On the Complexity of Disaster-Aware Network Extension Problems

Balázs Vass*†, Balázs Brányi*, Beáta Éva Nagy*, János Tapolcai*

* MTA-BME Future Internet Research Group and ELKH-BME Information Systems Research Group, High-Speed Networks Laboratory (HSN*Lab*), Department of Telecommunication and Media Informatics, Faculty of Electrical Engineering and Informatics (VIK), Budapest University of Technology and Economics (BME), Budapest University of Technology and Economics, {balazs.vass,balazs.branyi,beata.nagy,tapolcai}@tmit.bme.hu

† EIT Digital Industrial Doctoral School

***Abstract*—In this paper, we deal with the complexity of problems related to finding cost-efficient, disaster-aware cable routes. In particular, we compare two very different versions of the problem. In the first version, only the worst-case scenarios are considered, while in the second, the probability of the disasters is also known, and the aim is to find the most cost-efficient solution in terms of investment cost and risk of a network outage. Worst-case computations allow using efficient computational geometry algorithms, and even large networks can be analyzed. On the other hand, to find the cost-efficient soltuion, a lot of empirical hazard data must be processed, and heavy algorithms must be used. In this paper, we study the benefits and drawbacks of each model. In particular, we define multiple versions of the problem and investigate their complexity.**

I. Introduction

Several studies revealed how vulnerable the Internet backbone is to natural disasters, such as earthquakes, hurricanes, and tsunamis, which may destroy several nodes and links located in possibly a few hundred kilometers wide geographic areas [1]–[13]. This study focuses on the problems related to installing new fiber cables [14]–[16], which are effective only if the new cables are properly placed. The first studies used a simplified model, where the risk of a cable cut depends only on the location, and the possible failures are independent of each other.

Recently, two papers [17], [18] have been presented in the same section of a conference that already consider more sophisticated failure models, where natural disasters are modeled as a failure of a geographic area, and thus the correlation of failures is considered. Although the engineering problems are similar, the two studies use different mathematical models and result in slightly controversial conclusions. In particular, [17] mostly focuses on subproblems that can be solved with polynomial-time algorithms, while [18] shows even the minimalistic problem is NP-hard and focuses on heuristic algorithms. This paper first describes the differences in the mathematical models, then investigates their benefits and weaknesses. Our focus is mainly on algorithmic complexity aspects: defining a problem in one way, we have efficient polynomial algorithms, otherwise, we are facing a clearly NP-hard problem.

The paper is organized as follows. Sec. II provides an introduction and comparison of problem formulations of [17] and [18]. Sec. III presents the main contributions of the paper that is the NP-hardness and inapproximability proofs of several closely related network extension problem formulations, including the one tackled in [17]. In Sec. IV, a discussion on the model of [18] is presented. Finally, Sec. V and VI, presents our future research directions, and concludes the paper, respectively.

II. Models and Assumptions

The network is modeled as an undirected connected geometric graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ with $n = |\mathscr{V}|$ nodes and $m = |\mathscr{E}|$ edges[1]. The nodes of the graph are embedded as points in the Euclidean plane, and each edge is considered to be a finite sequence of line segments[2]. The task is to extend the network topology with new links such that its vulnerability against natural disasters decreases. The outcome of the algorithms is a set of new links (end nodes of each link) with the exact route of the cable (between the endpoints). The objective is to minimize the installation cost of the new cables which is related to the total cable length.

Both studies assume a disaster destroys some set of points in the plane, and if a link or node intersects with any of these disaster points, it is destroyed. The first key difference is that [17] assumes the points of **a disaster must be connected**, while in [18] there is no such restriction. Our first observation is that

Observation 1. *The problem defined in [18] is NP-hard [18, Theorem 1]; however, it is not known whether it remains NP-hard if every disaster has to be a connected set of points.*

Note that the construction in the NP-hardness proof in Theorem 1 of [18] builds on the existence of disasters that result in a simultaneous failure of multiple network elements very far from each other, see Sec. IV-B for more discussion. On the other hand, [17] assumes if multiple network elements fail simultaneously because of a disaster, then they must be in the same geographic area. The first key difference between the models is that such geometric constraints of the failure significantly simplify the underlying algorithmic

[1] [18] defines the network topology as a directed graph.

[2] [17] assumes the links are a line segment, but it can be easily extended.

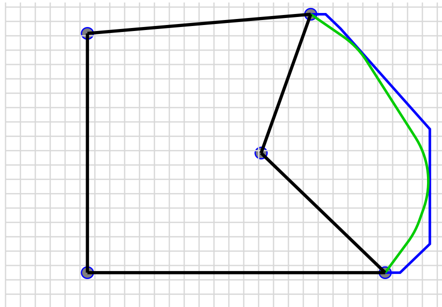


Fig. 1: The blue line segments illustrate the construction of a new edge in the cost-efficient [18] model: the edge is along a predefined grid, where it can take a vertical, horizontal, or diagonal line segment. The green curve demonstrates the shape of the added cables in the worst-case [17] model: it avoids a circular area to survive disk failures.

problem. When the aim is to prepare the network for worst-case scenarios, it is realistic to assume a disaster in practice damages network equipment in a connected geographic area. Otherwise, there might be regional failures, where multiple areas are damaged close to each other.

