Enhancing Seismic Resilience of Water Pipe Networks

Taoan Huang
University of Southern California
taoanhua@usc.edu

Bistra Dilkina
University of Southern California
dilkina@usc.edu

ABSTRACT

As disasters such as earthquakes and floods become more frequent and detrimental, it is increasingly important that water infrastructure resilience be strategically enhanced to support post-disaster functionality and recovery. In this paper, we focus on the problem of strategically building seismic-resilient pipe networks to ensure direct water supply to critical customers and certain proximity to water sources for residential areas, which we formalize as the Steiner Network Problem with Coverage Constraints. We provide complexity statements of the problem and present an efficient mixed-integer linear program encoding to solve the problem. We also investigate the problem of planning partial network installments to maximize efficiency over time and propose a fast and effective sequential planning algorithm to solve it. We evaluate our algorithms on synthetic water networks and also apply them to a case study on a water service zone in Los Angeles, which demonstrate the effectiveness of our methods for large-scale real-world applications.

CCS CONCEPTS

• **Social and professional topics** → **Sustainability**; • **Theory of computation** → **Network optimization**; • **Computing methodologies** → *Combinatorial algorithms*.

KEYWORDS

Water infrastructure; Network optimization

1 INTRODUCTION

Human-made and natural disasters such as earthquakes, floods, and oil spills are global concerns that have already caused widespread disruptions to infrastructures, individual lives, and environmental systems. Given current trends in environmental change and human population growth, such disasters are likely to increase in frequency, severity, and distribution. Hence, it is vital to have informed, effective, and efficient disaster mitigation, planning, and response capabilities. Enhancing infrastructure resilience is particularly crucial since it concerns energy and water supply, transportation, and communication. Indeed, to "build resilient infrastructure, promote sustainable industrialization and foster innovation" is one of the United Nation's Sustainable Development Goals [2].

Water infrastructure is especially critical as it provides access to drinking water, functioning fire departments, and healthcare services. Earthquakes can pose serious threats to water supply by damaging underground water pipes – e.g. the 1994 magnitude 6.7 Northridge earthquake extensively damaged water supply and transmission subsystems in Los Angeles, incurring $41 million in repair costs [5]. Hence, designing strategies to fortify water infrastructure in order to improve resilience to earthquakes is an important problem relevant to many cities across the USA and the globe. Los Angeles is one of the cities at the forefront of proactively prioritizing infrastructure resilience. In particular, fortifying the Los Angeles water pipe network to prepare for an earthquake is one of the primary actions the city is taking within the "Improved Infrastructure" goal to make a more resilient Los Angeles [11]. The Los Angeles Department of Water and Power (LADWP) serves about 4 million people over a 465 square mile service area and oversees 7,000 miles of pipes [7]. Due to the city's layout and its geographical conditions, after a major earthquake, subterranean pipes could be severely damaged, causing losses of water service in many areas. In 2013, LADWP began replacing some older pipes with earthquake-resistant ductile iron pipes that are designed to lock to keep pipes together and accommodate earthquake forces, so that in the event of an earthquake, the pipes are less likely to be damaged [21]. Placing seismically robust pipes at key locations helps to improve continuous water delivery service and to reduce the workload to restore water in areas suffering from water shortage after an earthquake. The City of Los Angeles and LADWP are planning to form a system-wide Seismic-Resilient Pipe Network (SRPN) to ensure sustained water delivery at critical locations after an earthquake occurs. In particular, delivering water to critical customers responsible for public health and safety, e.g., hospitals, evacuation centers, and fire departments, as well as customers providing community resilience services, e.g., schools and community centers, is of utmost importance [6]. Earthquake risk zones also cover large residential and commercial areas that are relatively isolated from the critical customers targeted by the SRPN. Hence, it is also important to ensure that all residents are 'covered' by being within a certain proximity to running water via SRPN. However, given the complexity of the water network, the spatially varied seismic risk and locations of critical customers, and the limited resources available for infrastructure upgrades, the planning problem becomes complex. In this paper, we develop optimization approaches to strategically identify where old pipes should be replaced by seismic resilient pipes, respecting connectivity constraints for critical customers and coverage constraints for residential areas.

To address the challenges in planning the SRPN, we consider an edge-weighted network design problem with multiple water sources and connectivity constraints on nodes and edges. Our problem is closely related to the well-known Steiner tree problem [15], where the goal is to find a tree with minimum cost on a given graph that connects a set of so-called Steiner nodes. We refer to our problem as edge-weighted Steiner Network Problem with Coverage Constraints (SNP-CC): given an edge-weighted graph, a set of Steiner nodes, a set of source nodes, and coverage constraints represented by a collection of subsets of edges, find the set of edges with minimum cost that connects each Steiner node to a source node while satisfying the coverage constraints by including at least one edge of each subset in the solution. We provide hardness and inapproximability statements of the problems and present an efficient flow-based mixed-integer linear program (MILP) encoding of the problem with only a linear number of variables and constraints.

However, even if we can obtain the optimal plan for the SRPN, it is not practical to complete the installation of earthquake-resistant pipes in a short period due to the limited budget and resources, the large scale of planned areas, and the difficulty of on-site constructions. Since Los Angeles started fortifying the water network with earthquake-resistant pipes, a total of 13,600 feet (2.58 miles) of them have been installed so far and they plan to upgrade about 14 miles of pipes in a three-year period starting in 2018 [21], in contrast to the total length of 7,000 miles of pipe. In this situation, it becomes an important task to effectively determine which part of the optimal plan should be selected for each period of installment. We introduce an optimization problem to plan for optimal partial installments with the objective of satisfying as many of the water provision requirements as soon as possible over the planning time horizon within sequential budget allotments. We show that this problem is strongly NP-Hard to solve and propose a dynamic programming-based sequential planning algorithm.

In experiments, we test our proposed approaches on synthetic water network graphs and a large-scale real-world water service zone. We compare the performance with a myopic baseline and demonstrate that our algorithms are able to handle instances with almost 8,000 threatened pipes and provide high-quality water pipe network designs as well as efficient installment plans.

