

The value of stochastic crowd resources and strategic location of mini-depots for last-mile delivery: A Benders decomposition approach

Santiago Nieto-Isaza^a, Pirmin Fontaine^b, Stefan Minner^a

^a*TUM School of Management, Technical University of Munich, Arcisstraße 21, Munich, 80333, Germany*

^b*Ingolstadt School of Management, Catholic University of Eichstätt-Ingolstadt, Auf der Schanz 49, Ingolstadt, 85049, Germany*

Abstract

Crowd-shipping is an emergent solution to avoid the negative effects caused by the growing demand for last-mile delivery services. Previous research has studied crowd-shipping typically at an operational planning level. However, the study of support infrastructure within a city logistics framework has been neglected, especially from a strategic perspective. We investigate a crowd-sourced last-mile parcel delivery system supported by a network of strategically located mini-depots and present a two-stage stochastic network design problem with stochastic time-dependent arc capacity to fulfill stochastic express deliveries. The first-stage decision is the location of mini-depots used for decoupling flows allowing more flexibility for crowd-demand matchings. The second stage of the problem is the demand allocation of crowd carriers or professional couriers for a finite set of scenarios. We propose an exact Benders decomposition algorithm embedded in a branch-and-cut framework. To enhance the algorithm, we use partial Benders decomposition, warm-start, and non-dominated cuts. We perform computational experiments on networks that were inspired by the public transportation network of Munich. The proposed solution method outperforms an off-the-shelf solver by solving instances in between 3% to 39 % of its run time. The results show the potential to exploit the stochastic crowd flows to deliver packages with deadlines of 3 or 8 hours. The crowd can transport 8.3% to 32.5% of the total demand by using between 4% to 24% of the crowd capacity and we observe average daily savings of 2.1% to 7.6% of the total expected operational cost. The results show values of the stochastic solution of at least 1% and up to 10%.

Keywords:

Crowd-shipping, Network Design, Two-stage Stochastic Programming, Benders Decomposition

1. Introduction

Making cities sustainable is a key challenge. Currently, 55% of the human population lives in urban areas and the expectation is that this number will increase to 68% by the year 2050 (United Nations, 2018). Moreover, e-commerce has been steadily increasing and the statistics show an annual growth rate of 7.3% (Striapunina, 2019). The demand for transporting people and freight in the cities is growing and becoming more complex in terms of customers' expectations, e.g. the desire for speed and price sensitivity (Savelsbergh and Van Woensel, 2016). Crowd-sourced last-mile delivery, often referred to as crowd-shipping or city crowd logistics (CCL), is an emerging strategy for last-mile deliveries considered in connection with *city logistics*, *hyper connected city logistics* and the *physical internet* (Montreuil et al., 2013; Savelsbergh and Van Woensel, 2016). The concept of CCL has been explored in Buldeo Rai et al. (2017) and Sampaio et al. (2019). Buldeo Rai et al. (2017) define crowd logistics as a marketplace that matches supply and demand for logistics services with an undefined and external crowd that has the capacity regarding time and/or space, participates voluntarily, and is compensated accordingly.

Earlier publications discuss potential positive impacts of CCL on city logistics (Rougès and Montreuil, 2014; Crainic and Montreuil, 2016; Savelsbergh and Van Woensel, 2016). Moreover, important logistics providers consider that CCL is one of the concepts of the sharing economy that will have a medium or high impact in less than five years (Chung et al., 2018). All of this attention that academia and practitioners have given to CCL is explained by the social welfare aspect for its potential to reduce the negative effects of the increasing demand for freight transportation in cities, i.e.: congestion, pollution, and greenhouse gas emissions. Another explanation is its potential when it comes to empowering communities, their social networks and, employment generation via digital work (Buldeo Rai et al., 2017; Le et al., 2019). Furthermore, from a business point of view, it can offer more personalized services (Kang et al., 2019) and same-day or even 2-hour deliveries (Savelsbergh and Van Woensel, 2016; Sampaio et al., 2019).

Nonetheless, many challenges remain for a successful implementation of CCL. According to Le et al. (2019), these challenges can be analyzed from three perspectives: *operations and management*, *supply*, and *demand*. From the operations and management perspective, supply-demand matching and routing is an important aspect in a CCL system (Le et al., 2019). Archetti et al. (2016) and Arslan et al. (2019) address matching and routing problems by solving variants of the vehicle routing problem with different features to account for the dynamic nature of the problem and time or capacity constraints. However, most of the business models that are considered support only point-to-point deliveries (Rougès and Montreuil, 2014). Point-to-point deliveries imply that crowd carriers must have origin-destination

pairs close to those of the packages to deliver. This may be difficult to achieve with the crowd due to a spatial-temporal mismatch between (crowd) supply and demand, and due to the stochastic and dynamic nature of the people’s mobility in cities, in particular, if the system intends to use existing transportation flows for a more sustainable system. Kafle et al. (2017) and Mousavi et al. (2019) have studied last-mile delivery systems with two tiers that make use of facilities to perform transshipments on crowd-shipping. Although two-tier systems may alleviate the problem of supply-demand mismatch, these may not be the best option to fulfill dynamic same-day deliveries. Moreover, a guarantee of service is an important success factor in CCL, which is hard to achieve in two-tier systems, especially where the second tier is executed entirely by the crowd.

Currently, there is a significant number of companies that provide on-demand deliveries by using crowd-shipping. Rougès and Montreuil (2014) survey 18 providers of CCL services and develop a taxonomy of different business models. Frehe et al. (2017), based on best practice information from 13 companies, provide a conceptual framework for new companies. The authors summarize these best practices in four steps: individualization of service to create value, long-term investment, the consideration of region-specific regulations, and the generation of a network. Dablanc et al. (2017) survey 96 last-mile delivery providers, more than 30 of which implement crowd-sourced on-demand deliveries. In their analysis, they discuss the potential gains for customers, retailers, and social welfare. Both surveys include strategic planning in city logistics within the success factors for the implementation of new business models.

We address the problem of shipping parcels in a last-mile system with express deliveries, i.e. same-day or short delivery time, by exploiting existing transportation flows of a crowd of potential carriers who travel across urban areas. This supports the use of the crowd potential for last-mile delivery and avoids the creation of new service flows. To overcome the difficulty of having to find perfect origin-destination matchings, we define a more flexible setting where a network of strategically located mini-depots, such as parcel lockers that act as automated service points, allow partial order-crowd matchings with cross-docking, which makes the (hard) constraint of finding similar origin-destination pairs unnecessary. Moreover, as a strategy towards guaranteeing service for the customer (shipper), we consider professional courier carriers to overcome the randomness of crowd availability.

This paper makes the following contributions. First, we investigate a CCL system at a strategic level, which determines an innovative crowd-shipping platform that enables both the concept of a physical internet and the sharing economy. Secondly, we simultaneously consider random demand and supply. This is an important feature, in particular, if the intention is to take advantage of existing transportation

flows in the crowd. We conduct numerical experiments to assess how a CCL system can exploit random crowd capacities with time-dependent profiles (non-stationarity). We take into account the dependency between crowd compensation, crowd supply, and demand in the experimental design. Methodologically, the contribution is twofold. We model the problem as a two-stage stochastic program by using a space-time network representation to use a multi-commodity flow formulation in the second stage. We propose enhanced Benders decomposition using the Branch-and-Benders-cut approach, partial Benders decomposition, a warm-start strategy, and non-dominated cuts.

The remainder of this paper is organized as follows. In Section 2, we present a literature review of crowd-sourced last-mile delivery and related literature on Benders decomposition. In Section 3, we define the network design problem. In Section 4, we present the solution algorithm. We conduct numerical experiments that show the algorithm’s performance and the value of the stochastic solution for different cost structures and (crowd) capacity in Section 5. In Section 6, we provide concluding remarks.

2. Literature review

In the first subsection, we review the related literature on Benders decomposition for two-stage stochastic programming and network design in city logistics. Next, we review the research on crowd-sourced last-mile delivery.

2.1. Benders decomposition for network design problems in city logistics

Network design problems arise in many strategic planning problems in different disciplines (Barnhart et al., 2009). The design of networks with dynamic traffic assignments (Fontaine and Minner, 2017), routing messages in computer networks (Magnanti et al., 1995), strategic transportation problems in city logistics (Crainic, 2000; Crainic et al., 2004; Andersen et al., 2009; Crainic et al., 2018) and facility location problems (Fischetti et al., 2017) are examples. Benders decomposition (Benders, 1962) is often used as an exact solution method for network design problems. Fontaine and Minner (2014) use Benders decomposition to solve large instances of a discrete design applied in traffic networks. The authors model the problem as a bilevel program that can be linearized by taking advantage of the structure of the decision variables in the leader allowing the application of Benders decomposition.

Enhancement techniques for Benders decomposition have been developed to overcome its slow convergence. Rahmaniani et al. (2017) review enhancement techniques that include common practices in

transportation problems like Pareto-optimal cuts (Magnanti and Wong, 1981) and explore complementary techniques, such as warm start strategies. Solution generation methods for the master problem are discussed along with decomposition techniques such as partial Benders decomposition to accelerate the convergence and thus reduce the number of optimality cuts. Fischetti et al. (2017) investigate the performance enhancement of Benders decomposition based on a cut loop stabilization technique at the root node of the relaxed master problem. Contreras et al. (2011) develop an algorithm for the uncapacitated hub location problem. They investigate enhancement strategies with Pareto-optimal cuts and multi-cut generation, taking advantage of the structure of the multi-commodity formulation by decomposing the cuts for each commodity. Fontaine et al. (2021) investigate a scheduled service network design for two-tier city logistics. They develop an exact algorithm where both the master and slave problems are integer programs. The algorithm’s performance is improved by implementing Pareto-optimal cuts, combinatorial cuts, and valid inequalities.