The second key difference is the final goal: whether the **worst-case scenarios** should be considered [17], or the probability of the disasters is also known, and the aim is to find a **cost-efficient** investment cost as a trade-off with the risk of a network outage [18]. Worst-case computations allow using very efficient computational geometry algorithms, and even large networks can be analyzed. On the other hand, to find a cost-efficient solution, a lot of historical hazard data must be processed, and heavy algorithms must be used. We note here that, in [18], the trade-off for the cost-efficient approach is defined through a parameter α representing the relative importance of high availability compared to the upgrade cost.

In the study [18], the set of disasters is predefined ($\sim 500\,000$ in the simulations), and each disaster can destroy an arbitrary set of points in the plane. For each predefined disaster, a probability is also assigned, such that these probabilities are summed up to 1 in the input. This failure model is similar to Probabilistic Shared Risk Link Group failures, where the set of failure events with their probability is known in network planning [13], [19]–[22]. In this disaster model, it is challenging to define whether a new cable is affected by a disaster if the cable can take any arbitrary route. To mitigate this problem, [18] limits the route of the cable to traverse along a 2D grid. More precisely, the cable should traverse the points of a predefined grid, where it can take a vertical, horizontal, or diagonal line segment, see the blue link on Fig. 1 as an example.

The problem input size strongly depends on the size of the cells. In the simulation, the authors evaluate 5.5km cells (0.05 degree) over Italy. Another drawback of mapping the routes to the grid is that it introduces an error in its length evaluation. Compared to Euclidean distances, the excess length can be as much as 8%, see Section IV-A.

However, in practice, the cost of implementing a cable strongly depends on the location, which can be easily included in the grid model. Note that the failure probabilities are computed using stochastic models of the disasters, where the failed points form circular disks in the simulation [13], [19].

On the other hand, [17] aims to provide a worst-case analysis, where the network should be prepared for an infinite number of possible disaster scenarios. Here the probability of each disaster is ignored, and the solution should survive any possible disaster. This is a fundamentally different concept, where the route of the cables should not be restricted to a grid, see the green curve on Fig. 1 as an example. It utilizes the *disk failure model with fixed radius*, where a regional failure occurs at a point known as the epicenter that corresponds to an area in the shape of a circular disk c of radius r. All network elements intersecting with the area might be affected, and all other network elements are unaffected. This worst-case model results in significantly smaller problem sizes. For worst-case analysis, the latter failure model boils down to preparing the network to survive the joint failure of the links intersecting a circular disk of the given radius, formally

Definition 1. *A circular disk* ***failure*** *c* ***hits*** *an edge e if e intersects the interior of disk c. Similarly, node v is hit by failure c if it is in the interior of c. Let $\mathcal{E}_c$ (and $\mathcal{V}_c$) denote the set of edges (and nodes) hit by disk c.*

Apart from the above differences, the two models adopt a similar objective. Both models focus on connectivity, thus, both ignore the routing. In other words, if the network remains connected after the failure, both assume the nodes can communicate. The study [18] is more general in two aspects,

1) defines an impact metric (denoted by M and similar to a traffic matrix), and
2) defines the cost of losing a connection between two nodes because of a disaster (denoted by α) relative to the cost of implementing 1 km optical fiber.

Note that the worst-case model is equivalent to $\alpha = \infty$. M defines how important it is to avoid losing a connection between a given source and destination node pair in case of disasters. In the worst-case model, $M = 0$ if all node pairs are connected, and $M > 0$ otherwise. Note that since in [17] $\alpha = \infty$, their value is not relevant. In the simulation of [18], M is the number of disconnected node pairs divided by the total amount of node pairs. Thus, $M = 0$ if all node pairs are connected, and $M = 1$ if all node pairs are disconnected.

Furthermore, [18] can adjust the cost of implementing new cables at each location, which depends on the terrain and other factors. This is uniform in [17], thus it corresponds to their total physical length.

Finally, the objective function is composed of two parts: the first is α times the impact of losing a connection because of a disaster, while the second is the cost of implementing new cables.

Note that the special case of $\alpha = \infty$ boils down to the following definition in [17].

Definition 2. *A **network survives** a circular disk failure c if the graph* $\mathscr{G}_c = (\mathscr{V} \setminus \mathscr{V}_c, \mathscr{E} \setminus \mathscr{E}_c)$ *is connected.*

In the following, multiple versions of the algorithmic problems are defined, depending on whether a single edge or multiple edges can be added. Table I summarizes the differences in the problem definitions investigated in the related papers.

For the worst case, the first problem concentrates on connecting two given nodes of the network after a circular disk failure. [17, Claim 4.] proves that it is possible to find the optimal solution in polynomial time, given that a single cable is added. The algorithm is based on computational geometry algorithms, such as offsetting areas bounded by line segments and circular arcs and the geometric Dijkstra algorithm. It is proved in [17, Theorem 1.] that the optimal solution contains either one or two newly added edges, depending on the length of the destroyed edge that previously connected them. If the length is shorter than 4 times the radius of the circular disk, it is more economical to deploy two shorter cables. On the other hand, the complexity of finding the optimal solution with two new cables is still an open question.