1.1 Related Work

Research related to seismic resilience of water infrastructure includes designing simulation frameworks for risk and resilience assessments [17, 23, 25] and optimizing post-earthquake water service recovery [3, 20]. Our work is closely related to studies in disaster response and pre-disaster planning, in which the goal is to optimize road network resilience and accessibility [13, 14, 24, 26], and facilitate emergency evacuation [10, 18, 22]. The line of research in network design and optimization for real-world applications and social good is also closely related. [9, 19] design wildlife conservation reserves that satisfy budget constraints and meet connectivity requirements by solving Steiner tree or network problems. [13] formulate the problem of maximizing road network resilience within a fixed budget as a budget-prize-collecting Steiner forest problem.

The Group Steiner Tree problem is related to our network design problem. In that problem, at least one node from each given group

of Steiner nodes should be included in the tree; while our design objective is a general network, we have constraints on subsets of edges as well as Steiner nodes. Theoretical results of hardness, as well as deterministic [4], randomized [12], primal-dual based [8] approximation algorithms for the Group Steiner Tree problem have been established. However, no practical tool or technique has been developed that we can borrow to solve our problem in large scales and meet real-world optimality requirements.

2 STEINER NETWORK PROBLEM WITH COVERAGE CONSTRAINTS

In this section, we first show how the SRPN planning problem can be formulated as a network design problem. Then, we formally define the edge-weighted Steiner Network Problem with Coverage Constraints (SNP-CC), present its computational complexity, and provide the MILP encoding to solve to problem.

To capture the SRPN planning problem as a network design problem, we model the water network as an undirected graph $G = (V, E)$, where pipes in the water network are edges and the intersections of pipes are nodes. The intersections of pipes are determined by pump stations, valves, hydrants, storages and other appurtenances in the network system that connect pipes. Each pipe e is associated with a non-negative replacement cost $c(e) \geq 0$. The set of water sources T is a subset of V, which corresponds to pumps, storages or the hoses of trunk pipes. Let C be the set of critical customers. Given the location of each critical customer, the customer is assigned to the closest node in the network from which it can extract water; therefore, we can assume $C \subseteq V$. Denote by R the set of housing areas. For each $r \in R$, let $S(r) \subseteq E$ be the set of pipes that are close enough to serve housing area r.

Formally, we define SNP-CC as follow:

Given: an undirected graph $G = (V, E)$, the sets of Steiner nodes C, source nodes T, coverage constraints R with $\{S(r) : r \in R\}$, a cost function on the edges $c : E \rightarrow \mathbb{R}^{\geq 0}$.

Find: a set of edges $E' \subseteq E$ with minimum cost $\sum_{e \in E'} c(e)$, such that all $u \in C$ and $r \in R$ are satisfied. We say that a Steiner node $u \in C$ is *satisfied*, if in the induced subgraph of E' u is connected to a source in T. We say r is *satisfied*, if there exists $e \in E' \cap S(r)$ connected to a source in T in the induced subgraph of E'.

W.l.o.g., we assume $T \cap C = \emptyset$ and $T \cap \{u : \exists (u, v) \in S(r)\} = \emptyset, \forall r \in R$.

2.1 Complexity Facts

Next, we present the following facts to show that this problem is hard to solve exactly and approximately.

Fact 2.1. SNP-CC *is NP-Hard to solve.*

Fact 2.1 follows since even without coverage constraints ($R = \emptyset$), SNP-CC is equivalent to the Steiner tree problem, which is known to be NP-Hard [16]. On the other hand, with only coverage constraints, the problem can be reduced to the weighted set cover problem, which is also a well-known NP-Hard problem [16].

Fact 2.2. SNP-CC *cannot be approximated to a factor of $o(\ln |R|)$ in polynomial time.*

Fact 2.2 follows from the fact that the weighted set cover problem with a set of elements R cannot be approximated to a factor of $o(\ln |R|)$. Consider an instance of the weighted set cover problem, where a set of elements R and a collection of subsets of R are given. The objective is to find the subsets with a minimum total weight that cover all elements in R. We build a star graph with a center node and edges representing each subset in the given collection. The cost of edges corresponds to the weight of subsets. For each element $r \in R$, let $S(r)$ be the set of edges corresponding to the sets that contain this element. Let T contains only the center node and $C = \emptyset$. If there exists a $o(\ln |R|)$-approximate algorithm for our problem, then the set cover problem can be $o(\ln |R|)$-approximated in polynomial time, leading to a contradiction.

2.2 MILP Formulation

In this section, we present our flow-based MILP formulation to solve the problem. At first glance of this problem, satisfying the constraints on C could be formulated using single commodity flow [9] and satisfying the coverage constraints for each $S(r)$ is equivalent to including an edge in the solution in any cut that separates $S(r)$ and T. This cut-based formulation shares the same ideas as the cut-based formulation for solving the Group Steiner Tree problem [12]. However, this formulation serves only for an expository purpose for theoretical analyses in previous work, which is formidable to use in practice since it consists of an exponential size of constraints. Following the primal-dual relationship between flow and cuts, the cut-based formulation can be translated to a flow-based formulation where each individual $r \in R$ requires a single commodity flow formulation on the graph, resulting in a total of $O(|R|(|E| + |V|))$ constraints and decision variables. Next, we present an efficient flow-based formulation that requires only $O(|V| + |E|)$ variables with $O(|V| + |E| + |R|)$ constraints.

We first transform the undirected graph G to a directed graph $\hat{G} = (V \cup \{0\}, \hat{E} \cup \{(0, t) : t \in T\})$ by replacing each undirected edge in E with two directed edges (collected in set $\hat{E}$) and adding a super source 0 that connects to all $t \in T$. Let $\delta^+(v)$ and $\delta^-(v)$ denote the sets of outgoing and incoming edges of vertex v in the transformed graph $\hat{G}$. Our overall Flow-Based MILP (FB-MILP) encoding of the problem is:

