A Tailored Benders Decomposition Approach for Last-mile Delivery with Autonomous Robots

Ivana Ljubić

ESSEC Business School

ICORES, 3 to 5 Feb, 2022

Based on the paper

European Journal of Operational Research 299 (2022) 510–525

Contents lists available at ScienceDirect

European Journal of Operational Research

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

ELSEVIER

Production, Manufacturing, Transportation and Logistics

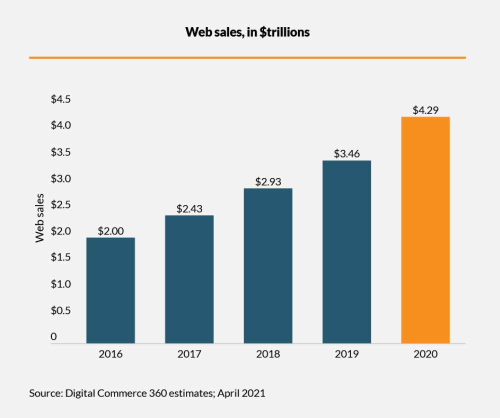
A tailored Benders decomposition approach for last-mile delivery with autonomous robots

Laurent Alfandari, Ivana Ljubić*, Marcos De Melo da Silva

ESSEC Business School, Cergy-Pontoise, France

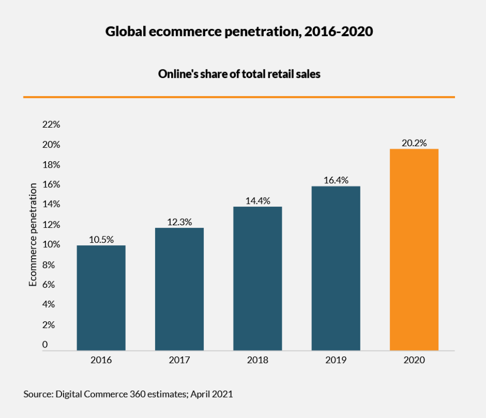
Context: E-commerce growth

- E-commerce sales worldwide grew sixfold in a decade
- From \$0.57 trillion in 2010 to some \$3.5 trillion in 2019.
- Increase of 24% in worldwide eCommerce revenue between pre- and post-COVID period.
- Food & Personal Care products show the most growth.



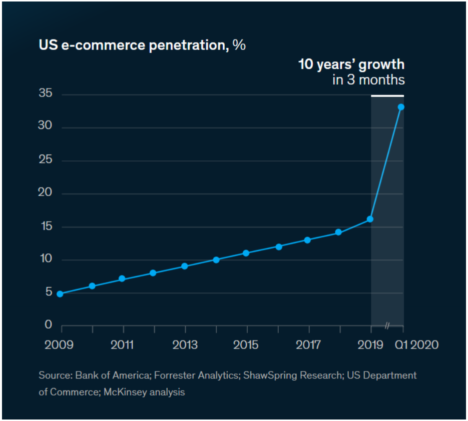
Context: E-commerce growth

- Paradigm shift due to COVID disruptions: drastic increase in online retail sales.



Context

- In the US, ten years' worth of growth took place within three months when the pandemic broke out.



Last-Mile Delivery

- Last leg of delivery accounts for more than 50% of shipping and transportation costs

(source: `advancedfleetmanagementconsulting.com`)

Last-Mile Delivery

Stakeholders:

- **Clients:** Expect swifter delivery times, but are not ready to pay for it
- **Retailers:** Increased volumes of goods $\Rightarrow$ slower delivery times, less flexibility in delivery time slots, and higher delivery costs
- **Citizens:** Delivery traffic rising, underutilized delivery vehicles, increased congestion and pollution

Bottleneck: **last-mile delivery systems**

Innovative last-mile concepts

- Constraints: same-day deliveries, due dates, etc.
- Innovative last-mile concepts:
  - pickup points networks
  - integrated public and freight transportation
  - deliveries directly into the customer car's trunk
  - crowdshipping
  - unmanned aerial vehicles (drones)
  - self-driving autonomous robots: multi- or single-deliveries (droids)