Meanwhile, if the aim is to find a cost-efficient solution, the problem of adding a single edge between two given nodes is already NP-hard [18, Theorem 1.]. As the former can be phrased as a special case of the more generic problem of computing the cost-efficient network augmentation, the generic problem of [18] is also NP-hard.[3]

The following four problems are based on the model of [17]. The aim is to find the minimum cost network topology addition in order to survive a single point or circular disk failure of radius r at any location. These connections are reached by either allowing only new edges, or extra nodes as well. Such extensions result in increased complexity: we will prove in the following, that the decision versions of the problems are NP-hard, and no Fully Polynomial Time Approximation Scheme exists for them (unless P=NP).

For each of the problem versions, the input is the same network topology:

Input : A network represented by an undirected geometric graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where the nodes are embedded as points, and the edges as line segments in the Euclidean plane, maximum radius r of a possible regional failure, and the limit on the total length of the curves Δ^r.

The problem outputs are given as follows:

Network Augmentation (NA) Problem

Output: Are there curves added as new edges to $\mathscr{E}$, such that the network survives any point failure, and the total length of the curves is at most Δ^r.

[3]We note that the route of fibers typically follows highways, railways, or gas pipelines [24]; neither the model of [17], nor that of [18] fully support this as they are presented. For more disaster models, see [25].

Geometric Network Augmentation (GNA) Problem

Output: Are there curves added as new edges to $\mathscr{E}$, such that the network survives any circular disk failure of radius r (see Def. 2), and the total length of the curves is at most Δ^r.

Network Extension (NE) Problem

Output: Are there nodes added to $\mathscr{V}$ as so-called Steiner nodes and curves through $\mathscr{V}$ and these Steiner nodes added as new edges to $\mathscr{E}$, such that the network survives any point failure, and the total length of the curves is at most Δ^r.

Geometric Network Extension (GNE) Problem

Output: Are there nodes added to $\mathscr{V}$ as so-called Steiner nodes and curves through $\mathscr{V}$ and these Steiner nodes added as new edges to $\mathscr{E}$, such that the network survives any circular disk failure of radius r, and the total length of the curves is at most Δ^r.

III. Complexity results

In this section, we present our main results, which are the NP-hardness proofs of the above problems (as listed in the last two lines of Table I). These proofs are inspired by the theoretical papers [23], [26]. For an introduction to complexity theory, and approximation algorithms, we refer the reader to [27] and [28], resp.

Very intuitively, all of out proofs are based on the fact that, in case of taking the nodes (without edges) of a *grid graph* (Def. 3) accompanied with a small-enough disaster radius, deciding whether a best solution is cheaper than a given threshold translates to answering the question if there extists a Hamiltonian cycle in the original grid graph, that is an NP-hard problem.

Our most important proposition is that the worst-case problem defined in [17] is NP-hard:

Theorem 1. *The GNA Problem is NP-hard.*

As defined, in the case of the GNA problem, we want the network to survive circular disk failures of a given radius. Compared to this, intuitively, it is easier to cope with single-point failures. Thus, in the following, we first prove the NP-hardness of the NA problem. In its proof of NP-hardness (just as in the upcoming proofs), we will rely on the NP-completeness of finding a Hamilton-path in *grid graphs*:

Definition 3. $G(V,E)$ *is a grid graph if its nodes are embedded in the plane with integer coordinates in a related Cartesian coordinate system, and there is an edge between two nodes u and v exactly if their distance* $d(u,v) = 1$.

Claim 4. *[26, Theorem 2.1.] The Hamilton circuit problem for grid graphs is NP-complete.*

We call the above problem as HCGG. See Fig. 2 for an example grid graph with a Hamilton circle.

TABLE I: The complexity of different versions of problems investigated

	Worst-case ($\alpha = \infty$, circular disk failures, uniform cable costs)		Cost-efficient
	The radius r is very small	Arbitrary radius r	
Add a single edge	$\in P$ [17], it is based on offsetting areas and the geometric Dijkstra algorithm		The problem is NP-hard [18]
Add two edges between a node-pair	Complexity not known		
Add multiple new edges	Problem NA is NP-hard (Lemma 5)	Problem GNA is NP-hard (Theorem 1)	
Add multiple new edges and nodes	Problem NE is NP-hard (Lemma 7) [23]	Problem GNE is NP-hard (Lemma 8)	

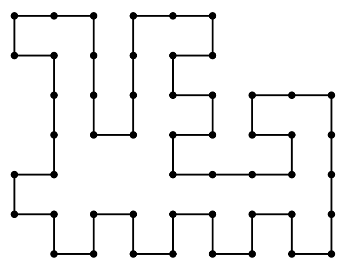


Fig. 2: An example of a grid graph that contains a Hamilton circle.

Lemma 5. *The Network Augmentation (NA) problem is NP-hard.*

Proof: We start with a set of nodes $\mathcal{V}$ that have integer coordinates according to a Cartesian coordinate system over the Euclidean plane. Let $\mathcal{E}$ be empty, i.e., we have no network links built yet. The cost Δ^r is set to n.

We first observe that, for surviving a single point failure, the network has to remain connected after the removal of any link or node, i.e., it has to be 2-connected. Thus, the degree of every node in $\mathcal{V}$ has to be at least 2. Since the minimum distance between nodes is 1, this means the cheapest solution to the NA problem has a cost of at least n. Since our budget Δ^r is just n, this means that if the NA problem has a solution, it has a cost of Δ^r. In addition, the degree of all nodes has to be 2, i.e., the resulting graph has to be a cycle.

