Aalto-yliopisto
Perustieteiden korkeakoulu

An alternative IP formulation for zone-based tariff design

Ellen Koivisto

13.06.2025

Instructor: Philine Schiewe

Supervisor: Philine Schiewe

Aalto-yliopisto
Perustieteiden korkeakoulu

Public transportation networks

- Tariff design is important when developing efficient and cost-effective public transportation networks.

- Smart zone-based tariff design eases the comprehensibility of the network for the customer and exploits their willingness to pay to maximize the profit for the transportation provider.

- The zone-based tariff design problem (ZTDP), defined by B. Otto *et al*. [1], offers diverse possibilities for further research.

Modelling the problem mathematically

- Integer programs (IP's) are mathematical optimization problems where some or all variables must be integers.
- From article [1], the generic form of ZTDP is the following:

$$R(\pi) = \sum_{c=1}^{C} f_c \times \pi_c \rightarrow max.$$

$$\left\{ \text{Connected / Ring} \right\} \text{zoning constraints}$$

$$\left\{ \text{Counting zones / Cumulative / Maximum} \right\} \text{pricing constraints}$$

- In the original ZTDP formulation [1], the zone selection is embedded within a mixed-integer program, where the constraints ensure a feasible zoning: each zone remains connected and follows the specified structure etc. .
- The process forces the solver to search through all valid groupings of stops which is a combinatorially explosive, NP-hard problem (Theorem 1 [1]) .

Goals and scope

- Goal: Improve the computational efficiency of solving the zone-based tariff design problem (ZTDP).

- Approach: Instead of searching across *all* possible zones, we start from a predefined pool of zones, reducing the search space and simplifying the optimization.

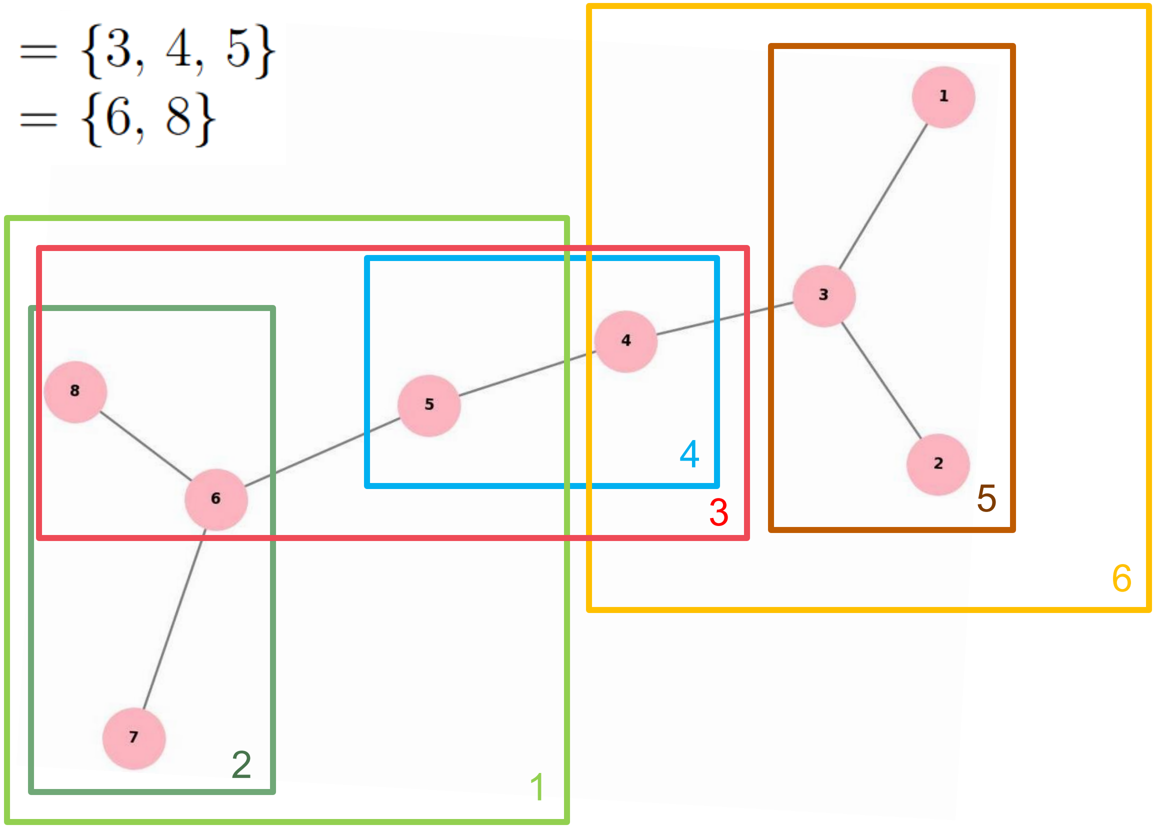
- Key findings: This simplification speeds up computation, while maintaining realistic zone structures.