Recent survey: Archetti, Bertazzi (Networks, 2021)

Last-Mile Delivery

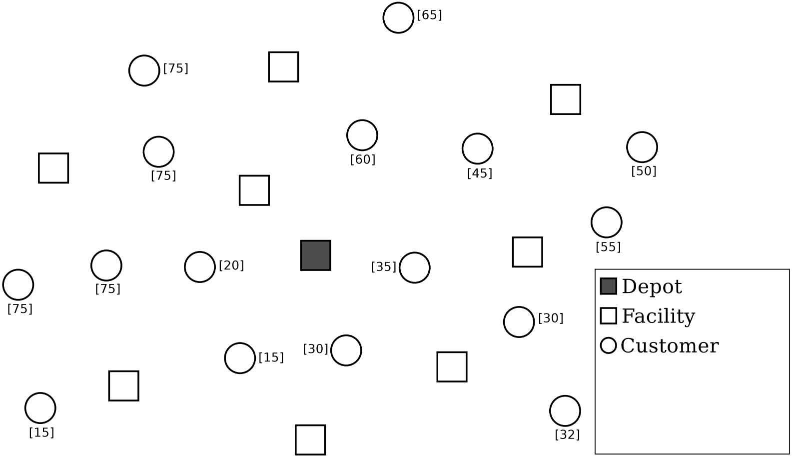
- **Location:** Bring products into pop-up facilities (smaller capacities) closer to customers, to reduce transportation costs.
- **Fleet:** porters, drones, droids. Autonomous vehicles could make 24/7 deliveries possible and overcome labor-shortages.
- **Efficient and environmentally friendly**

(a) Scout

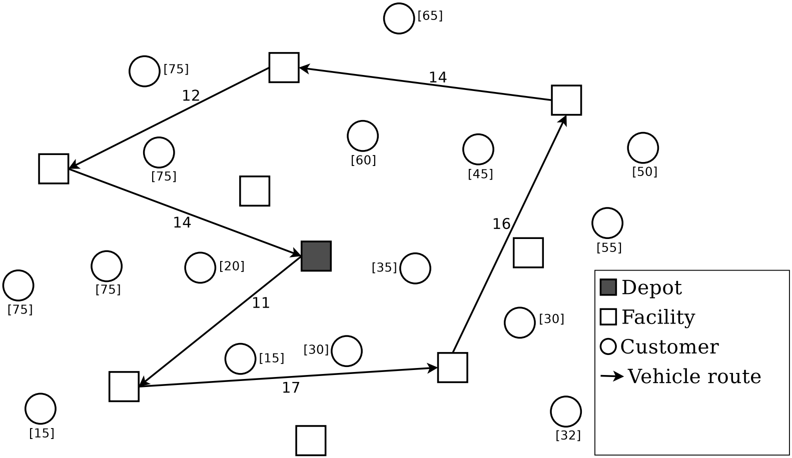
(b) Yape

(c) Starship

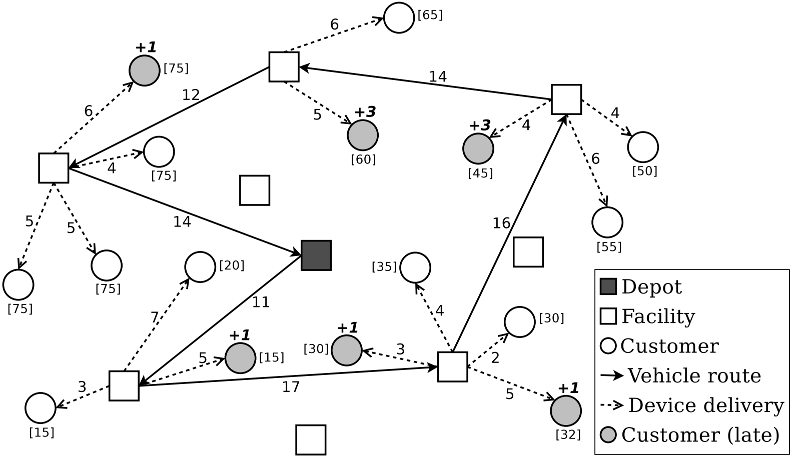
Setting


[65]
[75]
[75]
[60]
[45]
[50]
[55]
[20]
[35]
[75]
[75]
[30]
[15]
[30]
[15]
[32]
Depot
Facility
Customer

A subset of robot-facilities is chosen and a vehicle brings parcels closer to the clients



Droids are sent out from robot-facilities to end clients



Droids vs Drones?