We claim that any cycle over $\mathcal{V}$ that is not a subgraph of the grid graph induced by $\mathcal{V}$ has a length at least $n+\sqrt{2}-1$, since it necessarily contains at least one diagonal edge between two nodes with integer coordinates. Thus, if there exists a valid solution to our NA problem setting, it can be nothing else than a Hamilton circuit $H(\mathcal{V}, E_H)$ in the grid graph induced by $\mathcal{V}$. We note that H does not cross itself either in a geometric sense, thus it is a valid solution to the NA problem.

We conclude that the NA problem is NP-hard since the Hamilton circuit problem in grid graphs (HCGG) is NP-complete (Claim 4). ■

Now we turn to the proof of Theorem 1 stating the GNA problem is NP-hard:

Proof of Theorem 1: Let our GNA problem instance be the following. We only consider the problem on grid graphs, i.e., let the coordinates of every node $v \in \mathcal{V}$ be integers in a related Cartesian coordinate system in the plane. Let $\mathcal{E}$ be empty, i.e., we have no network links built yet. Let the disaster radius be $0 \le r < \frac{1}{n}\frac{\sqrt{2}-1}{4\pi}$. Finally, let the cost be $\Delta^r = \left(n+\sqrt{2}-1+n(4r\pi+1)\right)/2$. Note that $\Delta^r \in \left[n(4r\pi+1),\ n+\sqrt{2}-1\right)$.

Our first observation is that any solution S of the GNA problem is a solution of the NA problem too, since S has to remain connected in case of any point failure. This observation, combined with the proof of Lemma 5 yields that any solution $S(\mathcal{V}, E_S)$ of our GNA problem instance (that has to have a cost at most $\Delta^r < n+\sqrt{2}-1$) has to have the following properties:

1) S is a cycle,
2) every edge in E_S connects two nodes that are adjacent in the grid.

This means that if there exists a solution to our GNA problem setting, then (neglecting the geometric embedding for a moment,) it has to be a Hamilton circuit over the grid graph induced by $\mathcal{V}$. To complete our argument of NP-hardness of the GNA problem, we need to show the following: if there exists a Hamilton circuit on the grid graph induced by $\mathcal{V}$, then there exists a Hamilton-circuit-like solution of our problem setting with cost $\le \Delta^r$.

Next, we will show that we can use curves for the edges of the Hamilton circuit such that the network survives any disk failure, and for small enough r this solution has a length less than $n+\sqrt{2}-1$. The high-level idea is that in the vicinity of each node, the curve is a circular arc of radius r, and these arcs are connected by straight line segments, see Fig. 3. One can imagine it as if there are circular disks, (e.g., coins) attached to each side of the cable at each node, and we pull the cable to be as short as possible (as it was a rubber band) since the goal is to minimize the total cable length. In this case, the cable will traverse each node along a circular arc, otherwise, it consists of line segments.

In the next paragraph, we will show that the above setting is a valid solution, which is resistant to any circular disk failure. To complete the proof, we will also need to show that the length of the curves is at most $n(2r\pi + (2r\pi+1))$.

In this paragraph, we show that there is no circular disk of radius r whose failure will separate the network into multiple isolated components. We draw blue circular disks on Fig. 3 corresponding to the arcs for every node. We draw a red line through the node and the center of the corresponding circle. Assume there is a circular disk D failure close to node v. The center of D is either on the red line or on either side of it. In the first case, D cannot hit both edges incident to v, only if v is hit too. In the second case, if D does not hit v, it will be too far from the edge on the other side of the red dividing line, therefore only the closer edge is affected. This argument can be repeated for every node.

Finally, we give an upper estimate of the total length of the edges of the resulting graph. Considering the routes of the edges together as a close curve, it has at most n arcs, which cannot be longer than the circumference of the circle, thus, they contribute to the total length of no more than $2r\pi$ each. The curve also has n line segments, which have lengths at most $2r\pi + 1$ each. This latter estimate holds because the routes given between the end of the edges are no longer than $2r\pi + 1$, see the dashed red line in Fig. 3. As an example, line segment $\overline{wx}$ is shorter than path w-v-y-x (in red) because the former is the straight line segment between w and x. Nodes v and y were adjacent grid points thus their distance is 1. Arcs $\widehat{wv}$ and $\widehat{yx}$ are arcs of circles with radius r and neither of them can be longer than the half circumference. Together, $|\widehat{wv}| + |\widehat{yx}| \le 2 \cdot \frac{2r\pi}{2}$. Based on the former, we can see that the length of edge $\overline{wx}$ can not be more than $2r\pi + 1$. This can be shown similarly for every edge. We can conclude that the total length of the edges of the graph is at most $n(2r\pi + 2r\pi + 1)$. If $r < \frac{1}{n}\frac{\sqrt{2}-1}{4\pi}$, then the total cable length is strictly less than $n + \sqrt{2} - 1$.

Based on the above, we conclude that the GNA problem is NP-hard, since the Hamilton circuit problem in grid graphs is NP-complete (Claim 4). ■

The above problem versions assumed that only new cables could be installed. However, installing new network nodes along new cables is also an option. With this in mind, now we turn to present the proofs of NP-hardness of the NE and GNE problems. First, we present the following Claim 6 that is a rephrasal of [23, Lemma 2]. The original Lemma is tackling the *2-Connected Steiner Network Problem in the Plane (2SNPP)* [23], that, in case of no network edges (i.e., $\mathcal{E}$ is empty), is a special case of our Network Extension problem.

