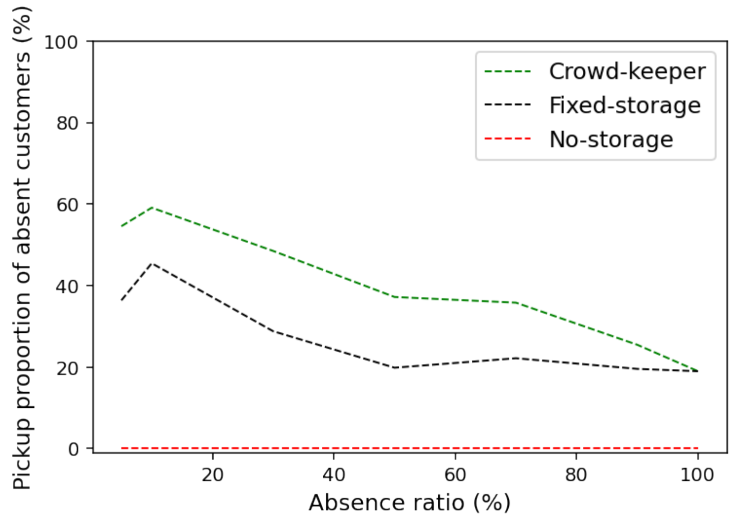
As shown in Figure 12(a), when the customer absence ratio increases, the pickup proportion of fixed-storages overall increases since more absent customers choose the pickup option and the supply is stable. The pickup proportion of crowd-keepers is stable above 20%, but it slightly fluctuates since the supply decreases when more customer keepers become unavailable. When the absence ratio reaches to 100%, the fixed-storage and crowd-keeper systems have the same pickup proportion since the scenarios coincide when all customers are absent and served by non-customer keepers. The crowd-keeper system always have a higher pickup proportion of all customer and that of absent customers (see Figure 12(a) and (b)), which means crowdkeeping has a better performance on consolidating deliveries and eliminating failed deliveries. In Figure 12(c) and (d), we observe that both the delivery time and platform profit decrease as the absence ratio increases. For the no-storage system, the delivery time and the platform profit decrease to zero when the absence ratio is 100%. That is, all absent customers have no choice but reschedule their deliveries. For the crowd-keeper and fixed-storage systems, in addition to the rescheduled delivery, absent customers can choose to be served by keepers or storages, therefore leading to a positive platform profit. The crowd-keeper system always yields a higher platform profit than the no-storage and fixed-storage systems, and has a lower delivery time than the fixed-storage system, implying an overall better performance. The no-storage system has a lowest delivery time when the absence ratio is higher than 52%, but this does not mean that the system causes less pollution, but just represents more rescheduled deliveries and higher future costs.

7 Conclusion

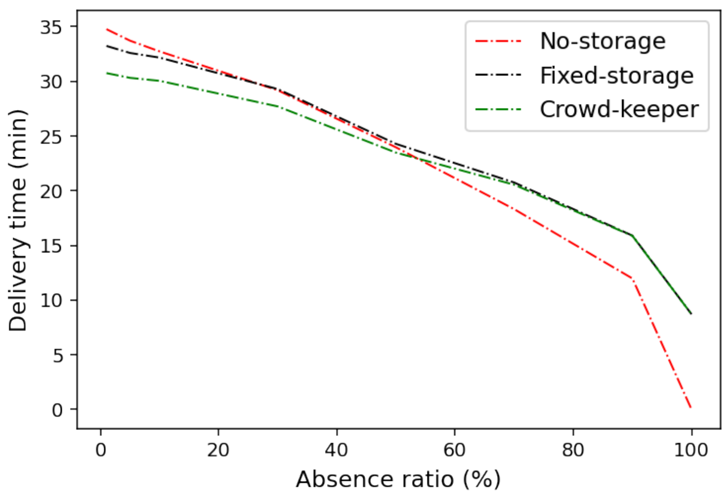
We have presented a new business model in last-mile deliveries. The key idea is to make use of the unused space owned by the crowd. Crowd-keepers have more flexibility, larger availability, and lower costs than what is offered by fixed-storages, and this leads to a more efficient and a more profitable system for