Example

A simple transportation network with the following example customers:

$$c_1\colon\ w_1 = 1,\ f_1 = 1,\ V_1 = \{3,\ 4,\ 5\}$$
$$c_2\colon\ w_2 = 3,\ f_2 = 1,\ V_2 = \{6,\ 8\}$$

Instead of letting the program form the zones, we provide a zone-pool of size 6 (|Z| = 6) from which the problem then chooses the zones:



IP-formulation: Zone choosing process

$$\sum_{i \in \{1,..,|Z|\}} z_i \leq N$$

$$\sum_{i \in \{1,...,|N|\}} a_{is} z_i = 1, \quad \forall s \in S$$

$$z_i \in \{0, 1\} \quad \forall i \in 1, ...N$$

Parameters	
Z	The zone-pool
S	Set of stops
N	Zone budget
a_{is}	Binary parameter: 1 if zone i contains stop s, 0 otherwise
Decision Variables	
z_i	Binary variable: 1 if zone $i \in Z$ is selected; 0 otherwise

Table 1: Explanation of model sets, parameters, and decision variables

IP-formulation: Pricing constraints

Counting zones pricing

$$0 \le p_1, \quad p_z \le p_{z+1} \quad \forall z \in Z \tag{20}$$

$$x_c = \sum_{i \in Z} \gamma_{ci} z_i \quad \forall c \in C \tag{21}$$

$$x_c = \sum_{j=1}^{N} v_c \times j \tag{22}$$

$$\pi_c \le p_z + (2 - u_c - v_{cz}) \times M \quad \forall c \in C \tag{23}$$

$$p_c \le \pi_z + (2 - u_c - v_{cz}) \times M \quad \forall c \in C \tag{24}$$

$$\pi_c \le u_c w_c \quad \forall c \in C \tag{25}$$

$$u_c, v_{cz} \in \{0, 1\} \quad \forall c \in C \quad z = 1, ..., N \tag{26}$$

Sets and Parameters	
C	Set of customers
γ_{cz}	Binary parameter: 1 if the path of customer c crosses zone z, 0 otherwise
u_c	Binary parameter: 1 if client c uses public transportation, 0 otherwise
v_{cz}	Binary parameter indicating the number z of zones used by customer c
M	A large constant
w_c	The willingness to pay of customer c
Decision Variables	
p_z	The price of using z zones
x_c	The number of zones used by customer c

IP-formulation: Pricing constraints

Cumulative pricing

$$0 \leq p_z \quad \forall z \in 1, ..., N \tag{5}$$

$$\pi_{cz} \leq \gamma_{cz} \times M \quad \forall c \in C \quad \forall z \in Z \tag{6}$$

$$\pi_{cz} \leq p_z + (2 - \gamma_{cz} - u_c) \cdot M, \quad \forall c \in C \quad z = 1, 2, \ldots, N \tag{7}$$

$$p_z \leq \pi_{cz} + (2 - \gamma_{cz} - u_c) \cdot M, \quad \forall c \in C \quad z = 1, 2, \ldots, N \tag{8}$$

$$\pi_c \leq u_c w_c \quad \forall c \in C \tag{9}$$

$$\sum_{z=1}^{N} \pi_{zc} = \pi_c \quad \forall c \in C \tag{10}$$

$$u_c \in \{0, 1\} \quad \forall c \in C \tag{11}$$

Sets and Parameters	
C	Set of customers
γ_{cz}	Binary parameter: 1 if the path of customer c crosses zone z, 0 otherwise
u_c	Binary parameter: 1 if client c uses public transportation, 0 otherwise
v_{cz}	Binary parameter indicating the number z of zones used by customer c
M	A large constant
w_c	The willingness to pay of customer c
Decision Variables	
π_{cz}	Auxiliary variable for the price paid by the customer c of zone z
p_z	The price of using z zones
x_c	The number of zones used by customer c