Claim 6 (rephrased Lemma 2 of [23])**.** *Let $\mathcal{V}$ be a set of n integer grid points, and let the set $\mathcal{E}$ of original links be empty. Then*

- *Every solution to the NE problem is of length at least n.*
- *The only solutions to the NE problem of length exactly n are Hamilton circuits such that every two successive points are adjacent in the grid.*
- *All other solutions are of length at least $n + \sqrt{2} - 1$.*

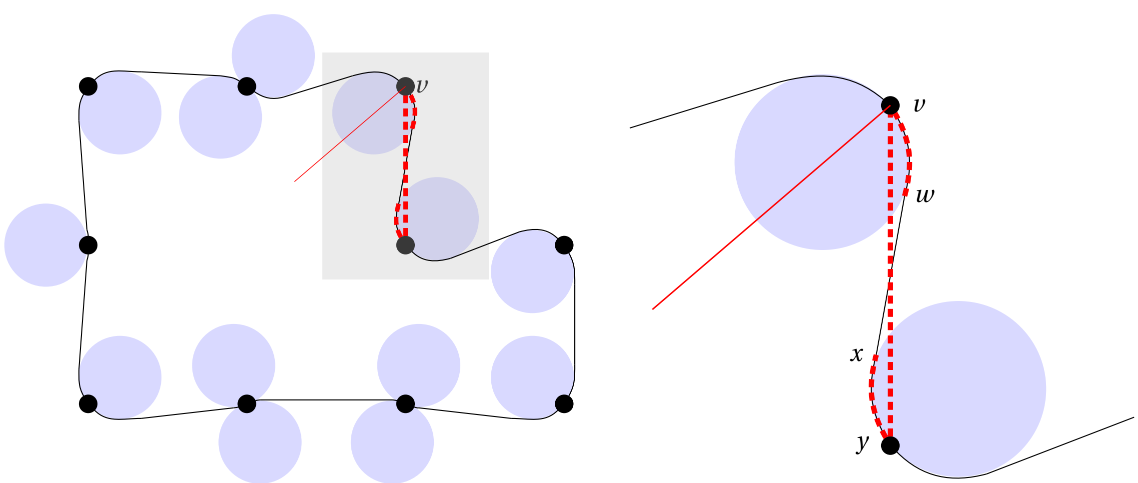
Lemma 7. *The Network Extension (NE) problem is NP-hard.*

Proof: We start with a set of nodes $\mathcal{V}$ that have integer coordinates according to a Cartesian coordinate system over the Euclidean plane. Let $\mathcal{E}$ be empty, i.e., we have no network links built yet. The cost Δ^r is set to n. By Claim 6, if there exists a solution to this Network Extension problem setting, it has to be a Hamilton circuit that is a subgraph of the grid graph induced by n. We conclude that the NE problem is NP-hard, since the HCGG problem is NP-complete (Claim 4). ■

Lemma 8. *The Geometric Network Extension (GNE) problem is NP-hard.*

Proof: Let the settings of our GNE problem instance be similar to those seen in the proof of Theorem 1. More precisely, we consider the following. We only consider the problem on grid graphs, i.e., let the coordinates of every node $v \in \mathcal{V}$ be integers in a related Cartesian coordinate system in the plane. Let $\mathcal{E}$ be empty, i.e., we have no network links built yet. Let the disaster radius be $0 \le r < \frac{1}{n}\frac{\sqrt{2}-1}{4\pi}$. Finally, let the cost be $\Delta^r = \left(n + \sqrt{2} - 1 + n(4r\pi + 1)\right)/2$. Note that $\Delta^r \in \left[n(4r\pi + 1), n + \sqrt{2} - 1\right)$.

We observe that a solution of the GNE problem is a solution to the NE problem too. This, combined with Claim 6 means that a valid solution of the GNE problem with cost $\Delta^r < n + \sqrt{2} - 1$ has to visit the vertices of n according to a Hamilton circuit on the grid graph induced by $\mathcal{V}$. Note that if a Hamilton-



(a) An example grid network with (near) optimal cable routes

(b) Part of Fig. 3a inside the gray rectangle magnified

Fig. 3: The illustration of the network used in the proof of Theorem 1

circuit-like solution described above exists, it has a cost of $n(4r\pi+1) < \Delta^r$.

We conclude that the GNE problem is NP-hard, since the Hamilton circuit problem in grid graphs is NP-complete (Claim 4). ■

As we have seen, all of the problems NA, GNA, NE, and GNE are NP-hard. Looking inside their proofs of NP-hardness, we can see that there is even no Fully Polynomial Time Approximation Scheme (FPTAS) for these problems.

Corollary 9. *For each of the NA, GNA, NE, and GNE problems, there is no Fully Polynomial Time Approximation Scheme (FPTAS), unless $P = NP$. More precisely, no polynomial-time $1 + 1/5n$ approximation exists for these problems, unless $P = NP$.*

Proof: An FPTAS takes as input an instance of a problem (e.g., NA, GNA, NE, or GNE) and a parameter $\epsilon > 0$. In the case of minimization problems, it returns as output a solution whose value is at most $1+\epsilon$ times the optimum within a runtime that is polynomial in the problem size and in $1/\epsilon$. A consequence of the following argument is that in the case of NA, GNA, NE, or GNE, no such algorithm exists for $\epsilon = 1/5n$, unless $P = NP$.