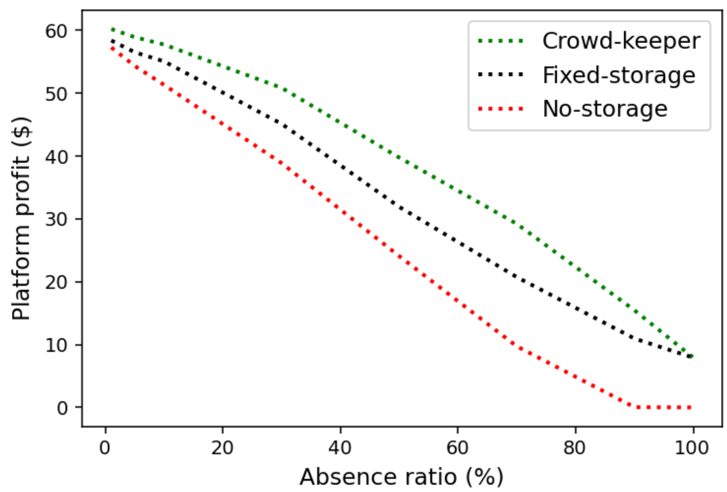
(a) Pickup proportion



(b) Pickup proportion of absent customers



(c) Delivery time



(d) Platform profit

Figure 12: The impact of the customer absence ratio

last-mile deliveries. We have constructed a bilevel program by considering customer preferences, keeper behaviors, and platform operations. We have used the strong duality to reformulate the bilevel program into an equivalent single-level program. This tractable formulation is a quadratic mixed-integer program with subtour elimination constraints and is solved to optimality using a row generation algorithm. To improve the efficiency of the solution procedure, we have developed an approximation model for the bilevel program by approximating the optimal travel time using linear regression, and have derived a more compact representation of the best response set of customers and keepers.

The numerical study is implemented on a real-world dataset provided by Amazon. The results show that the crowdkeeping delivery system benefits from economies of scale since the platform profit increases by serving more customers and the marginal cost of serving one more customer decreases. Additionally, both the platform and the system benefit from delivery consolidations. Specifically, the platform earns more profits and the system causes less pollution under the cases with a larger service range, lower pickup costs, higher delivery costs, lower keeping costs, and higher customer keeper availabilities, and these cases always accompany a higher pickup proportion implying that more deliveries are consolidated. Compared to the no-storage and fixed-storage systems, the crowd-keeper system is beneficial for all participants in the last-mile delivery system by improving the platform profits, reducing environment pollutions due to truck deliveries, and bringing about more savings for customers and extra earnings for keepers. The reason is that crowd-keepers are capable of consolidating deliveries and eliminating failed deliveries, and this capability is higher than fixed-storages due to its more flexibility and larger availability provided by the crowd.

The study can be extended in multiple directions. First, we assumed that the information is completely shared among all participants before optimizing the delivery operations. However, in reality the customer density may not be known when the pickup fee and the compensation decisions are made. Additionally, adapting the model to account for multi-period deliveries that serve dynamically-changing customer groups is a promising direction of future research. In this case, modeling time-windows is an important facet and can potentially better highlight the importance of crowd-keepers.

References

Agatz, N., Campbell, A., Fleischmann, M., and Savelsbergh, M. (2011). Time slot management in attended home delivery. *Transportation Science*, 45(3):435–449.

Alnaggar, A., Gzara, F., and Bookbinder, J. H. (2021). Crowdsourced delivery: A review of platforms and academic literature. *Omega*, 98:102139.

Applegate, D. L., Bixby, R. E., Chvátal, V., and Cook, W. J. (2007). *The Traveling Salesman Problem: A Computational Study*. Princeton University press, Princeton.

Archetti, C., Guerriero, F., and Macrina, G. (2021). The online vehicle routing problem with occasional drivers. *Computers & Operations Research*, 127:105144.

Arslan, A. M., Agatz, N., Kroon, L., and Zuidwijk, R. (2019). Crowdsourced delivery—a dynamic pickup and delivery problem with ad hoc drivers. *Transportation Science*, 53(1):222–235.

Arslan, O. (2021). The location-or-routing problem. *Transportation Research Part B: Methodological*, 147:1–21.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55(4):299–327.

Bruck, B. P., Castegini, F., Cordeau, J.-F., Iori, M., Poncemi, T., and Vezzali, D. (2020). A decision support system for attended home services. *INFORMS Journal on Applied Analytics*, 50(2):137–152.

Bruck, B. P., Cordeau, J.-F., and Iori, M. (2018). A practical time slot management and routing problem for attended home services. *Omega*, 81:208–219.

Campbell, A. M. and Savelsbergh, M. (2006). Incentive schemes for attended home delivery services. *Transportation Science*, 40(3):327–341.

Carbone, V., Rouquet, A., and Roussat, C. (2017). The rise of crowd logistics: a new way to co-create logistics value. *Journal of Business Logistics*, 38(4):238–252.

Cavdar, B. and Sokol, J. (2015). A distribution-free tsp tour length estimation model for random graphs. *European Journal of Operational Research*, 243(2):588–598.