- Regulation issues (slower adoption of drones due to safety concerns)
- Droids operate at low (pedestrian) speeds, they can safely share sidewalks or bike lanes with people
- Drones allow for unattended delivery, which is (currently) not possible with droids
- Drones travel at a higher speed, so that the trucks can collect them en-route and reuse them for later deliveries

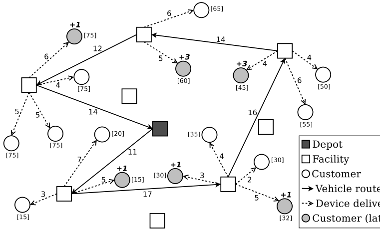
Assumptions (operational decision making)

- Single parcel delivery
- Attended delivery
- Due dates for customer delivery agreed beforehand
- Sufficient number of droids available at robot-facilities
- Droids have a coverage radius (speed/battery)

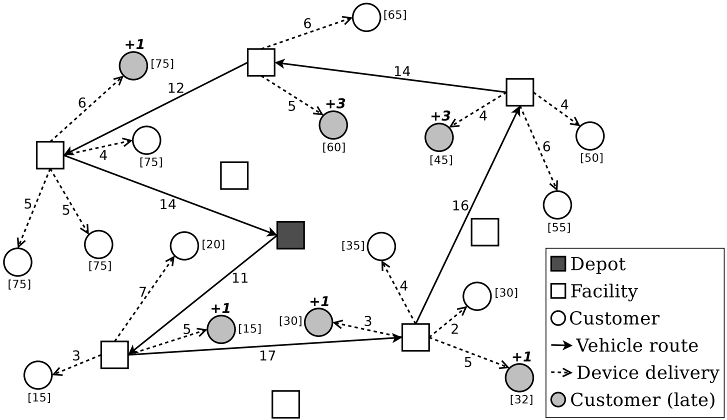
Objectives?

Change of paradigm:

- Distance-Minimization no longer an issue (e-powered vehicles)
- **New KPIs based on late deliveries**
- Minimization
  - the maximum tardiness (`min-max`),
  - the total tardiness (`min-sum`), or
  - the number of late deliveries (`min-num`)



Example



the maximum tardiness (`min-max`): 3

the total tardiness (`min-sum`): 10

the number of late deliveries (`min-num`): 6

Contribution

- Tardiness-based KPIs have been almost neglected in the last-mile delivery (mainly humanitarian logistics so far)
- What is the complexity of the underlying **routing-scheduling** optimization problems?
- We use MIP-based techniques to find (nearly) optimal solutions. Advanced decomposition technique (tailored Benders cuts)
- Managerial insights:
  - What is the impact of the coverage radius (or speed of droids) on the QoS?
  - How are the droid speed/radius affecting the environment (the distance traveled by the delivery truck, and CO2 emissions)
  - How is the structure of the optimal solution affected by the choice of the KPI?

The Uncapacitated Routing Scheduling Problem

The Uncapacitated Routing Scheduling Problem

- Given: F set of robot facilities, C set of customers (u_k due dates), central depot (0).
- Uncapacitated Routing Scheduling Problem (URSP):
  - Solution: Solution: a route starting from the depot 0 and visiting a subset of facilities so that all customers are delivered from their assigned (closest) facility by a droid.
  - Objective: minimize
    - max (weighted) tardiness, (`min-max`)
    - total (weighted) tardiness, (`min-sum`)
    - total (weighted) number of late deliveries, (`min-num`)

Positioning vs literature

Special case of Boysen et al. (2018). Their paper :

- truck loaded with both customer parcels and droids (this work: only parcels in truck)
- additional drop-off points from where droids in truck can be launched to perform deliveries (this work: droids only sent from facilities)
- truck can be reloaded with new droids at given stations
- heuristic method (this work: exact method, Benders)

For more bibliography: see our paper Alfandari, Ljubić, Melo, EJOR, 2022

URSP Model

- Multi-commodity flow MILP formulation.