For an instance of the Network Augmentation and the Network Extension problems, we have from Lemma 5 and 7, respectively that a solution to the problem instance that is a Hamilton circuit on the grid graph induced by $\mathcal{V}$, has a cost of n, while any other solution must have length at least $n + \sqrt{2} - 1 > n + 0.414$. Now for $\epsilon = 0.414/n$, an ϵ-approximation to the NA and NE problem, resp., will have a length less than $n + 0.414$ if and only if HCGG has a feasible solution. It follows that there can be no algorithm that finds an ϵ-approximation in time polynomial in $1/\epsilon = n/0.414$ unless $P = NP$.

Similarly, in case of the geometric problem versions, the Hamilton-circuit-like solutions have a cost of $n(4r\pi+1)$, that, for $r = \frac{1}{n}\frac{\sqrt{2}-1}{8\pi}$ is $n + \frac{\sqrt{2}-1}{2}$. All the other solutions cost at least $n + \sqrt{2} - 1$. Now for $\epsilon = \frac{\sqrt{2}-1}{2}/n \simeq 0.207/n$, an ϵ-approximation to the GNA and GNE problem, resp., will have length less than $n + 0.207$ if and only if there exists a Hamilton circuit in the grid graph induced by $\mathcal{V}$. It follows that there can be no algorithm that finds an ϵ-approximation in time polynomial in $1/\epsilon \simeq n/0.207$ unless $P = NP$. ■

We note that the inexistence of an FPTAS does not mean that big instances of the problem versions cannot be solved efficiently nearly optimally. In fact, indifferently of n, the best Hamilton-circuit-like solutions are no more than $\simeq 0.414$ shorter than any other type of solution, meaning only an additive gap that may be negligible in the case of large networks extending over vast areas.

IV. Discussion of the Cost-Efficient Model

In this section, we investigate the cost-efficient version of the problem. In particular, we evaluate the maximum error in the cost function because of limiting the routes of the new cables to connect a series of neighboring cell center points. We call this effect the discretization of the topology map.

Next, we discuss Obs. 1, namely, why the NP-hardness proof in Theorem 1 of [18] is not valid if a disaster must be a connected set of points.

A. The maximum error in the cost over a grid for uniform cable cost

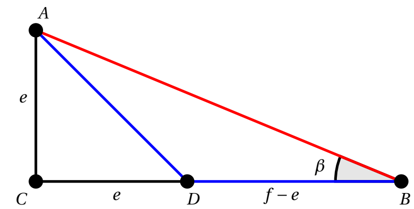


Fig. 4: Illustration for the proof of Lemma 10. The red line segment denotes the Euclidean shortest path between A and B, while the blue line segments denote a shortest grid-based cable route. The grid-based cable route can be up to ~ 8.24% longer than the Euclidean distance. The greatest error occurs at $\beta = \pi/8$.

Lemma 10. *The maximum error in the cost over a grid is ~ 8.24% for uniform cable cost.*

Proof: Suppose we have a grid, where the cells have unit side lengths, and each endpoint of the cable is placed in the center of a cell. Points can be connected using vertical, horizontal, or diagonal line segments that connect adjacent cells in the grid. Here, two cells are said to be adjacent if they share a common face or corner; consequently, each cell has 8 neighbors. Suppose that our starting point is A, and we want to reach B, placing line segments after each other. Let e be the vertical, and f be the horizontal distance between A and B.

Generally, if we place a horizontal and a vertical line segment after each other, the same endpoint can be reached by a shorter diagonal line segment, thus, as the goal is to minimize costs, cables cannot contain both horizontal and vertical line segments. We will use this fact to construct a shortest path.

Starting from A, first we take e vertical, then f horizontal steps towards B. This path connects A and B, but not in the shortest way possible. We take one horizontal and one vertical line segment, and substitute these two with one diagonal line segment, then repeat this as many times as possible. Afterwards we will have $\min(e,f)$ diagonal, and $|f-e|$ either horizontal or vertical line segments: horizontal if $e < f$, vertical if $e > f$, and neither if $e = f$. Starting from A, we place these line segments after each other, such that each added line segment brings the cable's endpoint closer to B, otherwise, the path would not be the shortest possible. This way, the last line segment successfully reaches B. Note that the order of the line segments is irrelevant, as either way, we reach the same endpoint B. We consider the construction where the diagonal line segments are placed first, forming one longer line segment, then the vertical/horizontal line segments are added, again forming one longer line segment. As an example, see the blue path in Fig. 4. Each cable that contains a vertical line segment can be converted to a cable with a horizontal line segment, by rotating it about point A, by $\pi/2$ - the length of the cable, and the Euclidean distance between A and B remains the same. Therefore, it is enough to consider solutions without a vertical line segment (just as in Fig. 4).