Dayarian, I. and Savelsbergh, M. (2020). Crowdshipping and same-day delivery: Employing in-store customers to deliver online orders. *Production and Operations Management*, 29(9):2153–2174.

Deloison, T., Hannon, E., Huber, A., Heid, B., Klink, C., Sahay, R., and Wolff, C. (2020). The future of the last-mile ecosystem. Technical report, World Economic Forum, Switzerland.

Devari, A., Nikolaev, A. G., and He, Q. (2017). Crowdsourcing the last mile delivery of online orders by exploiting the social networks of retail store customers. *Transportation Research Part E: Logistics and Transportation Review*, 105:105–122.

Franceschetti, A., Jabali, O., and Laporte, G. (2017). Continuous approximation models in freight distribution management. *Top*, 25(3):413–433.

Gendreau, M., Laporte, G., and Semet, F. (1997). The covering tour problem. *Operations Research*, 45(4):568–576.

Hasija, S., Shen, Z.-J. M., and Teo, C.-P. (2020). Smart city operations: Modeling challenges and opportunities. *Manufacturing & Service Operations Management*, 22(1):203–213.

Jacobs, K., Warner, S., Rietra, M., Mazza, L., Buvat, J., Khadikar, A., Cherian, S., and Khemka, Y. (2019). The last-mile delivery challenge. Technical report, Capgemini Research Institute, Paris.

Joerss, M., Schröder, J., Neuhaus, F., Klink, C., and Mann, F. (2016). Parcel delivery: The future of last mile. Technical report, McKinsey & Company, Chicago.

Jozefowiez, N. (2014). A branch-and-price algorithm for the multivehicle covering tour problem. *Networks*, 64(3):160–168.

Kartal, Z., Hasgul, S., and Ernst, A. T. (2017). Single allocation p-hub median location and routing problem with simultaneous pick-up and delivery. *Transportation Research Part E: Logistics and Transportation Review*, 108:141–159.

Klein, R., Mackert, J., Neugebauer, M., and Steinhardt, C. (2018). A model-based approximation of opportunity cost for dynamic pricing in attended home delivery. *OR Spectrum*, 40(4):969–996.

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

Koch, S. and Klein, R. (2020). Route-based approximate dynamic programming for dynamic pricing in attended home delivery. *European Journal of Operational Research*, 287(2):633–652.

Le, T. V., Stathopoulos, A., Van Woensel, T., and Ukkusuri, S. V. (2019). Supply, demand, operations, and management of crowd-shipping services: A review and empirical evidence. *Transportation Research Part C: Emerging Technologies*, 103:83–103.

Lin, Y. H., Wang, Y., He, D., and Lee, L. H. (2020). Last-mile delivery: Optimal locker location under multinomial logit choice model. *Transportation Research Part E: Logistics and Transportation Review*, 142:102059.

Martinez-Sykora, A., McLeod, F., Lamas-Fernandez, C., Bektaş, T., Cherrett, T., and Allen, J. (2020). Optimised solutions to the last-mile delivery problem in london using a combination of walking and driving. *Annals of Operations Research*, 295(2):645–693.

McKinnon, A. C. and Tallam, D. (2003). Unattended delivery to the home: an assessment of the security implications. *International Journal of Retail & Distribution Management*, 31(1):30–41.

Merchan, D., Pachon, J., Arora, J., Konduri, K., Winkenbach, M., Parks, S., and Noszek, J. (2021). Amazon last mile routing research challenge dataset. Accessed January 6, 2022.

Qi, W., Li, L., Liu, S., and Shen, Z.-J. M. (2018). Shared mobility for last-mile delivery: Design, operational prescriptions, and environmental impact. *Manufacturing & Service Operations Management*, 20(4):737–751.

Rohmer, S. and Gendron, B. (2020). A guide to parcel lockers in last mile distribution: Highlighting challenges and opportunities from an or perspective. Technical report, CIRRELT, Montreal.

Savelsbergh, M. and Van Woensel, T. (2016). 50th anniversary invited article—city logistics: Challenges and opportunities. *Transportation Science*, 50(2):579–590.

Schwerdfeger, S. and Boysen, N. (2020). Optimizing the changing locations of mobile parcel lockers in last-mile distribution. *European Journal of Operational Research*, 285(3):1077–1094.

Spliet, R. and Gabor, A. F. (2015). The time window assignment vehicle routing problem. *Transportation Science*, 49(4):721–731.

Toth, P. and Vigo, D. (2014). *Vehicle Routing: Problems, Methods, and Applications.* Society for Industrial and Applied Mathematics, Philadelphia.

Ulmer, M. and Savelsbergh, M. (2020). Workforce scheduling in the era of crowdsourced delivery. *Transportation Science*, 54(4):1113–1133.