Two-stage stochastic programs have a particular matrix useful to be exploited by the use of Benders decomposition. The application to two-stage stochastic programs or L-shaped method (Van Slyke and Wets, 1969) has been widely used for these problems. Crainic et al. (2021) and Rahmaniani et al. (2018) investigate partial Benders decomposition for two-stage stochastic programs by maintaining information of scenarios derived from the second stage in the master problem. Extensive computational experiments show the advantages of this method for two-stage stochastic programs in network design. We take advantage of a space-time network formulation and adapt several enhancement techniques.

2.2. Crowd-sourced last-mile delivery

Operations planning in CCL often uses variants of the vehicle routing problem. Archetti et al. (2016) address crowd-shipping by solving a variant of the vehicle routing problem with occasional drivers and investigate the option of using crowd-sourced deliveries besides professional drivers. They present an integer programming formulation and a multi-start heuristic. Arslan et al. (2019) consider a pickup and delivery system managed by a crowd-shipping platform with ad-hoc drivers. This work considers a peer-to-peer platform that matches demand and supply. The authors develop an exact solution approach in a rolling horizon scheme. Computational experiments show the benefits of using ad-hoc drivers for cost reduction and environmental sustainability.

Learning-based techniques have been considered to account for the dynamical nature of crowdsourced pickup and delivery problems. Kang et al. (2019) investigate operational planning and control algorithms

for a CCL system. The authors implement an order acceptance algorithm that uses reinforcement learning and artificial neural networks. For scheduling and routing, they use a continuous-variable feedback control algorithm minimizing fuel consumption.

The item-sharing problem arises from peer-to-peer platforms and is related to CCL. Behrend et al. (2019) investigate the assignment and shipping problem where the decisions of the fulfillment of demands and the allocation of capacitated crowd-shippers are integrated. They model the problem as a deterministic detour routing problem and propose an exact solution method based on a set packing formulation. They study the impact of having crowd-shippers with higher capacities and show that having larger capacities has a positive impact in the platform’s profit and that the exact method outperforms procedures that consider only single capacities. Behrend et al. (2021) study a stochastic multi-period variant of the item-sharing and crowd-shipping problem where items can serve multiple requests before they are returned to the original location. They show that including stochastic information improves the performance of such a system and that look-ahead policies are suitable in smaller instances.

The literature reviewed up to this point considers only point-to-point deliveries. However, finding carriers from the crowd may be difficult if the CCL system only supports crowd-order matchings with similar origin-destination pairs. Chen et al. (2017) study the problem of collection of the returned goods in *e-commerce* by using a fleet of taxis that are already scheduled for passenger transportation and allow cross-docking operations. The problem is to assign vehicles to pick up parcels from collection centers and later drop them off at a retailer shop while considering the total detour of the scheduled taxi for that task. Their results show that only 1% of packages can be delivered without cross-docking and around 75% of them are delivered using three transport legs.

Although crowd capacity is stochastic, few research on matching and routing consider this characteristic explicitly. Moreover, investigating the design and utilization of transport infrastructure in crowd-shipping is one of the research directions proposed in the literature (Rougès and Montreuil, 2014; Frehe et al., 2017). Two-tier systems have been studied in city logistics as a strategy to reduce negative impacts of last-mile delivery in large urban areas. Kafle et al. (2017) investigate a deterministic two-tier location routing problem with bid selection. They propose a mixed-integer non-linear program and decompose the problem into two sub-problems. First, they solve a winner determination for the bid selection problem, afterwards, a simultaneous pickup and delivery problem is solved by Tabu Search heuristic. Mousavi et al. (2019) consider uncertainty in crowd capacity and the decision of using support transport infrastructure. They propose a stochastic two-tier crowd-shipping system similar to Kafle et al. (2017)

but deciding the location of mobile depots to allow more flexibility. Their computational experiments show the benefit of the stochastic approach over the deterministic version. These two works have studied the installation of support infrastructure, e.g. intermediate facilities, our work is similar in this regard. However, our work differs from these in the fact that they model a two-tier system where the first tier is executed by a fleet of regular trucks. On one hand, this setting is not suitable to respond to dynamic and stochastic demands which is the characteristic of express deliveries. On the other hand, Kafle et al. (2017) consider a business model where the crowd capacity is not stochastic but is obtained via a bidding system. Moreover, Mousavi et al. (2019) do not use professional couriers as a recourse in the second tier but they penalize unattended demands.

Raviv and Tenzer (2018) study a crowd-shipping system that allows intermediate transfer points based on a network of automatic service points, which is similar to our problem setting. The problem is modeled as a stochastic dynamic program complemented with heuristic policies and the computational study shows the advantage of using the network of service points to achieve short delivery times. However, they only consider a routing policy from an operational planning perspective, and they do not take into account either stochastic demands or the option of professional couriers to give a guarantee of service to the customer. Dayarian and Savelsbergh (2020) is the only work in crowd-shipping that considers both stochastic capacity of the crowd and stochastic arrival of demand, which is common for express deliveries. However, they study a system where the crowd carriers are in-store customers that agree to deliver on-line orders. This type of business model has limitations in the application of crowd-shipping from a broader city logistics perspective. Nevertheless, they do show the benefits of taking into account random crowd capacity and demand simultaneously.

The review of research on crowd-sourced last-mile delivery shows that many of the contributions deal with operational problems e.g. matching and routing, by solving deterministic or dynamic vehicle routing problems with occasional drivers. Only a few consider the stochastic capacity or demand in crowd-shipping and only the work by Dayarian and Savelsbergh (2020) consider these two features at the same time. However, the strategic aspect of a network design to support crowd-shipping platforms needs to be studied for different business models. We intend to fill this literature gap by incorporating strategic decisions in infrastructure and uncertainty in crowd capacity and demand for express delivery services while providing a guarantee of service to the customer with professional couriers.

3. The CCL network design problem

Goods flow across a multi-segment, multi-modal transportation network using *mini-depots* as support infrastructure for decoupling transportation flows. A set of physical nodes \mathcal{N}' represents locations in the city. These nodes attract demand or crowd supply (origin and/or destination for parcels or crowd carriers) and potential locations for *mini-depots*, i.e. we assume that mini-depots can be installed at any node. The nodes may also represent the location of the tramway, bus, or subway stations, etc. \mathcal{A}' is the set of all feasible connections between the physical nodes. $\mathcal{G}' = (\mathcal{N}', \mathcal{A}')$ defines the *physical network* that supports the *CCL* system.

We model a discrete-time planning horizon $\mathcal{T} = \{1, 2, \dots, T_{max}\}$ using a space-time network representation $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ of \mathcal{G}' . \mathcal{N} denotes the set of physical nodes at each time period, i.e. each specific location is replicated $|\mathcal{T}|$ times. \mathcal{A} defines the set of possible space-time connections that are subdivided into two disjunctive subsets \mathcal{H} and \mathcal{A}_s . \mathcal{A}_s denotes the subset of possible connections between any pair (i', j') of physical nodes. Therefore, the arc $(i, j) \in \mathcal{A}_s$ represents a space-time connection for a (physical) node i' starting at time period t_i and ending at (physical) node j' at time period t_j ($t_i < t_j$), i.e. $\tau_{ij} = t_j - t_i$ is the traveling time between the two physical nodes. We assume that travel times are integer. The arcs $(i, j) \in \mathcal{H}$ represent connections of the same physical node in consecutive time periods. These are holding arcs used to model that parcels wait at the same location across periods of time. The sets $N^+(i) = \{j \in \mathcal{N} : (i, j) \in \mathcal{A}\}$ and $N^-(i) = \{j \in \mathcal{N} : (j, i) \in \mathcal{A}\}$ represent successor and predecessor nodes, respectively.

The crowd provides capacity for the flow of commodities, which we model in the arcs $(i, j) \in \mathcal{A}_s$ as a random variable U_{ij} with an expected value λ_{ij} . Figure 1 (a) shows a space-time network representation. The arrows represent the connections between the physical nodes 1 and 3 with a travel time of 1 period. The capacity of each service arc depends on the realization of the random variable U_{ij} , Figure 1 (b) shows the expected value of the random variables U_{ij} for each service arc that connects the nodes 1 and 3 at each period. The demand of the last-mile delivery system must be shipped from an origin to a destination node with a specific release time and a promised delivery time. We define \mathcal{P} as the set of commodities (shipments) representing demand and denote $o_p \in \mathcal{N}$ and $d_p \in \mathcal{N}$ as the origin and destination of commodity $p \in \mathcal{P}$ in the space-time representation. W_p is the random number of parcels of commodity $p \in \mathcal{P}$.