To find the greatest possible error, we need to maximize the ratio between the grid-based and Euclidean cable lengths:

$$L = \frac{\sqrt{2} \cdot e + (f - e)}{\sqrt{e^2 + f^2}} \tag{1}$$

Let β denote the angle between $\overline{AB}$ and the horizontal distance between A and B ($\overline{CB}$). As $e = \tan(\beta) \cdot f$, we can substitute e to $\tan(\beta) \cdot f$ in the above expression:

$$L = \frac{((\sqrt{2} - 1)\tan\beta + 1) \cdot f}{\sqrt{(\tan^2\beta + 1) \cdot f^2}} \tag{2}$$

As $f > 0$ and $\cos\beta > 0$ for all possible angles in a triangle, thus $\sqrt{\tan^2\beta + 1} = \frac{1}{\cos\beta}$, the value of L is the following:

$$L = (\sqrt{2} - 1)\sin\beta + \cos\beta \tag{3}$$

To find the maximum value of L, we need to find β such that L', the value of the first derivative with respect to β is 0:

$$L' = (\sqrt{2} - 1)\cos\beta - \sin\beta = 0 \tag{4}$$

Knowing that $0 < \beta < \pi/2$, the only solution in this domain is exactly at $\beta = \pi/8$.

From ACD triangle we know that CDA angle is $\pi/4$, therefore ADB angle is $3 \cdot \pi/4$. As $\beta = \pi/8$, DAB angle also has to be $\pi/8$. Therefore ABD is an isosceles triangle, and $f = (\sqrt{2} + 1) \cdot e$. Thus, if $\beta = \pi/8$,

$$L = (\sqrt{2} - 1)\sin\pi/8 + \cos\pi/8 \approx 1.0824 \tag{5}$$

Note that e and f are integers because they connect the center points of grid cells. There are no possible integer values for $f = (\sqrt{2} + 1) \cdot e$, therefore, in practice, the exact maximum (more precisely, supremum) error cannot be reached. On the other hand, for instance, if $f = 169$ and $e = 70$, conversion to the grid results in a close approximation to the maximum error, with only $\sim 6.042 \cdot 10^{-11}$ difference. ■

We note that, as also indicated in [18], this maximum error can be decreased by enabling additional cable directions, however, it further complicates the optimization problem.

B. Adding a single edge is NP-hard if the disaster can be any set of points

In this subsection, we will show that the NP-hardness proof in Theorem 1 of [18] is not valid if a disaster must be a connected set of points. The proof is based on a Karp reduction from a Boolean satisfiability problem (3-SAT) instance. The 3-SAT instance is a Boolean formula of k clauses, where each clause contains exactly three literals. For each variable x of the 3-SAT instance, two disasters are assigned, one corresponds to literal x, and the other is literal $\neg x$. The problem is mapped to a grid with 3 rows where each clause corresponds to a column. The 3 cells of each column correspond to 3 disaster sub-regions assigned to the corresponding literals. For example, the set of disaster sub-regions assigned to literal x_2 on Fig. 2. of [18] is the set of cells named x_2, which are isolated cells. We do not see any modification to the proof to ensure that the set of disaster sub-regions are connected.

V. Future work

The NP-hardness proofs presented in this paper do not work if the original graph is not empty, we exclude grid graphs or 'small' disaster radii. Thus, for a better understanding of the complexity landscape of the disaster resilient network extension problems, among others, we will investigate the complexity of problem formulations of [18], [16], and of [17], including variants NA, GNA, NE, and GNE, in case of different graph classes and a wider possible range of disaster radii.

VI. Conclusions

This paper focuses on the problems of how to extend a network topology to become more resilient against disasters. We have investigated different mathematical models of the problem, where each captures a unique aspect. In the first model, the aim was to provide a worst-case analysis of extending the topology with minimal total cable length. Here several subproblems can be solved with a polynomial-time algorithm. We have shown that the problem is NP-hard if links can be added, and remains NP-hard if nodes can be added besides the new links. We have also discussed the limitations of a more complex version of the problem, where the goal is to find the cost-optimal extension as a compromise with the level of protection against disasters. In this case, very detailed input data is needed, such as the historical data on the disasters and the cost of implementing new cable routes at each location. We have shown that no matter how precise the input data is, in the case of uniform cable implementation cost, we have to deal with an at most 8% of imprecision of the result because of the discretization of the topology map. This paper is a further step in understanding what type of network extending problems can be solved efficiently.

Acknowledgements

The authors are thankful to Jorik Oostenbrink and Fernando Kuipers for the fruitful discussions. This work was supported in part by Projects no. 135606, 134604, and 128062 that have been implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, financed under the ANN_20, FK_20, and K_18 funding schemes, respectively. Supported by the ÚNKP-21-4 New National Excellence Program of the Ministry of Innovation and Technology from the source of the National Research, Development and Innovation Fund.

REFERENCES

[1] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, "Assessing the vulnerability of the fiber infrastructure to disasters," *IEEE/ACM Trans. Netw.*, vol. 19, no. 6, pp. 1610–1623, 2011.

[2] O. Gerstel, M. Jinno, A. Lord, and S. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?" *Communications Magazine, IEEE*, vol. 50, no. 2, pp. s12–s20, 2012.

[3] M. F. Habib, M. Tornatore, M. De Leenheer, F. Dikbiyik, and B. Mukherjee, "Design of disaster-resilient optical datacenter networks," *J. Lightw. Technol.*, vol. 30, no. 16, pp. 2563–2573, 2012.

[4] J. Heidemann, L. Quan, and Y. Pradkin, *A preliminary analysis of network outages during hurricane Sandy*. University of Southern California, Information Sciences Institute, 2012.

[5] F. Dikbiyik, M. Tornatore, and B. Mukherjee, "Minimizing the risk from disaster failures in optical backbone networks," *J. Lightw. Technol.*, vol. 32, no. 18, pp. 3175–3183, 2014.