FB-MILP :

$$\min_{x,y,z} \sum_{(i,j) \in E} c(i,j)(x_{i,j} + x_{j,i})$$

$$\sum_{e \in \delta^-(v)} y_e = \mathbf{1}_{[v \in C]} + \sum_{e \in \delta^+(v)} (y_e + x_e) \qquad \forall v \in V \quad (1)$$

$$x_{i,j} + x_{j,i} \leq 1 \qquad \forall (i,j) \in E \quad (2)$$

$$\sum_{(i,j) \in S(r)} x_{i,j} + x_{j,i} \geq 1 \qquad \forall r \in R \quad (3)$$

$$0 \leq y_e \leq (|\hat{E}| + |V|)x_e \qquad \forall e \in \hat{E} \quad (4)$$

$$z + \sum_{t \in T} y_{0,t} = |\hat{E}| + |V| \qquad (5)$$

$$\sum_{t \in T} y_{0,t} = |C| + \sum_{e \in \hat{E}} x_e \qquad (6)$$

$$x_e \in \{0, 1\} \qquad \forall e \in \hat{E} \quad (7)$$

For each $e \in \hat{E}$, we introduce a binary variable x_e representing whether e is selected ($x_e = 1$) or not ($x_e = 0$) and a non-negative continuous variable y_e representing the number of units of flow on edge e. The main idea is that we take the nodes in T as the sources of the flow network and span the graph by sending flow out of nodes in T. We let each edge e consume 1 unit of flow if e is used in the solution, i.e., $y_e > 0 \Rightarrow x_e = 1$, which is ensured by Constraint (4). Constraint (1) models the flow conservation. In Constraint (1), a node will absorb 1 unit of flow if v is a Steiner node. Constraint (3) makes sure at least one edge from $S(r)$ is selected. We give the system a total of $|\hat{E}| + |V|$ units of flow since the edges and nodes can consume at most that many units of flow. Constraint (5) states that the residual flow z plus the total flow injected to the sources corresponds to the total flow, where $y_{0,t}$ ($t \in T$) represents flow injected to source t. Constraint (6) enforces that the flow consumed by the system corresponds to the total flow injected to sources.

3 OPTIMAL PARTIAL NETWORK INSTALLMENTS

Even if given the optimal plan to the SRPN planning problem, it is not practical to complete the whole plan in a short time due to the limited budget, the high cost of seismic materials, and the large scale of the planning area. An alternative approach for the water department is to split the plan into several installments, each given a certain amount of budget. Therefore, the key question raised given this setting is what pipes should be selected from the given global plan for each installment to maximize efficiency over time?

In this section, we first formally define the optimization problem for maximizing the efficiency of partial installments. Then, we show that the problem is strongly NP-Hard to solve and develop an efficient sequential algorithm for a computationally tractable case of the problem, which is also a realistic case in practice.

Suppose that given an instance of SNP-CC, we have already obtained the optimal solution represented by a set of edges E_{OPT}. We extend the cost function on edges to a set function for any subsets of E, i.e., let $c(E') = \sum_{e \in E'} c(e) \; \forall E' \subseteq E$. Denote by $B = c(E_{\text{OPT}})$ the total cost of the plan. We want to split the total cost B across n installments. Suppose the time horizon is $[0, 1]$ and the i-th installment is planned to be done at time $\frac{i-1}{n}$ with a budget of B_i allocated. Formally, we describe our problem as follow:

Given: an instance of SNP-CC and its optimal solution E_{OPT}, n time steps and $(B_1, \ldots, B_n)$ the budget allocation of the plan where $\sum_{i=1}^{n} B_i = B = c(E_{\text{OPT}})$, $U : 2^{E_{\text{OPT}}} \to \mathbb{R}^{\geq 0}$ an utility function that evaluates the efficiency of any partial plan.

Find: the installment plan $(E_1, \ldots, E_n)$ such that $\cup_{i \leq n} E_i = E_{\text{OPT}}$, $c(\cup_{j \leq i} E_j) \leq \sum_{j \leq i} B_j \; \forall i \leq n$, that maximizes the accumulated efficiency over time

$$\text{EFF} = \frac{1}{n} \sum_{i=1}^{n} U(\cup_{j \leq i} E_j).$$

The objective EFF can be seen as the integral of the time-efficiency function eff $: [0, 1) \to \mathbb{R}^{\geq 0}$ defined as eff$(t) := U(\cup_{j \leq i} E_j)$ if $t \in [\frac{i-1}{n}, \frac{i}{n})$ where $i \in [n]$. For an expository purpose, we define $U(E')$ to be the number of Steiner nodes satisfied by E' in the induced subgraph of E_{OPT} in the rest of this paper. This definition of

$U(\cdot)$ corresponds to the number of satisfied critical customers in the SRPN. However, we will see that our algorithm could accommodate a variety of utility functions, including any *non-negative additive set function $U(\cdot)$* that quantifies efficiency in practice, e.g., the number of leaky pipes fixed by the plan.

Next, we show the hardness of solving this problem in the following theorem.

THEOREM 3.1. *Finding the optimal installment plan is strongly NP-Hard.*

PROOF. We will show a reduction from the 3-partition problem (3PART), which is a strongly NP-Hard problem. In an instance of 3PART, we are given $3n$ positive integers $a_1, \ldots, a_{3n}$ and B_0, such that $B_0/4 < a_i < B_0/2$ and $\sum_{i=1}^{3n} a_i = nB_0$. B_0 is polynomially bounded, i.e., $B_0 \leq \text{poly}(n)$. We want to determine whether $\{a_i\}$ can be partitioned into n groups, each having 3 numbers and a sum of B_0.

Given the instance, we construct a graph $G = (V, E)$ where $V = \{v_i : 0 \leq i \leq 3n\}$ and $E = \{(v_0, v_i) : 1 \leq i \leq 3n\}$. The cost function on edges is given by $c(v_0, v_i) = a_i$. Let $E_{\text{OPT}} = E$ and the utility function is defined as $U(E) = c(E)$ for simplicity, which can be interpreted as having a_i Steiner nodes attached to node v_i with no extra cost. The budget allocation is given by $(B_1, \ldots, B_n) = (B_0, \ldots, B_0)$. We want to show that the answer is true to the 3PART instance iff. there exists a plan with EFF $= (n + 1)B_0/2$. Suppose there exists a 3-partition, then we let $E_i = \{(v_0, v_x), (v_0, v_y), (v_0, v_z)\}$ if the i-th group in the partition is $\{a_x, a_y, a_z\}$. Then we can easily verify the efficiency EFF $= (n + 1)B_0/2$. Conversely, since EFF $= (n + 1)B_0/2$ where the equality holds iff. $U(E_i) = B_0 \, \forall i \leq n$, and $B_0/2 < a_i < B_0/4$, we can infer that each E_i contains exactly 3 edges $\{(v_0, v_x), (v_0, v_y), (v_0, v_z)\}$ corresponding to (a_x, a_y, a_z) the i-th group in the 3-partition solution. Since B_0 and a_i are polynomially bounded, solving the problem is NP-Hard in the strong sense. □

To circumvent the computational hardness, we provide a sequential planning algorithm (SeqPlan), as shown in Algorithm 1. In Algorithm 1, we first initialize $E' = \emptyset$ and EFF $= 0$ (line 1) and plan for time step $1, \ldots, n$ sequentially (line 2). Given the current plan $E' = \cup_{j<i} E_j$ at each time step i, we greedily choose E_i that leads to the largest increment in utility. In other words, we define

$$\text{PartialPlan}(E', B') = \arg\max_{E'_i : c(E'_i \cup E') \leq B'} U(E'_i \cup E')$$

and let $E_i = \text{PartialPlan}(E', \sum_{j \leq i} B_i)$ (line 4). Then, we add E_i to the current plan (line 5) and update EFF (line 6).