Transportation costs depend on the traveled physical arc (i', j') , the travel period t_i and the mode of transportation, i.e. *professional couriers* or *crowd carriers*. We define unitary (per parcel) transportation

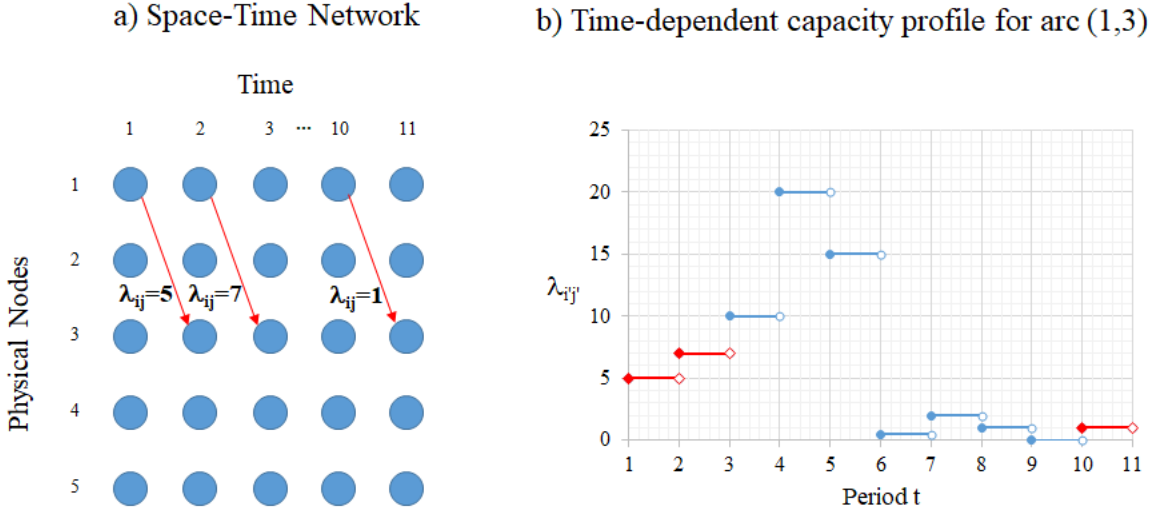


Figure 1: Space-time network representation.

costs for *professional couriers* c_{ij} for each arc $(i, j) \in \mathcal{A}$. We assume that the compensation paid to the crowd is lower than the price paid to the *professional couriers*, therefore, we define a unitary cost fraction $\zeta < 1$. $\phi_{i'}$ are costs for installing a *mini-depot* at candidate location $i' \in \mathcal{N}'$. These costs may include rental costs, purchase or leasing costs, maintenance, and operating costs, e.g. software, and internet connection. The installation costs are computed proportionally to the planning horizon of one operation day. The network design problem determines the number and locations of *mini-depots*, and the flows of goods across the network to minimize the total expected cost. We can summarize the assumptions of the problem as follows. First, we assume that the crowd is cheaper than professional couriers and it provides stochastic capacity. Second, we assume that professional couriers are available on-demand without capacity restrictions. Third, all demand must be fulfilled and arrive at the destination node before the specified deadline.

3.1. Mathematical formulation

The problem is modeled as a two-stage stochastic network design problem for multi-commodity flows with both stochastic demands and arc capacities. The first stage determines the binary decisions $\Omega_{i'}$ that locate the *mini-depots* at candidate locations $i \in \mathcal{N}'$. We model the second stage as a multi-modal, multi-commodity flow problem. Let ξ represent the array containing all the random variables in the

transportation network U_{ij} for $(i, j) \in \mathcal{A}_s$ and W_p for $p \in \mathcal{P}$. Let \mathcal{S} be the sample space defined by ξ and $S \subset \mathcal{S}$ a set of scenarios given by possible realizations of ξ . The second stage decides the allocation of commodity flows to the transportation network to fulfill the demand. These decisions are recourse decisions that depend on a specific realization ξ^s of ξ , which occurs with probability π_s .

u_{ij}^s denotes the realization of the random variable U_{ij} that determines the capacity provided by the crowd at arc $(i, j) \in \mathcal{A}_s$ for the scenario $s \in S$. w^{ps} denotes the realization of the demand random variable W_p . The variables x_{ij}^{ps} and y_{ij}^{ps} represent the fraction of the demand w^{ps} of commodity $p \in \mathcal{P}$ that is transported by *crowd carriers* and *professional couriers* on arc $(i, j) \in \mathcal{A}$ in the space-time network after the realization of scenario $s \in S$. We formulate the Sample Average Approximation (SAA) of the two-stage Stochastic Program as follows:

$$\min \sum_{i' \in \mathcal{N}'} \phi_{i'} \Omega_{i'} + \pi_s \sum_{s \in S} \left[\sum_{(i,j) \in \mathcal{A}} \sum_{p \in \mathcal{P}} \zeta c_{ij} \cdot w^{ps} x_{ij}^{ps} + \sum_{(i,j) \in \mathcal{A}} \sum_{p \in \mathcal{P}} c_{ij} \cdot w^{ps} y_{ij}^{ps} \right] \quad (1)$$

s.t.

$$\sum_{j \in \mathcal{N}^+(i)} (x_{ij}^{ps} + y_{ij}^{ps}) - \sum_{j \in \mathcal{N}^-(i)} (x_{ji}^{ps} + y_{ji}^{ps}) = \begin{cases} 1 & \text{if } i = o_p \\ 0 & \text{if } i \neq o_p, d_p \\ -1 & \text{if } i = d_p \end{cases} \quad \forall s \in S, p \in \mathcal{P}, i \in \mathcal{N} \quad (2)$$

$$\sum_{p \in \mathcal{P}} w^{ps} x_{ij}^{ps} \leq u_{ij}^s \Omega_{j'} \quad \forall s \in S, (i, j) \in \mathcal{A}_s \quad (3)$$

$$x_{ij}^{ps} \leq \Omega_{j'} \quad \forall s \in S, p \in \mathcal{P}, (i, j) \in \mathcal{A}_s \quad (4)$$

$$x_{ij}^{ps} \leq \Omega_{i'} \quad \forall s \in S, p \in \mathcal{P}, (i, j) \in \mathcal{H} : i \neq o^p \quad (5)$$

$$y_{ij}^{ps} \leq \Omega_{i'} \quad \forall s \in S, p \in \mathcal{P}, (i, j) \in \mathcal{H} : i \neq o^p \quad (6)$$

$$x_{ij}^{ps}, y_{ij}^{ps} \geq 0 \quad \forall s \in S, p \in \mathcal{P}, (i, j) \in \mathcal{A} \quad (7)$$

$$\Omega_{i'} \in \{0, 1\} \quad \forall i' \in \mathcal{N}' \quad (8)$$

The objective function (1) minimizes the total expected cost of installation of *mini-depots*, and transportation. (2) are the flow conservation constraints that ensure demand fulfillment for every commodity transported from its origin to its destination. (3) defines the arc capacity constraints of each scenario

$s \in \mathcal{S}$. (4) forbid cross-docking of crowd flows when there is no *mini-depot* installed at that location. (5) and (6) prevent commodities from waiting in a node (using a holding arc) if there is no *mini-depot* installed.

4. Solution method

We use a classic decomposition for two-stage stochastic programs, i.e. we define the first stage of the problem as the relaxed master problem (RMP), using $\Omega_{i'}$, as the complicating variables, and we define the Benders subproblem as the dual of the recourse function.

4.1. Benders subproblems

Let $\bar{\Omega}$ be a vector of the (integer) solution of the master problem. Let $G(\Omega, \xi^s)$ be the multi-commodity flow for the realization ξ^s with objective function $\sum_{(i,j) \in \mathcal{A}} \sum_{p \in \mathcal{P}} \zeta c_{ij} \cdot w^{ps} x_{ij}^{ps} + \sum_{(i,j) \in \mathcal{A}} \sum_{p \in \mathcal{P}} c_{ij} \cdot w^{ps} y_{ij}^{ps}$ and subject to (2) – (6). We define the primal subproblem for a given $\bar{\Omega}$ as: $\text{SP}(\bar{\Omega}) := \mathbb{E} [G(\bar{\Omega}, \xi)]$. Furthermore, since each realization ξ_s is independent, we separate $\text{SP}(\bar{\Omega})$ into $|\mathcal{S}|$ subproblems and solve the multi-commodity flow problem for each scenario $s \in \mathcal{S}$. To this end, we define the primal subproblems for a given scenario as: $\text{SP}_s(\bar{\Omega}) := \mathbb{E} [G(\bar{\Omega}, \xi_s)]$. To derive the Benders cuts, we define the dual subproblems for each scenario $s \in \mathcal{S}$, denoted by $\text{DSP}_s(\bar{\Omega})$. The $\text{DSP}_s(\bar{\Omega})$ reads as follows:

$$\begin{aligned} \text{DSP}_s(\bar{\Omega}) : \quad & \max \sum_{(i) \in \mathcal{N}} \sum_{p \in \mathcal{P}} b_i^p \alpha_i^{ps} + \sum_{(i,j) \in \mathcal{A}_s} u_{ij}^s \bar{\Omega}_{j'} \beta_{ij}^s + \sum_{(i,j) \in \mathcal{A}_s} \sum_{p \in \mathcal{P}} \bar{\Omega}_{j'} \gamma_{ij}^{ps} \\ & + \sum_{(i,j) \in \mathcal{H}} \sum_{p \in \mathcal{P}} \bar{\Omega}_{i'} \varepsilon_{x(ij)}^{ps} + \sum_{(i,j) \in \mathcal{H}} \sum_{p \in \mathcal{P}} \bar{\Omega}_{i'} \varepsilon_{y(ij)}^{ps} \end{aligned} \quad (9)$$

s.t.

$$\alpha_i^{ps} - \alpha_j^{ps} + w_p \beta_{ij}^s + \gamma_{ij}^{ps} + \varepsilon_{x(ij)}^{ps} \geq \zeta c_{ij} \cdot w^p \pi_s \quad \forall (i, j) \in \mathcal{A}, p \in \mathcal{P} \quad (10)$$

$$\alpha_i^{ps} - \alpha_j^{ps} + \varepsilon_{y(ij)}^{ps} \geq c_{ij}^{ps} w^p \pi_s \quad \forall (i, j) \in \mathcal{A}, p \in \mathcal{P} \quad (11)$$

$$\alpha_i^{ps} \in \mathbb{R} \quad \forall \quad i \in \mathcal{N}, p \in \mathcal{P} \quad (12)$$

$$\beta_{ij}^s \leq 0 \quad \forall \quad (i, j) \in \mathcal{A} \quad (13)$$

$$\gamma_{ij}^{ps}, \varepsilon_{x(ij)}^{ps}, \varepsilon_{y(ij)}^{ps} \leq 0 \quad \forall \quad (i, j) \in \mathcal{A}, p \in \mathcal{P} \quad (14)$$

Where $b_i^p = 1$ if $i = o_p$, $b_i^p = -1$ if $i = d_p$ or $b_i^p = 0$, otherwise. Note that the $\text{DSP}_s(\bar{\Omega})$ has five sets of decision variables as the primal problem $\text{SP}_s(\bar{\Omega})$ has five sets of constraints defined similarly by (2) to (6). All of them, however, are applicable for one $s \in S$. α_i^{ps} are the dual variables of the flow conservation constraints (2), β_{ij}^s are the dual variables of the arc capacity constraints (3), γ_{ij}^{ps} are the dual variables of the constraints (4), and $\varepsilon_{x(ij)}^{ps}, \varepsilon_{y(ij)}^{ps}$ are the dual variables of constraints (5) and (6), respectively.