Notation

- Network is modelled by direct graph $G(V, A)$:
  - $V = (C \cup F_0)$ is the set of vertices.
  - $A = A_F \cup A_C$, is the set of arcs.
  - $F_0 = (F \cup \{0\})$ is the set of facilities + depot.
  - $A_F = (F_0 \times F_0)$ and $A_C = (F \times C)$.
  - t_{ij} is the time required for travelling from i to j.
  - Customer $k \in C$ has due date u_k.

Problem Complexity

Central facility policy: Truck delivers all parcels to a single facility, from which all self-driving robots are sent.

When is the central-facility policy optimal?

The central-facility policy is optimal for all three variants of the URSP if droids are at least as fast as the truck and all customers can be reached from any facility.
In this case, the solution can be found in $O(|F||C|)$ time.

Otherwise...

If travel times do not satisfy the triangle inequality, all three problems are NP-hard, in general.

By reduction from: metric Shortest Hamiltonian Path problem (`min-max` and `min-num`), the Traveling Repairman Problem (TRP) (`min-sum`).

URSP Model - Variables

Variables

- $f_{ij}^k = \begin{cases} 1, & \text{if arc } (i,j) \in A_F \text{ is on path to customer } k \in C \\ 0, & \text{otherwise} \end{cases}$
- $z_{ik} = \begin{cases} 1, & \text{if customer } k \in C \text{ is served by facility } i \in F(k) \\ 0, & \text{otherwise} \end{cases}$
- $x_{ij} = \begin{cases} 1, & \text{if arc } (i,j) \in A_F \text{ is on the vehicle tour} \\ 0, & \text{otherwise} \end{cases}$
- Auxiliary variables $s_k \geq 0$: number of late units when serving customer $k \in C$.

URSP Model - Constraints 01

Tour Constraints

$$x(\delta^-(i)) = x(\delta^+(i)) \qquad i \in F \tag{1}$$

$$x(\delta^-(i)) \leq 1 \qquad i \in F \tag{2}$$

$$x(\delta^+(0)) = x(\delta^-(0)) = 1 \tag{3}$$

$$x_{ij} \in \{0, 1\} \qquad (i, j) \in A_F \tag{4}$$

Subtour-elimination constraints

$$x(A(S)) \leq |S| - 1 \qquad S \subseteq F, |S| \geq 2 \tag{5}$$

URSP Model - Constraints 02

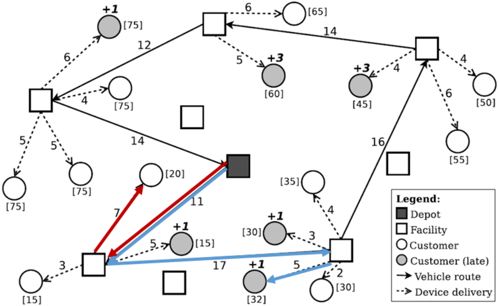
Facility Assignment Constraints

$$z_{ik} \leq x(\delta^{-}(i)) \qquad k \in C,\ i \in F(k) \tag{6}$$

$$\sum_{i \in F(k)} z_{ik} = 1 \qquad k \in C \tag{7}$$

$$0 \leq z_{ik} \leq 1 \qquad k \in C, i \in F(k) \tag{8}$$

How to model travel time?



we need to access the unique path between the depot and a customer, to calculate the arrival time

URSP Model - Constraints 02

Flow Constraints

$$\sum_{j \in F_0} f_{ji}^k - \sum_{j \in F_0} f_{ij}^k = \begin{cases} -1, & \text{if } i = 0 \\ z_{ik}, & \text{if } i \in F(k) \\ 0, & \text{otherwise} \end{cases} \qquad i \in F,\ k \in C \quad (9)$$

$$0 \leq f_{ij}^k \leq x_{ij} \qquad (i,j) \in A_F,\ k \in C \quad (10)$$

Travel-time Constraints: s_k is #(late units)

$$\sum_{(i,j) \in A_F} t_{ij} f_{ij}^k + \sum_{i \in F(k)} t_{ik} z_{ik} \leq u_k + s_k \qquad k \in C \quad (11)$$