IP-formulation: Pricing constraints

Maximum pricing

$$x_c = \sum_{i \in Z} \gamma_{ci} z_i \quad \forall c \in C \tag{12}$$

$$x_c = \sum_{j=1}^{N} v_c \times j \tag{13}$$

$$\pi_c \leq p_z + (2 - u_c - v_{cz}) \times M \quad \forall c \in C \tag{14}$$

$$p_c \leq \pi_z + (2 - u_c - v_{cz}) \times M \quad \forall c \in C \tag{15}$$

$$\pi_c \leq u_c w_c \quad \forall c \in C \tag{16}$$

$$v_{cz} \leq \gamma_{cz} \forall c \in C \quad z = 1, 2, ..., N \tag{17}$$

$$p_{z'} \leq p_z + (2 - \gamma_{cz} - v_{cz}) \times M \quad \forall c \in C \quad z, z' = 1, 2, ..., N \tag{18}$$

$$u_c, v_{cz} \in \{0, 1\} \quad \forall c \in C \quad z = 1, ..., N \tag{19}$$

Sets and Parameters	
C	Set of customers
γ_{cz}	Binary parameter: 1 if the path of customer c crosses zone z, 0 otherwise
u_c	Binary parameter: 1 if client c uses public transportation, 0 otherwise
v_{cz}	Binary parameter indicating the number z of zones used by customer c
M	A large constant
w_c	The willingness to pay of customer c
Decision Variables	
p_z	The price of using zone z
x_c	The number of zones used by customer c

IP-formulation: Full model

$$\max \quad R(\pi) = \sum_{c \in C} \pi_c \times f_c$$

$$\sum_{i \in \{1,..,|Z|\}} z_i \leq N$$

$$\sum_{i \in \{1,\ldots,|N|\}} a_{is} z_i = 1, \quad \forall s \in S$$

$$z_i \in \{0,1\} \quad \forall i \in 1, \ldots N$$

$$0 \leq p_1, \quad p_z \leq p_{z+1} \quad \forall z \in Z$$

$$x_c = \sum_{i \in Z} \gamma_{ci} z_i \quad \forall c \in C$$

$$x_c = \sum_{j=1}^{N} v_c \times j$$

$$\pi_c \leq p_z + (2 - u_c - v_{cz}) \times M \quad \forall c \in C$$

$$p_c \leq \pi_z + (2 - u_c - v_{cz}) \times M \quad \forall c \in C$$

$$\pi_c \leq u_c w_c \quad \forall c \in C$$

$$u_c, v_{cz}, z_i \in \{0,1\} \quad \forall c \in C \quad z = 1, \ldots, N$$

Methods / Tools for evaluation

- First the complexity of the algorithm is analyzed mathematically, and it is shown to be NP-complete.

- Then the results of computational experiments conducted with the LinTim software [1] are briefly discussed.

Set up for the computational study

- We use the Mandl network [1], a Swiss case study by Christopher Mandl with 15 stops, 21 edges, and 15,570 ‘passengers’.
- The zone-pools were computed using Python package networkx.
- The customers’ willingness to pay was computed using a function implemented in the LinTim [2] environment, which considered the Beeline distances between the nodes.

Set up for the computational study

- The generation of the connected zones - zone-pool will be done with the definition of connected zoning from [1]:

  > A set of pairwise disjoint zones Z = {Z1, Z2, ... , ZN } is a connected zoning, if each zone induces a connected subgraph of G

- However, this is only done to simplify the testing process. Any other type of zoning is still compatible with the process here.