Ulmer, M. W. (2020). Dynamic pricing and routing for same-day delivery. *Transportation Science*, 54(4):1016–1033.

Vidal, T., Laporte, G., and Matl, P. (2020). A concise guide to existing and emerging problem variants. *European Journal of Operational Research*, 286(2):401–416.

Yang, X. and Strauss, A. K. (2017). An approximate dynamic programming approach to attended home delivery management. *European Journal of Operational Research*, 263(3):935–945.

Yildiz, B. and Savelsbergh, M. (2020). Pricing for delivery time flexibility. *Transportation Research Part B: Methodological*, 133:230–256.

Appendices

A Delivery systems

We develop models for the no-storage system, the fixed-storage system, and the crowd-delivering-keeper system, and demonstrate their main differences.

A.1 The No-storage System

Before the existing of pickup options, the distribution company has to deliver all the parcels to customers home address. If customers are absent, failed deliveries happen and a second delivery is necessary. Here, we consider the best case that customers can delay and reschedule their deliveries.

Suppose that customers $i \in \mathcal{N}$ should be visited, but some of them $i \in \mathcal{N}'$ are absent. In this case, since there is no keeper or storage in the system, those deliveries of absent customers have to be rescheduled, and the profit will not be captured. Then, the optimization model with the objective function of maximizing the profit for visiting customers in $\mathcal{N} \setminus \mathcal{N}'$ is

$$\max_{x} \quad f^d |\mathcal{N} \setminus \mathcal{N}'| - c^d \sum_{i \in \mathcal{N} \setminus \mathcal{N}'} \sum_{j \in \mathcal{N} \setminus \mathcal{N}'} t_{ij} x_{ij} \tag{11a}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{N} \setminus \mathcal{N}'} x_{ij} = 1, \forall j \in \mathcal{N} \setminus \mathcal{N}' \tag{11b}$$

$$\sum_{i \in \mathcal{N} \setminus N'} x_{ji} = 1, \forall j \in \mathcal{N} \setminus \mathcal{N}' \tag{11c}$$

$$\sum_{i,j \in \mathcal{S}} x_{ij} \leq |\mathcal{S}| - 1, \forall \mathcal{S} \subset \mathcal{N} \setminus \mathcal{N}', 2 \leq |\mathcal{S}| \leq \left|\mathcal{N} \setminus \mathcal{N}'\right| - 2 \tag{11d}$$

$$x_{ij} \in \{0, 1\}, \forall i, j \in \mathcal{N} \setminus \mathcal{N}', \tag{11e}$$

where $|\mathcal{N} \setminus \mathcal{N}'|$ is the number nodes to be visited, and f^d, c^d, t_{ij}, and x_{ij} are presented in Table 1.

A.2 The Fixed-Storage System

The fixed-storage, including the automated locker and pickup location, is safe and efficient, and has been widely used in some countries. If customers are not available, their parcels would be kept in storages, from where customers can pick up at their convenience. However, their locations are fixed, their capacities are limited and the setup cost could be expensive. To emphasize the flexibility of crowd-keepers, we assume that the setup cost of fixed-storages is zero but instead focus on their fixed-location property. That is, the main difference between different storages here we consider is the flexibility of crowd-keepers and the stability of fixed-storages. Therefore, the best case of the fixed-storages is modeled without considering their high setup cost. In this case, customers still have the choice of picking up from fixed-storages by paying a lower pickup fee than the standard delivery fee (e.g., Amazon, IKEA, and Zara). The optimization of the compensation is also necessary since in some cases, fixed-storages are stores who work as storages for compensation and have fixed locations. Therefore, the main difference on modeling between the fixed-storage and crowd-keeper delivery systems is that in the former system, customers cannot work as keepers and therefore are not able to keep parcels for others. To get the performance outcomes of the fixed-storage system, we just need to add the following constraint into BP,

$$w_i = 0, \forall i \in \mathcal{N}.$$

Then, the optimal solutions can be obtained with the same way presented in Section 5.