URSP Model - Objective Functions

min-max

$$\text{Minimize} \quad t$$
$$t \geq w_k s_k \qquad k \in C$$

min-sum

$$\text{Minimize} \quad \sum_{k \in C} w_k s_k$$

min-num

$$\text{Minimize} \quad \sum_{k \in C} w_k \ell_k$$
$$s_k \leq (M_k - u_k)\ell_k \qquad k \in C$$
$$\ell_k \in \{0,1\} \qquad k \in C$$

Benders Decomposition

Generic Benders

- Structure of the problem fits on the Benders Decomposition framework:
  - Master: "difficult" routing constraints (binary variables).
  - Subproblems: "easy" flow constrains (continuous variables).
- *General*: Applicable to all three objective functions
- *Decomposable*: Benders subproblem is separable, one subproblem per customer.

URSP Benders - Master

Master Problem

$$\text{Minimize} \quad t \tag{12}$$

$$
\begin{aligned}
s.t.: \qquad\qquad t &\geq w_k s_k \quad k \in C \\
\theta_k(x) &\leq s_k + u_k \quad k \in C \quad \text{/*Benders cuts*/} \\
x &\text{ ``is a route''} \\
\sum_{i \in F(k)} x(\delta^-(i)) &\geq 1 \quad k \in C \\
t, s_k &\geq 0 \quad k \in C
\end{aligned}
$$

We are projecting out flow (f) and assignment (z) variables.
Adding Benders cuts on-the-fly in a branch-and-cut fashion.

To generate Benders cuts...

- Given a route x and a "guess" of the lateness-times s_k: is there a feasible solution such that
  - each client $k \in C$ can be reached from the depot within the time $u_k + s_k$?

$$\theta_k(x) \leq s_k + u_k \quad k \in C \quad \text{/*Benders cuts*/}$$

- $\theta_k(x)$ has to evaluate what is the time needed to reach k?
- For a given $x \Rightarrow$ the shortest path problem in the support graph induced by x.

Non-standard Benders approach

- **Textbook implementation:** generate Benders feasibility cuts, add a cut for an extreme ray, etc.
- **Problems:** numerically unstable, slow convergence, choice of rays,...

- **Our approach:**
  - **Normalization:** Instead of checking the unboundedness of the dual of the LP, we are solving a well-posed shortest-path-like problem (optimal solution always exists).
  - **Separation:** instead of solving an LP, we have a combinatorial way of separating **normalized Benders cuts** (labeling algorithm, runs in linear time).
  - Our cuts are **numerically stable** and **sparse**.

Two major theoretical results - Part 1

Theorem [Alfandari, L., Melo da Silva, 2022]

For a master solution $x^* \in \{0,1\}^{|A_F|}$ and customer $k \in C$, coefficients (α^*, β^*) of the the Benders cut

$$\alpha_k^* - \alpha_0^* - \sum_{(i,j)\in A_F} \beta_{ij}^* x_{ij} \leq s_k + u_k$$

can be computed as: $\alpha_k^* = t_T^*(0,k) = \min_{i \in T}\{t_T(0,i) + t_{ik}\}$ and for facility nodes $i \in F$:

(i) $\alpha_i^* = t_T^*(0,i)$, if $i \in F_T$, $i \in P_T^*(0,k)$

(ii) $\alpha_i^* = \min(t_T^*(0,i),\ \alpha_k^* - t_G^*(i,k))$, if $i \in F_T$, $i \notin P_T^*(0,k)$

(iii) $\alpha_i^* = \alpha_k^* - t_G^*(i,k)$, if $i \in F_{\bar{T}} = F \setminus T$

and finally

$$\beta_{ij}^* = \max(0,\ \alpha_j^* - \alpha_i^* - t_{ij}),\ \forall (i,j) \in A_F$$

Two major theoretical results - Part 2, Sparsity

Theorem [Alfandari, L., Melo da Silva, 2019]