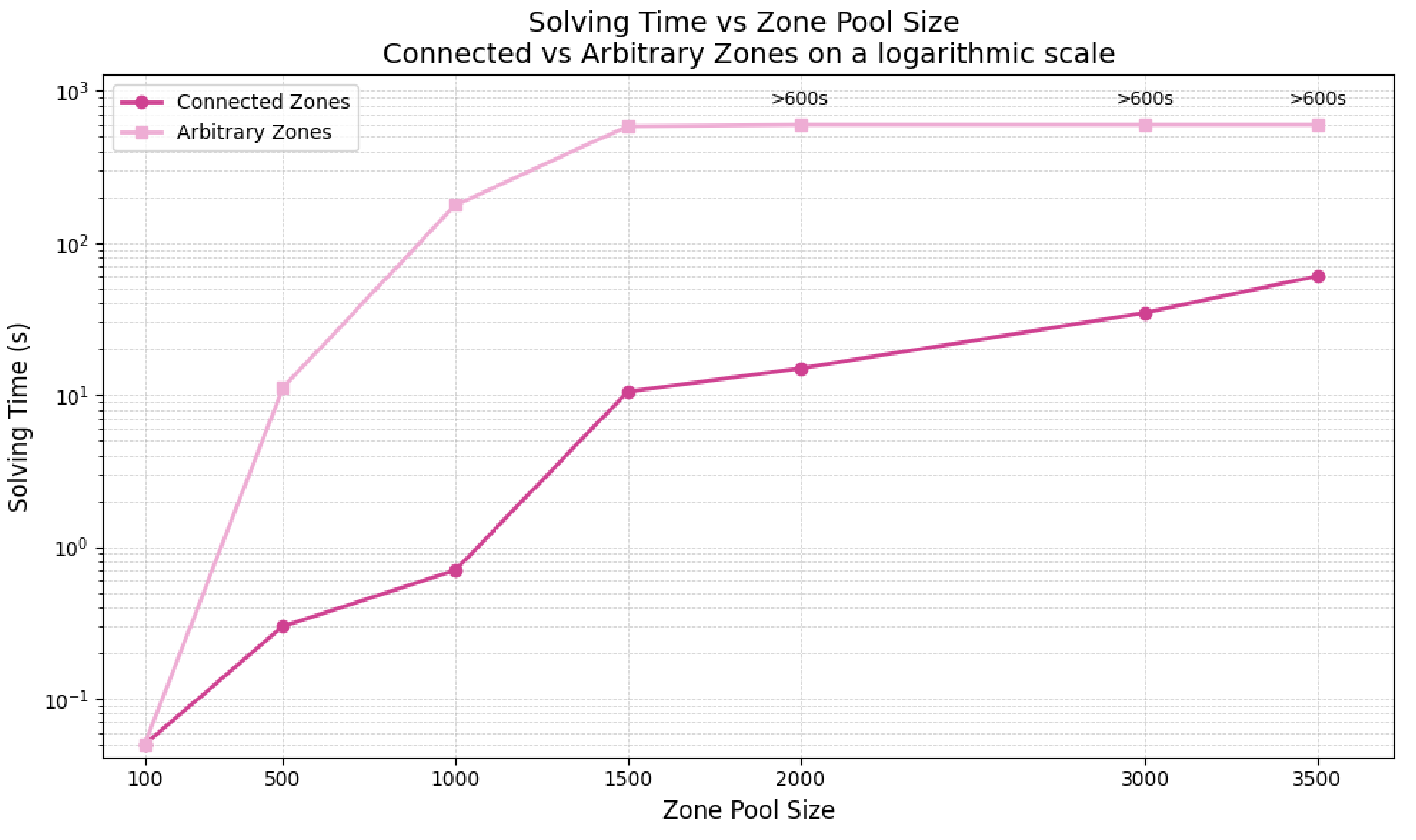
Computational study: Zone-pool size comparison

Connected zones			Arbitrary zones	
Zone Pool	Time [s]		Zone Pool	Time [s]
100	<0.1		100	<0.1
500	0.3		500	11.1
1000	0.7		1000	176.9
1500	10.5		1500	585.9
2000	14.9		2000	>600.0
3000	34.7		3000	>600.0
3500	60.4		3500	>600.0

Table 6: Solving Times for Different Zone Pool Sizes under Two Zoning Strategies

- Each solving time is the average of solving 5 instances with different zone pools chosen as a random sample from the connected zones (3551 in total) and the arbitrary zones (32 752 in total) of the Mandl set. The zone budget was kept constant (N = 5) for each iteration and the solving time limit was 600.0 s.