PROPOSITION 3.2. SeqPlan *is optimal when $n = 2$.*

Next, we give the definition of non-overlapping coverage constraints and explore its property in Proposition 3.4.

DEFINITION 3.3. *(Non-overlapping coverage constraints) For any $r \in R$, let $V(r) = \{u : \exists(u, v) \in S(r)\}$ be the set of nodes in the induced subgraph of $S(r)$. We have non-overlapping coverage constraints if $V(r_i) \cap V(r_j) = \emptyset \, \forall r_i, r_j \in R$.*

PROPOSITION 3.4. *Assuming non-overlapping coverage constraints, there exists an optimal solution to SNP-CC that forms a forest, i.e.,*

Algorithm 1 Sequential Planning Algorithm

1: $E' = \emptyset$, EFF $= 0$
2: **for** $1 \leq i \leq n$ **do**
3:      $B' = \sum_{j=1}^{i} B_j$
4:      $E_i = \text{PartialPlan}(E', B')$
5:      $E' = E' \cup E_i$
6:      EFF $= \text{EFF} + \frac{1}{n} U(E')$
7: **return** $(E_1, \ldots, E_n)$, EFF

---

the induced graph of E_{OPT} is a forest. Each component in the forest contains exactly one node in T.

In the rest of this section, we will focus on the case assuming non-overlapping coverage constraints and show how we find the *optimal* PartialPlan(E', B') using dynamic programming (DP). For the utility function, since E_{OPT} forms a forest, we can let $U(e) = 1$ if the child node of e is a Steiner node and $U(e) = 0$ if not. To find PartialPlan(E', B'), w.l.o.g. we assume that all edges in E' are connected to a source in T; otherwise, we could always discard those edges. Then clearly, the additive set function $U(E') = \sum_{e \in E'} U(e)$ gives the number of Steiner nodes satisfied by E'. Next, we contract all the components connected by E' in the induced graph of E_{OPT}, and the contracted graph forms a forest with each component containing one source. We can transform the contracted graph into a rooted tree by adding edges between a virtual root and nodes that contain a source. The optimal solution can be computed using a DP, which essentially solves a knapsack problem on a tree. Specifically, assuming v is already satisfied, we can compute $\text{OPT}(v, b)$ defined to be the maximum utility can be achieved in the subtree of v, respecting the budget constraint b. Since the computation of $\text{OPT}(v, b)$ is standard, we omit the details.

Remark 3.5. $\text{OPT}(v, b)$ can be computed in polynomial time using DP since $v \leq U(E_{\text{OPT}}) \leq |E_{\text{OPT}}|$ is polynomially bounded and discretized. This DP-based approach to find PartialPlan(E', B') can be extended to accommodate any non-negative addtive set function $U(\cdot)$ under the assumption of non-overlapping coverage constraints. Even under this assumption, finding PartialPlan(E', B') is NP-Hard in general if $U(\cdot)$ and $c(\cdot)$ are non-integer. However, we can find approximately optimum solutions by discretizing the utility on edges.

4 EXPERIMENT

In experiments, we test our algorithms on synthetic water pipe networks utilizing open-source road data and conduct a case study on a water service zone in Los Angeles. We provide extensive results to demonstrate the effectiveness of our proposed methods.

4.1 Data Description and Preprocessing

We apply our approaches on synthetic water pipe networks and the Service Zone 1134 in Los Angeles, where we utilize three pieces of data: i) road network data as a surrogate for water pipe networks from OpenStreetMap [1] for the synthetic setting and a GIS representation of the water pipe network in Zone 1134, including connectivity information, geographical data and basic features of water pipes and joints; ii) geographical data of critical customers;
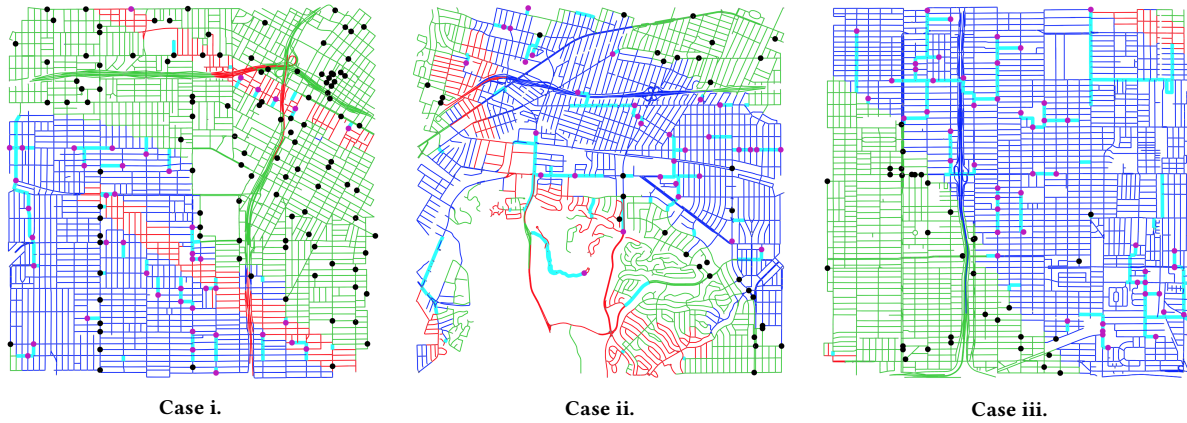
Figure 1: Visualized solutions to 4 **miles**$\times 4$ **miles instances in Cases i, ii, and iii using roads as surrogates for water pipes. Pipes chosen by our algorithms are highlighted using thick cyan lines. Red and blue pipes are threatened pipes crossing fault zones and within liquefaction areas, respectively. Thin green pipes are non-threatened pipes. Safe customers, who are connected to water sources via safe pipes and located near to a water source, are marked as black dots; threatened customers are marked as pink dots. Water sources are not displayed in the figures.**
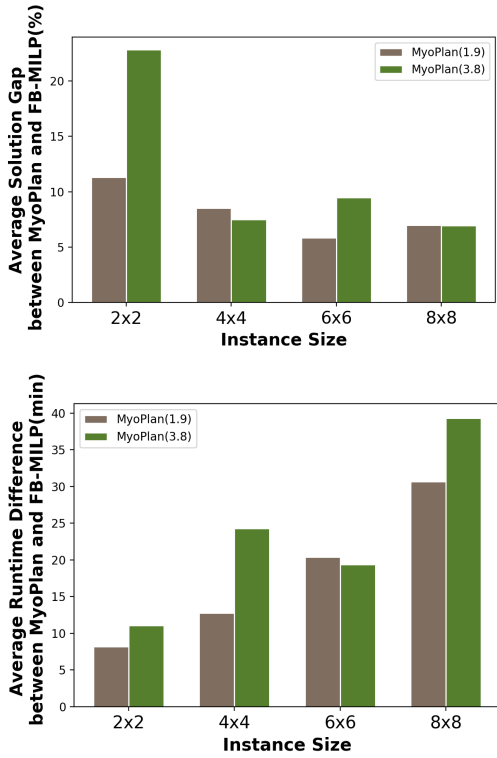