For the coefficients (α^*, β^*) of the Benders cut computed as above, and the proportion of facility nodes including 0 that are in tour T $\rho = |F_T|/|F_0| \in [\frac{2}{m}, 1]$), where $m = |F_0|$), the fraction of variables β^*_{ij} equal to zero in the Benders cut is at least

$$g(\rho) = 1 - \rho + \frac{\rho^2}{2} + \frac{1,5\rho - 1}{m} - \frac{1}{m^2} \geq \frac{1}{2}. \qquad (13)$$

Moreover, we have $\lim_{\rho \to 2/m} g(\rho) = 1 - o(1/m)$.

- At least half of the coefficients in the cut are zero
- All coefficients are integer (for t_{ij} integers)

Heuristics to initialize Branch-and-Cut

- A **greedy heuristic** to build initial tour T':
  - Greedy insertion of facilities following the most-urgent-deadline-first policy. We apply the *best insertion policy with respect to the total tardiness criterion*.
- **Local search** applied to T': node re-insertions and swaps in a Variable Neighborhood Descent fashion, and finally remove unused facilities.

Computational Study

Experiments - Implementation and Instances

- Algorithms were implemented in C/C++.
- Machine: Intel Core™ i7 4.00 GHz, 64.0 GB of RAM, running under GNU/Linux Arch.
- LP/MIP Solver: CPLEX 12.8:
  - Time limit: One hour (3600s).
  - Comparison against: CPLEX Automatic Benders, CPLEX compact model, our Benders with LP-separation.

Experiments - Implementation and Instances

- Facilities and customers coordinates: uniformly generated on a 10×10 km square grid.
- Customer due dates computed as in Boysen et al. (2018).

- Two sets of instances based on devices coverage area:
  1. All customers within facility's coverage radius (comparison of methods).
  2. Coverage radius limited (managerial study).

- A greedy heuristic + local search was implemented for computing feasible solutions.
- Connectivity cuts separated for integer master solutions.

Parameters considered

Table 1: Parameters with $*$ used for the managerial study.

Parameters	Values
`Number of Customers` ($\|C\|$)	25, 50*, 75, 100
`Number of Facilities` ($\|F\|$)	10, 15, 20*, 25
`Device speed` (km/h)	5*, 6, 10, 15
`Device coverage radius` (R min.)	30, 35, 40, 45, 60*
`Objective function`	`min-max`*, `min-sum`, `min-num`

(*) default values.

Starship (6km coverage radius, 6km/h speed).
e-novia (80km radius, 6-20km/h speed)

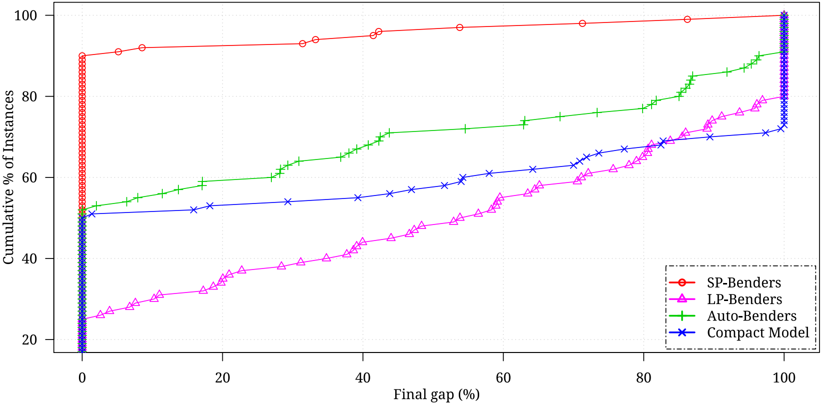
Methods Comparison: 20 facilities, up to 100 customers, min-max

	Compact Model (Cplex)			Auto-Benders (Cplex)		
Instances	#Solved	Time (s)	Gap (%)	#Solved	Time (s)	Gap (%)
rsp_20_25	23	626.50	1.36	15	1568.36	35.84
rsp_20_50	17	1714.56	20.65	14	1643.19	28.58
rsp_20_75	10	3005.53	46.10	13	1932.53	22.24
rsp_20_100	0	3600.00	95.60	9	2617.63	35.64
Total/Avg.	50		41.06	51		30.58

