A Series of ILP Models for the Optimization of Water Distribution Networks

NIKHIL HOODA[1,*], OM DAMANI[1] and ASHUTOSH MAHAJAN[2]

[1] Department of Computer Science and Engineering, Indian Institute of Technology, Bombay
[2] Department of Industrial Engineering and Operations Research, Indian Institute of Technology, Bombay

Abstract. The design of rural drinking water schemes consists of optimization of several network components like pipes, tanks, pumps and valves. The sizing and configuration of these network configurations needs to be such that the water requirements are met while at the same time being cost efficient so as to be within government norms. We developed the JalTantra system to design such water distribution networks. The Integer Linear Program (ILP) model used in JalTantra and described in our previous work solved the problem optimally, but took a significant amount of time for larger networks, an hour for a network with 100 nodes. In this current work we describe a series of three improvements of the model. We prove that these improvements result in tighter models, i.e. the set of points of linear relaxation is strictly smaller than the linear relaxation for the initial model. We test the series of three improved models along with the initial model over eight networks of various sizes and show a distinct improvement in performance. The 100 node network now takes only 49 seconds to solve. These changes have been implemented in JalTantra, resulting in a system that can solve the optimization of real world rural drinking water networks in a matter of seconds. The JalTantra system is free to use, and is available at https://www.cse.iitb.ac.in/jaltantra/.

MS received ; revised ; accepted

Keywords. Water Distribution, Optimization, Integer Linear Program, Pipe Diameter Selection, Tank Configuration Selection

1 Introduction

Piped water distribution networks are used to transport drinking water from common water sources to several demand areas. Therefore, the design of such networks is an important problem and has been studied in various forms over several decades.

A typical piped water distribution network, as shown in figure 1, consists of several infrastructure components like pipes, tanks, pumps and valves. The location and sizing of these components are determined as part of the network design. The network consists of one or more sources of water and several demand nodes. Each of these demand nodes are described by their elevations, demand and minimum pressure requirements. These nodes are connected by several links along which pipes have to be laid out to transport the water from the source to each of the nodes. The network layout can be looped/cyclic (typically urban) or branched/acyclic (typically rural). As the water flows through the pipes, the water pressure head reduces due to frictional losses. This loss, commonly referred to as headloss, depends on various factors like the diameter, roughness, flow and length of the pipe.

The pipe diameter selection problem consists of assigning pipe diameters to each link in the network. This selection is to be made from a discrete set of commercially available pipe diameters. Each link can be broken up into multiple segments, each consisting of pipes of different diameters or each link can be restricted to just one pipe diameter. In the most basic problem formulation other components like tanks, pumps and valves are not considered. Several approaches have been considered over the years, ranging from traditional optimization techniques like linear programming (LP) [7] [4], non linear programming (NLP) [9], integer linear programming (ILP) [8] to meta-heuristic techniques like genetic programming [10], tabu search [1], shuffled frogs [2] etc.

Networks, however, do not consist of pipes and nodes exclusively. Other components like tanks, pumps and valves are also part of any network design. Typically, rural water networks are gravity fed, i.e. water from the source flows downstream to the various demand nodes in the network. The water head along the links decreases gradually due to headloss. For certain nodes, there might not be enough water head in the system to ensure their demands are satisfied. In such cases, pumps are used to provide additional head to the system. Depending on the network configuration pumps can be installed at the source or at various points in the network, as per requirement. Though the use of pumps might decrease the cost of pipes (since pipes with smaller diameters would be required), and in some cases their use may be unavoidable, they cause a significant burden on the network operation since pumping requires electric supply. Therefore, apart from a one-time capital cost of infrastructure, there is now an additional operational cost of running the water network.

Conversely, in certain networks the source might be at

*For correspondence

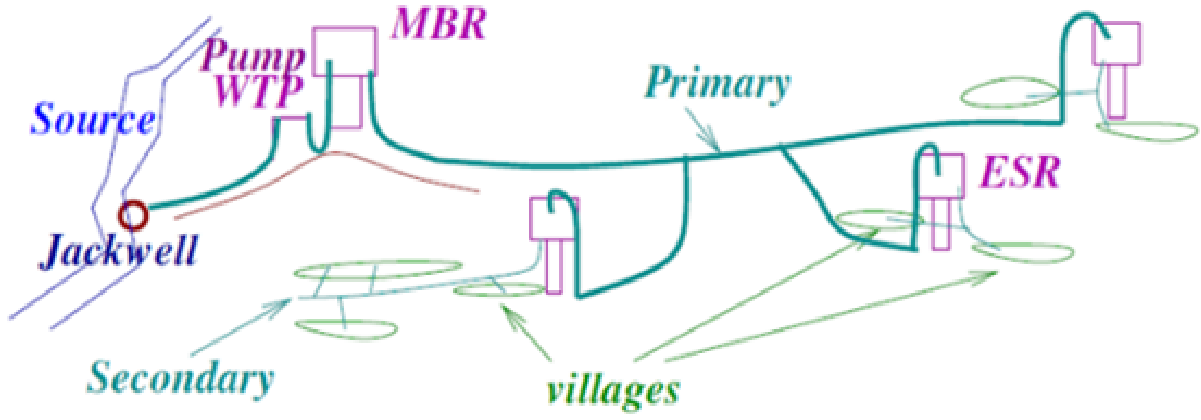

Figure 1. Components of a typical Rural Piped Water Scheme. Water is pumped from the Source to the Water Treatment Plant (WTP) and then to the Mass Balancing Reservoir (MBR). The Primary Network then transports water from the MBR to the Tanks/Elevated Storage Reservoirs (ESRs), and then finally the Secondary Network connects the Tanks/ESRs to individual villages. (courtesy: CTARA, IITB)

a significantly higher elevation than the rest of the network. This would result in excess pressure throughout the network which may cause pipes to burst. Therefore pipes with higher pressure rating would need to be installed, causing a significant increase in capital cost. In such cases, pressure reducing valves maybe employed to artificially reduce the excess pressure in the system. Valves may also be installed to restrict flow through certain pipes for maintenance and operational purposes.

Tanks help provide buffer capacity to the network. Since demand varies with time, tanks can be filled during low demand periods and provide water in times of higher demands. They can also be used to manage distribution of water and act as intermediary sources. This is particularly relevant in the case of areas where water is scarce. Tanks are filled from the source and they in turn act as secondary sources to the final demand nodes. The inflow and outflow of the tanks is managed to ensure equitable and timely distribution of water. In the absence of such a system, upstream nodes with high pressures will draw majority of the water from the source leading to insufficient supply to downstream nodes.

Pumps were the earliest component to be considered during network optimization, in addition to the selection of pipe diameters, although they were restricted to a single pump at the source [11]. Tanks and valves were incorporated within meta heuristic frameworks [1] [2] [10]. The networks considered in these studies are urban, where the role of tanks is to act as buffers to be used during periods of high demand. The choice to be made is the location, size and height of the tank. The number of tanks to be installed is fixed. But as mentioned earlier, water scarce areas use tanks as secondary sources rather than buffers. The demand nodes are partitioned and allocated to individual tanks. The source supplies water to the tanks in a primary network. The tanks in turn supply water to their allocated nodes in secondary networks. The cost of the scheme can vary significantly depending upon the number of tanks and the partition of nodes to these tanks.

Indian government bodies use software like WaterGems [14] and BRANCH [6] to design water distribution networks ([3], [12], [13], [16]). These only consist of pipe diameter optimization. Other components i.e pumps, tanks and valves are considered manually by the design engineer, relying on his/her experience and intuition [12]. We have implemented a network deign software, JalTantra, which includes tanks, pumps and valves in addition to the pipe diameter selection problem. JalTantra is a web based application that is free to use and available at https://www.cse.iitb.ac.in/jaltantra/ . JalTantra has been officially adopted by the government of Maharashtra as a tool in the design process of their drinking water schemes.

In [4] we presented the first version of JalTantra that included just the pipe diameter optimization for branched networks (typical in the case of rural areas). It used a LP model and thus solved the problem quickly and optimally. This allowed even networks of a thousand nodes to be solved in a couple of seconds. In [5] we extended the model to include tanks. The added complexity of considering both primary and secondary networks simultaneously, required an ILP model. Although still optimal in terms of cost, the time taken was significantly worse. In the present work we describe three significant improvements that were made to the model. These improvements reduced the time taken to optimize the larger networks by orders of magnitude. The time taken to optimize a 150 node network has gone from over 40 minutes to 5 seconds, and a 200 node network which could not be solved within 24 hours now takes just 70 seconds.

The improvements consist of tightening the set of constraints used to describe the ILP model. Consider the example shown in figure 2. The points represent the integer points over which we are trying to optimize. The lines represent

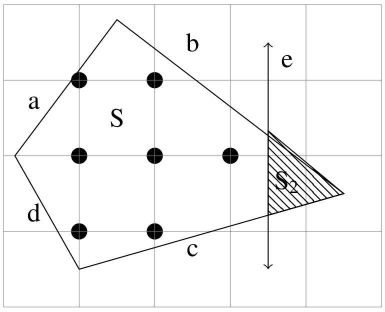

Figure 2. Constraints a, b, c and d describe the area S which represents the linear relaxation of the set of 7 integer points in two dimensions. Introducing the constraint e cuts off the area S_2 from the linear relaxation while still maintaining the same set of integer points.

the constraints that encompass those integer points. When solving the linear point (LP) relaxation, the entire set S is considered. By introducing the constraint e, we can still capture the same integer points while cutting off a part (S_2) of the linear relaxation. Since a smaller solution space is now considered while solving the LP relaxation, this speeds up the optimization. For each of the three improvements presented, we prove that the newer set of constraints have a linear relaxation that is a strict subset of the linear relaxation of the older set, while maintaining the same set of integer points. In particular, for the tank configuration improvement we show that the newer subset of constraints is as tight as possible, i.e. the linear relaxation has no fractional points. Since the overall model is complex, while discussing each improvement, we only consider a small subset of relevant constraints at a time.

The rest of the paper is structured as follows: In section 2 we describe the optimization problem formulation and the initial model used to solve the problem. In sections 3, 4 and 5 we describe the three improvements. For each improvement, we first repeat the relevant subset of constraints from the initial model, then provide the new set of constraints of the improved model and then finally prove that the improved model is strictly better than the initial model. In section 6 we describe an initial attempt at an alternative edge based approach to modelling the problem. The initial model and the three post improvement models were tested on eight networks of varying sizes. Section 7 provides the performance details of these tests. Finally we provide our concluding thoughts in section 8.

2 Initial Model (Model 1)

As discussed above, drinking water distribution networks consists of various components. To optimize the cost of such networks, several inputs must be considered and for each component several parameters must be determined. We first explicitly formulate the problem that we are attempting to solve. Then in section 2.2 we provide details of the initial ILP model used to solve the problem.