Figure 2: Average solution gaps and runtime differences between FB-MILP **and** MyoPlan **for different instance sizes on synthetic water pipe networks derived from Table 1.**

iii) GIS representations of fault zones and liquefaction areas in Los Angeles. In the following, we provide a brief description of each

piece of data as well as how we extract input parameters to our problem.

The connectivity data of the service zone describes how pipes are linked together by pumps, valves, storages, hydrants, etc. In addition, we have access to the latitude-longitude coordinates and features (e.g., the age and diameter of pipes, the type of valves) of all the elements in the network. From the data, we construct the undirected graph representation $G = (V, E)$ of the water pipe network, with edges E representing pipe segments and vertices V representing coordinates of the joints. In our study area, pipes with diameter no less than 24 inches are identified as trunk pipes and are the main water sources of Zone 1134. Besides, we utilize geospatial data made available by OpenStreetMap [1] for synthetic representations of water pipe network in Los Angeles. We use the road network to approximate the water pipe networks, where we use the road segments as a surrogate for pipes location. We crop the map using L miles $\times L$ miles bounding boxes at different locations and use the obtained road networks within the boxes as substitutes of G.

The data of critical facilities provides latitude-longitude coordinates of critical customers. We assign each customer to its closest node in V and collect those nodes as the set of critical customers C.

To determine the set of R, we use a downsampling method. We divide the bounding box of Zone 1134 into 12×14 grids, for each grid, find the node closest to its center, and collect them as R. For each $r \in R$, $S(r)$ is the set of edges within 3-hop neighborhood of r in the graph G. For synthetic water pipe network graphs, we divide the L miles $\times L$ miles region into $\left\lceil \frac{L}{0.7} \right\rceil \times \left\lceil \frac{L}{0.7} \right\rceil$ grids and apply the same method. Ideally, we want to ensure that every node in the graph is within a certain proximity to a water supply. We provide justifications for our downsampling method in the following: assume the grid size is $a \times b$, covering the set of edges in the center of the grid guarantees the largest distance between any node and the earthquake-resistant pipe closest to the node is not larger than

		2 x 2				4 x 4			
	Scale	(16, 345, 214, 367)				(44, 1529, 864, 1541)			
		Cost (Gap)	Time	EFF		Cost (Gap)	Time	EFF	
				MyoPlan	SeqPlan			MyoPlan	SeqPlan
	FB-MILP	**2.71(0)**	0	MyoPlan	SeqPlan	**7.43(0)**	1.93	MyoPlan	SeqPlan
	MyoPlan(1.9)	3.00(10.9%)	0.5	**14.0**	13.5	8.44(13.6%)	17.4	34.0	**34.8**
	MyoPlan(3.8)	3.27(21.0%)	0	**16.0**	**16.0**	7.74(4.17%)	32.4	36.5	**37**
i		6 x 6				8 x 8			
	Scale	(87, 3371, 1943, 3427)				(128, 5738, 3370, 5770)			
		Cost (Gap)	Time	EFF		Cost (Gap)	Time	EFF	
	FB-MILP	**16.19(0.78%)**	90	MyoPlan	SeqPlan	**24.32(1.91%)**	120	MyoPlan	SeqPlan
	MyoPlan(1.9)	16.75(4.27%)	101	60.0	**62.7**	26.02(9.04%)	145	87.43	**90.2**
	MyoPlan(3.8)	18.37(14.4%)	97.7	66.8	**69.2**	26.19(9.79%)	133	91.57	**95.43**
		2 x 2				4 x 4			
	Scale	(14, 502, 378, 550)				(34, 2021, 1335, 2087)			
		Cost (Gap)	Time	EFF		Cost (Gap)	Time	EFF	
	FB-MILP	**4.47(0)**	0.3	MyoPlan	SeqPlan	**8.38(0)**	1.36	MyoPlan	SeqPlan
	MyoPlan(1.9)	4.60(3.00%)	0.71	11.0	**12.0**	8.63(3.01%)	15.6	25.4	**26.9**
	MyoPlan(3.8)	4.56(2.14%)	3.78	**13.5**	12.0	8.87(5.83%)	28.7	29	**31.0**
ii		6 x 6				8 x 8			
	Scale	(57, 3944, 2367, 3912)				(102, 6187, 3626, 6127)			
		Cost (Gap)	Time	EFF		Cost (Gap)	Time	EFF	
	FB-MILP	**13.29(0)**	9.43	MyoPlan	SeqPlan	**21.93(0)**	74.07	MyoPlan	SeqPlan
	MyoPlan(1.9)	14.42(8.58%)	47.5	41.13	**44.63**	23.88(8.89%)	119	72.38	**76.46**
	MyoPlan(3.8)	14.31(7.65%)	42.9	44.75	**47.75**	22.83(4.10%)	147	74.17	**78.67**
		2 x 2				4 x 4			
	Scale	(20, 678, 401, 722)				(55, 2473, 1590, 2679)			
		Cost (Gap)	Time	EFF		Cost (Gap)	Time	EFF	
	FB-MILP	**4.22(0)**	0.6	MyoPlan	SeqPlan	**15.24(7.17%)**	60	MyoPlan	SeqPlan
	MyoPlan(1.9)	5.06(20.0%)	24.2	16.33	**17.33**	16.61(16.8%)	68.5	36.77	**38.77**
	MyoPlan(3.8)	5.63(33.4%)	30.2	19.0	**19.5**	17.15(20.6%)	75.0	42.0	**44.0**
iii		6 x 6				8 x 8			
	Scale	(82, 4557, 2897, 4853)				(99, 7485, 4680, 7861)			
		Cost	Time	EFF		Cost	Time	EFF	
	FB-MILP	**24.51(9.08%)**	90	MyoPlan	SeqPlan	**36.99(10.05%)**	120	MyoPlan	SeqPlan
	MyoPlan(1.9)	25.86(15.1%)	102	53.21	**58.93**	38.85(15.6%)	142	72.33	**79.24**
	MyoPlan(3.8)	26.29(17.0%)	107	57.43	**62.43**	40.34(20.0%)	152	76.73	**84.10**