A.3 The Crowd-Delivering-Keeper System

Crowd-delivering-keepers can keep and also deliver parcels to customers' doorstep. In this case, customers do not have to pick up from their specified keepers, so they do not have to make any decisions, but pay the delivery fee and receive their parcels at their convenience. Crowd-delivering-keepers need to determine if they are willing to work as a keeper (w) and if they are about to accept each request for visiting the customer assigned to them (v). A bilevel program is constructed when both the platform and keepers maximize their profits. The model for the crowd-delivering-keeper $j \in \mathcal{M} \cup \mathcal{N}$ is

$$G_j(c, \bar{v}) \triangleq \max_{v_{:j}, w_j} \quad \left(\sum_{i \in \mathcal{N}} (c - c^p t_{ij}) v_{ij} - c^k \right) w_j \tag{12a}$$

$$\text{s.t.} \quad w_j \leq 1 - a_j, \tag{12b}$$

$$v_{ij} \leq \bar{v}_{ij}, \forall i \in \mathcal{N} \tag{12c}$$

$$v_{ij} \leq w_j, \forall i \in \mathcal{N} \tag{12d}$$

$$v_{ij}, w_j \in \{0, 1\}, \forall i \in \mathcal{N}, \tag{12e}$$

where $v_{:j}$ denotes a column vector. Objective function (12a) states that keepers can accept one request only when the to-be-earned compensation is higher than the inconvenient cost for keeping and delivering services, and keepers are active only when the total profit is positive. The constraint (12b) states that the crowd can be keepers only when they are available. The constraint (12c) specifies that keepers can choose to serve customer i or not that is assigned to them by the platform. The constraint (12d) states that only active keepers can serve customers. The constraints (12e) are domain restrictions.

The platform model is

$$
\begin{aligned}
G_p \triangleq \max_{c,\bar{v},x,y,z} \quad & nf^d - c\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} - c^d \sum_{i\in\mathcal{M}\cup\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} t_{ij}x_{ij} - c^r\sum_{i\in\mathcal{N}} z_i && \text{(13a)}\\
\text{s.t.} \quad & \sum_{i\in\mathcal{N}} \bar{v}_{ij} \le b_j w_j, \forall j \in \mathcal{M}\cup\mathcal{N} && \text{(13b)}\\
& y_j \ge w_j, \forall j \in \mathcal{M}\cup\mathcal{N} && \text{(13c)}\\
& y_i + \sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} + z_i = 1, \forall i \in \mathcal{N} && \text{(13d)}\\
& y_i \le 1 - a_i, \forall i \in \mathcal{N} && \text{(13e)}\\
& z_i \le a_i, \forall i \in \mathcal{N} && \text{(13f)}\\
& \bar{v}_{ij} \le r_{ij}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\cup\mathcal{N} && \text{(13g)}\\
& \sum_{i\in\mathcal{M}\cup\mathcal{N}} x_{ij} = y_j, \forall j \in \mathcal{M}\cup\mathcal{N} && \text{(13h)}\\
& \sum_{i\in\mathcal{M}\cup\mathcal{N}} x_{ji} = y_j, \forall j \in \mathcal{M}\cup\mathcal{N} && \text{(13i)}\\
& \sum_{i,j\in\mathcal{S}} x_{ij} \le |\mathcal{S}| - 1, \forall \mathcal{S} \subset \mathcal{M}\cup\mathcal{N}, 2 \le |\mathcal{S}| \le |\mathcal{M}\cup\mathcal{N}| - 2 && \text{(13j)}\\
& x_{ij} \in \{0,1\}, \forall i \in \mathcal{M}\cup\mathcal{N}, \forall j \in \mathcal{M}\cup\mathcal{N} && \text{(13k)}\\
& \bar{v}_{ij}, y_j, z_i \in \{0,1\}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\cup\mathcal{N} && \text{(13l)}\\
& c \ge 0 && \text{(13m)}
\end{aligned}
$$

The parameters $(f^d, c^d, c^r, t_{ij}, a_i, b_j, r_{ij})$ and decisions $(c, \bar{v}, x, y, z)$ are described in the same way as shown in Table 1. Objective function (13a) shows that the platform revenue is generated from the delivery fee paid by all n customers, and that the cost is due to the compensation, due to visiting all active nodes, and due to rescheduled deliveries. The constraint (13b) is the capacity constraint and (13c) is used to get active keepers. The constraint (13d) ensures that all customers have to be served by the direct delivery, by the rescheduled delivery, or by one keeper. The constraints (13e) and (13f) state that absent customers cannot be served by the direct delivery, and that available customers should not be served by the rescheduled delivery. The constraint (13g) represents that customers can only be assigned to keepers in the same zone and within the acceptable walk time. The constraints (13h)−(13m) are degree constraints, subtour elimination constraints, and domain restrictions for visiting all active customers and keepers.