4.2. Benders cuts

From the dual formulation for each $\text{SP}_s(\bar{\Omega})$, we derive the classical Benders cuts. Since the flows determined by the decision variables y_{ij}^{ps} are not constrained by the arcs' capacity, the demand can always be fulfilled even in the worst case (no capacity from the crowd and no *mini-depots*) by *professional carriers* for all the scenarios $s \in S$. For this reason, the feasibility cuts are not necessary in the CCL network design problem. Let $\mathcal{E}(\text{DSP}_s)$ be the set of extreme points of each DSP_s and let $(\bar{\alpha}^s, \bar{\beta}^s, \bar{\gamma}^s, \bar{\varepsilon}_x^s, \bar{\varepsilon}_y^s)$ be one extreme point of DSP_s . We compute optimality cuts for the $\text{DSP}_s(\Omega)$, for each $s \in S$ as follows:

$$\begin{aligned} & \sum_{(i) \in \mathcal{N}} \sum_{p \in \mathcal{P}} b_i^p \bar{\alpha}_i^{ps} + \sum_{(i,j) \in \mathcal{A}_s} u_{ij}^s \Omega_{j'} \bar{\beta}_{ij}^s + \sum_{(i,j) \in \mathcal{A}_s} \sum_{p \in \mathcal{P}} \Omega_{j'} \bar{\gamma}_{ij}^{ps} \\ & + \sum_{(i,j) \in \mathcal{H}} \sum_{p \in \mathcal{P}} \Omega_{i'} \bar{\varepsilon}_{x(ij)}^{ps} + \sum_{(i,j) \in \mathcal{H}} \sum_{p \in \mathcal{P}} \Omega_{i'} \bar{\varepsilon}_{y(ij)}^{ps} \leq z_s \quad \forall (\bar{\alpha}^s, \bar{\beta}^s, \bar{\gamma}^s, \bar{\varepsilon}_x^s, \bar{\varepsilon}_y^s) \in \mathcal{E}(\text{DSP}_s) \end{aligned} \quad (15)$$

4.3. Master problem

The master problem is a relaxation of the complete two-stage stochastic problem defined in Subsection 3.1. The Benders decomposition algorithm takes the RMP, which begins with the first stage of the stochastic program, and iteratively adds cuts that construct the polytope converging in a finite number of steps to the optimal solution (or a solution within the desired optimality gap).

Let $\mathcal{C}_s^{\mathcal{E}} \subset \mathcal{E}(\text{DSP}_s)$ be the current set of optimality cuts and let $c_s^{\mathcal{E}} \in \mathcal{C}_s^{\mathcal{E}}$ be a single optimality cut as defined in (15) and generated by solving $\text{DSP}_s(\bar{\Omega})$. Taking advantage of the block structure of the matrix that defines the second stage of the problem $\text{SP}(\bar{\Omega})$, we can add tighter cuts to the relaxed master problem, thereby giving computational advantages and speeding up the convergence. This is done by solving each $\text{SP}_s(\bar{\Omega})$ on each iteration of the algorithm. We formulate the master problem or relaxed master problem as follows:

$$\min \sum_{i' \in \mathcal{N}'} \phi_{i'} \Omega_{i'} + \sum_{s \in S} z_s \quad (16)$$

s.t.

$$\Omega_{i'} \in \{0; 1\} \quad \forall i' \in \mathcal{N}' \quad (17)$$

$$c_s^{\mathcal{E}} \quad \forall c_s^{\mathcal{E}} \in \mathcal{C}_s^{\mathcal{E}}, s \in S \quad (18)$$

$$z_s \in \mathbb{R} \quad \forall s \in S \quad (19)$$

4.4. Algorithmic enhancements for Benders decomposition

We solve the two-stage stochastic network design problem by using several enhancements previously investigated and referred to in the literature review. We implement a *Branch-and-Benders-cut* algorithm (Crainic et al., 2021). This technique has proven to be more efficient because, instead of solving the complete RMP to optimality (generating a different B&B tree at each iteration) it generates the cuts within a single tree whenever a new integer or a new incumbent solution is found. Moreover, we enhance the algorithm's performance by using a warm-start on the first stage variables, by applying partial Benders decomposition, and improving the classic optimality cuts by computing Pareto-optimal cuts.

The concept of Pareto-optimal or non-dominated cuts (Magnanti and Wong, 1981) is a common strategy for enhancing the performance of the Benders decomposition for transportation problems. Let Ω^0 be a core point in the interior of the RMP space and z_s^* be the objective value of $\text{DSP}_s(\Omega^0)$ that defines the optimality cut $c_s^{\mathcal{E}} \in \mathcal{E}(\text{DSP}_s)$. The Pareto-optimal cut is computed by solving the linear program $\text{DSP}_s(\Omega^0)$ with the additional constraint:

$$\begin{aligned} & \sum_{(i) \in \mathcal{N}} \sum_{p \in \mathcal{P}} b_i^p \alpha_i^{ps} + \sum_{(i,j) \in \mathcal{A}_s} u_{ij}^s \Omega_{j'}^0 \beta_{ij}^s + \sum_{(i,j) \in \mathcal{A}_s} \sum_{p \in \mathcal{P}} \Omega_{j'}^0 \gamma_{ij}^{ps} \\ & + \sum_{(i,j) \in \mathcal{H}} \sum_{p \in \mathcal{P}} \Omega_{i'}^0 \varepsilon_{x(ij)}^{ps} + \sum_{(i,j) \in \mathcal{H}} \sum_{p \in \mathcal{P}} \Omega_{i'}^0 \varepsilon_{y(ij)}^{ps} = z_s^* \end{aligned} \quad (20)$$

In each iteration, we update the core point Ω^0 using: $\Omega_{updated}^0 = \frac{\Omega^0 + \bar{\Omega}}{2}$.

To accelerate the convergence of the algorithm, we keep information from the second stage defined by the SP_s problems. Let s_{ev} be a scenario created by replacing all realizations of the random variables in ξ with their expected values, which occurs with probability $\pi_{s_{ev}}$. We reformulate the RMP as follows:

$$\min \sum_{i' \in \mathcal{N}'} \phi_{i'} \Omega_{i'} + z_{s_{ev}} + \sum_{s \in S} z_s \quad (21)$$

s.t.

$$\pi_{s_{ev}} \left[\sum_{(i,j) \in \mathcal{A}} \sum_{p \in \mathcal{P}} \zeta c_{ij} \cdot w^p x_{ij}^{ps_{ev}} + \sum_{(i,j) \in \mathcal{A}} \sum_{p \in \mathcal{P}} c_{ij} \cdot w^p y_{ij}^{ps_{ev}} \right] \leq z_{s_{ev}} \quad (22)$$

$$\sum_{j \in \mathcal{N}^+(i)} (x_{ij}^{ps_{ev}} + y_{ij}^{ps_{ev}}) - \sum_{j \in \mathcal{N}^-(i)} (x_{ji}^{ps} + y_{ji}^{ps}) = \begin{cases} 1 & \text{if } i = o_p \\ 0 & \text{if } i \neq o_p, d_p \\ -1 & \text{if } i = d_p \end{cases} \quad \forall p \in \mathcal{P}, i \in \mathcal{N} \quad (23)$$

$$\sum_{p \in \mathcal{P}} w^{ps_{ev}} x_{ij}^{ps_{ev}} \leq u_{ij}^{s_{ev}} \Omega_{j'} \quad \forall (i, j) \in \mathcal{A}_s \quad (24)$$

$$x_{ij}^{ps_{ev}} \leq \Omega_{j'} \quad \forall p \in \mathcal{P}, (i, j) \in \mathcal{A}_s \quad (25)$$

$$x_{ij}^{ps_{ev}} \leq \Omega_{i'} \quad \forall p \in \mathcal{P}, (i, j) \in \mathcal{H} : i \neq o^p \quad (26)$$

$$y_{ij}^{ps_{ev}} \leq \Omega_{i'} \quad \forall p \in \mathcal{P}, (i, j) \in \mathcal{H} : i \neq o^p \quad (27)$$

$$x_{ij}^{ps_{ev}}, y_{ij}^{ps_{ev}} \geq 0 \quad \forall p \in \mathcal{P}, (i, j) \in \mathcal{A} \quad (28)$$

$$\Omega_{i'} \in \{0; 1\} \quad \forall i' \in \mathcal{N}' \quad (29)$$

$$c_s^{\mathcal{E}} \quad \forall c_s^{\mathcal{E}} \in \mathcal{C}_s^{\mathcal{E}}, s \in S \quad (30)$$

$$z_s \in \mathbb{R} \quad \forall s \in S \cup \{s_{ev}\} \quad (31)$$

Note that (23) – (27) are the same constraints (2) – (6) but defined only for the scenario s_{ev} , since this is the only realization we include in the RMP.

5. Numerical experiments

We conduct numerical experiments to assess four aspects: first, the capability to exploit the crowd supply, second, the required infrastructure to support the operation of the system, third, the value of the stochastic solution, and fourth, the performance of the proposed algorithm. To conduct the experiments, we determine the topology of the space-time network and the cost structure of the transportation system. We define the demand and supply processes for the CCL system and the sampling procedures to generate the instances. The experiments are designed taking into account the dependency between the compensation of the crowd, supply, and demand. The numerical experiments are solved with Gurobi-Python 9.1.0

on an Intel(R) XEON(R) 2.60GHz with 128 GB RAM using in a single core for a fair comparison with the built-in branch-and-cut algorithm of the commercial solver. All instances are run with a time limit of 8 hours.