Computational study: Zone-pool size comparison



Computational study: Testing with limited zone-pool instances

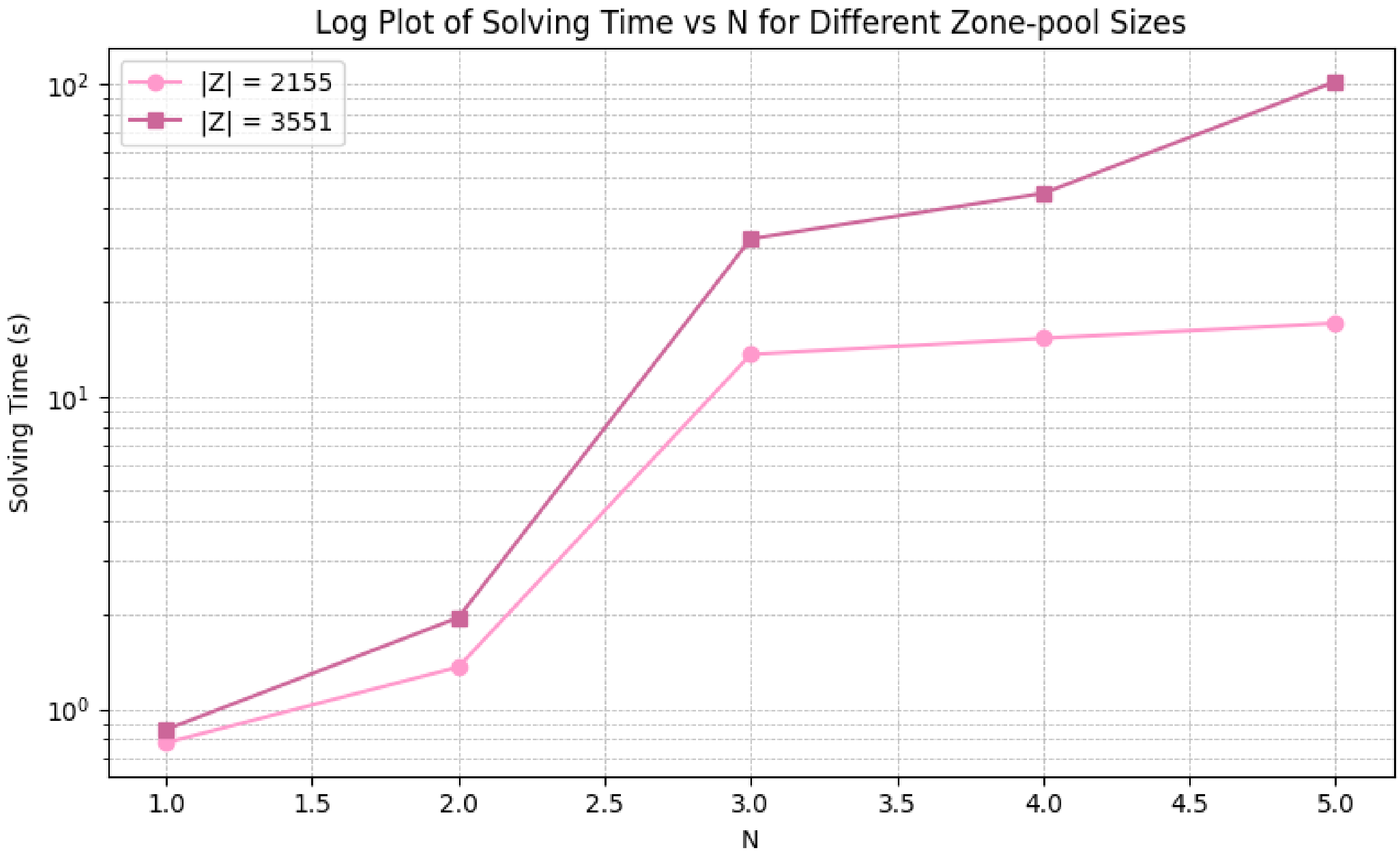
- There are total of 3551 connected subgraphs of the Mandl set. Adding constraints to the size of the subsets, the zone pool was limited to 2155 zones. The results of testing with that zone-pool:

N	Time [s]	
	$\lvert Z \rvert = 2155$	$\lvert Z \rvert = 3551$
1	0.370	0.860
2	1.362	1.960
3	13.622	31.918
4	15.334	44.581
5	17.145	101.550