Table 1: Experiment results on synthetic water pipe networks in Cases i, ii and iii. The cost and budget are in miles, the runtime is in minutes. The gaps are the optimality gaps. The problem scale is described by a tuple, where the four figures are the numbers of threatened critical customers, threatened pipes, nodes and edges in the contracted graph. MyoPlan **with budget B_0 is denoted by** MyoPlan(B_0)**.**

$\sqrt{a^2 + b^2}/2$ approximately. Equivalently, we can divide the region into smaller grids of size $\frac{a}{2} \times \frac{b}{2}$, and define R to be the set of grids and $S(r)$ be all the pipes in the corresponding grid $r \in R$. However, the latter method leads to a larger number of coverage constraints and larger sizes of $S(r)$, which empirically leads to a 14% increase in cost for our case study on Zone 1134. Our method guarantees non-overlapping coverage constraints and 0.75-mile proximity to water supply for all nodes empirically in all solutions generated by baselines and our algorithms in experiments.

In GIS representations, the fault zones are represented by segments, and the liquefaction areas are represented by polygons. We say that a pipe is threatened if it touches a liquefaction area or is within 500 feet of a fault zone, which is very likely to be damaged

during an earthquake. All trunk pipes are assumed to be earthquake-resistant and not considered as threatened pipes. The cost function on edges represents the replacement cost of the corresponding pipes. Assuming only threatened pipes will fail and need replacement, we assign a cost equal to its lengths for each threatened pipe and a zero cost for the others.

We provide statistics about Zone 1134 in Table 2a. The numbers of nodes and pipes are the size of V and E, respectively. Safe pipes are guaranteed connection to water sources even if all threatened pipes fail. Besides safe pipes and threatened pipes, there are 449 pipes connected to water sources via at least one threatened pipe and will be isolated from the water supply in the worst case. In addition, we count the number of non-covered nodes and threatened

critical customers. Non-covered nodes are nodes not within 1 mile proximity to any safe pipe or water source. Threatened critical customers are those connected to water sources via at least one threatened pipe and will lose direct water supply in the worst case.

4.1.1 Graph Contraction. Typically, the areas impacted by fault zones and liquefaction areas form continuous regions on the map, which correspond to components of threatened pipes with non-zero cost in graph G. The rest of the graph consists of large components of pipes with zero replacement cost. Given such component $G_C = (V_C, E_C)$ where $V_C \subseteq V$ and $E_C \subseteq E$, to speed up computation, we can contract V_C and replace it with a new node v_C that represents component G_C. If G_C contains a source, i.e., $V_C \cap T \neq \emptyset$, then Steiner nodes within V_C and coverage constraints that overlap with E_C are satisfied for free. All edges in E_C can be removed since we have contracted V_C, and for all $v \in V \setminus V_C$ that has edges connecting to nodes in V_C, we replace them with an edge (v, v_C) with cost $\min_{(v,v') \in E: v' \in V_C} c(v, v')$. Note that after graph contraction, not only the number of nodes but also the number of edges with non-zero cost reduces.

4.2 A Myopic Baseline

When given a specific limit of budget, decision-makers may plan greedily based on what has been planned and installed so far, without the guidance of the optimal global solution. Thus, we propose a myopic planning algorithm (MyoPlan) that works as follow: i) initially, $E_{\text{MyoPlan}} = \emptyset$; ii) Given budget B_0, extend E_{MyoPlan} by optimizing the number of newly-satisfied customers and housing areas, such that the total cost of newly added edges does not exceed B_0; iii) if not all customers and housing areas are satisfied, continue to ii), otherwise, return E_{MyoPlan} and its total cost. Explicitly, Step ii) in MyoPlan can be computed using BC-MILP, a budget-constraint version of the flow-based MILP. For simplicity, we provide the formulation for the first iteration (i.e., $E_{\text{MyoPlan}} = \emptyset$; we can contract edges in E_{MyoPlan} after each iteration and easily reapply the same MILP on the contracted graph):

BC-MILP :

$$\max_{x,y,z,w} \sum_{v \in C} w_v + \sum_{r \in R} w_r$$

Constraint (2),(4),(5),(6),(7)

$$\sum_{e \in \delta^-(v)} y_e = \sum_{e \in \delta^+(v)} (y_e + x_e) \qquad \forall v \in V$$

$$\sum_{(i,j) \in E} c(i, j)(x_{i,j} + x_{j,i}) \leq B_0$$

$$w_v \leq \sum_{e \in \delta^+(v)} x_e, w_v \in \{0, 1\} \qquad \forall v \in C$$

$$w_r \leq \sum_{(i,j) \in S(r)} x_{i,j} + x_{j,i}, w_r \in \{0, 1\} \qquad \forall r \in R$$

In the formulation above, we introduce a set of new binary variables w_v (w_r) to represent whether customer $v \in V$ (housing area $r \in R$) is satisfied.