2.1 *Problem Formulation*

Input:

- General: primary/secondary supply hours, minimum/maximum headloss per km, maximum water speed
- Source node: head
- Node: elevation, water demand, minimum pressure requirement
- Link: start/end node, length
- Existing Pipes: start/end node, length, diameter, parallel allowed, roughness
- Commercial pipe diameter: roughness, cost per unit length
- Tanks: maximum tank heights, tank capacity factor, nodes that must/must not have tanks, capital cost table
- Pumps: minimum pump size, efficiency, design lifetime, capital/energy cost, discount/interest rate, pipes that cannot have pumps
- Valves: location, pressure rating

Output:

- Length and diameter of pipe segments for each link
- Partitioning the set of links into primary and secondary network
- Location, height and size of Tanks
- Set of nodes being served by each Tank
- Location and power of Pumps

Objective:

- Minimize total capital cost (pipe + tank + pump) and total energy cost (pump)

Constraints:

- Pressure at each node must exceed minimum pressure specified
- Water demand must be met at each node

2.2 *Model Details*

The pipe diameter selection in the model is represented by the continuous variable l_{ij} which represents the length of the j^{th} pipe diameter component of the i^{th} link in the network. This determines the capital cost of the pipes. The tank allocation is represented by the binary variable s_{nm} which is true if the tank at the n^{th} node in the network provides water to the m^{th} node in the network. The choice of tank allocation variables, fixes the total demand that each tank serves i.e. the variable d_n. This in turn determines the capital cost of the tanks. Apart from the cost considerations, each node n must also have its minimum pressure constraint satisfied. The head at each node, h_n is dependent on the headloss hl_i in the links of the network. This headloss depends on the pipe variables l_{ij} and the tank variables s_{nm} mentioned earlier. In addition, the introduction of pumps/valves increases/decreases the headloss respectively. The details of the parameters, variables, objective function and constraints of the model are as follows:

Parameters:

- NL: Number of links in the network
- NP: Number of commercial pipe diameters
- D_j: Diameter of j^{th} commercial pipe diameter
- C_j: Cost per unit length of j^{th} commercial pipe diameter
- NN: Number of nodes in the network
- NE: Number of rows in the tank cost table
- B_k: Base cost of the k^{th} row of the tank cost table
- UN_k: Unit cost of the k^{th} row of the tank cost table
- UP_k: Upper limit capacity for the k^{th} row of the tank cost table
- LO_k: Lower limit capacity for the k^{th} row of the tank cost table
- CP: Capital cost of pumps per unit kW
- EP: Energy cost of pumps per unit kWh
- DF: Discount factor for the energy cost over the entire scheme lifetime
- PH: Number of hours of water supply in the primary network
- SH: Number of hours of water supply in the secondary network
- Y: Lifetime of scheme in years
- $INFR$: Inflation rate
- $INTR$: Interest rate
- L_i: Length of the i^{th} link
- P_n: Minimum pressure required at node n
- E_n: Elevation of the n^{th} node
- DE_n: Water demand of the n^{th} node
- DE: The total water demand of the network
- VH_i: Head reduction by valve in i^{th} link
- HL^p_{ij}: Headloss for the j^{th} diameter of the i^{th} link, if i is part of the primary network
- HL^s_{ij}: Headloss for the j^{th} diameter of the i^{th} link, if i is part of the secondary network
- FL^p_i: Flow in i^{th} link if i is part of the primary network
- FL^s_i: Flow in i^{th} link if i is part of the secondary network
- R_j: Roughness of j^{th} commercial pipe diameter
- T_{min}: Minimum tank height allowed
- T_{max}: Maximum tank height allowed
- ρ: Density of water
- g: Acceleration due to gravity
- η: Efficiency of pump
- PP_{min}: Minimum pump power allowed
- PP_{max}: Maximum pump power allowed
- A_n: Set of nodes that are ancestors of node n
- D_n: Set of nodes that are descendants of node n
- C_n: Set of nodes that are children of node n
- P_n: Parent node of node n
- I_n: Incoming link for node n
- O_n: Set of outgoing links from node n

Continuous Variables:

- l_{ij}: Length of the j^{th} pipe component of the i^{th} link
- l^p_{ij}: Length of the j^{th} pipe component of the i^{th} link, if link i is part of the primary network
- hl_i: Total headloss across link i
- d_n: Total demand served by tank at node n
- z_{nk}: Total demand served by tank at node n, if costed by the k^{th} row of the tank cost table
- p_i: Power of pump installed at link i
- p^p_i: Power of pump installed at link i, if link i is part of primary network
- p^s_i: Power of pump installed at link i, if link i is part of secondary network
- ph_i: Head provided by pump at link i
- h_n: Water head at node n
- t_n: Height of tank at node n
- h'_{ni}: Effective head provided to link i by its starting node n

Binary Variables:

- e_{nk}: 1 if tank at n^{th} node is costed by the k^{th} row of tank cost table
- f_i: 1 if link i is part of the primary network, 0 if part of the secondary network
- es_{ni}: 1 if source for link i is its starting node n
- s_{nm}: 1 if tank at node n is source for node m
- pe_i: 1 if a pump is installed at link i

Objective Function: The objective function is simply the sum of capital cost of the pipes, tanks, pumps and valves used in the network. In addition, we also have the operational cost of the pumps. This operational cost is computed as the present value of the total cost over the scheme lifetime.

$$\begin{aligned} O(.) = & \sum_{i=1}^{NL}\sum_{j=1}^{NP} C_j(D_j) l_{ij} \quad + \quad \sum_{i=1}^{NL} CP * p_i \\ + & \sum_{n=1}^{NN}\sum_{k=1}^{NE} e_{nk} * (B_k + UN_k * (d_n - LO_k)) \qquad (1) \\ + & \; EP * DF * \left(\sum_{i=1}^{NL} PH * p_i^p + \sum_{i=1}^{NL} SH * p_i^s \right) \end{aligned}$$

$$where \quad DF = \sum_{n=1}^{Y} \left(\frac{1 + INFR}{1 + INTR} \right)^{n-1}$$

Constraints:

- The total length of the pipe diameter segments must equal to the total link length:

$$L_i = \sum_{j=1}^{NP} l_{ij}, \quad i = 1 \ldots NL \qquad (2)$$

- The pressure at each node must exceed the minimum pressure required:

$$P_n \leq h_n - (E_n + t_n), \quad n = 1 \ldots NN \qquad (3)$$

- Across every link i there is headloss hl_i. This headloss depends on the flow, length and diameter of the pipe diameter component. We use the Hazen-Williams equation [15] to calculate the headloss. The headloss across a link also depends on the pump and valve installed across it, if any. The valves are simply input parameters to the model since they are manually fixed. The constraints related to the pump head ph_i are described further below. The flow through the link depends on whether the link is part of the primary or secondary network:

$$hl_i = \sum_{j=1}^{NP} (HL_{ij}^p l_{ij}^p + HL_{ij}^s (l_{ij} - l_{ij}^p)) - ph_i + VH_i, \quad i = 1 \ldots NL \qquad (4)$$

$$HL_{ij}^p = \frac{10.68 * \left(\frac{FL_i^p}{R_j}\right)^{1.852}}{D_j^{4.87}}, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP \qquad (5)$$

$$HL_{ij}^s = \frac{10.68 * \left(\frac{FL_i^s}{R_j}\right)^{1.852}}{D_j^{4.87}}, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP \qquad (6)$$

$$FL_i^s = FL_i^p * \frac{PH}{SH}, \quad i = 1 \ldots NL \qquad (7)$$

- The head h_n at each node n is calculated by the effective head h'_{mi} provided by its parent node m and the headloss hl_i across the link connecting two nodes. The effective head in turn depends on whether the link i has the tank at the starting node m as its source. This is represented by the Boolean variable es_{mi}:

$$h_n = h'_{mi} - hl_i, \quad n = 1 \ldots NN, \quad m = P_m, \quad i = I_n \qquad (8)$$

$$h'_{mi} = (t_m + E_m) * es_{mi} + h_m * (1 - es_{mi}), \quad m = 1 \ldots NN, \quad i \in O_m \qquad (9)$$

$$es_{mi} = s_{mm} * (1 - f_i), \quad m = 1 \ldots NN, \quad i \in O_m \qquad (10)$$

- Next, we look at the constraints relating to the tank allocation. The first tank constraint is to ensure that every tank height is between parameters T_{min} and T_{max}.

$$T_{min} \leq t_n \leq T_{max} \qquad (11)$$

- We then look at the constraints that deal with allocation of demand nodes to tanks. s_{nm} is 1 if tank at node n serves the demand of node m. If a node n does not serve its own demand i.e. it is part of a secondary network, then all its downstream nodes will also be part of a secondary network.

$$s_{mm} \leq s_{nn}, \quad n = 1 \ldots NN, \quad m \in D_n \qquad (12)$$

- If a node n does not serve its own demand, then it cannot serve the demand of its downstream nodes.

$$s_{nm} \leq s_{nn}, \quad n = 1 \ldots NN, \quad m \in D_n \qquad (13)$$

- For every node n, only one upstream node m can serve its demand.

$$\sum_m s_{mn} = 1, \quad n = 1 \ldots NN, \quad m \in A_n \cup \{n\} \qquad (14)$$

- The total demand d_n served by node n is the sum of the demands of the downstream nodes that it serves i.e. all m such that $s_{nm} = 1$.

$$d_n = \sum_m s_{nm} * DE_m, \quad n = 1 \ldots NN, \quad m \in D_n \cup \{n\} \qquad (15)$$

- For a node n, its incoming pipe i will have primary flow only if the node serves itself.

$$f_i = s_{nn}, \quad n = 1 \ldots NN, \quad i = I_n \tag{16}$$

- If a node n serves node m i.e. $s_{ij} = 1$, each node o in the path from n to m belongs to a secondary network and therefore cannot serve itself.

$$s_{nm} \leq 1 - s_{oo}, n = 1 \ldots NN, m \in D_n, o \in D_n \cup A_m \tag{17}$$

- Next, we have the constraints that relate the demand that a tank serves to its cost variables e_{nk}. Note that we require z_{nk} in our objective function:

$$z_{nk} = e_{nk} \times d_n, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE \tag{18}$$

- Since every tank can be costed using exactly one row, the sum of e_{nk} for a given n must be 1:

$$\sum_{k=1}^{NE} e_{nk} = 1, \quad n = 1 \ldots NN \tag{19}$$

- Next we have constraints that make sure that the tank capacity d_n lies between the minimum and maximum capacity of the selected row of the cost table:

$$For \quad n = 1 \ldots NN, \quad k = 1 \ldots NE :$$

$$LO_k e_{nk} \leq d_n \tag{20}$$

$$DE * e_{nk} + d_n \leq UP_k + DE \tag{21}$$

- Next, we look at constraints related to pumps. The pump power p_i relates to the pump head ph_i in the following way:

$$p_i = \quad p_i^p + p_i^s, \quad i = 1 \ldots NL, \tag{22}$$

$$p_i^p = \frac{(\rho * g * FL_i^p * ph_i)}{\eta} * f_i, \quad i = 1 \ldots NL, \tag{23}$$

$$p_i^s = \frac{(\rho * g * FL_i^s * ph_i)}{\eta} * (1 - f_i), \ i = 1 \ldots NL \tag{24}$$

- Finally, the pump power for each pump must lie between minimum and maximum allowed pump power. This is implemented using the binary variable pe_i.

$$PP_{min} * pe_i \leq p_i \leq PP_{max} * pe_i, \quad i = 1 \ldots NL \tag{25}$$

This completes the description of the initial model. Although this model provides optimal results in terms of capital cost, the time taken to solve networks rises rapidly with increased network size. In the next three sections we go over three improvements made iteratively to this initial model. For each improvement, we first describe the subset of variables and constraints from the initial model that are being considered. We next provide the improved set of constraints. Finally, we prove how the linear relaxation of the improved set is a strict subset of the linear relaxation of the initial set.

3 Pipe Headloss Improvement

3.1 *Initial Model*

We describe a part of the model whose purpose is to determine the pipe diameters chosen for each link in the network. Each link can consist of multiple pipe diameters. Also, each link can be part of the primary network or the secondary network. The headloss across the link depends on these choice of pipe diameters and whether it belongs to the primary or secondary network. The set of variables and data used for this purpose are defined as follows. Consider a network of NL links. Let NP be the number of pipe diameters available.