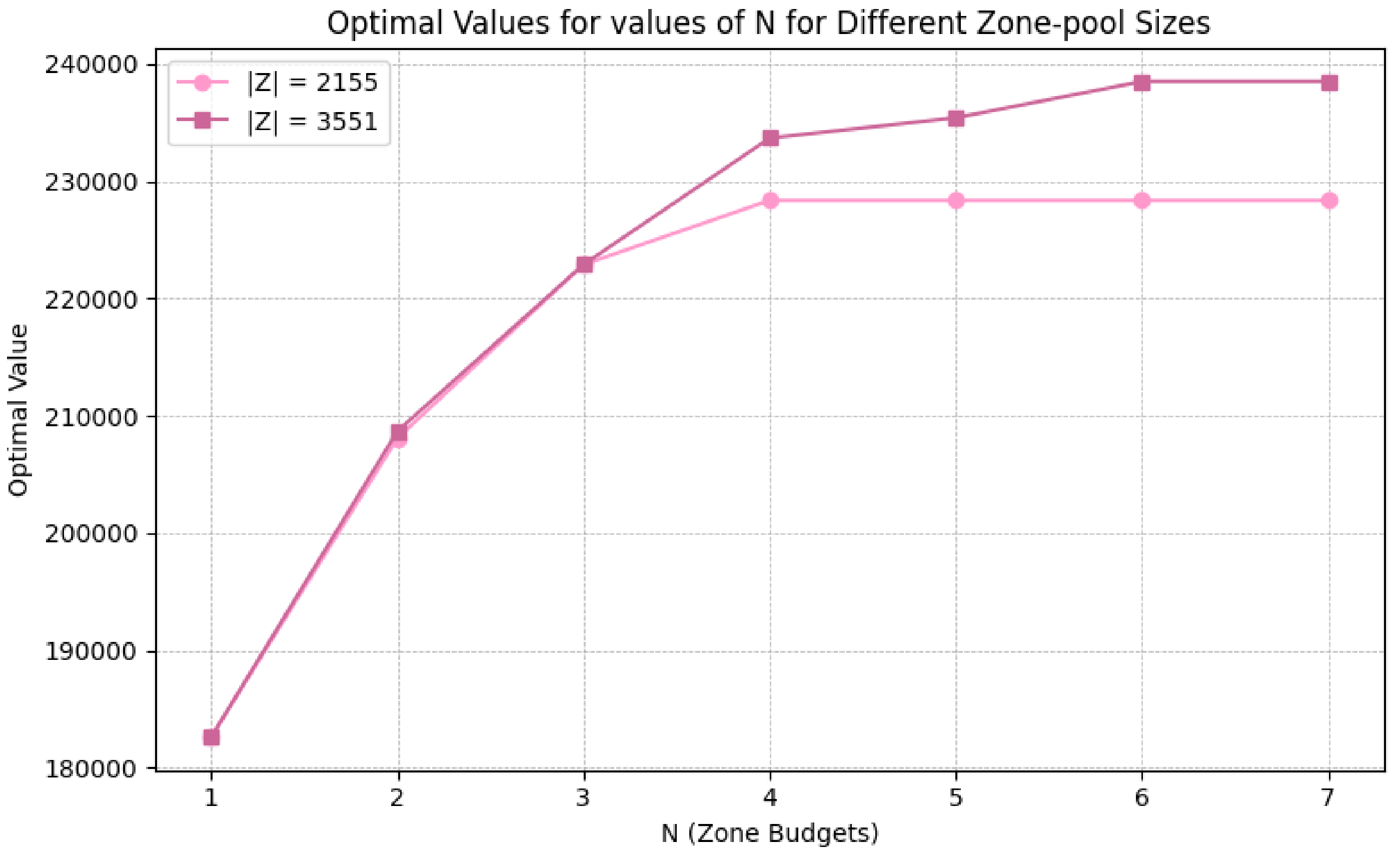
Table 8: Solution time averages for different values of N using two zone pools: 2155 (limited with $2 < |Z_i| < 10$) and 3551 (full connected set).

		ZTDP-C#	
N	C	Time [s]	Opt. [#]
1	10	<0.1	25
	30	<0.1	25
	50	0.2	25
2	10	<0.1	25
	30	1	25
	50	12.5	25
3	10	0.2	25
	30	14.7	25
	50	185.6	21
4	10	0.7	25
	30	189.7	20
	50	300	0
5	10	2.3	25
	30	300	0
	50	300	0