The bilevel program is

$$
\begin{aligned}
\max_{c,\bar{v},x,y,z,w,v} \quad & nf^d - c\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} - c^d \sum_{i\in\mathcal{M}\cup\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} t_{ij}x_{ij} - c^r\sum_{i\in\mathcal{N}} z_i\\
\text{s.t.} \quad & (13b) - 13m\\
& (v_{:j}, w_j) \in \arg\max G_j(c, \bar{v}), \forall j \in \mathcal{M}\cup\mathcal{N}
\end{aligned}
$$

Then, we reformulate the bilevel program into an equivalent single-level program.

$$\begin{aligned}
\max_{c,\bar{v},x,y,z,w,v} \quad & nf^d - c\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} v_{ij} - c^d \sum_{i\in\mathcal{M}\cup\mathcal{N}}\sum_{j\in\mathcal{M}\cup\mathcal{N}} t_{ij}x_{ij} \\
\text{s.t.} \quad & (13b)-(13m),(12b)-(12e) \\
& c\sum_{i\in\mathcal{N}} v_{ij} \geq \left(c^k + \sum_{i\in\mathcal{N}} c^p t_{ij} v_{ij}\right) w_j, \forall j \in \mathcal{M}\cup\mathcal{N} \\
& c^p t_{ij} v_{ij} \leq c, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\cup\mathcal{N}
\end{aligned}$$

This model is a reduced form of the BP without considering customer preferences and the pricing decision of pickup fee. Similar to the solution procedure in Section 5.2, this model can be solved to optimality. We find that this model yields the same performance as BP with the same delivery time, the same platform profit, and the same system cost.

A.4 Figures of Delivery Systems

To demonstrate the main differences of different delivery systems more explicitly, we show the graph for each system when serving the same customer sample. (a) In the no-storage system, all customers in each period are visited by a closed route. (b) In the fixed-storage system, the consolidation is provided by storages, from where customers pick up their parcels. However, storage locations are always fixed in different periods, such as automated lockers. (c) In the crowd-keeper system, in addition to the consolidation, storage locations can change and adapt to the different customer locations in different periods, leading to a higher flexibility than fixed-storages. (d) In the crowd-delivery-keeper system, instead of letting customers pick up from keepers, keepers make deliveries to customers. The main difference between Figure(c) and (d) is the direction of the dashed arrows. That is, (c) represents that customers pick up parcels from keepers and (d) represents that keepers deliver parcels to customers.

B Proofs

B.1 Proof of Proposition 1

Let H_i^R be the model formed by relaxing the integrality requirements in model (1). Suppose that all optimal solutions to H_i^R have fractional entries.

(1) When $e_i = 1$, let $(u_i, v_{i:}, w_i, z_i)$ be an optimal solution to H_i^R with $0 < u_i < 1$. Due to the constraints $(1e)$ and $(1f)$ and the fact that $u_i > 0$, we have $a_i = 0$ and $z_i = 0$. This implies that $\sum_j v_{ij} = 1 - u_i > 0$ due to the constraint (1b). However, without increasing the objective function value, we can always set $u_i = 1$ and $\sum_j v_{ij} = 0$ when $f_d \leq f_p + c^p \min_{j:\bar{v}_{ij}r_{ij}=1} t_{ij}$, or set $u_i = 0$ and $v_{ik} = 1$ ($k = \arg\min_{j:\bar{v}_{ij}r_{ij}=1} t_{ij}$) when $f_d > f_p + c^p \min_{j:\bar{v}_{ij}r_{ij}=1} t_{ij}$. Similar arguments also apply to $v_{i:}$, w_i, and z_i. Therefore, for any optimal solution with fractional entries, we can always find another optimal solution that has binary entries and yields the same optimal value. This contradicts the hypothesis that