Variables:

l_{ij} = length of the j^{th} pipe diameter component of link i, $\quad i = 1 \ldots NL, \quad j = 1 \ldots NP,$

l_{ij}^p = length of the j^{th} pipe diameter component of link i, if link i is part of the primary network, $\quad i = 1 \ldots NL,$ $j = 1 \ldots NP,$

hl_i = headloss across link i, $\quad i = 1 \ldots NL,$

$f_i = 1$ of link i is part of the primary network, 0 if it is part of the secondary network, $\quad i = 1 \ldots NL.$

Data:

L_i = Length of link i, $\quad i = 1 \ldots NL,$

HL_{ij}^p = Unit headloss for the j^{th} pipe diameter component of link i, if i is part of primary network, $\quad i = 1 \ldots NL, \quad j = 1 \ldots NP,$

HL_{ij}^s = Unit headloss for the j^{th} pipe diameter component of link i, if i is part of secondary network, $\quad i = 1 \ldots NL, \quad j = 1 \ldots NP.$

Constraints:

The first constraint captures l_{ij}^p as a product of l_{ij} and f_i:

$$l_{ij}^p = l_{ij} \times f_i, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{26}$$

Equation (26) consists of a product of two variables, and is therefore a non-linear equation. Fortunately since f_i is a binary variable, we can linearize the equation using the following inequalities:

$$0 \leq l_{ij}^p, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{27}$$

$$l_{ij}^p \leq L_i f_i, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{28}$$

$$l_{ij} - L_i(1 - f_i) \leq l_{ij}^p, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{29}$$

$$l_{ij}^p \leq l_{ij}, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{30}$$

The sum of all pipe diameter components must equal the link length:

$$\sum_{j=1}^{NP} l_{ij} = L_i, \quad i = 1 \ldots NL, \tag{31}$$

Next we have the equation for hl_i, which is the sum all headloss components contributed by the different pipe diameter components of link i :

$$hl_i = \sum_{j=1}^{NP} P_{ij} l_{ij}^p + \sum_{j=1}^{NP} S_{ij}(l_{ij} - l_{ij}^p), \quad i = 1 \ldots NL. \tag{32}$$

Finally we have constraints that relate to the bounds for the variables:

$$l_{ij} \geq 0, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{33}$$

$$f_i \in \{0, 1\}, \quad i = 1 \ldots NL, \tag{34}$$

Since there exists a l_{ij}^p for each link and pipe diameter combination in the network, a large number of linear decompositions of equation (26) need to be done. In the next section we show an improved model that has the same feasible 0-1 set of values but with a tighter LP bound resulting in better performance.

3.2 *Improved Model* ***(Model 2)***

In order to decompose the product of variables in (26), a large number of constraints needed to be added. This is avoided in the new model by not explicitly defining l_{ij}^p. Instead its relation to l_{ij} and f_i is implicit. In the next section we show that new model is better, in that it has a tighter LP bound than the old model.

Variables:
We introduce one new variable, which is similar to l_{ij}^p but for the secondary network:
l_{ij}^s = length of the j^{th} pipe diameter component of link i, if link i is part of the secondary network, and 0 if link i is part of the primary network $\quad i = 1 \ldots NL, \quad j = 1 \ldots NP.$

Constraints:
The first constraint simply states that l_{ij} is the sum of the primary and secondary components, i.e l_{ij}^p and l_{ij}^s respectively:

$$l_{ij} = l_{ij}^p + l_{ij}^s, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{35}$$

For a given link i, either all l_{ij}^p are 0 or all l_{ij}^s are 0, depending on the value of f_i. And the sum of the non-zero components must equal the length of the link L_i. The first two inequalities of the new model capture this:

$$\sum_{j=1}^{NP} l_{ij}^p = L_i f_i, \quad i = 1 \ldots NL, \tag{36}$$

$$\sum_{j=1}^{NP} l_{ij}^s = L_i(1 - f_i), \quad i = 1 \ldots NL. \tag{37}$$

Next we have the equation for hl_i, which is the sum all headloss components contributed by the different pipe diameter components of link i. For the new model we equivalently use l_{ij}^s instead of $l_{ij} - l_{ij}^p$ due to equation (35):

$$hl_i = \sum_{j=1}^{NP} P_{ij} l_{ij}^p + \sum_{j=1}^{NP} S_{ij} l_{ij}^s, \quad i = 1 \ldots NL. \tag{38}$$

Finally as before we have the bounds for the variables:

$$l_{ij} \geq 0, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{33}$$

$$l_{ij}^p \geq 0, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{27}$$

$$l_{ij}^s \geq 0, \quad i = 1 \ldots NL, \quad j = 1 \ldots NP, \tag{39}$$

$$f_i \in \{0, 1\}, \quad i = 1 \ldots NL. \tag{34}$$

We now prove that the improved model is tighter than the initial model, that is the linear relaxation of the improved model is a strict subset of the linear relaxation of the initial model. Let S_1 be the set of points belonging to the initial model and S_2 be the set of points belonging to the improved model. Let R_1 and R_2 be the set of points corresponding to the LP relaxations of S_1 and S_2 respectively. Both R_1 and R_2 are defined by the same set of constraints that describe the initial sets S_1 and S_2, except for the constraint (34) which refers to the binary nature of f_i. Instead, the continuous bounds for f_i is defined as follows:

$$0 \leq f_i \leq 1, \quad i = 1 \ldots NL. \tag{40}$$

Proposition 1. *R_2 is a strict subset of R_1 i.e. $R_2 \subset R_1$.*

We prove R_2 is a strict subset of R_1 in two steps. First we show that R_2 is a subset of R_1 and then we show that R_2 is not equal to R_1.

Claim 1.1. *$R_2 \subseteq R_1$, i.e. for every point P, $P \in R_2 \Rightarrow P \in R_1$.*

Proof. Consider a point $P \in R_2$. It satisfies the constraints (27), (33) and (35) - (40). We prove that it also lies in R_1 by showing that it satisfies the constraints (27)-(33) and (40). Constraints (27), (33) and (40) are trivially satisfied since they are common for both sets.

$$For \quad i = 1 \ldots NL, \quad j = 1 \ldots NP$$

Proving (28) : $\quad l_{ij}^p \leq L_i f_i$

$$\sum_{j=1}^{NP} l_{ij}^p = L_i f_i \tag{36}$$

$\equiv \quad$ {using $l_{ij}^p \geq 0$ (27)}

$l_{ij}^p \leq L_i f_i$

Hence satisfied.

Proving (29) : $\quad l_{ij} - L_i(1 - f_i) \leq l_{ij}^p$

$$\sum_{j=1}^{NP} l_{ij}^s = L_i(1 - f_i) \tag{37}$$

$\Rightarrow \quad$ {using $l_{ij}^s \geq 0$ (39)}

$l^s_{ij} \leq L_i(1 - f_i)$

$\equiv$ {using $l_{ij} = l^p_{ij} + l^s_{ij}$ (35)}

$l_{ij} - l^p_{ij} \leq L_i(1 - f_i)$

$\equiv$ {rearranging}

$l_{ij} - L_i(1 - f_i) \leq l^p_{ij}$

Hence satisfied.

Proving (30) : $l^p_{ij} \leq l_{ij}$

$$0 \leq l^s_{ij} \tag{39}$$

$\equiv$ {using $l_{ij} = l^p_{ij} + l^s_{ij}$ (35)}

$0 \leq l_{ij} - l^p_{ij}$

$\equiv$ {rearranging}

$l^p_{ij} \leq l_{ij}$

Hence satisfied.

Proving (31) : $\sum_{j=1}^{NP} l_{ij} = L_i$

$$\sum_{j=1}^{NP} l^p_{ij} = L_i f_i \tag{36}$$

$$\sum_{j=1}^{NP} l^s_{ij} = L_i(1 - f_i) \tag{37}$$

$\equiv$ {adding equations}

$$\sum_{j=1}^{NP} (l^p_{ij} + l^s_{ij}) = L_i f_i + L_i(1 - f_i)$$

$\equiv$ {using $l_{ij} = l^p_{ij} + l^s_{ij}$ (35) and simplifying}

$$\sum_{j=1}^{NP} l_{ij} = L_i$$

Hence satisfied.

Proving (32) : $hl_i = \sum_{j=1}^{NP} P_{ij} l^p_{ij} + \sum_{j=1}^{NE} S_{ij}(l_{ij} - l^s_{ij})$

$$hl_i = \sum_{j=1}^{NP} P_{ij} l^p_{ij} + \sum_{j=1}^{NP} S_{ij} l^s_{ij} \tag{38}$$

$\equiv$ {using $l_{ij} = l^p_{ij} + l^s_{ij}$ (35)}

$$hl_i = \sum_{j=1}^{NP} P_{ij} l^p_{ij} + \sum_{j=1}^{NE} S_{ij}(l_{ij} - l^s_{ij})$$

Hence satisfied.

Therefore point $P \in R_1$, since it satisfies the constraints (27)-(33) and (40). Therefore $R_2 \subseteq R_1$. □

Claim 1.2. *There exists a point Q such that $Q \in R_1$ and $Q \notin R_2$.*

Proof. Take point $Q(l, l^p, l^s, hl, f) = ([L/2, L/2], [L/2, L/2], [0, 0], L, 1/2)$. Here $(n, m) = (1, 2)$ and $(L, P, S) = (L, [1, 1], [1, 1])$ where $L \geq 0$. We show $Q \in R_1$ since it satisfies all the constraints.

27 :

$0 \leq l^p_{11}$, replacing values we get $0 \leq L/2$

$0 \leq l^p_{12}$, replacing values we get $0 \leq L/2$

28 :

$l^p_{11} \leq L_1 f_1$, replacing values we get $L/2 \leq L/2$

$l^p_{12} \leq L_1 f_1$, replacing values we get $L/2 \leq L/2$

29 :

$l_{11} - L_1(1 - f_1) \leq l^p_{11}$, replacing values we get $L/2 - L(1 - 1/2) \leq L/2$

$l_{12} - L_1(1 - f_1) \leq l^p_{12}$, replacing values we get $L/2 - L(1 - 1/2) \leq L/2$

30 :

$l^p_{11} \leq l_{11}$, replacing values we get $L/2 \leq L/2$

$l^p_{12} \leq l_{12}$, replacing values we get $L/2 \leq L/2$

31 :

$l_{11} + l_{12} = L_1$, replacing values we get $L/2 + L/2 = L$

32 :

$hl_1 = P_{11} l^p_{11} + P_{12} l^p_{12} + S_{11}(l_{11} - l^p_{11}) + S_{12}(l_{12} - l^p_{12})$, replacing values we get $L = L/2 + L/2 + 0 + 0$

33 :

$l_{11} \geq 0$, replacing values we get $L/2 \geq 0$

$l_{12} \geq 0$, replacing values we get $L/2 \geq 0$

40 :

$0 \leq f_1 \leq 1$, replacing values we get $0 \leq 1/2 \leq 1/2$

Therefore point $Q \in R_1$. To show that $Q \notin R_2$ consider equation 37:

$l^s_{11} + l^s_{12} = L_1(1 - f_1)$, replacing values we get $0 + 0 = L/2$

Therefore $Q \notin R_2$. □

Claim 1.1 shows that $R_2 \subseteq R_1$. Claim 1.2 shows that $R_2 \neq R_1$. These two together imply $R_2 \subset R_1$. Proposition 1 therefore shows that the LP relaxation of the new model (R_2) has a tighter bound than the LP relaxation of the old model (R_1).

4 Tank Cost Improvement

4.1 *Initial Model*

We describe a part of the model whose purpose is to determine the capital cost of each tank in the network. Since the tank cost is a piece-wise linear function, we need to determine which row in the tank cost table does the tank capacity fall in. Each row in the tank cost table has minimum and maximum capacity values. If the tank capacity is within these values, then that row is used to compute the tank's cost. Binary variables are used to capture for each tank, the row in the cost table that is chosen to compute the cost. The set of variables and data used for this purpose are defined as follows. Consider a network of n locations. Let m be the number of linear components of the piecewise linear cost of construction of a tank.

Variables:

e_{nk} = 1 if the tank at location n is costed using the k^{th} row of the tank cost table, $n = 1 \ldots NN$, $k = 1 \ldots NE$,

z_{nk} = capacity of the tank at location n if it is costed using the k^{th} row of the tank cost table, 0 otherwise, $n = 1 \ldots NN$, $k = 1 \ldots NE$,

d_n = capacity of the tank at location n, $n = 1 \ldots NN$.