4.3 Experiment Result on Synthetic Water Pipe Networks

We run three sets of experiments (labeled as case i, ii and iii) on each of the three locations we picked on the map of Los Angeles. For each location, we use its coordinate as the center to crop the road networks from OpenStreetMap using $L \times L$ ($L = 2, 4, 6, 8$ miles) bounding boxes and approximate the water pipe network G using the corresponding road network graph. For each L, we run FB-MILP and MyoPlan ($B_0 = 1.9, 3.8$ miles) and compare their runtime and solution quality. The cutoff time for FB-MILP to solve a $L \times L$ instance is set to $15L$ minutes. For MyoPlan, we first estimate the time step needed as $n' = \lceil c(E_{OPT})/B_0 \rceil$ and set the cutoff time to $18L/n'$ minutes for each solve of BC-MILP. We also run SeqPlan and compare the plan efficiency with MyoPlan. To set up SeqPlan given B_0 the budget for each installment, we plan for n installments over time horizon $[0, 1]$ where n is set to the number of time steps required by MyoPlan, i.e., $n = \lceil \frac{c(E_{\text{MyoPlan}})}{B_0} \rceil$. For the budget allocation, we simply set $B_i = B_0$ if $iB_0 \leq c(E_{OPT})$ otherwise, $B_i = \max\{c(E_{OPT}) - (i-1)B_0, 0\}$. Results are shown in Table 1, and we highlight the solution gap and runtime difference between MyoPlan and FB-MILP in Figure 2. FB-MILP outperforms MyoPlan in all cases in terms of runtime and solution quality. The EFF of MyoPlan can be calculated in linear time after MyoPlan finds its solution. The EFF of SeqPlan takes less than two seconds to compute in all instances. SeqPlan outperforms MyoPlan on most instances, except on two 2×2 instances in case i and ii. Since we only need to plan for two time steps in these cases, MyoPlan only needs to outperform in the first step to beat SeqPlan. This exception does not contradict with Proposition 3.2 since MyoPlan is not restricted to choose edges from E_{OPT}. We visualize the solutions of FB-MILP to the 4 miles×4 miles instances in Figure 1.

4.4 Case Study on Zone 1134

4.4.1 The Global Optimal Plan. Given the instance to SNP-CC extracted from data, we run FB-MILP to compute the optimal solution. We also run MyoPlan with different budgets. In Table 2b, we compare the solution quality and runtime with MyoPlan. FB-MILP is able to find the optimal solution with cost $c(E_{OPT}) = 23.47$ in 18.05 minutes. While in MyoPlan, we need to run BC-MILP several times, and it takes BC-MILP a relatively long time to find the optimal solution. Thus we set the cutoff time for each solve of BC-MILP to 150 seconds for $B_0 = 1.9, 2.5$ miles, 200 seconds for $B_0 = 3.2, 3.8$ miles, and 250 seconds for $B_0 = 4.8$ miles. We can see that MyoPlan with different budgets finds sub-optimal solutions that cost 3.03% to 4.53% higher than E_{OPT}.

4.4.2 The Optimal Plan for Partial Installment. Given the global optimal plan E_{OPT}, we investigate how we should plan for partial installments. We show the efficiency of SeqPlan and compare it with MyoPlan that plans without the guidance of the optimal solution. The setup of SeqPlan is the same as in previous experiments. We set $B_0 = 1.9, 2.5, 3.2, 3.8, 4.8$ miles and show the solution quality in Table 2b. The EFF of SeqPlan takes less than 2 seconds to compute in all instances. We can see that in all cases, SeqPlan that plans upon the optimal solution dominates MyoPlan in plan efficiency. In Figure 3a and 3b, we show the utility curves for $B_0 = 3.1$ miles and

Pipes	34,462
Safe Pipes	25,609
Threatened Pipes	8,434
Nodes	31,674
Threatened Nodes	232
Critical Customers	298
Threatened Critical Customers	93
Nodes in Contracted Graph	7607
Edges in Contracted Graph	7844

(a) Statistics about Zone 1134.

	Cost (Gap)	Time/min	EFF	
			MyoPlan	SeqPlan
FB-MILP	**23.47 (0)**	**18.05**		
MyoPlan(1.9)	24.37 (3.83%)	31.76	59.30	**59.50**
MyoPlan(2.5)	24.53(4.53%)	29.38	57.40	**58.30**
MyoPlan(3.2)	24.18 (3.03%)	23.35	57.38	**58.00**
MyoPlan(3.8)	24.22 (3.21%)	22.55	56.8	**57.71**
MyoPlan(4.8)	24.30 (3.53%)	20.88	57.80	**60.16**

(b) Solution quality of FB-MILP and MyoPlan, and plan efficiency of SeqPlan and MyoPlan. The SeqPlan column is computed based on the solution of FB-MILP using Algorithm 1.

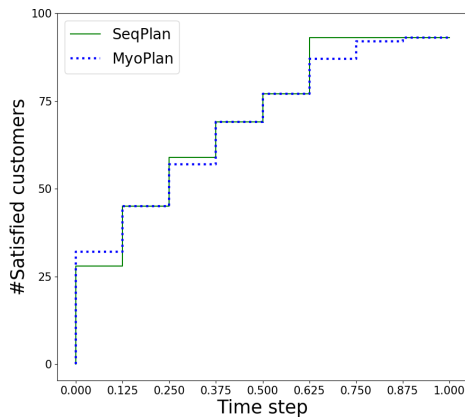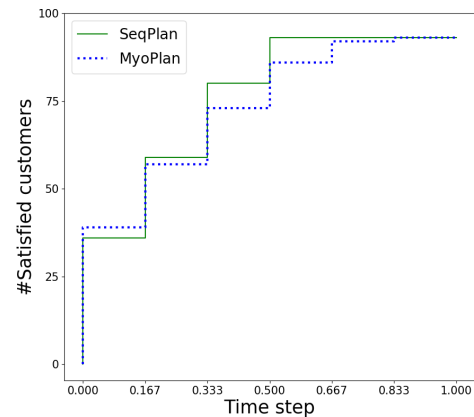**Table 2: Statistics and results of Zone 1134.**

(a) The utility curve for $B_0 = 3.2$ miles.