The space-time network topology is inspired by the public transportation network of Munich. The size of the network is determined by the number of physical nodes, corresponding to subway stations that are expanded in time across 16 time periods of 30 minutes to represent the network in one operation day of 8 hours, from 11:00 to 19:00. We conduct experiments on different sizes of the network. We use networks of 14 to 24 physical nodes. From these sets, up to four nodes are exclusively origin locations, meaning that no shipments are dropped off at those.

5.1. Demand and Supply

We use *Poisson* distributions with expected values λ_p^d to model the demand process and to model the crowd capacity, *Poisson-Gamma* processes with shape and scale parameters to guarantee expected values λ_{ij} and a coefficient of variation CV. We consider special deliveries, i.e. same-day deliveries that may appear early in the day (until 11:00, morning release time) or in the middle of the day (until 16:00, afternoon release time) with short (3 hours) or medium deadlines (8 hours). We study demand scenarios with total expected demands of $\lambda^d = 1000, 3000, \text{ and } 5000$ shipments. The shipments are mapped by allocating the total demand among origin-destination pairs across the space-time network. Each shipment is mapped to an origin node with a release time and a destination node with a deadline.

Let $p(i', j', r, d) \in \mathcal{P}$ be a shipment with origin node (i'), destination node (j'), release time (r) and deadline (d) and $\mathcal{D} \subset \mathcal{N}'$ be the set of physical nodes that are exclusively destination locations. Now let $\mathcal{P}(i', r, d) = \bigcup_{j' \in \mathcal{D}} p(i', j', r, d)$ be a subset of all shipments departing from (i') with release time (r) and deadline (d). Table 1 shows the partition of the set of shipments in \mathcal{P} for the networks with 10 or 20 destination nodes, with different origin, release time and deadline. In Table 1 ($r1$) is the release time at 11:00, ($r2$) the release time at 16:00, ($d1$) is the deadline at 14:00 and ($d2$) the deadline at 19:00. Let $O \subset \mathcal{N}'$ be the set of physical nodes that are exclusively origin locations. Let $f_{j'}^d$ be the fraction of the demand distributed across the destination nodes, given an origin node. These factors are shown in Table 2. The computation of λ_p^d is done by allocating the total expected demand according the expression: $\lambda_p^d = \lambda^d \cdot \frac{1}{|O|} \cdot 0.5 \cdot f_{j'}^d$ if the release time of the shipment is until 16:00 or $\lambda_p^d = \lambda^d \cdot \frac{1}{|O|} \cdot 0.25 \cdot f_{j'}^d$ if the release time is until 11:00 with deadline at 14:00 or 19:00. This means that half of the demand has release time until 16:00 and half of the demand is released until 11:00. Half of the demand that is released until

11:00 has a 3-hour deadline (and half an 8-hour deadline) and all of the demand that is released until 16:00 has a 3-hour deadline.

Table 1: Mapping of Shipments

Origin	Release Time	Deadline	10 Destinations	20 Destinations
O1	11:00	14:00	$\mathcal{P}(o1, r1, d1)$	$\mathcal{P}(o1, r1, d1)$
		19:00	$\mathcal{P}(o1, r1, d2)$	$\mathcal{P}(o1, r1, d2)$
	16:00	19:00	$\mathcal{P}(o1, r2, d2)$	$\mathcal{P}(o1, r2, d2)$
O2	11:00	14:00	$\mathcal{P}(o2, r1, d1)$	$\mathcal{P}(o2, r1, d1)$
		19:00	$\mathcal{P}(o2, r1, d2)$	$\mathcal{P}(o2, r2, d2)$
	16:00	19:00	$\mathcal{P}(o2, r2, d2)$	$\mathcal{P}(o2, r2, d2)$
O3	11:00	14:00	$\mathcal{P}(o3, r1, d1)$	$\mathcal{P}(o3, r1, d1)$
		19:00	$\mathcal{P}(o3, r1, d2)$	$\mathcal{P}(o3, r1, d2)$
	16:00	19:00	$\mathcal{P}(o3, r2, d2)$	$\mathcal{P}(o3, r2, d2)$
O4	11:00	14:00	$\mathcal{P}(o4, r1, d1)$	$\mathcal{P}(o4, r1, d1)$
		19:00	$\mathcal{P}(o4, r1, d2)$	$\mathcal{P}(o4, r1, d2)$
	16:00	19:00	$\mathcal{P}(o4, r2, d2)$	$\mathcal{P}(o4, r2, d2)$
		$ \mathcal{P} $	120 shipments	240 shipments

We define two main parameters that determine the capacity generation process depending on the total expected capacity.

1. *Crowd profiles* (f_t^c). Are period-dependent factors that change the generation rate depending on the time of the day. The values are shown in Figure 2.
2. *Generation-attraction fractions array* (A^c). An array of size $|\mathcal{N}'| \times 2$ that defines the distribution of the crowd capacity across the network. Let $a_{i'0}^c$ and $a_{j'1}^c$ be elements of the matrix A^c that denote the fraction of the crowd that travel with origin i' and destination j' , where i', j' represent any physical node. Then $b_{i',j'}^c = a_{i'0}^c \cdot a_{j'1}^c$ is the fraction of the crowd that travels across the physical arc (i', j') . The values of these parameters are shown in Table 2.

The expected capacity on space-time arc $(i, j) \in \mathcal{A}_s$ is: $\lambda_{ij} = \frac{\lambda^c}{|\mathcal{T}|} \cdot f_{t_i}^c \cdot b_{i',j'}^c$. Figure 2 (a) depicts the crowd profiles f_t^c and (b) the expected capacity of arc (i, j) where $\lambda^c = 1000$ and $b_{i',j'}^c = 0.012$, therefore $\lambda_{i',j'} = \frac{1000}{16} \cdot 0.012 \cdot f_t^c$.

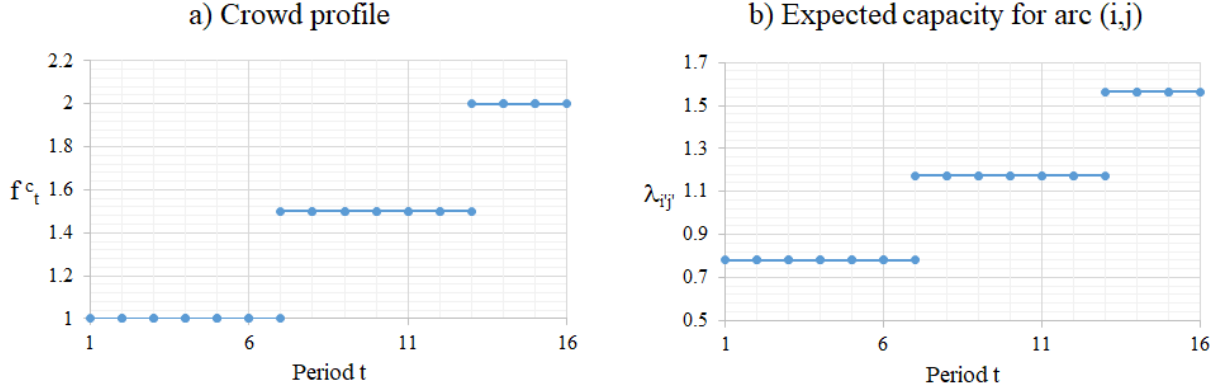


Figure 2: Crowd profiles and expected arc capacity.

We sample the random variables W_p from a Poisson distribution with expected value λ_p^d to compute the parameters w^{ps} . To compute the parameters u_{ij}^s , we sample the random variables U_{ij} . We use a *Negative Binomial* distribution, resulting from the Poisson-Gamma process, to account for more variability (different CV) in the crowd capacity that cannot be modeled with Poisson distributions. We define (*ratio*) as the ratio between the total demand λ^d and the total expected capacity λ^c . Then the total expected capacity i.e. the total expected number of crowd members willing to carry a parcel across the whole network is: $\lambda^c = \text{ratio} \cdot \lambda^d$. Both demand and supply are sampled with the Latin hypercube sampling method. The sample size $|S|$ is set to 50 to guarantee an estimation error of at most 10% with a level of confidence of 0.95 according to the estimates for the upper and lower bounds of the true problem defined in Shapiro (2014).

The size of the set \mathcal{P} can make the model intractable. For this reason, we sample these shipments to reduce the size of the model on larger instances. The number of shipments shown in Table 1 is defined to have a full network representation of the demand. Let ξ_d^s be a realization of the vector $\xi_d \subset \xi$ containing all random variables $W_p : p \in \mathcal{P}$. We sample a subset \mathcal{P}' with a desired size $|\mathcal{P}'|$ by sorting the vector ξ_d^s in ascending order and selecting the shipments corresponding to the $(k - 0.5)/|\mathcal{P}'|$ percentile, with $k = 1, 2, \dots, |\mathcal{P}'|$. By sampling shipments, the total demand is reduced. That reduction affects the capacity-demand ratio, the ratio of installation costs to transportation costs, and thus, the first-stage solution. To reduce this impact, we scale down the sampled crowd capacity on each arc $(i, j) \in \mathcal{A}_s$ by multiplying each sampled value by the fraction (η) of the total sampled demand determined by the

sampled shipments. Moreover, we scale up the objective function by multiplying the recourse function (expected transportation costs) by $1/\eta$.

5.2. General and experimental parameters

The cost parameters c_{ij} are based on the distance traveled through each arc, such that $c_{ij} = c \cdot d_{i'j'}$, where $d_{i'j'}$ is the distance between the physical nodes (i', j') . The distances are computed based on the physical connections determined by the public transportation network. The unitary transportation cost parameter c is set to 1.875 €/Km per transported parcel. This cost parameter is calculated based on typical prices for 1 day (next day) delivery in Germany ¹ (average of 7.5 € per parcel) and an average distance trip of 4 km in the city of Munich.