Data:

LO_k = minimum capacity that the k^{th} row of the tank cost table can satisfy, $k = 1 \ldots NE$,

UP_k = maximum capacity that the k^{th} row of the tank cost table can satisfy, $k = 1 \ldots NE$,

DE = value of the total water demand in the network.

Constraints:

The first constraint relates the tank capacity corresponding to the k^{th} row as a product of the tank capacity and the binary choice variable e_{nk} :

$$z_{nk} = e_{nk} \times d_n, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{41}$$

Since equation (41) consists of a product of two variables, it is a non-linear equation. Fortunately since e_{nk} is a binary variable, we can linearize the equation using the following inequalities:

$$0 \le z_{nk}, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{41.a}$$

$$z_{nk} \le DEe_{nk}, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{41.b}$$

$$d_n - DE(1 - e_{nk}) \le z_{nk}, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{41.c}$$

$$z_{nk} \le d_n, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{41.d}$$

Since every tank can be costed using exactly one row, the sum of e_{nk} for a given n must be 1:

$$\sum_{k=1}^{NE} e_{nk} = 1, \quad n = 1 \ldots NN, \tag{42}$$

Next we have constraints that make sure that the tank capacity d_n lies between the minimum and maximum capacity of the selected row of the cost table:

$$LO_k e_{nk} \le d_n, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{43}$$

$$DEe_{nk} + d_n \le UP_n + D, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{44}$$

Finally we have constraints that relate to the bounds for the variables:

$$DE \ge d_n, \quad n = 1 \ldots NN, \tag{45}$$

$$DE \ge UP_k, \quad k = 1 \ldots NE. \tag{46}$$

$$e_{nk} \in \{0, 1\}, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{47}$$

$$d_n \ge 0, \quad n = 1 \ldots NN. \tag{48}$$

Since there exists a z_{nk} for each tank and row of cost table combination, a large number of linear decompositions of equation (41) need to be done. This results in poor performance of the model. In the next section we show an improved model that has the same feasible 0-1 set of values but with a tighter LP bound resulting in better performance. We then show that the improved model has a tighter bound.

4.2 *Improved Model **(Model 3)***

As discussed in the previous section, the principal issue with the old model was equation (41), where $z_n k$ is expressed as a product of two variables. In order to decompose the variables, a large number of constraints needed to be added. This is avoided in the new model by not explicitly defining z_{nk}. Instead its relation to e_{nk} and d_n is implicit. In the next section we first show that new model is better in that it has a tighter LP bound than the old model and then we go on to show that the LP for the new model has tight solutions.

The variables remain same for the new model. The first two inequalities of the model provide the bounds for z_{nk} in terms of e_{nk} and the minimum(LO_k) and maximum(UP_k) capacities for each row of the cost table:

$$LO_k e_{nk} \le z_{nk}, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{49}$$

$$z_{nk} \le UP_k e_{nk}, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{50}$$

The next equation for the model remains unchanged, it represents the fact that each row of the cost table is chosen exactly once for each tank:

$$\sum_{k=1}^{NE} e_{nk} = 1, \quad n = 1 \ldots NN, \tag{42}$$

Next, we have a similar equation but this time related to the variable z_{nk}. The sum of all z_{nk} values for a given tank must equal d_n:

$$\sum_{k=1}^{NE} z_{nk} = d_n, \quad n = 1 \ldots NN, \tag{51}$$

In fact along with the previous three equations of the model, one can imply that exactly one of the z_{nk} values will be non zero for a specific tank and therefore will be equal to d_n. This therefore captures the non-linear constraint that equation 41 of the old model captured. The remaining constraints relate to the bounds for the variables:

$$DE \geq d_n, \quad n = 1 \ldots NN, \tag{45}$$
$$DE \geq UP_k, \quad k = 1 \ldots NE. \tag{46}$$
$$e_{nk} \in \{0, 1\}, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE, \tag{47}$$
$$d_n \geq 0, \quad n = 1 \ldots NN, \tag{48}$$
$$z_{nk} \geq 0, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE. \tag{41.a}$$

Let S_1 be the set of points belonging to the old model and S_2 be the set of points belonging to the new model. Let R_1 and R_2 be the set of points corresponding to the LP relaxations of S_1 and S_2 respectively. Both R_1 and R_2 are defined by the same set of constraints that describe the initial sets S_1 and S_2, except for the constraints (47) which refers to the binary nature of e_{nk}. Instead, the continuous bounds for e_{nk} is defined as follows:

$$0 \leq e_{nk} \leq 1, \quad n = 1 \ldots NN, \quad k = 1 \ldots NE. \tag{52}$$

Similar to section 3, we prove that the LP relaxation of the new model is tighter than the LP relaxation of the old model. We do so by showing that R_2 is a strict subset of R_1. We then go on to show that R_2 has no fractional corner points and thus cannot be tightened further.

Proposition 2. *R_2 is a strict subset of R_1 i.e. $R_2 \subset R_1$*

As in section 3, we prove R_2 is a strict subset of R_1 in two steps. First we show that R_2 is a subset of R_1 and then we show that R_2 is not equal to R_1.

Claim 2.1. *$R_2 \subseteq R_1$, i.e. for every point P, $P \in R_2 \Rightarrow P \in R_1$.*

Proof. Consider a point $P \in R_2$. It satisfies the constraints (41.a), (42) and (45)-(52). We prove that it also lies in R_1 by showing that it satisfies the constraints (41.a)-(48) and (52). Constraints (41.a), (42), (45)-(48) and (52) are trivially satisfied since they are common for both sets.

$$For \quad n = 1 \ldots NN, \quad k = 1 \ldots NE$$

Proving $(41.b)$: $\quad z_{nk} \leq DEe_{nk}$

$$z_{nk} \leq UP_k e_{nk} \tag{50}$$

$\Rightarrow$ {using $DE \geq UP_k$ (46)}

$$z_{nk} \leq DEe_{nk}$$

Hence satisfied.

Proving $(41.c)$: $\quad d_n - DE(1 - e_{nk}) \leq z_{nk}$

$$\sum_{k'=1}^{NE} z_{nk'} = d_n \tag{51}$$

$\equiv$ {splitting sum}

$$z_{nk} + \sum_{k'=1, k' \neq k}^{NE} z_{nk'} = d_n$$

$\equiv$ {rearranging}

$$\sum_{k'=1, k' \neq k}^{NE} z_{nk'} = d_n - z_{nk} \tag{53}$$

$$\sum_{k'=1}^{NE} e_{nk'} = 1, \tag{42}$$

$\equiv$ {splitting sum}

$$e_{nk} + \sum_{k'=1, k' \neq k}^{NE} e_{nk'} = 1$$

$\equiv$ {rearranging}

$$\sum_{k'=1, k' \neq k}^{NE} e_{nk'} = 1 - e_{nk}, \tag{54}$$

$$z_{nk} \leq DEe_{nk} \tag{41.b}$$

$\Rightarrow$ {sum over k'}

$$\sum_{k'=1, k' \neq k}^{NE} z_{nk'} \leq DE \sum_{k'=1, k' \neq k}^{NE} e_{nk'}$$

$\equiv$ {using (53), (54)}

$$d_n - z_{nk} \leq DE(1 - e_{nk})$$

$\equiv$ {rearranging}

$$d_n - DE(1 - e_{nk}) \leq z_{nk}$$

Hence satisfied.

Proving $(41.d)$: $\quad z_{nk} \leq d_n$

$$\sum_{k=1}^{NE} z_{nk} = d_n, \tag{51}$$

$\Rightarrow$ {using $0 \leq z_{nk}$ (41.a)}

$$z_{nk} \leq d_n$$

Hence satisfied.

Proving (43) : $\quad LO_k e_{nk} \leq d_n$

$$LO_k e_{nk} \leq z_{nk} \tag{49}$$

$\Rightarrow$ {using $z_{nk} \leq d_n$ (41.d)}

$$LO_k e_{nk} \leq d_n$$

Hence satisfied.

Proving (44) : $\quad d_n + DEe_{nk} \leq UP_k + DE$

$$
\begin{aligned}
& d_n - DE(1 - e_{nk}) \leq z_{nk} && (41.c)\\
\Rightarrow \quad & \{\text{using } z_{nk} \leq UP_k e_{nk} \ (50)\}\\
& d_n - DE + DEe_{nk} \leq UP_k e_{nk}\\
\Rightarrow \quad & \{\text{using } 0 \leq e_{nk} \leq 1 \ (52)\}\\
& d_n - DE + DEe_{nk} \leq UP_k\\
\equiv \quad & \{\text{rearranging}\}\\
& d_n + DEe_{nk} \leq UP_k + DE
\end{aligned}
$$

Hence satisfied.

Therefore point $P \in R_1$, since it satisfies all the constraints. Therefore $R_2 \subseteq R_1$. □

Claim 2.2. *There exists a point Q such that $Q \in R_1$ and $Q \notin R_2$.*

Proof. Take a point $Q(z, e, d) = ([d, d], [1/2, 1/2], d)$. Here $(n, m) = (1, 2)$, $(LO_k, UP_k, DE) = ([0, d], [d, 2d], 2d)$ where $d \geq 0$. We show $Q \in R_1$ since it satisfies all the constraints.

$41.a$:

$0 \leq z_{11}$, replacing values we get $0 \leq d$

$0 \leq z_{12}$, replacing values we get $0 \leq d$

$41.b$:

$z_{11} \leq DEe_{11}$, replacing values we get $d \leq 2d/2$

$z_{12} \leq DEe_{12}$, replacing values we get $d \leq 2d/2$

$41.c$:

$d_1 - DE(1 - e_{11}) \leq z_{11}$, replacing values we get $d - 2d(1 - 1/2) \leq d$

$d_1 - DE(1 - e_{12}) \leq z_{12}$, replacing values we get $d - 2d(1 - 1/2) \leq d$

$41.d$:

$z_{11} \leq d_1$, replacing values we get $d \leq d$

$z_{12} \leq d_1$, replacing values we get $d \leq d$

42 :

$e_{11} + e_{12} = 1$, replacing values we get $1/2 + 1/2 = 1$

43 :

$LO_1 e_{11} \leq d_1$, replacing values we get $0 \times 1/2 \leq d$

$LO_2 e_{12} \leq d_1$, replacing values we get $d \times 1/2 \leq d$

44 :

$DEe_{11} + d_1 \leq UP_1 + DE$, replacing values we get $2d \times 1/2 + d \leq d + 2d$

$DEe_{12} + d_1 \leq UP_2 + DE$, replacing values we get $2d \times 1/2 + d \leq 2d + 2d$

46 :

$DE \geq UP_1$, replacing values we get $2d \geq d$

$DE \geq UP_2$, replacing values we get $2d \geq 2d$

45 :

$DE \geq d_1$, replacing values we get $2d \geq d$

52 :

$0 \leq e_{11} \leq 1$, replacing values we get $0 \leq 1/2 \leq 1$

$0 \leq e_{12} \leq 1$, replacing values we get $0 \leq 1/2 \leq 1$

Therefore point $Q \in R_1$. To show that $Q \notin R_2$ consider equation 51:

$$z_{11} + z_{12} = d, \text{ replacing values we get } d + d = 2d$$

Therefore $Q \notin R_2$. □

Proposition 2.1 shows that $R_2 \subseteq R_1$. Proposition 2.2 shows that $R_2 \neq R_1$. These two together imply $R_2 \subset R_1$. Proposition 2 therefore shows that the LP relaxation of the new model (R_2) has a tighter bound than the LP relaxation of the old model (R_1). We next show that in fact R_2 has the tightest bound possible by showing that a point with fraction value for e_{nk} will never be a corner point.