	LP-Benders			Tailored Benders		
Instances	#Solved	Time (s)	Gap (%)	#Solved	Time (s)	Gap (%)
rsp_20_25	10	2317.31	40.55	25	253.65	0.00
rsp_20_50	10	2844.46	47.45	21	669.23	8.31
rsp_20_75	3	3272.03	51.52	23	525.43	5.14
rsp_20_100	2	3454.41	62.68	21	1020.25	5.48
Total/Avg.	25		50.55	90		4.73

URSP Computational Results - Methods Comparison

- No coverage radius instances:
  - $|F| = 20$, $|C| = \{25, 50, 75, 100\}$.
  - Vehicle speed: $30\ km/h$. Device speed: $5\ km/h$.
  - 25 instances in each group.
- Objective function: `min-max`.



Varying Robots' Speed

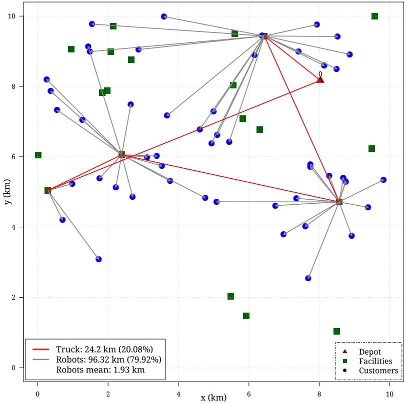
Table 2: Solution properties.

	Varying Robots' Speed			
Speed (km/h)	Avg. tour #Facilities	Avg. truck dist. (km)	Avg. all robots dist. (km)	Avg. single robot dist. (km)
5	6.06	27.26	183.29	3.67
6	5.46	24.17	202.12	4.04
10	4.28	19.68	264.10	5.28
15	2.90	16.01	360.67	7.21

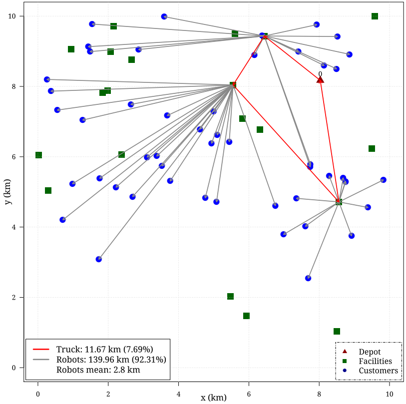
Increasing the speed of robots from 5km/h to 15km/h $\Rightarrow$ annual savings of ≈ 0.7 tons of CO_2 (for a single urban area of 10 km^2).

URSP Computational Results - Cost-benefit Analysis

- URSP selected solutions for robot speed $\in \{5, 15\}$ km/h



(a) Robot speed $= 5$ km/h, Max. Tardiness $= 11.33$ min.



(b) Robot speed $= 15$ km/h, Max. Tardiness $= 0.46$ min.

Varying Coverage Radius

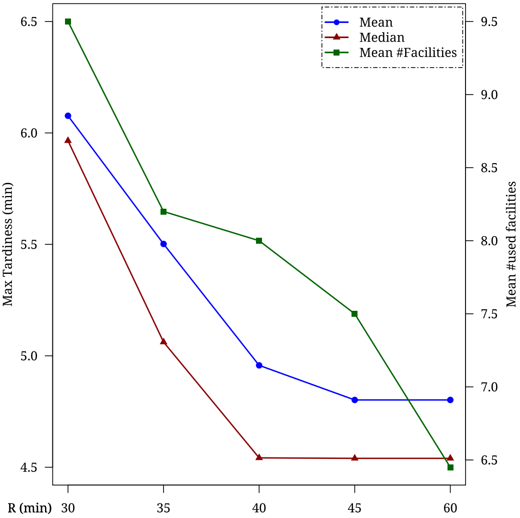
Table 3: Solution properties.