The installation costs are based on average operation costs for a local company that provides the service and administration of parcel lockers. This costs include leasing costs (50 €), software (100 €), electric power supply (48 €), cleaning (128 €) and rental area cost (520 €) on a monthly basis (Herrmann and Kunze, 2019). The total cost for *mini-depots* is divided by 20 operation days in a month, assuming five days per week. This gives a fixed (daily) installation cost ϕ per installed *mini-depot*.

Given the general setting, computational experiments are run with different levels for the experimental parameters. As mentioned before, we consider the correlation between the experimental parameters, therefore we define their levels and combinations accordingly.

1. *Capacity-demand ratio*. The total expected capacity λ^c is 40%, 70% or 100% of the total expected demand in the case of 10 node instances, or 140%, 200% or 280% of the total expected demand in the case of instances with 20 nodes. i.e. three ratio levels: 0.4, 0.7 and 1.0 or 1.4, 2.0 and 2.8 for instances with 10 nodes and 20 nodes, respectively. The reason why we use higher capacity-demand ratios on larger networks is that we consider that more people from the crowd can be attracted if more destinations (nodes) are offered. At each run, the sampled random capacity is replicated, whereas the sampling procedure for the demand is done with a fixed seed to reduce variance in the results.
2. *Crowd compensation*. Three levels of compensation are defined by the parameter ζ : The crowd compensation is either 30%, 40% or 50% of the price of a professional courier, i.e. $\zeta_1 = 0.3$, $\zeta_2 = 0.4$ or $\zeta_3 = 0.5$

¹Amazon General Delivery Information. Delivery within Germany, <https://www.amazon.de/gp/help/customer/display.html?nodeId=201708680>. Accessed: 25.03.2021

3. *Installation costs for mini-depots.* We consider parcel lockers of different costs ϕ depending on the potential location, capacity-demand ratio, and crowd compensation levels. We define three levels of ϕ , based on the values shown above, with increasing base installation costs. This cost increment can be explained by an increase in the required capacity, rental area cost, or operational costs. We denote three settings for the installation costs ϕ_1 , ϕ_2 and ϕ_3 . These values are summarized in Table 2.

We report the results of 10 instances of each type determined by a combination of the experimental parameters defined by the tuple $(CV, \lambda^d, ratio, \phi, \zeta)$. These tuples are defined in such a way that the dependency between the experimental parameters is represented, i.e. instances with higher compensation of the crowd have a higher expected capacity but up to some point, that is why for instances with $\lambda^d = 5000$, the capacity-demand ratio is not increased for higher compensations. We assess the following key performance indicators (KPI's). Total expected cost: which is the value of the objective function as defined by the SAA. Crowd usage: defined as the average fraction of the crowd capacity that is used across all the scenarios. Crowd utilization: defined as the average fraction of the total demand that is transported by the crowd across any service arc and all the scenarios. Value of the stochastic solution (VSS): defined as the difference between the expected value of the *expected value problem (EVP)* and the *recourse problem (RP)*. Value of the Stochastic Solution in Terms of Transportation Costs (TVSS): defined as the difference in the total expected transportation costs (recourse function) between the EVP and the RP. The purpose of including this measure is to get insights on the benefits of the stochastic solution compared to the deterministic version of the problem in terms of transportation costs. Finally, we consider the cost reduction achieved by using the crowd and the number of installed mini-depots. The experiments are performed on space-time networks of two sizes as shown in Table 1. For the smaller instances (10+4 physical nodes), we use the full set \mathcal{P} of shipments ($|\mathcal{P}| = 120$). For the larger instances (20+4 physical nodes), we use a sample \mathcal{P}' of size $|\mathcal{P}'| = 120$ from the full set \mathcal{P} ($|\mathcal{P}| = 240$).

5.3. Enhanced Benders decomposition performance

We compare the run time performance with the solution provided by Gurobi. We conduct experiments on networks of a smaller size compared to the instances used to assess the performance of the CCL system. The networks have two (physical) origin nodes, 15 or 20 destination nodes, and are solved for 30 shipments with 8-hour or 3-hour deadlines, two levels of capacity-demand ratio, two levels of installation costs, and

Table 2: Supply, demand and installation cost parameters

physical node	$a_{i'0}$	$a_{i'1}$	$f_{i'}^d$	ϕ_1	ϕ_2	ϕ_3
n1	0.07	0.14	0.03	60	90	120
n2	0.03	0.08	0.03	60	60	60
n3	0.03	0.01	0.03	40	40	40
n4	0.01	0.08	0.03	60	60	60
n5	0.02	0.01	0.04	40	40	40
n6	0.07	0.01	0.07	40	40	40
n7	0.05	0.01	0.06	40	40	40
n8	0.03	0.01	0.04	40	40	40
n9	0.02	0.08	0.04	60	60	60
n10	0.04	0.01	0.09	40	40	40
n11	0.05	0.01	0.06	40	40	40
n12	0.03	0.06	0.03	60	60	60
n13	0.05	0.01	0.09	40	40	40
n14	0.03	0.01	0.06	40	40	40
n15	0.02	0.06	0.05	60	60	60
n16	0.02	0.01	0.03	40	40	40
n17	0.02	0.01	0.03	40	40	40
n18	0.07	0.06	0.01	60	60	60
n19	0.09	0.11	0.01	60	90	120
n20	0.07	0.06	0.06	60	60	60
n21	0.04	0.01	0.06	40	40	40
n22	0.04	0.01	0.03	40	40	40
n23	0.07	0.14	0.01	60	90	120
n24	0.03	0.06	0.01	60	60	60

three levels of crowd compensation. We report the average results across 10 replications of each instance type in Table 3.

Table 3: Run time performance

Instance	3-hour deadline			8-hour deadline		
	MIP		BD	MIP		BD
	Run	Gap(%)	Run	Run	Gap(%)	Run
	time		time	time		time
	(sec)		(sec)	(sec)		(sec)
(15N, $\phi_1, \zeta_1, \text{low}$)	20991	0.5	647	3156	0.0	1214
(15N, $\phi_1, \zeta_2, \text{low}$)	17103	0.6	685	2871	0.7	1109
(15N, $\phi_2, \zeta_3, \text{high}$)	10996	0.8	742	2331	0.9	847
(20N, $\phi_1, \zeta_1, \text{low}$)	35210	> 100	1541	5743	0.0	1751
(20N, $\phi_1, \zeta_2, \text{low}$)	31148	> 100	1911	5074	0.7	1814
(20N, $\phi_2, \zeta_3, \text{high}$)	30421	1.4	1607	5623	0.9	1684
Average	24311	0.8	1189	4133	0.5	1403

The results show that BD algorithm outperforms Gurobi in every case. For the case of same-day deliveries, both Gurobi and BD can find optimal solutions within the time limit. However, BD takes between 29.9% to 38.6% of the time taken by Gurobi. In the case of deliveries with time windows, Gurobi takes significantly more time to find optimal solutions and cannot prove optimality in two cases for the larger instances. BD performs better in terms of execution times compared to the same-day delivery instances. The reason for this is because the algorithm converges in between 3 to 5 iterations in every case, but for the instances with time windows, the Pareto-optimal cuts can be computed faster.

5.4. CCL performance

We solve the instances defined in Subsection 5.2 by using the proposed methodology presented in Section 4. All instances were solved to optimality. Table 4 shows the average results on networks with 10 (10N) and 20 (20N) physical nodes.