(b) The utility curve for $B_0 = 4.8$ miles.

Figure 3: Utility curves.

4.8 miles respectively. In both cases, MyoPlan achieves a higher utility for the first time step. However, after the second time step, both curves of MyoPlan are dominated by the curves of SeqPlan and take two more time steps than SeqPlan to cover all customers.

5 CONCLUSION AND FUTURE WORK

In this paper, we addressed the problem of strategically planning SRPN and introduce the Steiner Network Problem with Coverage Constraints. To solve the network design problem, we provided an efficient flow-based MILP. To maximize the efficiency of installing the plan, we proposed a sequential planning algorithm to find the optimal plan for partial network installments. In experiments, we showed that our methods performed better than a myopic baseline on both synthetic and real-world water pipe networks in terms of solution quality and computation time. Importantly, our methods demonstrated the potential to scale well to solve large practical water pipe network design and planning problems, and the results can serve as useful guidance for decision-makers when planning SRPN.

To further evaluate the social impact of our work, we could compare our solution to what has been done since the city started fortifying the network and test it on main network failures related to prior earthquakes. In our model, the benefits from replacing pipes are only accrued when a critical customer is fully connected to a source via seismic resilient pipes. For future work, we could have stochastic costs on pipes and develop algorithms to minimize the risk of a water pipe network failure.

REFERENCES

[1] 2020. OpenStreetMap. https://www.openstreetmap.org/. Accessed: 2020-03-01.
[2] 2020. Sustainable Development Goals. https://sustainabledevelopment.un.org/. Accessed: 2020-03-01.
[3] Xavier Bellagamba, Brendon A Bradley, Liam M Wotherspoon, and Walter D Lagrava. 2019. A Decision-Support Algorithm for Post-Earthquake Water Services Recovery and Its Application to the 22 February 2011 Mw 6.2 Christchurch Earthquake. *Earthquake Spectra* 35, 3 (2019), 1397–1420.
[4] Moses Charikar, Chandra Chekuri, Ashish Goel, and Sudipto Guha. 1998. Rounding via trees: deterministic approximation algorithms for group Steiner trees and k-median. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing.* 114–123.
[5] CA Davis, TD O'Rourke, ML Adams, and MA Rho. 2012. Case study: Los Angeles water services restoration following the 1994 Northridge earthquake. In *15th*

World Conference on Earthquake Engineering. 24–28.

[6] Craig A Davis. [n.d.]. Developing a seismic resilient pipe network using performance based seismic design procedures.

[7] Craig A Davis. 2017. Developing a seismic resilient pipe network using performance based seismic design procedures. https://www.ncree.org/conference/UserData/0/I20171018A/Papers/WSSC10_P-03.pdf.

[8] Erik D Demaine, MohammadTaghi Hajiaghayi, and Philip N Klein. 2009. Node-weighted Steiner tree and group Steiner tree in planar graphs. In *International Colloquium on Automata, Languages, and Programming*. Springer, 328–340.

[9] Bistra Dilkina and Carla P Gomes. 2010. Solving connected subgraph problems in wildlife conservation. In *International Conference on Integration of Artificial Intelligence (AI) and Operations Research (OR) Techniques in Constraint Programming*. Springer, 102–116.

[10] Caroline Even, Victor Pillac, and Pascal Van Hentenryck. 2015. Convergent plans for large-scale evacuations. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

[11] Eric Garcetti. 2018. Resilient Los Angeles. https://www.lamayor.org/sites/g/files/wph446/f/page/file/Resilient%20Los%20Angeles.pdf.

[12] Naveen Garg, Goran Konjevod, and R Ravi. 2000. A polylogarithmic approximation algorithm for the group Steiner tree problem. *Journal of Algorithms* 37, 1 (2000), 66–84.

[13] Amrita Gupta and Bistra Dilkina. 2019. Budget-Constrained Demand-Weighted Network Design for Resilient Infrastructure. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 456–463.

[14] Amrita Gupta, Caleb Robinson, and Bistra Dilkina. 2018. Infrastructure resilience for climate adaptation. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. 1–8.

[15] Frank K Hwang and Dana S Richards. 1992. Steiner tree problems. *Networks* 22, 1 (1992), 55–89.

[16] Richard M Karp. 1972. Reducibility among combinatorial problems. In *Complexity of computer computations*. Springer, 85–103.

[17] Katherine A Klise, Michael Bynum, Dylan Moriarty, and Regan Murray. 2017. A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. *Environmental Modelling & Software* 95 (2017), 420–431.

[18] Kunal Kumar, Julia Romanski, and Pascal Van Hentenryck. 2016. Optimizing infrastructure enhancements for evacuation planning. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[19] Ronan Le Bras, Bistra Dilkina, Yexiang Xue, Carla Gomes, Kevin McKelvey, Michael Schwartz, and Claire Montgomery. 2013. Robust network design for multispecies conservation. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

[20] Milad Memarzadeh and Matteo Pozzi. 2019. Model-free reinforcement learning with model-based safe exploration: Optimizing adaptive recovery process of infrastructure systems. *Structural Safety* 80 (2019), 46–55.

[21] Los Angeles Department of Water and Power (LADWP). 2017. Briefing book: 2017-2018. https://s3-us-west-2.amazonaws.com/ladwp-jtti/wp-content/uploads/sites/3/2017/09/08143247/Briefing-Book-Rolling-PDF.pdf.

[22] Julia Romanski and Pascal Van Hentenryck. 2016. Benders decomposition for large-scale prescriptive evacuations. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[23] Taronne Tabucchi, Rachel Davidson, and Susan Brink. 2010. Simulation of post-earthquake water supply system restoration. *Civil Engineering and Environmental Systems* 27, 4 (2010), 263–279.

[24] Xiaojian Wu, Daniel Sheldon, and Shlomo Zilberstein. 2016. Optimizing resilience in large scale networks. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[25] Sungsik Yoon, Young-Joo Lee, and Hyung-Jo Jung. 2018. A comprehensive framework for seismic risk assessment of urban water transmission networks. *International journal of disaster risk reduction* 31 (2018), 983–994.

[26] Eda Yücel, F Sibel Salman, and Idil Arsik. 2018. Improving post-disaster road network accessibility by strengthening links against failures. *European Journal of Operational Research* 269, 2 (2018), 406–422.