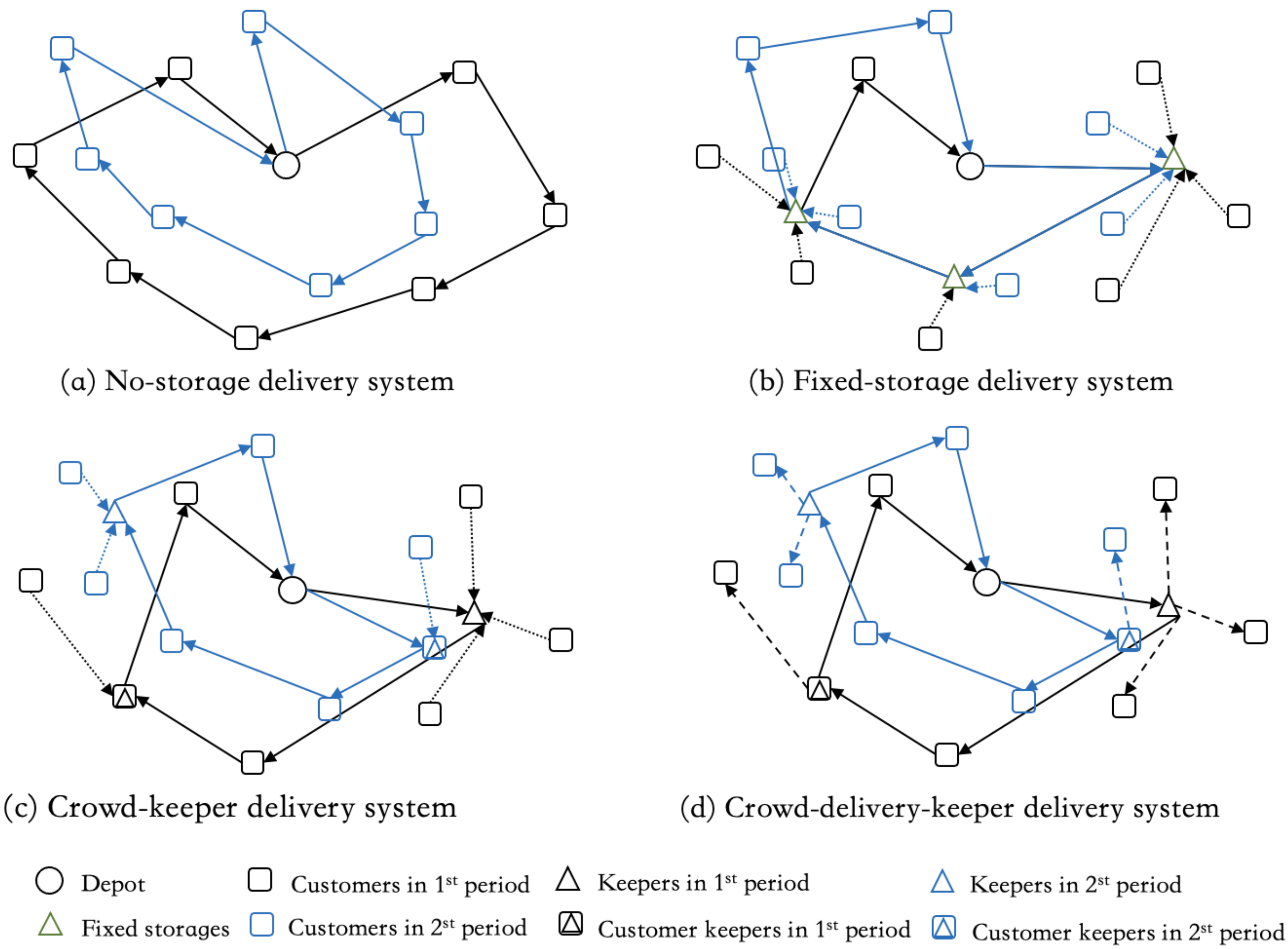


Figure 13: Different systems.

all optimal solutions had fractional entries.

(2) When $e_i = 0$ (i.e. model (2)), let w_i with $0 < w_i < 1$ be an optimal solution. We can set $w_i = 1$ if $c^k - c\sum_{k\in\mathcal{N}} \hat{v}_{ki} \leq 0$, or set $w_i = 0$ if $c^k - c\sum_{k\in\mathcal{N}} \hat{v}_{ki} > 0$, without increasing the objective function value. Again, we observe the same contradiction. $\square$

B.2 Proof of Proposition 2

We need to prove that the two conditions characterize the optimal solution set of model (1). Letting η_i^* be the optimal value of problem (1), the optimal solution set is characterized by:

$$\left\{(u_i, v_{i:}, w_i, z_i) \middle| (1b) - (1h), f^d(u_i + z_i) + \left(c^k - \mathfrak{c}\sum_{k\in\mathcal{N}} \hat{v}_{ki}\right) w_i + \sum_{j\in\mathcal{M}\cup\mathcal{N}} (f^p + c^p t_{ij})v_{ij} \leq \eta_i^* \right\}.$$

We are therefore left the task of showing that condition (2) is satisfied if and only if:

$$f^d(u_i + z_i) + \left(c^k - \mathfrak{c}\sum_{k\in\mathcal{N}} \hat{v}_{ki}\right) w_i + \sum_{j\in\mathcal{M}\cup\mathcal{N}} (f^p + c^p t_{ij})v_{ij} \leq \eta_i^*,$$

which can equivalently be done by showing that :

$$\eta_i^* = \min\left\{f^d + (1 - a_i)\left(c^k - \mathfrak{c}\sum_{k\in\mathcal{N}} \hat{v}_{ki}\right),\ f^d,\ \min_{j\in\mathcal{M}\cup\mathcal{N}\setminus i} f^p + c^p t_{ij} + M_i(2 - \bar{v}_{ij} - r_{ij}\hat{w}_j)\right\},$$

for nodes $i \in \mathcal{N}$ and

$$\eta_i^* = \min\left\{0,\ c^k - \mathfrak{c}\sum_{k\in\mathcal{N}} \hat{v}_{ki}\right\},$$

for $i \in \mathcal{M}$.