Table 4: KPIs results on 10 and 20 nodes networks

$(\mathbf{CV}, \lambda^d, \mathbf{ratio}, \phi, \zeta)$	Total Cost		#		Depots		Avg Util%		Avg Usage%		VSS		TVSS		EVPI	
	10N	20N	10N	20N	10N	20N	10N	20N	10N	20N	10N	20N	10N	20N	10N	20N
$(0.4, 1000, 0.4, \phi_1, \zeta_1)$	6.36	6.97	4.0	10.0	10.1	20.2	20.2	16.0	9.8	3.2	12.6	10.1	5.4	3.8		
$(0.7, 1000, 0.4, \phi_1, \zeta_1)$	6.43	7.04	4.0	9.3	8.3	16.9	21.0	14.7	8.6	2.2	11.4	8.4	5.1	4.4		
$(1.0, 1000, 0.4, \phi_1, \zeta_1)$	6.44	7.08	4.0	8.9	10.6	18.9	15.4	10.1	8.4	1.6	11.2	7.4	4.4	5.4		
$(0.4, 1000, 0.7, \phi_2, \zeta_2)$	6.35	6.97	5.0	10.0	14.6	24.9	21.0	14.4	10.0	3.2	14.0	10.4	5.4	3.5		
$(0.7, 1000, 0.7, \phi_2, \zeta_2)$	6.42	7.05	4.0	9.0	11.1	20.2	20.2	13.4	8.9	2.1	12.1	8.5	5.0	4.2		
$(1.0, 1000, 0.7, \phi_2, \zeta_2)$	6.44	7.10	4.2	8.1	12.5	18.9	17.4	9.6	8.4	1.4	11.8	7.0	4.1	5.3		
$(0.4, 1000, 1.0, \phi_2, \zeta_3)$	6.31	6.98	5.0	9.4	18.9	27.6	19.6	11.0	8.8	2.4	11.6	8.7	5.5	3.4		
$(0.7, 1000, 1.0, \phi_2, \zeta_3)$	6.38	7.05	5.0	9.0	16.2	23.8	20.6	11.4	9.4	2.1	13.3	8.4	5.1	4.0		
$(1.0, 1000, 1.0, \phi_2, \zeta_3)$	6.43	7.10	5.0	5.3	17.3	11.2	18.0	3.7	8.7	1.3	12.6	4.6	4.2	4.8		
$(0.4, 3000, 0.4, \phi_2, \zeta_2)$	19.74	19.69	7.0	19.0	11.8	30.7	23.0	22.2	1.8	2.5	2.7	5.6	0.8	1.6		
$(0.7, 3000, 0.4, \phi_2, \zeta_2)$	19.92	20.07	6.0	16.8	9.8	24.3	22.4	22.1	4.8	5.3	6.1	9.8	0.7	2.0		
$(1.0, 3000, 0.4, \phi_2, \zeta_2)$	19.97	20.17	6.7	16.6	11.9	26.2	20.7	17.8	5.1	4.8	6.7	9.2	0.2	2.8		
$(0.4, 3000, 0.7, \phi_3, \zeta_3)$	19.62	19.81	8.0	17.0	17.0	32.5	24.0	16.9	2.1	2.0	3.3	4.5	0.8	1.6		
$(0.7, 3000, 0.7, \phi_3, \zeta_3)$	19.81	20.13	7.0	15.7	14.2	26.9	23.3	17.4	3.3	5.1	4.7	9.2	0.7	1.9		
$(1.0, 3000, 0.7, \phi_3, \zeta_3)$	19.91	20.27	7.1	15.2	15.9	27.8	20.1	14.1	5.3	4.3	7.3	8.3	0.1	2.7		
$(0.4, 5000, 0.4, \phi_3, \zeta_2)$	32.74	31.37	10.0	20.0	12.7	32.2	24.7	23.4	1.9	1.6	2.8	3.0	1.1	1.4		
$(0.7, 5000, 0.4, \phi_3, \zeta_2)$	33.05	31.96	9.0	20.0	11.0	27.8	24.3	24.3	2.2	4.8	3.1	7.3	1.0	1.3		
$(1.0, 5000, 0.4, \phi_3, \zeta_2)$	33.12	32.17	10.0	20.0	13.5	29.3	22.5	19.8	6.1	7.1	7.7	10.3	0.7	0.7		
$(0.4, 5000, 0.4, \phi_3, \zeta_3)$	33.16	32.51	8.3	18.6	12.1	28.2	23.7	20.0	1.6	1.1	2.2	2.9	1.1	1.3		
$(0.7, 5000, 0.4, \phi_3, \zeta_3)$	33.41	32.92	7.9	17.0	10.5	22.8	23.6	19.6	1.9	4.2	2.6	6.8	1.0	1.2		
$(1.0, 5000, 0.4, \phi_3, \zeta_3)$	33.50	33.07	8.2	17.4	12.5	24.8	21.1	16.5	5.0	4.2	6.3	7.0	0.7	0.6		

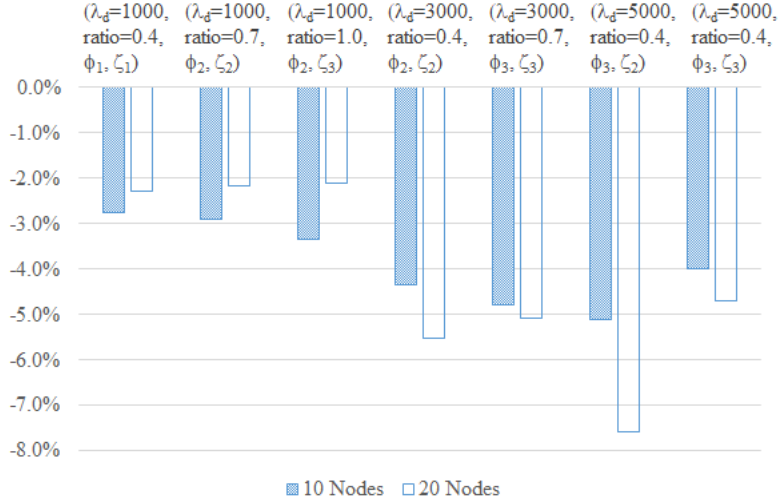


Figure 3: Total average cost reduction with the crowd.

The results show that when the expected demand is higher, the CCL network has a higher number of installed mini-depots. The number of selected locations to install mini-depots is between 40-80 % of all potential locations. However, we observe that an increment in the compensation paid to the crowd slightly affects the required infrastructure (number of mini-depots). The utilization of the crowd is between 8.3% and 18.9% in the case of the instances with 10 nodes and between 11.2% and 32.5% for instances with 20 nodes. The usage of the crowd is between 15.4% and 24.7% for the instances with 10 nodes. In the case of the networks with 20 nodes, the crowd utilization is significantly higher, however, we observe that even though the expected capacity is higher, the utilization of the crowd does not increase drastically compared to the instances with 10 nodes. On the other hand, the average usage of the crowd is lower and more variable. We observe values of the average crowd usage between 3.7% and 24.3%, which is expected given that the capacity is higher and the utilization increases slightly. As shown in Figure 3, the cost savings when using the crowd are 2.8% - 5.1% on 10 node instances and 2.1% - 7.6% on 20 node instances. These values correspond to daily savings between 151 € and 220 €, for instances with $\lambda^d = 1000$, between 901 € and 1168 €, for instances with $\lambda^d = 3000$, and between 1387 € and 2613 €, for instances with $\lambda^d = 5000$.

By analyzing the transportation flows of the demand allocated to the crowd, we can deduce the value of the mini-depots as a support infrastructure to use the stochastic crowd capacity. Figure 4 shows partial shipments, i.e. fractions of a shipment (commodity) $p \in \mathcal{P}$, that are allocated to the crowd or



Figure 4: Transportation flows allocated to the crowd.

professional couriers. These partial shipments are classified into four categories depending on their origin and destination (physical nodes). These four categories are direct shipments: same origin and destination as the commodity; first-leg: same origin with a mini-depot as the destination; mid-leg: both origin and destination are mini-depots; and last-leg: the origin is a mini-depot and the destination is the destination of the commodity. The results reported in the figure are averages across scenarios and commodities.

Figure 4 (a), (b), and (c) describes the average number of shipments, and demand allocation on instances with 10 (destination) nodes. Figure 4 (d), (e), and (f) shows the same description but on instances with 20 (destination) nodes. Figure 4 (a) shows that direct crowd shipments occur more often compared to shipments that use mini-depots as cross-docking platforms (first, mid or, last-leg). Direct shipments are between 0.4 to 1.4 on average, whereas first and last-leg shipments are between 0.1 to 0.2 on average. On the other hand, we observe that mid-leg shipments are close to zero on average, which shows that transshipments between mini-depots occur only occasionally. There is also an increasing trend in the number of direct shipments with the total demand. The latter can be explained because as the demand increases, the expected crowd capacity per arc is higher, thus, direct shipments may occur more often. Figure 4 (b) shows that professional couriers perform direct shipments almost exclusively, this means that cross-docking operations are always completed (and started) by the crowd. Figure 4 (c) shows a comparison between the volume allocated to the crowd that is transported with direct shipments and cross-docking. Only the first-leg cross-docking shipments are taken into account here. We observe that between 11% to 20% of the demand volume allocated to the crowd is transported using mini-depots as cross-docking platforms.

The results reported in Figure 4 (d), (e), and (f) show that the cross-docking operations are used more on the instances with 20 nodes. In general, we can see that both direct and cross-docking shipments are used with a similar frequency. Also, cross-docking operations are not performed by professional couriers as almost all of these shipments are direct. Moreover, we observe that between 25% to 46% of the demand volume allocated to the crowd is transported using mini-depots as cross-docking platforms. For these reasons, we can say that the mini-depots are required not only to store parcels temporally but to perform cross-docking operations that allow using the existing transportation flows from the crowd.

The Value of the Stochastic Solution (VSS) between 1.1% and 10% for networks with 10 nodes and between 1.1% and 7.1% for 20 nodes networks. However, the relative VSS decreases on instances with a larger expected demand on the 10 node instances and we do not observe this trend on the 20 node instances. The latter is because the total expected costs are higher and the crowd utilization and usage

do not change considerably in the instances with 10 nodes. The VSS is sensitive to the variance of the crowd capacity (measured by its CV) on both network sizes. The results show that the VSS tends to decrease at a higher variation when the total expected crowd capacity is at the lowest level ($\lambda^d = 1000$). On the other hand, it shows the opposite trend at higher expected demands ($\lambda^d = 3000$ and $\lambda^d = 5000$). This behavior is explained when the capacities on each space-time arc are analyzed in detail. After the distribution of the capacity on instances with $\lambda^d = 1000$, the sample presents scenarios with a higher fraction of space-time arcs with capacity zero when the CV increases, which harms the VSS. Similarly, on instances with $\lambda^d \geq 3000$, at a higher CV, the sampled scenarios have more space-time arcs with capacity zero but the observed average capacity on "non zero arcs" is higher. The latter increases the potential to exploit crowd flows, which results in a lower total (expected) cost with the same (or similar) infrastructure which at the same time has a positive impact on the VSS.

6. Conclusions

We present a transportation network for a crowd-sourced last-mile delivery system where professional courier services operate to give a guarantee of demand fulfillment. A multi-modal, multi-segment network, represents the concept of hyper-connected city logistics, which allows more flexibility to match supply and demand. The capacity of the network is dependent on the stochastic crowd's willingness to participate. The problem is formulated as a two-stage stochastic multi-commodity network design problem and is solved using an enhanced Benders decomposition. Numerical experiments are conducted on instances of 14 nodes to 24 nodes for a planning horizon of 8 hours of operation. The experiments are focused on the assessment of the CCL system to fulfill a demand of special (express) deliveries with short time windows that are difficult to fulfill with standard last-mile delivery services.

The results show that a network of mini-depots is needed to exploit the existing flows of people in the crowd. All of the instances require a network of mini-depots to either exploit the stochastic crowd capacity using cross-docking or to temporarily store parcels. More importantly, we show the importance of the use of the mini-depots to support cross-docking between crowd flows as an important fraction of the crowd shipments are routed using cross-docking. The experiments also show that the network design may be robust to changes in the cost structure and capacity-demand ratios. The use of the crowd for last-mile delivery yields important daily (average) savings of between 2% and 7.6% of the total expected operational costs.