Proposition 3. *If point $P \in R_2$ has a fractional value for e_{nk}, P cannot be a corner point of R_2.*

Proof. Consider a point $P(z, e, d) \in R_2$ with at least one fractional value for e_{nk} i.e. $0 < e_{n'k'} < 1$ for some n', k'. Let $e_{n'k'} = t$. Construct another point P_1 that has the same components of P for $n \neq n'$. For $n = n'$ take (z, e, d) as follows:

$$For \quad k = 1 \ldots NE$$

$$
\begin{aligned}
z_{n'k'} &= 0\\
z_{n'k} &= \frac{z^P_{n'k}}{1 - t}, \quad for\ k \neq k'\\
e_{n'k'} &= 0\\
e_{n'k} &= \frac{e^P_{n'k}}{1 - t}, \quad for\ k \neq k'\\
d_{k'} &= \frac{d^P_{n'} - z^P_{n'k'}}{1 - t}
\end{aligned}
$$

Here $z^P_{n'k'}$, $e^P_{n'k'}$, $d^P_{n'}$ are the corresponding values of point P.

We show that $P_1 \in R_2$ since it satisfies all the constraints:

49 :

$$LO_{k'} e_{n'k'} \leq z_{n'k'}$$
$$\equiv LO_{k'} \times 0 \leq 0 \qquad \text{(definition)}$$

$$LO_k e_{n'k} \leq z_{n'k}, \quad k \neq k'$$
$$\equiv LO_k \frac{e^P_{n'k}}{1-t} \leq \frac{z^P_{n'k}}{1-t}, \quad k \neq k' \qquad \text{(definition)}$$
$$\equiv LO_k e^P_{n'k} \leq z^P_{n'k}, \quad k \neq k' \qquad (0 < t < 1)$$

Satisfied since $P \in R_2$.

50 :

$$z_{n'k'} \leq UP_{k'} e_{n'k'}$$
$$\equiv 0 \leq UP_{k'} \times 0 \qquad \text{(definition)}$$

$$z_{n'k} \leq UP_k e_{n'k}, \quad k \neq k'$$
$$\equiv \frac{z^P_{n'k}}{1-t} \leq UP_k \frac{e^P_{n'k}}{1-t}, \quad k \neq k' \qquad \text{(definition)}$$
$$\equiv z^P_{n'k} \leq UP_k e^P_{n'k}, \quad k \neq k' \qquad (0 < t < 1)$$

Satisfied since $P \in R_2$.

42 :

$$\sum_{k=1}^{NE} e_{n'k}$$
$$= 0 + \sum_{k=1,k\neq k'}^{NE} e_{n'k} \qquad \text{(splitting sum)}$$
$$= \sum_{k=1,k\neq k'}^{NE} \frac{e^P_{n'k}}{1-t} \qquad \text{(definition)}$$
$$= \frac{1 - e^P_{n'k'}}{1-t} \qquad (\sum_{k=1}^{NE} e^P_{n'k} = 1)$$
$$= \frac{1-t}{1-t} \qquad \text{(definition)}$$
$$= 1 \qquad (0 < t < 1)$$

51 :

$$\sum_{k=1}^{NE} z_{n'k}$$
$$= 0 + \sum_{k=1,k\neq k'}^{NE} z_{n'k} \qquad \text{(splitting sum)}$$
$$= \sum_{k=1,k\neq k'}^{NE} \frac{z^P_{n'k}}{1-t} \qquad \text{(definition)}$$
$$= \frac{d^P_{n'} - e^P_{n'k'}}{1-t} \qquad (\sum_{k=1}^{NE} z^P_{n'k} = d^P_{n'})$$
$$= d_{n'} \qquad \text{(definition)}$$

52 :

$$\sum_{k=1}^{NE} e^P_{n'k} = 1 \qquad (P \in R_2)$$
$$\equiv e^P_{n'k'} + \sum_{k=1,k\neq k'}^{NE} e^P_{n'k} = 1 \qquad \text{(splitting sum)}$$
$$\equiv t + \sum_{k=1,k\neq k'}^{NE} e^P_{n'k} = 1 \qquad \text{(definition)}$$
$$\equiv \sum_{k=1,k\neq k'}^{NE} e^P_{n'k} = 1 - t \qquad \text{(rearranging)}$$
$$\equiv 0 \leq e^P_{n'k} \leq 1 - t, \quad k \neq k' \qquad (e^P_{n'k} \geq 0)$$
$$\equiv 0 \leq \frac{e^P_{n'k}}{1-t} \leq 1, \quad k \neq k' \qquad (1 - t > 0)$$
$$\equiv 0 \leq e_{n'k} \leq 1, \quad k \neq k' \qquad \text{(definition)}$$

48 :

$$\sum_{k=1}^{NE} z^P_{n'k} = d^P_{n'} \qquad (P \in R_2)$$
$$\equiv z^P_{n'k'} + \sum_{k=1,k\neq k'}^{NE} z^P_{n'k} = d^P_{n'} \qquad \text{(splitting sum)}$$
$$\equiv \sum_{k=1,k\neq k'}^{NE} z^P_{n'k} = d^P_{n'} - z^P_{n'k'} \qquad \text{(rearranging)}$$
$$\equiv d^P_{n'} - z^P_{n'k'} \geq 0 \qquad (z^P_{n'k} \geq 0)$$
$$\equiv \frac{d^P_{n'} - z^P_{n'k'}}{1-t} \geq 0 \qquad (1 - t > 0)$$
$$\equiv d_{n'} \geq 0 \qquad \text{(definition)}$$

41.a :

$$z^P_{n'k} \geq 0, \quad k \neq k' \qquad (P \in R_2)$$
$$\equiv \frac{z^P_{n'k}}{1-t} \geq 0, \quad k \neq k' \qquad (1 - t > 0)$$
$$\equiv z_{n'k} \geq 0, \quad k \neq k' \qquad \text{(definition)}$$

Therefore $P_1 \in R_2$.

Similar to P_1 construct point P_2 having the same components as P for $n \neq n'$. For $n = n'$ take (z, e, d) as follows:

$$For \quad k = 1 \ldots NE$$

$$z_{n'k'} = \frac{z^P_{n'k'}}{t}$$
$$z_{n'k} = 0 \quad for\ k \neq k'$$
$$e_{n'k'} = 1$$
$$e_{n'k} = 0 \quad for\ k \neq k'$$

$$d_{n'} = \frac{z^P_{n'k'}}{t}$$

As before $z^P_{n'k'}$, $e^P_{n'k'}$, $d^P_{n'}$ are the corresponding values of point P.

We show that $P_2 \in R_2$ since it satisfies all the constraints:

49 :

$$\begin{aligned}
& LO_{k'} e^P_{n'k'} \le z^P_{n'k'} && (P \in R_2)\\
\equiv\ & LO_{k'} \times t \le z^P_{n'k'} && \text{(definition)}\\
\equiv\ & LO_{k'} \le \frac{z^P_{n'k'}}{t} && (0 < t)\\
\equiv\ & LO_{k'} \le z_{n'k'} && \text{(definition)}\\
\equiv\ & LO_{k'} \times 1 \le z_{n'k'} && \\
\equiv\ & LO_{k'} e_{n'k'} \le z_{n'k'} && \text{(definition)}
\end{aligned}$$

$$\begin{aligned}
& LO_k e_{n'k} \le z_{n'k}, \quad k \ne k' && \\
\equiv\ & LO_k \times 0 \le 0, \quad k \ne k' && (definition)
\end{aligned}$$

50 :

$$\begin{aligned}
& z^P_{n'k'} \le UP_{k'} e^P_{n'k'} && (P \in R_2)\\
\equiv\ & z^P_{n'k'} \le UP_{k'} \times t && \text{(definition)}\\
\equiv\ & \frac{z^P_{n'k'}}{t} \le UP_{k'} && (0 < t)\\
\equiv\ & z_{n'k'} \le UP_{k'} \times 1 && \text{(definition)}\\
\equiv\ & z_{n'k'} \le UP_{k'} e_{n'k'} && \text{(definition)}
\end{aligned}$$

$$\begin{aligned}
& z_{n'k} \le UP_k e_{n'k}, \quad k \ne k' && \\
\equiv\ & 0 \le UP_k \times 0, \quad k \ne k' && \text{(definition)}
\end{aligned}$$

42 :

$$\begin{aligned}
& \sum_{k=1}^{NE} e_{n'k} && \\
& = e_{n'k'} + \sum_{k=1,k\ne k'}^{NE} e_{n'k} && \text{(splitting sum)}\\
& = 1 + 0 && \text{(definition)}
\end{aligned}$$

51 :

$$\begin{aligned}
& \sum_{k=1}^{NE} z_{n'k} && \\
& = z_{n'k'} + \sum_{k=1,k\ne k'}^{NE} z_{n'k} && \text{(splitting sum)}\\
& = \frac{z^P_{n'k'}}{t} + 0 && \text{(definition)}\\
& = d_{n'} && \text{(definition)}
\end{aligned}$$

52 :

$$\begin{aligned}
& e_{n'k'} = 1 && \text{(definition)}\\
& e_{n'k} = 0, \quad k \ne k' && \text{(definition)}
\end{aligned}$$

48 :

$$\begin{aligned}
& z^P_{n'k'} \ge 0 && (P \in R_2)\\
\equiv\ & \frac{z^P_{n'k'}}{t} \ge 0 && (0 < t)\\
\equiv\ & d_{n'} \ge 0 && \text{(definition)}
\end{aligned}$$

41.a :

$$\begin{aligned}
& z^P_{n'k'} \ge 0 && (P \in R_2)\\
\equiv\ & \frac{z^P_{n'k'}}{t} \ge 0 && (0 < t)\\
\equiv\ & z_{n'k'} \ge 0 && \text{(definition)}
\end{aligned}$$

$$\begin{aligned}
& z_{n'k} \ge 0, \quad k \ne k' && \\
\equiv\ & 0 \ge 0, \quad k \ne k' && \text{(definition)}
\end{aligned}$$

Therefore $P_2 \in R_2$.

$$\begin{aligned}
P_1 &= \Big(\big(\frac{z^P_{n'k}}{1-t}, 0\big), \big(\frac{e^P_{n'k}}{1-t}, 0\big), \frac{d'^P_i - z^P_{n'k'}}{1-t}\Big)\\
P_2 &= \Big(\big(0, \frac{z^P_{n'k'}}{t}\big), (0, 1), \frac{z^P_{n'k'}}{t}\Big)\\
P_1(1-t) + P_2 t &= \big((z^P_{n'k}, z^P_{n'k'}), (e^P_{n'k}, t), d'^P_i\big)\\
&= P
\end{aligned}$$

Since P is a point with fractional e_{ik} that can be represented as a linear combination of two other points belonging to R_2, P cannot be a corner point of R_2. This implies that LP relaxation (R_2) of the new model will provide only integer solutions. Therefore the new model has the tightest bound possible. □

5 Tank Configuration Improvement

For a given network of nodes and links, one aspect of the problem is to determine the location of tanks and the set of downstream nodes that are to be served by each tank. We need a set of constraints to model a valid network configuration. For a given branched network layout with a single source, a valid network configuration is one in which:

1. Each node needs to be provided water, by exactly one of its ancestors (including itself).
2. If a node n provides water to itself i.e. it has a tank, only then can it provide water to its descendants.
3. If a node n gets water from another tank, then all its descendants cannot get water from themselves.

4. If node n provides water to one of its descendants k, then the nodes along the path connecting them cannot serve themselves.

In the following section we repeat the set of constraints that model such a network as laid out in section 2. We then show that the model is not tight, i.e. its linear relaxation is not guaranteed to have integral corner points. In section 5.2 we then describe an improved model and prove its tightness. In section 6 we describe an alternate edge based approach to modelling the network.

5.1 *Initial Model*

Consider a tree network of n nodes.