Computational study: Testing with limited zone-pool instances



Computational study: Testing with limited zone-pool instances



Computational study: Number of Customers

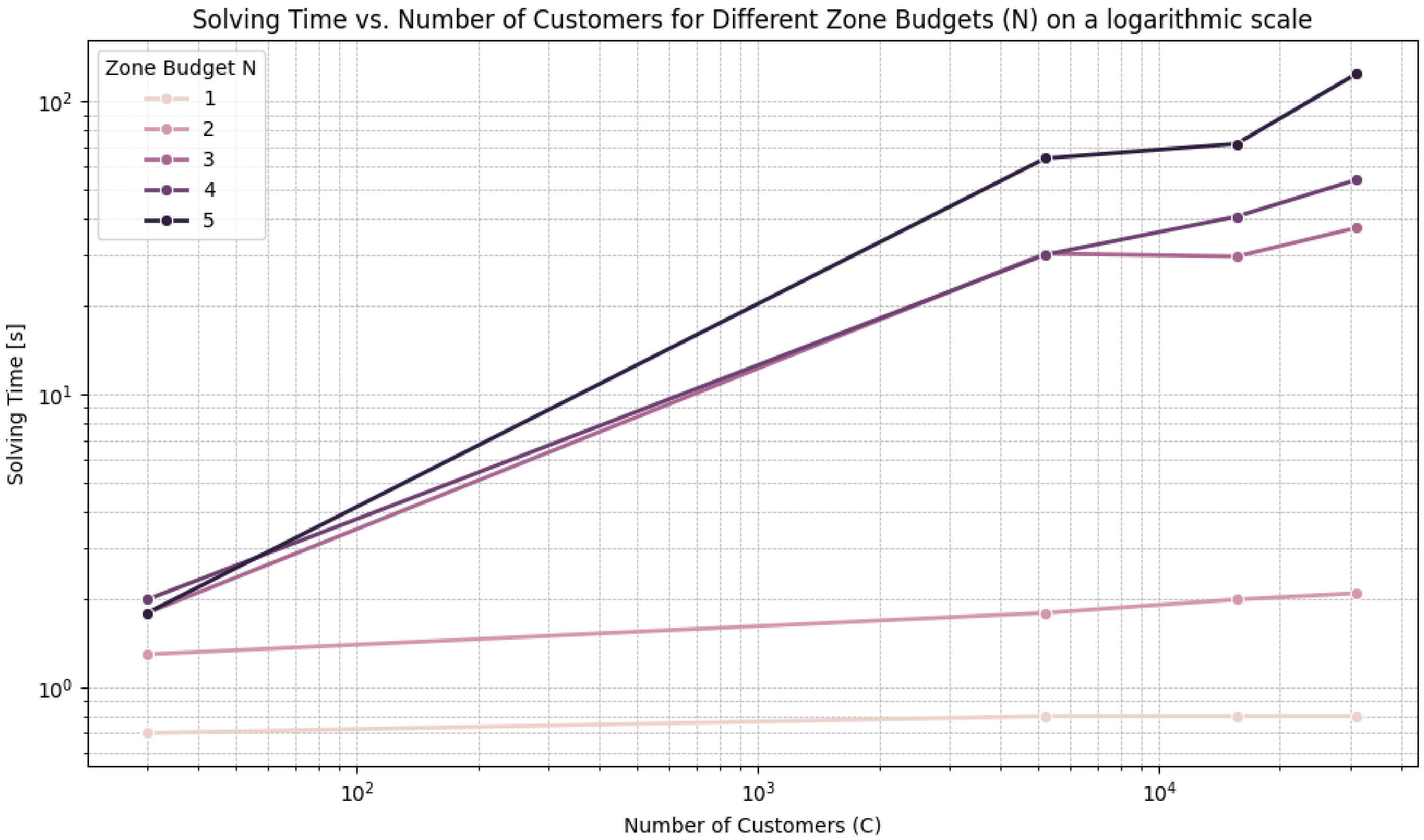
N	C	Time [s]
1	30	0.7
	5200	0.8
	15600	0.8
	31100	0.8
2	30	1.3
	5200	1.8
	15600	2.0
	31100	2.1
3	30	1.8
	5200	30.3
	15600	29.6
	31100	37.1
4	30	2.0
	5200	30.0
	15600	40.5
	31100	54.1
5	30	1.8
	5200	64.0
	15600	71.7
	31100	124.6

Table 2 Solving times (in seconds) for different customer values C and zone-budgets N, using the alternative ZTPD model with zone set size $|Z| = 3551$ (Mandl connected subgraphs).