Radius (minutes)	Varying Coverage Radius: Avg. tour #Facilities	Avg. truck dist. (km)	Avg. all robots dist. (km)	Avg. single robot dist. (km)
30	9.08	39.79	143.04	2.86
35	7.90	35.14	154.56	3.09
40	7.46	32.46	161.21	3.22
45	6.92	30.33	169.29	3.39
60	6.06	27.26	183.29	3.67

Increasing the coverage radius from 30 minutes to 60 minutes $\Rightarrow$ annual savings of ≈ 0.75 tons of CO_2 (for a single urban area of 10 km^2).

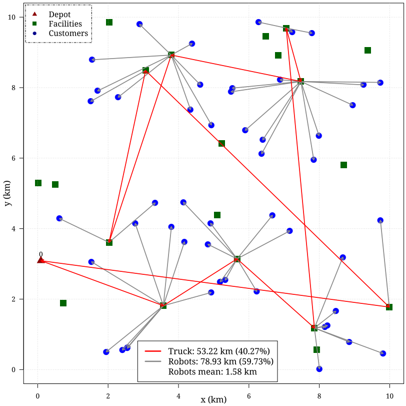
URSP Computational Results - Cost-benefit Analysis

- Effect of increasing the coverage radius.

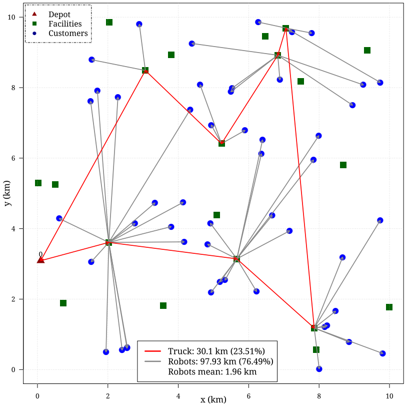


URSP Computational Results - Cost-benefit Analysis

- URSP selected solutions for coverage radius $R \in \{30, 60\}$ min



(a) R $= 30$ min, Max. Tardiness $= 6.84$ min.



(b) R $= 60$ min, Max. Tardiness $= 5.81$ min.

Comparing the three different KPIs.

Table 4: Each row: optimal solutions for 25 instances obtained by one of the three objective functions. Each column: KPI evaluation of these solutions.

	max		sum		num	
Objective	Median	Mean	Median	Mean	Median	Mean
`min-max`	(3.57)	(4.74)	3.57	5.99	14.60	13.06
`min-sum`	5.40	15.63	(5.40)	(11.69)	15.79	20.75
`min-num`	3.00	3.72	3.00	2.72	(2.00)	(1.72)

- `min-sum` and `min-max` solutions are similar, but can be quite different from `min-num` solutions.
- If the decision-maker wishes to control both criteria (the number of late customers and the total/max tardiness), both of them should be included in the decision model.
- Bi-objective optimization: ϵ-constrained methods, linear combination, etc.

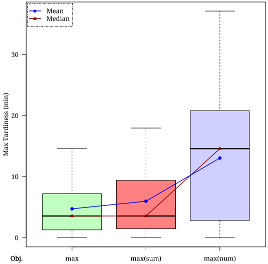
Directions for Future Work

- Capacitated RSP: capacities at the facilities
- Single stop for the truck, but multiple trips for self-driving robots
- Optimization under uncertainty: uncertain travel times (robust vs stochastic)
- Dynamic and/or stochastic arrival of the packages at the central depot
- Methodologies: Matheuristics, Monte-Carlo, Reinforcement Learning, Bi-objective optimization

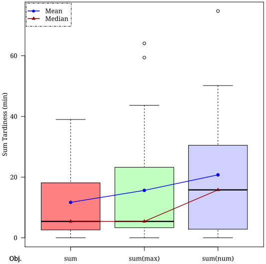
Thank you for your attention!!!

URSP Computational Results - Cost-benefit Analysis

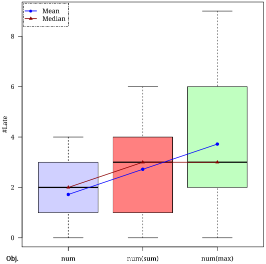
- Comparing the three key performance indicators



(a) `min-max` objective



(b) `min-sum` objective



(c) `min-num` objective