Data:

A_n = Nodes that are ancestors of node n, $n = 1 \dots NN$.

D_n = Nodes that are descendants of node n, $n = 1 \dots NN$.

C_n = Nodes that are children of node n, $n = 1 \dots NN$.

P_n = Parent node of node n, $n = 1 \dots NN$.

Variables:

$s_{nm} = 1$ if tank at n^{th} node serves the demand of m^{th} node, n = 1 ... NN, $m \in D_n \cup \{n\}$.

Constraints:

Then we can use the following set of constraints to describe the set of valid network configurations as described earlier:

$$s_{mm} \leq s_{nn}, \quad n = 1 \dots NN, \quad m \in D_n, \tag{55}$$

$$s_{nm} \leq s_{nn}, \quad n = 1 \dots NN, \quad m \in D_n, \tag{56}$$

$$\sum_m s_{mn} = 1, \quad n = 1 \dots NN, \quad m \in A_n \cup \{n\}, \tag{57}$$

$$s_{nm} \leq 1 - s_{oo}, \quad n = 1 \dots NN, \quad m \in D_n, \quad o \in D_n \cup A_m, \tag{58}$$

$$s_{nm} \in \{0, 1\}, \quad n = 1 \dots NN, \quad m \in D_n \cup \{n\}. \tag{59}$$

Proposition 4. *The linear relaxation of S is not tight.*

Proof. Let the linear relaxation of set S be S'. Instead of constraint 59 we will have the following constraint:

$$0 \leq s_{nm} \leq 1, \quad n = 1 \dots NN, \quad m \in D_n \cup \{n\}. \tag{60}$$

Consider a small network of 3 nodes with node 1 as root and node 2 child of node 1, and node 3 child of node 2. For a point to belong to S', the following constraints must be met:

$$s_{22} \leq s_{11}, \tag{55.a}$$
$$s_{33} \leq s_{11}, \tag{55.b}$$
$$s_{33} \leq s_{22}, \tag{55.c}$$
$$s_{12} \leq s_{11}, \tag{56.a}$$
$$s_{13} \leq s_{11}, \tag{56.b}$$
$$s_{23} \leq s_{22}, \tag{56.c}$$
$$s_{11} = 1, \tag{57.a}$$
$$s_{12} + s_{22} = 1, \tag{57.b}$$
$$s_{13} + s_{23} + s_{33} = 1, \tag{57.c}$$
$$s_{13} \leq 1 - s_{22}, \tag{58.a}$$
$$0 \leq s_{11}, s_{12}, s_{13}, s_{22}, s_{23}, s_{33} \leq 1, \tag{60}$$

Since $s_{11} = 1$, we replace its value in the constraints and replace repeating constraints to get the following set:

$$s_{33} \leq s_{22}, \tag{55.c}$$
$$s_{23} \leq s_{22}, \tag{56.c}$$
$$s_{11} = 1, \tag{57.a}$$
$$s_{12} + s_{22} = 1, \tag{57.b}$$
$$s_{13} + s_{23} + s_{33} = 1, \tag{57.c}$$
$$s_{13} \leq 1 - s_{22}, \tag{58.a}$$
$$0 \leq s_{12}, s_{13}, s_{22}, s_{23}, s_{33} \leq 1, \tag{60}$$

Consider a point P defined as:
$P\{s_{11}, s_{12}, s_{13}, s_{22}, s_{23}, s_{33}\} = \{1, \frac{1}{2}, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$. Since it satisfies all the constraints, $P \in S'$. We now show that P cannot be described as a linear combination of two distinct points that belong to S'.
Consider two points $Q, R \in S'$ such that:

$$P = tQ + (1-t)R, \quad 0 < t < 1$$

$$s^Q_{11} = s^R_{11} = 1 \qquad \{57.a\}$$

$$\begin{aligned} & s^P_{13} = 0 && \{\text{definition}\} \\ \Rightarrow & s^Q_{13} = s^R_{13} = 0 && \{60.c\} \end{aligned} \tag{61}$$

$$\begin{aligned} & s_{33} + s_{23} \leq 2s_{22} && \{\text{adding 55.c and 56.c}\} \\ \Rightarrow & 1 - s^Q_{13} \leq 2s^Q_{22} && \{57.c\} \\ \Rightarrow & \frac{1}{2} \leq s^Q_{22} && \{61\} \\ \Rightarrow & \frac{1}{2} \leq s^R_{22} && \{\text{Similarly}\} \\ \Rightarrow & s^Q_{22} = s^R_{22} = \frac{1}{2} && \{s^P_{22} = \frac{1}{2}\} \end{aligned} \tag{62}$$

$$\begin{aligned} & s_{12} + s_{22} = 1 && \{57.b\} \\ \Rightarrow & s^Q_{12} = s^R_{12} = \frac{1}{2} && \{62\} \end{aligned}$$

$$\begin{aligned} & s_{33} \leq s_{22} && \{55.c\} \\ \Rightarrow & s^Q_{33} \leq \frac{1}{2} && \{62\} \\ & s_{23} \leq s_{22} && \{56.c\} \\ \Rightarrow & s^Q_{23} \leq \frac{1}{2} && \{62\} \end{aligned}$$

$$\Rightarrow s^Q_{23} = s^Q_{33} = \frac{1}{2} \qquad \{57.c\}$$

$$\Rightarrow s^R_{23} = s^R_{33} = \frac{1}{2} \qquad \{\text{Similarly}\}$$

Therefore $P = Q = R$. Since P cannot be expressed as a linear combination of two distinct points, P is a corner point of S'. And since P contains non-integral values for s_{nm}, set S' is not tight. □

5.2 *Improved Model* ***(Model 4)***

A new model is proposed which although maintains the same structure as the initial model, it does so using tighter constraints. The primary insight about the structure is expressed in the second constraint mentioned below. A node i serves its child j if and only if it serves all the nodes downstream of j. Consider set $S2$ defined by the following set of constraints:

$$\sum_m s_{mn} = 1, \quad n = 1\ldots NN, \quad m \in A_n \cup \{n\}, \tag{57}$$

$$s_{nm} = s_{nk}, \quad n = 1\ldots NN, \quad m \in C_n, \quad k \in D_m, \tag{63}$$

$$s_{nm} \in \{0,1\}, \quad n = 1\ldots NN, \quad m \in D_n \cup \{n\}. \tag{59}$$

Proposition 5. *The linear relaxation of $S2$ is tight.*

Proof. Let the linear relaxation of set $S2$ be $S2'$. Instead of constraint 59 we will have the following constraint:

$$0 \le s_{nm} \le 1, \quad n = 1\ldots NN, \quad m \in D_n \cup \{n\}. \tag{64}$$

We will show that $S2'$ is tight by showing any point P, with a non-integer component can be expressed as a linear combination of two distinct points from $S2'$.
Consider a point $P \in S2'$ with $0 < s_{n'n'} = t < 1$ for some n'. Let n' be the first such node in the path from root.

Claim 5.1. $s_{nn} = 1, \quad n \in A_{n'}$

Proof. s_{nn} cannot be fractional since n' is the first such node from root by definition. Assume $s_{nn} = 0$ for some $n \in A_{n'}$. Let $E_{nn'} = (D_n \cup \{n\}) \cap (A'_n \cup \{n'\})$.

$$s_{nn} = 0$$

$$\equiv \quad \{\text{using} \sum_m s_{mn} = 1 \ (57)\}$$

$$\sum_m s_{mn} = 1 \qquad m \in A_n$$

$$\sum_m s_{mn'} = 1 \qquad m \in A'_n \cup \{n'\}$$

$$\equiv \quad \{\text{splitting sum}\}$$

$$\sum_m s_{mn'} + \sum_k s_{kn'} = 1 \qquad m \in A_n, k \in E_{nn'}$$

$$\equiv \quad \{\text{using } s_{nm} = s_{nk} \ (63)\}$$

$$\sum_m s_{mn} + \sum_k s_{kn'} = 1 \qquad m \in A_n, k \in E_{nn'}$$

$$\equiv \quad \{\text{using} \sum_m s_{mn} = 1 \text{ from above}\}$$

$$1 + \sum_k s_{kn'} = 1 \qquad k \in E_{nn'}$$

$$\equiv \quad \{\text{simplifying}\}$$

$$\sum_k s_{kn'} = 0 \qquad k \in E_{nn'}$$

$$\equiv \quad \{\text{using } 0 \le s_{nm} \le 1 \ (64)\}$$

$$s_{kn'} = 0 \qquad k \in E_{nn'}$$

$$\Rightarrow \quad \{\text{since } n' \in E_{nn'}\}$$

$$s_{n'n'} = 0$$

But this is a contradiction since we know $s_{n'n'}$ is fractional. Therefore s_{nn} cannot be fractional and it cannot be 0.

$$s_{nn} = 1, \quad n \in A_{n'} \tag{65}$$

□

Claim 5.2. $s_{nm} = 0, \quad n \in A_{p'}, \quad m \in D_n, \quad p' = P_{n'}$

Proof.

$$s_{mm} = 1, \qquad m \in A_{n'}$$

$$\equiv \quad \{\text{using} \sum_m s_{mn} = 1 \ (57)\}$$

$$s_{nm} = 0, \qquad n \in A_{p'}, \ m \in A_{n'}, \ j \in D_n, \ p' = P_{n'}$$

$$\equiv \quad \{\text{using } s_{nm} = s_{nk} \ (63)\}$$

$$s_{nm} = 0, \qquad n \in A_{p'}, \ m \in D_i, \ p' = P_{n'} \tag{66}$$

□

Consider a point Q with $s_{n'n'} = 0$:

$$s_{nm} = s^P_{nm}, \quad m \notin (A_{n'} \cup D_{n'} \cup \{n'\}) \tag{67}$$

$$s_{nm} = \frac{s^P_{nm}}{1-t}, \quad n \in A_{n'}, j \in D_n \tag{68}$$

$$s_{nm} = 0, \quad n \in (D_{n'} \cup \{n'\}), m \in D_n \tag{69}$$

Claim 5.3. *Point $Q \in S2'$*

Proof. We prove that point Q belongs to S2' by showing it satisfies the constraints (57), (63) and (64).

For nodes that are not downstream or upstream of n', s_{nm} values are same as that of point P. Therefore they satisfy the constraints since P belongs to $S2'$.

For the rest of the nodes:

For $n \in A_{n'}$:

$$\text{Proving (57)}: \quad \sum_m s_{mn} = 1$$

$$\{\text{using } s_{nn} = 1 \ (65)\}$$

$$s_{nn} = 1, \qquad n \in A_{n'}$$

$$\{\text{using } s_{nm} = 0 \ (66)\}$$

$$s_{mn} = 0, \qquad n \in A_{n'}, \ m \in A_n$$
$\equiv$ {summing over m}
$$\sum_m s_{mn} = 1, \qquad n \in A_{n'}$$
Hence satisfied.

Proving (63) : $\quad s_{nm} = s_{nk}$

{using $s_{nm} = s_{nk}$ (63)}
$$s^P_{nm} = s^P_{nk}, \qquad n \in A_{n'}, \ m \in C_n, \ k \in D_m$$
$\equiv$ {dividing by $(1-t)$ since $t \neq 1$ }
$$\frac{s^P_{nm}}{1-t} = \frac{s^P_{nk}}{1-t}, \qquad n \in A_{n'}, \ m \in C_n, \ k \in D_m$$
$\equiv$ {using $s_{nm} = \dfrac{s^P_{nm}}{1-t}$ (68)}
$$s_{nm} = s_{nk}, \qquad n \in A_{n'}, \ m \in C_n, \ k \in D_m$$
Hence satisfied.