We start with the case $i \in \mathcal{N}$ and exploit the fact that there are at most $|\mathcal{M}| + |\mathcal{N}| + 2$ feasible solutions to model (1). Namely, the feasible solutions are:

$$\begin{aligned}\mathcal{X} := &\{(u_i = 1 - a_i, v_{i:} = 0, w_i = 1 - a_i, z_i = a_i),\ (u_i = 1 - a_i, v_{i:} = 0, w_i = 0, z_i = a_i)\}\cup\\ &\{(u_1 = 0, v_{i:} = \mathfrak{e}_j, w_i = 0, z_i = 0)\}_{j:\min(\bar{v}_{ij}, r_{ij}w_j)=1},\end{aligned}$$

where $\mathfrak{e}_j$ is the j-th row of the identity matrix. Hence,

$$\begin{aligned}
\eta^* &= \min_{(u_i, v_{i:}, w_i, z_i) \in \mathcal{X}} f^d(u_i + z_i) + \left(c^k - \mathfrak{c} \sum_{k \in \mathcal{N}} \hat{v}_{ki}\right) w_i + \sum_{j \in \mathcal{M} \cup \mathcal{N}} (f^p + c^p t_{ij}) v_{ij} \\
&= \min \left\{ f^d(1 - a_i + a_i) + \left(c^k - \mathfrak{c} \sum_{k \in \mathcal{N}} \hat{v}_{ki}\right)(1 - a_i),\ f^d,\ \min_{j \in \mathcal{M} \cup \mathcal{N}} f^p + c^p t_{ij} + M_i(2 - \bar{v}_{ij} - r_{ij}\hat{w}_j) \right\} \\
&= \min \left\{ f^d + \left(c^k - \mathfrak{c} \sum_{k \in \mathcal{N}} \hat{v}_{ki}\right)(1 - a_i),\ f^d,\ \min_{j \in \mathcal{M} \cup \mathcal{N} \setminus i} f^p + c^p t_{ij} + M_i(2 - \bar{v}_{ij} - r_{ij}\hat{w}_j) \right\}.
\end{aligned}$$

The second equation holds due to the impossibility of i serving i with $r_{ii} = 0$. M_i is large enough so that when either $\bar{v}_{ij} = 0$ or $r_{ij}\hat{w}_j = 0$, $f^p + c^p t_{ij} + M_i$ is either larger than $f^p + c^p t_{ij'}$ for all $j' \in \mathcal{M} \cup \mathcal{N} \setminus j$ or larger than one of the two other terms in the outer minimum operator: $f^d + \left(c^k - \mathfrak{c} \sum_{k \in \mathcal{N}} \hat{v}_{ki}\right)(1 - a_i)$ and f^d. Such a M_i can take the value of $\min(f^d, \max_{j \in \mathcal{M} \cup \mathcal{N}} c^p t_{ij}) - \min_{j \in \mathcal{M} \cup \mathcal{N}} c^p t_{ij}$ since if $f^d \leq \max_{j \in \mathcal{M} \cup \mathcal{N}} c^p t_{ij}$, then

$$f^p + c^p t_{ij} + M_i = f^p + c^p t_{ij} + f^d - \min_{j \in \mathcal{M} \cup \mathcal{N}} c^p t_{ij} \geq f^d,$$

while if $f^d > \max_{j \in \mathcal{M} \cup \mathcal{N}} c^p t_{ij}$, then

$$f^p + c^p t_{ij} + M_i = f^p + c^p t_{ij} + \max_{j \in \mathcal{M} \cup \mathcal{N}} c^p t_{ij} - \min_{j \in \mathcal{M} \cup \mathcal{N}} c^p t_{ij} \geq f^p + \max_{j \in \mathcal{M} \cup \mathcal{N}} c^p t_{ij} \geq f^p + c^p t_{ij'},\ \forall j' \in \mathcal{M} \cup \mathcal{N} \setminus j\,.$$

In the case that $i \in \mathcal{M}$, the argument is similar and relies on the feasible solution set taking the form:

$$\mathcal{X} := \{(u_i = 0, v_{i:} = 0, w_i = 0, z_i = 0),\ (u_i = 0, v_{i:} = 0, w_i = 1, z_i = 0)\}\,.$$

Hence,

$$\eta^* = \min \left\{ 0,\ c^k - \mathfrak{c} \sum_{k \in \mathcal{N}} \hat{v}_{ki} \right\}. \quad \square$$