The value of the stochastic solution is significant on instances with different demand, crowd capacity, and cost structure. Moreover, we identify an important effect of the variability of the crowd capacity. This gives insights into the need for the reduction of the uncertainty in the estimation of probability distributions or the need for incentives to reduce the uncertainty on the crowd's willingness to participate, especially in scenarios with lower demand. Our experiments support the use of professional courier services as a recourse to fulfill the demand with a guarantee of service. This is shown by the results on the utilization of the crowd since professional couriers still need to move the majority of the demand.

Future research can extend our work on city crowd logistics. First, we take into account the influence of the compensation of the crowd on the expected capacity in our experimental design but we do not consider this explicitly in the model. Future work in this field could study in more detail the compensation as a decision. Second, we take into account the operational level of decisions in the second stage. However, dynamic algorithms can be implemented to make the online decision of routing parcels in the CCL network taking into account the stochasticity of the system and the capacity of the mini-depots.

References

- Andersen, J., Crainic, T.G., Christiansen, M., 2009. Service network design with asset management: Formulations and comparative analyses. *Transportation Research Part C: Emerging Technologies* 17, 197–207. doi:10.1016/j.trc.2008.10.005.
- Archetti, C., Savelsbergh, M., Speranza, M.G., 2016. The vehicle routing problem with occasional drivers. *European Journal of Operational Research* 254, 472–480. doi:10.1016/j.ejor.2016.03.049.
- Arslan, A.M., Agatz, N., Kroon, L., Zuidwijk, R., 2019. Crowdsourced delivery a dynamic pickup and delivery problem with ad hoc drivers. *Transportation Science* 53, 222–235. doi:10.1287/trsc.2017.0803.
- Barnhart, C., Krishnan, N., Vance, P.H., 2009. *Multicommodity flow problems*. Springer US, Boston, MA. pp. 2354–2362. doi:10.1007/978-0-387-74759-0_407.
- Behrend, M., Meisel, F., Fagerholt, K., Andersson, H., 2019. An exact solution method for the capacitated item-sharing and crowdshipping problem. *European Journal of Operational Research* 279, 589–604. doi:10.1016/j.ejor.2019.05.026.

- Behrend, M., Meisel, F., Fagerholt, K., Andersson, H., 2021. A multi-period analysis of the integrated item-sharing and crowdshipping problem. *European Journal of Operational Research* 292, 483–499. doi:10.1016/j.ejor.2020.10.034.
- Benders, J.F., 1962. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik* 4, 238–252.
- Buldeo Rai, H., Verlinde, S., Merckx, J., Macharis, C., 2017. Crowd logistics: an opportunity for more sustainable urban freight transport? *European Transport Research Review* 9, 39. doi:10.1007/s12544-017-0256-6.
- Chen, C., Pan, S., Wang, Z., Zhong, R.Y., 2017. Using taxis to collect citywide e-commerce reverse flows: a crowdsourcing solution. *International Journal of Production Research* 55, 1833–1844. doi:10.1080/00207543.2016.1173258.
- Chung, G., Geising, B., Chaturvedi, K., Bodenbenner, P., 2018. Logistics trend radar—delivering insight today. Creating value tomorrow. . DHL, DHL Customer Solutions and Innovation.
- Contreras, I., Cordeau, J.F., Laporte, G., 2011. Benders decomposition for large-scale uncapacitated hub location. *Operations Research* 59, 1477–1490. doi:10.1287/opre.1110.0965.
- Crainic, T.G., 2000. Service network design in freight transportation. *European Journal of Operational Research* 122, 272 – 288. doi:10.1016/S0377-2217(99)00233-7.
- Crainic, T.G., Hewitt, M., Maggioni, F., Rei, W., 2021. Partial benders decomposition: General methodology and application to stochastic network design. *Transportation Science* 55, 414–435. doi:10.1287/trsc.2020.1022.
- Crainic, T.G., Hewitt, M., Toulouse, M., Vu, D.M., 2018. Scheduled service network design with resource acquisition and management. *EURO Journal on Transportation and Logistics* 7, 277–309. doi:10.1007/s13676-017-0103-x.
- Crainic, T.G., Montreuil, B., 2016. Physical internet enabled hyperconnected city logistics. *Transportation Research Procedia* 12, 383–398. doi:10.1016/j.trpro.2016.02.074.
- Crainic, T.G., Ricciardi, N., Storchi, G., 2004. Advanced freight transportation systems for congested urban areas. *Transportation Research Part C: Emerging Technologies* 12, 119–137. doi:10.1016/j.trc.2004.07.002.

- Dabblanc, L., Morganti, E., Arvidsson, N., Woxenius, J., Browne, M., Saidi, N., 2017. The rise of on-demand instant deliveries in european cities. *Supply Chain Forum: An International Journal* 18, 203–217. doi:10.1080/16258312.2017.1375375.
- Dayarian, I., Savelsbergh, M., 2020. Crowdsipping and same-day delivery: Employing in-store customers to deliver online orders. *Production and Operations Management* 29, 2153–2174. doi:10.1111/poms.13219.
- Fischetti, M., Ljubi, I., Sinnl, M., 2017. Redesigning benders decomposition for large-scale facility location. *Management Science* 63, 2146–2162. doi:10.1287/mnsc.2016.2461.
- Fontaine, P., Crainic, T.G., Jabali, O., Rei, W., 2021. Scheduled service network design with resource management for two-tier multimodal city logistics. *European Journal of Operational Research* doi:10.1016/j.ejor.2021.02.009.
- Fontaine, P., Minner, S., 2014. Benders decomposition for discrete-continuous linear bilevel problems with application to traffic network design. *Transportation Research Part B: Methodological* 70, 163–172. doi:10.1016/j.trb.2014.09.007.
- Fontaine, P., Minner, S., 2017. A dynamic discrete network design problem for maintenance planning in traffic networks. *Annals of Operations Research* 253, 757–772. doi:10.1007/s10479-016-2171-y.
- Frehe, V., Mehmman, J., Teuteberg, F., 2017. Understanding and assessing crowd logistics business models—using everyday people for last mile delivery. *Journal of Business & Industrial Marketing* 32, 75–97.
- Herrmann, E., Kunze, O., 2019. Facility location problems in city crowd logistics. *Transportation Research Procedia* 41, 117–134.
- Kaffe, N., Zou, B., Lin, J., 2017. Design and modeling of a crowdsourcing-enabled system for urban parcel relay and delivery. *Transportation Research Part B: Methodological* 99, 62 – 82. doi:10.1016/j.trb.2016.12.022.
- Kang, Y., Lee, S., Chung, B.D., 2019. Learning-based logistics planning and scheduling for crowdsourced parcel delivery. *Computers & Industrial Engineering* 132, 271 – 279. doi:10.1016/j.cie.2019.04.044.

- Le, T.V., Stathopoulos, A., Van Woensel, T., 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. doi:10.1016/j.trc.2019.03.023.
- Magnanti, T.L., Mirchandani, P., Vachani, R., 1995. Modeling and solving the two-facility capacitated network loading problem. *Operations Research* 43, 142–157. doi:10.1287/opre.43.1.142.
- Magnanti, T.L., Wong, R.T., 1981. Accelerating benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research* 29, 464–484. doi:10.1287/opre.29.3.464.
- Montreuil, B., Meller, R.D., Ballot, E., 2013. *Physical Internet Foundations*. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 151–166. doi:10.1007/978-3-642-35852-4_10.
- Mousavi, K., Bodur, M., Roorda, M.J., 2019. Stochastic last-mile delivery with crowd-shipping and mobile depots. URL: http://www.optimization-online.org/DB_FILE/2020/02/7619.pdf. preprint.
- Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W., 2017. The benders decomposition algorithm: A literature review. *European Journal of Operational Research* 259, 801–817. doi:10.1016/j.ejor.2016.12.005.
- Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W., 2018. Accelerating the benders decomposition method: Application to stochastic network design problems. *SIAM Journal on Optimization* 28, 875–903. doi:10.1137/17M1128204.
- Raviv, T., Tenzer, E.Z., 2018. Crowd-shipping of small parcels in a physical internet. URL: <http://www.eng.tau.ac.il/~talraviv/Publications/Crowd-shipping%20of%20small%20parcels%20in%20a%20physical%20internet.pdf>. preprint.
- Rougès, J.F., Montreuil, B., 2014. Crowdsourcing delivery: New interconnected business models to reinvent delivery, in: 1st international physical internet conference, pp. 1–19. URL: <https://www.cirrelet.ca/ipic2014/pdf/1027a.pdf>.
- Sampaio, A., Savelsbergh, M., Veelenturf, L., van Woensel, T., 2019. Chapter 15 - crowd-based city logistics, in: Faulin, J., Grasman, S.E., Juan, A.A., Hirsch, P. (Eds.), *Sustainable Transportation and Smart Logistics*. Elsevier, pp. 381–400. doi:10.1016/B978-0-12-814242-4.00015-6.

- Savelsbergh, M., Van Woensel, T., 2016. 50th anniversary invited article city logistics: Challenges and opportunities. *Transportation Science* 50, 579–590. doi:10.1287/trsc.2016.0675.
- Shapiro, A., 2014. Chapter 5: Statistical Inference. chapter 5. pp. 163–269. doi:10.1137/1.9781611973433.ch5.
- Striapunina, K., 2019. eCommerce Report 2019. Technical Report. Statista, Statista Digital Market Outlook - Market Report.
- United Nations, 2018. Revision of world urbanization prospects. UN Department of Economic and Social Affairs 16.
- Van Slyke, R.M., Wets, R., 1969. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics* 17, 638–663. doi:10.1137/0117061.