Proving (64) : $\quad 0 \leq s_{mn'} \leq 1$

{using $\sum_m s_{mn} = 1$ (57)}
$$\sum_m s^P_{mn'} = 1, \qquad m \in A_{n'} \cup \{n'\}$$
$\equiv$ {splitting sum}
$$\sum_m s^P_{mn'} + s^P_{n'n'} = 1, \qquad m \in A_{n'}$$
$\equiv$ {using $s^P_{n'n'} = t$}
$$\sum_m s^P_{mn'} = 1 - t, \qquad m \in A_{n'}$$
$\equiv$ {using $s^P_{mn'} \geq 0$ (64)}
$$0 \leq s^P_{mn'} \leq 1 - t, \qquad m \in A_{n'}$$
$\equiv$ {dividing by $(1-t)$ since $t \neq 1$}
$$0 \leq \frac{s^P_{mn'}}{1-t} \leq 1, \qquad m \in A_{n'}$$
$\equiv$ {using $s_{nm} = \dfrac{s^P_{nm}}{1-t}$ (68)}
$$0 \leq s_{mn'} \leq 1, \qquad m \in A_{n'}$$
Hence satisfied.

For $n \in D_{n'} \cup \{n'\}$:

Proving (57) : $\quad \sum_m s_{mn} = 1$

{using $\sum_m s_{mn} = 1$ (57)}
$$\sum_m s^P_{mn'} = 1, \qquad m \in A_{n'} \cup \{m'\}$$
{using $s_{nm} = 0$ (66)}
$$s^P_{kn'} + s^P_{n'n'} = 1, \qquad k = P_{n'}$$
$\equiv$ {using $s^P_{n'n'} = t$}
$$s^P_{kn'} = 1 - t, \qquad k = P_{n'}$$
$\equiv$ {using $s_{nm} = \dfrac{s^P_{nm}}{1-t}$ (68)}
$$s_{kn'} = 1, \qquad k = P_{n'}$$
$\equiv$ {using $s_{nm} = 0$ (69)}
$$\sum_m s_{mn} = 1, \qquad n \in D_{n'} \cup \{n'\}, m \in A_n$$
Hence satisfied.

Proving (63) : $\quad s_{nm} = s_{nk}$

{using $s_{nm} = 0$ (69)}
$$s_{nm} = 0 \qquad n \in D_{n'} \cup \{n'\}, m \in D_n$$
$\equiv$
$$s_{nm} = s_{nk} \qquad n \in D_{n'} \cup \{n'\}, m \in D_n, k \in C_n$$
Hence satisfied.

Proving (64) : $\quad 0 \leq s_{nm} \leq 1$

{using $s_{nm} = 0$ (69)}
$$s_{nm} = 0 \qquad n \in D_{n'} \cup \{n'\}, m \in D_n$$
Hence satisfied.

Therefore point $Q \in S2'$.

□

Similarly consider point R with $s_{n'n'} = 1$:

$$s_{nm} = s^P_{nm}, \quad m \notin (A_{n'} \cup D_{n'} \cup \{n'\}) \tag{70}$$
$$s_{nm} = 0, \quad n \in A_{n'}, m \in D_n \tag{71}$$
$$s_{nm} = \frac{s^P_{nm}}{t}, \quad n \in (D_{n'} \cup \{n'\}), m \in D_n \cup \{n\} \tag{72}$$

Claim 5.4. *Point $R \in S2'$*

Proof. We prove that point R belongs to S2' by showing it satisfies the constraints (57), (63) and (64)

For nodes that are not downstream or upstream of n', s_{nm} values are same as that of point P. Therefore they satisfy the constraints since P belongs to $S2'$.

For the rest of the nodes:

For $n \in A_{n'}$:

Proving (57) : $\quad \sum_m s_{mn} = 1$

{using $s_{nn} = 1$ (65)}
$$s_{nn} = 1, \qquad n \in A_{n'}$$
{using $s_{nm} = 0$ (66)}
$$s_{mn} = 0, \qquad n \in A_{n'}, \ m \in A_n$$
$\equiv$ {summing over m}
$$\sum_m s_{mn} = 1, \qquad n \in A_{n'}$$
Hence satisfied.

Proving (63) : $s_{nm} = s_{nk}$

{using $s_{nm} = 0$ (71)}

$$s_{nm} = 0 \qquad n \in A_{n'}, m \in D_n$$

$\equiv$

$$s_{nm} = s_{nk} \qquad n \in A_{n'}, m \in D_n, k \in C_n$$

Hence satisfied.

Proving (64) : $0 \leq s_{nm} \leq 1$

{using $s_{nm} = 0$ (71)}

$$s_{nm} = 0 \qquad n \in A_{n'}, m \in D_n$$

Hence satisfied.

For $n \in D_{n'} \cup \{n'\}$:

Proving (57) : $\sum_m s_{mn} = 1$

{using $\sum_m s_{mn} = 1$ (57)}

$$\sum_m s^P_{mn} = 1, \qquad m \in A_n \cup \{n\}$$

$\equiv$ {splitting sum}

$$\sum_m s^P_{mn} + \sum_k s^P_{kn} = 1, \qquad m \in D_{n'} \cup \{n'\}, k \in A_{n'}$$

$\equiv$ {using $\sum_k s^P_{kn} = 1 - t$ }

$$\sum_m s^P_{mn} = t, \qquad m \in D_{n'} \cup \{n'\}$$

$\equiv$ {dividing by t since $t \neq 0$ }

$$\sum_m \frac{s^P_{mn}}{t} = 1, \qquad m \in D_{n'} \cup \{n'\}$$

$\equiv$ {using $s_{nm} = \frac{s^P_{nm}}{t}$ (72)}

$$\sum_m s_{mn} = 1, \qquad m \in D_{n'} \cup \{n'\}$$

Hence satisfied.

Proving (63) : $s_{nm} = s_{nk}$

{using $s_{nm} = s_{nk}$ (63)}

$$s^P_{nm} = s^P_{nk}, \qquad m \in C_n,\ k \in D_m$$

$\equiv$ {dividing by t since $t \neq 0$ }

$$\frac{s^P_{nm}}{t} = \frac{s^P_{nk}}{t}, \qquad m \in C_n,\ k \in D_m$$

$\equiv$ {using $s_{nm} = \frac{s^P_{nm}}{t}$ (72)}

$$s_{nm} = s_{nk}, \qquad m \in C_n,\ k \in D_m$$

Hence satisfied.

Proving (64) : $0 \leq s_{nm} \leq 1$

{using $\sum_m s_{mn} = 1$ (57)}

$$\sum_m s^P_{mn} = 1, \qquad m \in A_n \cup \{n\}$$

$\equiv$ {splitting sum}

$$\sum_m s^P_{mn} + \sum_k s^P_{kn} = 1, \qquad m \in D_{n'} \cup \{n'\}, k \in A_{n'}$$

$\equiv$ {using $\sum_k s^P_{kn} = 1 - t$ }

$$\sum_m s^P_{mn} = t, \qquad m \in D_{n'} \cup \{n'\}$$

{using $0 \leq s^P_{mn}$ (64)}

$$0 \leq s^P_{mn} \leq t, \qquad m \in D_{n'} \cup \{n'\}$$

$\equiv$ {dividing by t since $t \neq 0$ }

$$0 \leq \frac{s^P_{mn}}{t} \leq 1, \qquad m \in D_{n'} \cup \{n'\}$$

$\equiv$ {using $s_{nm} = \frac{s^P_{nm}}{t}$ (72)}

$$0 \leq s_{mn} \leq 1, \qquad m \in D_{n'} \cup \{n'\}$$

Hence satisfied.

Therefore point $R \in S2'$. □

Claim 5.5. *P is a linear combination of points Q and R i.e. P = (1-t)Q + tR*

Proof. For $m \notin (A_{n'} \cup D_{n'} \cup \{n'\})$:

{using $s^P_{nm} = s^Q_{nm}$ (67) and $s^P_{nm} = s^R_{nm}$ (70)}

$$s^P_{nm} = s^Q_{nm} = s^R_{nm}$$

$\Rightarrow$

$$s^P_{nm} = (1-t)s^Q_{nm} + t * s^R_{nm}$$

For $n \in A_{n'}, m \in D_n$:

$$s^Q_{nm} = \frac{s^P_{nm}}{1-t} \qquad \{68\}$$

$$s^R_{nm} = 0 \qquad \{71\}$$

$\Rightarrow$

$$s^P_{nm} = (1-t)s^Q_{nm} + t * s^R_{nm}$$

For $n \in D_{n'}, m \in D_n$:

$$s^Q_{nm} = 0 \qquad \{69\}$$

$$s^R_{nm} = \frac{s^P_{nm}}{t} \qquad \{72\}$$

$\Rightarrow$

$$s^P_{nm} = (1-t)s^Q_{nm} + t * s^R_{nm}$$

Therefore P is a linear combination of points Q and R □

Since any general point P with a fractional component can be expressed as linear combination of two other points in the set S2', it implies that such a point P cannot be a corner point and therefore set S2' is tight. □

This concludes the discussion on the three improvements made to the initial ILP model. Experimental results of the performance of the model after each improvement is presented in section 7. Although we have shown the tightness of various subsets of the improved model, the overall set of constraints of the model is still not tight. As such, there remains room for further improvements to the model. In the next section we describe an initial attempt at an alternative approach to the problem. Instead of using node variables s_{ij} to partition the primary and secondary network, purely edge based variables are used.

6 Edge Based Model

An alternative approach to the node based representation of the network is to have an edge based representation. Here instead of the focus being which tank serves which node, the focus is on which pipes in the network are part of the primary network and which pipes are part of the secondary network. Consider a tree network of NE edges:

Data:

C_i = Set of pipes that are immediately downstream of pipe i, i = 1 . . . NE.

U_i = Set of pipes that are upstream of pipe i, i = 1 . . . NE.

D_i = Set of pipes that are downstream of pipe i, i = 1 . . . NE.

S = Set of pipes that are immediately downstream of the source.

Variables:

f_i = 1 if i^{th} edge belongs to the primary network and = 0 if i^{th} edge belongs to the secondary network, i = 1 . . . e

The primary network connects the source to the tanks, and the secondary network connects the tanks to downstream nodes. Therefore pipes starting from the source must belong to the primary network. Also, secondary pipes must be downstream of the primary pipes. And once a pipe is secondary, then any pipes downstream can no longer be primary. We can use the following set of constraints to describe the set $S3$ of valid network configurations:

$$f_i = 1, \quad i \in S, \tag{73}$$

$$f_j \leq f_i, \quad i = 1 \ldots NE, \quad j \in C_i, \tag{74}$$

$$f_i \in \{0, 1\}, \quad i = 1 \ldots NE. \tag{75}$$

Proposition 6. *The linear relaxation of $S3$ is tight.*

Proof. Let the linear relaxation of set $S3$ be $S3'$. Instead of constraint 75 we will have the following constraint:

$$0 \leq f_i \leq 1, \quad i = 1 \ldots NE. \tag{76}$$

We will show that $S3'$ is tight by showing any point P, with a non-integer component can be expressed as a linear combination of two distinct points from $S3'$.
Consider a point $P \in S3'$ with $0 < f_i = t < 1$ for some i'. Let i' be the first such edge in the path from source.