		ZTDP-C#	
N	C	Time [s]	Opt. [#]
1	10	<0.1	25
	30	<0.1	25
	50	0.2	25
2	10	<0.1	25
	30	1	25
	50	12.5	25
3	10	0.2	25
	30	14.7	25
	50	185.6	21
4	10	0.7	25
	30	189.7	20
	50	300	0
5	10	2.3	25
	30	300	0
	50	300	0

Table 1: The results of testing the ZTPD-C# (ZTPD with connected zoning and counting zones pricing) model in article [1]

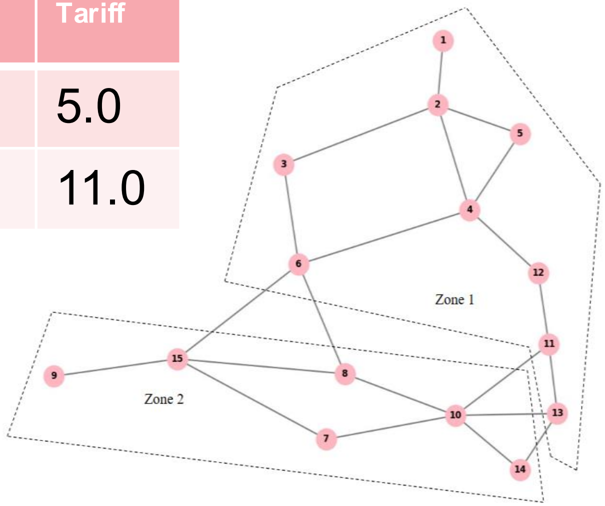
Computational study: Number of Customers



Aalto-yliopisto
Perustieteiden
korkeakoulu

Computational study: Zone design

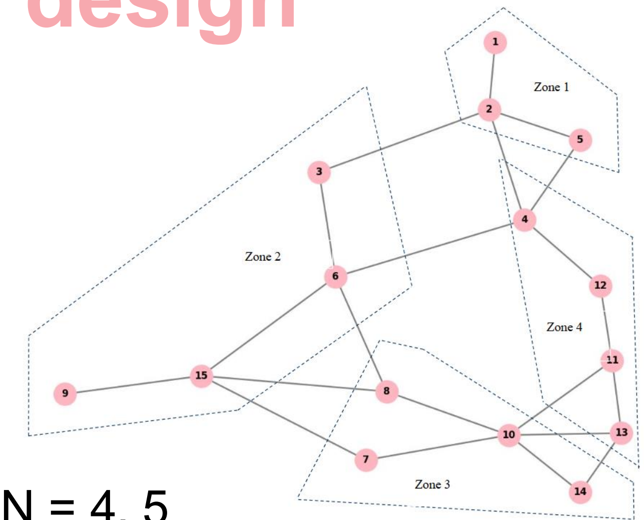
$\sum z$	Tariff
1	5.0
2	11.0



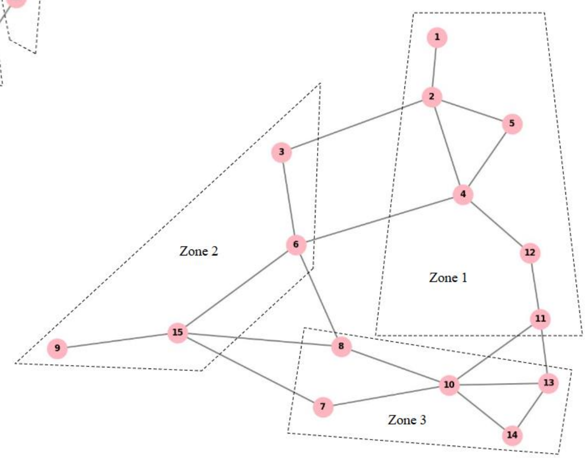
N = 2

$\sum z$	Tariff
1	5.0
2	11.0
3	19.0
4	27.0

N = 4, 5



$\sum z$	Tariff
1	5.0
2	11.0
3	19.0



N = 3

Conclusion and Further research

- Some further research could include:
  - Testing the scalability of the model by applying it to larger instances to evaluate the practical usefulness
  - Enhancing the zone-pool selection process by developing heuristics and making the process over all faster
- In conclusion:
  - An alternative version of the ZTPD model from [1] was formulated, which simplified the zone choosing process by introducing a zone-pool
  - The solving times were significantly reduced, enhancing the scalability of the optimization

Aalto-yliopisto
Perustieteiden
korkeakoulu

Thank you!

Aalto-yliopisto
Perustieteiden
korkeakoulu

Sources

1) Benjamin Otto and Nils Boysen: Zone-Based Tariff Design in Public Transportation Networks (2017)
2) P. Schiewe, A. Schöbel, S. Jäger, S. Albert, C. Biedinger, T. Dahlheimer, V. Grafe, O. Herrala, K. Hoffmann, S. Roth, A. Schiewe, M. Stinzendörfer, and R. Urban. LinTim - Integrated Optimization in Public Transportation. Homepage. https://www.lintim.net