[6] I. B. B. Harter, D. Schupke, M. Hoffmann, G. Carle *et al.*, "Network virtualization for disaster resilience of cloud services," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 88–95, 2014.

[7] X. Long, D. Tipper, and T. Gomes, "Measuring the survivability of networks to geographic correlated failures," *Optical Switching and Networking*, vol. 14, pp. 117–133, 2014.

[8] B. Mukherjee, M. Habib, and F. Dikbiyik, "Network adaptability from disaster disruptions and cascading failures," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 230–238, 2014.

[9] R. Souza Couto, S. Secci, M. Mitre Campista, K. Costa, and L. Maciel, "Network design requirements for disaster resilience in IaaS clouds," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 52–58, 2014.

[10] S. A. Jyothi, "Solar superstorms: planning for an internet apocalypse," in *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, 2021, pp. 692–704.

[11] T. Gomes, J. Tapolcai, C. Esposito, D. Hutchison, F. Kuipers, J. Rak, A. De Sousa, A. Iossifides, R. Travanca, J. André *et al.*, "A survey of strategies for communication networks to protect against large-scale natural disasters," in *RNDM*. IEEE, 2016, pp. 11–22.

[12] J. Rak, D. Hutchison, J. Tapolcai, M. T. Rasa Bruzgiene, C. M. Machuca, M. Furdek, and P. Smith, "Fundamentals of communication networks resilience to disasters and massive disruptions," in *Guide to Disaster-Resilient Communication Networks*, J. Rak and D. Hutchison, Eds. Springer, 2020, ch. 1.

[13] B. Vass, J. Tapolcai, Z. Heszberger, J. Bíró, D. Hay, F. A. Kuipers, J. Oostenbrink, A. Valentini, and L. Rónyai, "Probabilistic Shared Risk Link Groups Modeling Correlated Resource Failures Caused by Disasters," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 9, pp. 2672–2687, 2021.

[21] P. K. Agarwal, A. Efrat, S. K. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, "The resilience of WDM networks to probabilistic geographical failures," *IEEE/ACM Trans. Netw.*, vol. 21, no. 5, pp. 1525–1538, 2013.

[14] D. L. Msongaleli, F. Dikbiyik, M. Zukerman, and B. Mukherjee, "Disaster-aware submarine fiber-optic cable deployment for mesh networks," *Journal of Lightwave Technology*, vol. 34, no. 18, pp. 4293–4303, 2016.

[15] Q. Wang, J. Guo, Z. Wang, E. Tahchi, X. Wang, B. Moran, and M. Zukerman, "Cost-effective path planning for submarine cable network extension," *IEEE Access*, vol. 7, pp. 61 883–61 895, 2019.

[16] P. N. Tran and H. Saito, "Geographical route design of physical networks using earthquake risk information," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 131–137, 2016.

[17] J. Tapolcai, Z. L. Hajdú, A. Pašić, P.-H. Ho, and L. Rónyai, "On network topology augmentation for global connectivity under regional failures," in *Proc. IEEE INFOCOM*, Vancouver, Canada, May 2021.

[18] J. Oostenbrink and F. Kuipers, "Going the extra mile with disaster-aware network augmentation," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021, pp. 1–10.

[19] A. Valentini, B. Vass, J. Oostenbrink, L. Csák, F. A. Kuipers, B. Pace, D. Hay, and J. Tapolcai, "Network resiliency against earthquakes," in *Int. Workshop on Resilient Networks Design and Modeling (RNDM)*, Nicosia, Cyprus, Oct. 2019.

[20] J. Oostenbrink and F. Kuipers, "Computing the impact of disasters on networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 45, no. 2, pp. 107–110, 2017.

[22] H.-W. Lee, E. Modiano, and K. Lee, "Diverse routing in networks with probabilistic failures," *IEEE/ACM Transactions on Networking*, vol. 18, no. 6, pp. 1895–1907, 2010.

[23] E. L. Luebke and J. S. Provan, "On the structure and complexity of the 2-connected steiner network problem in the plane," *Operations Research Letters*, vol. 26, no. 3, pp. 111–116, 2000.

[24] E. K. Çetinkaya, M. J. F. Alenazi, A. M. Peck, J. P. Rohrer, and J. P. G. Sterbenz, "Multilevel resilience analysis of transportation and communication networks," *Telecommunication Systems*, vol. 60, no. 4, pp. 515–537, 2015. [Online]. Available: http://dx.doi.org/10.1007/s11235-015-9991-y

[25] E. K. Cetinkaya, D. Broyles, A. Dandekar, S. Srinivasan, and J. P. Sterbenz, "Modelling communication network challenges for future internet resilience, survivability, and disruption tolerance: A simulation-based approach," *Telecommunication Systems*, vol. 52, no. 2, pp. 751–766, 2013.

[26] A. Itai, C. H. Papadimitriou, and J. L. Szwarcfiter, "Hamilton paths in grid graphs," *SIAM Journal on Computing*, vol. 11, no. 4, pp. 676–686, 1982.

[27] M. R. Garey and D. S. Johnson, "Computers and intractability: A guide to the theory of NP-completeness," 1978.

[28] V. V. Vazirani, *Approximation algorithms*. Springer, 2001, vol. 1.