Claim 6.1. $f_i = 1, \quad i \in U_{i'}$

Proof. f_i cannot be fractional since i' is the first such edge from source by definition. If $f_i = 0$, then by 74 for all its downstream edges j, $f_i = 0$. But i' is downstream of i and $f_{i'} = t \neq 0$. Therefore f_i cannot be fractional and it cannot be 0.

$$f_i = 1, \quad i \in U_{i'} \tag{77}$$

□

Consider a point Q with $f_{i'} = 0$:

$$f_i = f_i^P, \quad i \notin (D_{i'} \cup \{i'\}) \tag{78}$$

$$f_i = 0, \quad i \in (D_{i'} \cup \{i'\}) \tag{79}$$

Claim 6.2. *Point $Q \in S3'$*

Proof. For all the edges not downstream of i', constraints 73, 74 and 76 are satisfied since the values are same as point P and $P \in S3'$. Setting $f_i = 0$ for all downstream i also maintains the constraints trivially. Therefore point $Q \in S3'$. □

Similarly consider point R with $f_{i'} = 1$:

$$f_i = f_i^P, \quad i \notin (D_{i'} \cup \{i'\}) \tag{80}$$

$$f_i = \frac{f_i^P}{t}, \quad i \in (D_{i'} \cup \{i'\}) \tag{81}$$

Claim 6.3. *Point $R \in S3'$*

Proof. We prove that point R belongs to S3' by showing it satisfies the constraints (73), (74) and (76). For edges that are not downstream of i', f_i values are same as that of point P. Therefore they satisfy the constraints since $P \in S3'$. For the rest of the edges:

For $i \in (D_{i'} \cup \{i'\})$: (73) is trivially true since i' (and its downstream edges) cannot be connected to the source since for point P, $f_i \neq 0$.

Proving (74) : $f_j \leq f_i$

$$
\begin{aligned}
&\{\text{using } f_j^P \leq f_i^P \text{ (74)}\} \\
&f_j^P \leq f_i^P, \qquad i \in (D_{i'} \cup \{i'\}), j \in C_i \\
\equiv\ &\{\text{dividing by } t \text{ since } t \neq 0\} \\
&\frac{f_j^P}{t} \leq \frac{f_i^P}{t}, \qquad i \in (D_{i'} \cup \{i'\}), j \in C_i \\
\equiv\ &\{\text{using } f_i = \frac{f_i^P}{t} \text{ (81)}\} \\
&f_j \leq f_i, \qquad i \in (D_{i'} \cup \{i'\}), j \in C_i
\end{aligned}
$$

Hence satisfied.

Proving (76) : $0 \leq f_i \leq 1$

{using $f_i^P \leq f_{i'}^P$ (74) and $0 \leq f_i^P$ (76)}

$$0 \leq f_i^P \leq f_{i'}^P \qquad i \in (D_{i'} \cup \{i'\})$$

$\equiv$ {using $f_i^P = t$ }

$$0 \leq f_i^P \leq t \qquad i \in (D_{i'} \cup \{i'\})$$

$\equiv$ {dividing by t since $t \neq 0$ }

$$0 \leq \frac{f_i^P}{t} \leq 1 \qquad i \in (D_{i'} \cup \{i'\})$$

$\equiv$ {using $f_i = \frac{f_i^P}{t}$ (81)}

$$0 \leq f_i \leq 1 \qquad i \in (D_{i'} \cup \{i'\})$$

Hence satisfied.

Therefore point $R \in S3'$. □

Claim 6.4. *P is a linear combination of points Q and R i.e. P = (1-t)Q + tR*

Proof. For $i \notin (D_{i'} \cup \{i'\})$:

$$f_i^P = f_i^Q = f_i^R \qquad \{78, 80\}$$

$\Rightarrow$

$$f_i^P = (1-t)f_i^Q + t * f_i^R$$

For $i \in (D_{i'} \cup \{i'\})$:

$$f_i^Q = 0 \qquad \{79\}$$

$$f_i^R = \frac{f_i^P}{t} \qquad \{81\}$$

$\Rightarrow$

$$f_i^P = (1-t)f_i^Q + t * f_i^R$$

Therefore P is a linear combination of points Q and R □

Since any general point P with a fractional component can be expressed as linear combination of two other points in the set S3', it implies that such a point P cannot be a corner point and therefore set S3' is tight.

□

The performance of the edge based model is worse than model 4. Although we prove that the LP relaxation of the set of constraints described by S3 is tight, the LP relaxation objective for the overall model is worse. This is due to changes in other constraints of the model, that are required since in this model only edge based variables are considered.

7 Results

The three pipe cost/tank cost/tank allocation improvements were applied sequentially to the initial model (model 1) to give model 2/model 3/model 4 respectively. These 4 models were tested over eight different networks of varying sizes in order to test their performance and scalability:

- Real World Networks: Three of the networks, Khardi, Shahpur and Mokhada are real life networks from Maharashtra state in India. These regions consist of tribal villages that regularly face extreme water stress during summer months and as a result have to be provided water using tankers.
- Artificial Networks: The other five networks are artificially created to test the performance of the models across different network sizes (10 to 200). Each of them is a randomly generated branched network. Ranges for the node and link properties are as follows:
  - Number of children nodes: 1 to 5,
  - Elevation (in metres): 100 to 300,
  - Demand (in litres per second): 0.01 to 5,
  - Length of links (in metres): 500 to 5000.

For all four models, the problem statement remains the same, to optimize the total pipe and tank cost of the network. The number of binary and continuous variables scale with the size of the network. Since all four models solve the problem optimally, the final capital cost of the pipes and tanks is the same. The performance of each model is measured in terms of three metrics: the total time taken in seconds, the size of the branch and bound tree and the objective value of the LP relaxation. For each of the eight networks, the time taken improves with each model, resulting in model 4 providing the best performance. Typically the time taken scales with the size of the network. However, this need not always be the case. For example, although gen50 has more nodes (50 nodes) than Mokhada (37 nodes), it is solved in lesser amount of time. This is because apart from the number of nodes being a factor, the network configuration also matters while solving the model.

8 Conclusion

In the present work we looked at the cost optimization of rural drinking water schemes. These schemes consist of several network components like pipes, tanks, pumps and valves. We first describe an initial ILP model that was used to solve the optimization. Although optimal, the model took a significant amount of time for larger networks, an hour for a network with 100 nodes. We then describe a series of three improvements of the model. For each improvement we prove that the improved model is tighter than the initial model. We then finally present the performance results of the three improved models along with the initial model over eight networks of various sizes. The 100 node network now takes only 49 seconds to solve.

Thus we show that tightening the ILP model can result in significant improvements in terms of performance. This enables practitioners to consider greater number of iterations of the design for large networks, since each iteration can be optimized in a matter of seconds. And although improvements were made in several constraints, the overall model is

Table 1. Performance of the various models on the eight networks. Size of the networks is represented by the number of nodes in each network. The objective is the minimum value of capital cost of pipes and tanks for that network, measured in rupees. Performance is measured across three metrics: time taken in seconds, the branch and bound tree size, and the objective value in rupees of the LP relaxation.

Network	Nodes	Binary Variables	Continuous Variables	Objective (Rs)	Metric	Model 1	Model 2	Model 3	Model 4
gen10	10	228	527	2.18E7	Time(s)	1.71	0.96	0.47	**0.32**
					B&B Tree Size	8	14	**2**	10
					LP Obj. (Rs)	7.89E6	7.89E6	**1.02E7**	**1.02E7**
Khardi	11	262	584	4.48E7	Time(s)	0.74	0.36	0.31	**0.24**
					B&B Tree Size	2	**0**	**0**	**0**
					LP Obj. (Rs)	1.14E07	1.16E07	**2.90E07**	**2.90E07**
Shahpur	21	712	1154	4.54E7	Time(s)	11.66	3.36	1.90	**0.72**
					B&B Tree Size	148	27	**4**	**4**
					LP Obj. (Rs)	2.00E7	2.02E7	**2.69E7**	**2.69E7**
Mokhada	37	1848	2066	4.94E7	Time(s)	28.78	7.91	3.48	**2.02**
					B&B Tree Size	965	188	68	**30**
					LP Obj. (Rs)	2.16E7	2.16E7	**3.05E7**	**3.05E7**
gen50	50	3148	2807	9.82E7	Time(s)	373.49	39.31	2.02	**1.27**
					B&B Tree Size	15228	1240	8	**6**
					LP Obj. (Rs)	4.00E7	4.00E7	**5.58E7**	**5.58E7**
gen100	100	11298	5657	2.47E8	Time(s)	197.89	129.75	9.73	**4.18**
					B&B Tree Size	1735	1958	120	**32**
					LP Obj. (Rs)	9.34E7	9.34E7	**1.27E8**	**1.27E8**
gen150	150	24448	8507	3.48E8	Time(s)	2582.35	887.68	18.67	**5.25**
					B&B Tree Size	22401	15498	148	**8**
					LP Obj. (Rs)	8.50E7	8.53E7	**1.35E8**	**1.35E8**
gen200	200	42598	11357	4.55E8	Time(s)	timeout*	timeout*	523.97	**69.80**
					B&B Tree Size	-	-	5034	**858**
					LP Obj. (Rs)	1.12E8	1.12E8	**1.83E8**	**1.83E8**

*timed out after running for 24 hours

still not tight. Further improvements are possible to increase performance even further. Additionally, the model can be extended in the future to include looped networks and/or consider scheduling of the water distribution.

References

[1] da Conceicao Cunha, Maria, and Luisa Ribeiro. ”Tabu search algorithms for water network optimization.” European Journal of Operational Research 157.3 (2004): 746-758.

[2] Eusuff, Muzaffar M., and Kevin E. Lansey. ”Optimization of water distribution network design using the shuffled frog leaping algorithm.” Journal of Water Resources Planning and Management 129.3 (2003): 210-225.

[3] Nikhil Hooda, Rajaram Desai, and Om P. Damani. Design and Optimization of Piped Water Network for Tanker Fed Villages in Mokhada Taluka Technical Report No. TR-CSE-2013e-55, 2013, Dept. of Computer Science and Engineering, IIT Bombay.

[4] Nikhil Hooda, and Om Damani, A System for Optimal Design of Pressure Constrained Branched Piped Water Networks, 18th Conference on Water Distribution System Analysis, WDSA (2016).

[5] Nikhil Hooda, and Om Damani, Inclusion of Tank Configurations as a Variable in the Cost Optimization of Branched Piped Water Networks, 14th Conference on Computing and Control for the Water Industry, CCWI (2016).

[6] Prasad M. Modak, Juzer Dhoonia, A Computer Program in Quick BASIC for the Least Cost Design of Branched Water Distribution Networks. UNDP/ World Bank, Asia Water Supply and Sanitation Sector Development Project, RAS/86/160, December 1991.

[7] Alperovits, Elyahu, and Uri Shamir. ”Design of optimal water distribution systems.” Water resources research 13.6 (1977): 885-900.

[8] Samani, Hossein MV, and Alireza Mottaghi. ”Optimization of water distribution networks using integer

linear programming." Journal of Hydraulic Engineering 132.5 (2006): 501-509.

[9] Lansey, Kevin E., and Larry W. Mays. "Optimization model for water distribution system design." Journal of Hydraulic Engineering 115.10 (1989): 1401-1418.

[10] Savic, Dragan A., and Godfrey A. Walters. "Genetic algorithms for least-cost design of water distribution networks." Journal of Water Resources Planning and Management 123.2 (1997): 67-77.

[11] Swamee, P. K., Virendra Kumar, and P. Khanna. "Optimization of dead end water distribution systems." Journal of the Environmental Engineering Division 99.2 (1973): 123-134.

[12] V. Choudhary, O. Damani, R. Desai, A. Joshi, M. Kanwat, M. Kaushal, S. Kharole ,Kharole, Y. Pawde, P. Rathore, and M. Sohoni. Redesigning Khardi Rural Piped Water Network Scheme for Sustainability Technical Report No. TR-CSE-2013-56, 2013, Dept. of Computer Science and Engineering, IIT Bombay.

[13] Vyas, Janki H., Narendra J. Shrimali, and Mukesh A. Modi. Optimization of Dhrafad Regional Water Supply Scheme using Branch 3.0., IJIRSET 2014

[14] WaterGEMS, Bentley Systems, 2006, Bentley Systems Incorporated, Exton, PA.

[15] Williams, Gardner Stewart, and Allen Hazen. "Hydraulic tables." (1933).

[16] Yugandhara Lad, J S Main, and S D Chawathe. Optimization of Hydraulic Design of Water Supply Tree Network Based on Present Worth Analysis. Journal of Indian Water Works Association. 2012