Survivable Paths in Multilayer Networks

by

Marzieh Parandehgheibi

Submitted to the Sloan School of Management and the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Master of Science in Operations Research and
Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

© Massachusetts Institute of Technology 2012. All rights reserved.

Author . . . .
Sloan School of Management and the
Department of Electrical Engineering and Computer Science
May 18, 2012

Certified by . . . .
Professor Eytan Modiano
Thesis Supervisor

Accepted by . . . .
Professor Patrick Jaillet
Codirector, Operations Research Center

Accepted by . . . .
Professor Leslie Kolodziejski
Chairman, Department Committee on Graduate Students
Department of Electrical Engineering and Computer Science

Survivable Paths in Multilayer Networks

by

Marzieh Parandehgheibi

Submitted to the Sloan School of Management and the
Department of Electrical Engineering and Computer Science
on May 18, 2012, in partial fulfillment of the
requirements for the degrees of
Master of Science in Operations Research and
Master of Science in Electrical Engineering and Computer Science

Abstract

We consider the problem of protection in multilayer networks. In single-layer networks, a pair of disjoint paths can be used to provide protection for a source-destination pair. However, this approach cannot be directly applied to layered networks where disjoint paths may not always exist. In this thesis, we take a new approach which is based on finding a *set of paths* that may not be disjoint but together will survive any single physical link failure. First, we consider the problem of finding the minimum number of survivable paths. In particular, we focus on two versions of this problem: one where the length of a path is restricted, and the other where the number of paths sharing a fiber is restricted. We prove that in general, finding the minimum survivable path set is NP-hard, whereas both of the restricted versions of the problem can be solved in polynomial time. We formulate the problem as Integer Linear Programs (ILPs), and use these formulations to develop heuristics and approximation algorithms. Next, we consider the problem of finding a set of survivable paths that uses the minimum number of fibers. We show that this problem is NP-hard in general, and develop heuristics and approximation algorithms with provable approximation bounds. We also model the dependency of communication networks on the power grid as a layered network, and investigate the survivability of communication networks in this layered setting. Finally, we present simulation results comparing the different algorithms.

Thesis Supervisor: Professor Eytan Modiano

Acknowledgments

I wish to express my gratitude to my advisor Prof. Eytan Modiano for his support, guidance and help. I feel fortunate to have the opportunity to work with him. I also would like to acknowledge his dedication in reading line by line and draft after draft of my thesis.

I also wish to thank Dr. Hyang-won Lee who provided helpful insights and comments throughout my research.

I would like to thank my fiancé, Hamed, for being with me in every second of my life through sadness and happiness, and for encouraging me during my study.

I would like to thank my brother Ali, and his wife Shabnam, for being so supportive and helpful; and above all, I would like to thank my parents for their life-long support in any respect. I couldnt achieve what I have now without their guidance, patience, and love.

Contents

List of Figures

Chapter 1

Introduction

Multilayer network architectures such as IP-over-WDM have played an important role in advancing modern communication networks. Typically, a layered network is constructed by embedding a logical topology onto a physical topology such that each logical link is routed using a path in the physical topology. While such a layering approach enables to take advantage of the flexibility of upper layer technology (e.g., IP) and the high data rates of the lower layer technology (e.g., WDM), it raises a number of challenges for efficient and reliable operations. In this thesis, we focus on the issue of providing protection in layered networks.

1.1 Background on Network Survivability

Optical communication networks are an increasingly popular technology for high-speed Wide Area Networks. This is due to the fact that fiber optic cable has a large bandwidth and by using Wavelength Division Multiplexing (WDM), this bandwidth can be shared among different channels (wavelengths). In IP-over-WDM networks, the IP network is the logical topology which is mapped on top of the physical topology of the optical network, such that each logical link (also known as lightpath) is routed on a path of fibers in the physical topology. Moreover, with WDM technology, each fiber can carry multiple logical links using different wavelengths. Although, this layered network has a very high capacity to transfer data, it is also very vulnerable to disruptions. This is due to the fact that in the case of a physical fiber's failure, all of the lightpaths traversing the failed fiber will be disrupted; so a fiber cut can lead to tremendous traffic loss. Due to the tremendous traffic loss that a failure may cause, network survivability becomes a critical concern in network design and its real-time operation [2, 3, 4].

Most research work on survivability in WDM networks focus on the recovery from a single link or node failure, where one failure is repaired before another failure is assumed to occur in the network, since single failures are the predominant form of failures in optical networks [4].

The protection problem in single-layer networks is rather straightforward; namely, providing a pair of disjoint paths (one for primary and one for backup) guarantees a route between two nodes against any single link failure. However, this approach cannot be directly applied to layered networks; because a pair of seemingly disjoint paths at the logical layer may share a physical link and thus simultaneously fail in the event of a physical link failure. To address this issue, Bhandari in [5] introduced the notion of *physically disjoint* logical paths. In fact, he showed that the requirement for primary path and protection path is that they have to be diversely routed so that at least one path can survive a single failure in the network.

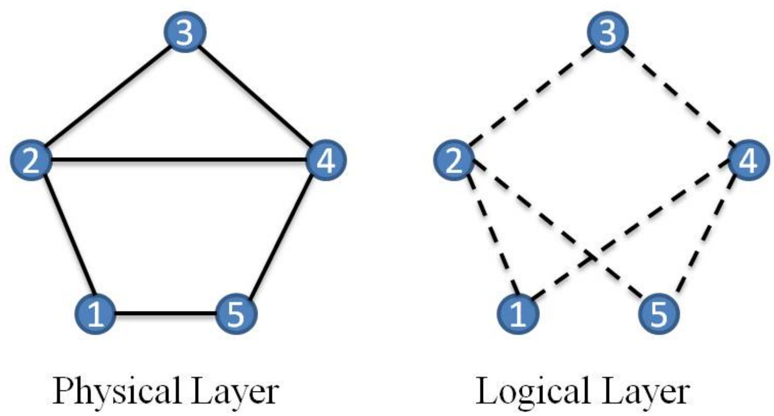
The notion of Shared Risk Link Group (SRLG) was introduced in [6], which refers to a group of links sharing the same risk (e.g., fiber and conduit). Hu generalized the diverse routing problem in optical mesh networks [7]. He proved that finding two risk disjoint paths in a two layered network is NP-complete using a reduction from the set-splitting problem. He also proved the same hardness result for a special case of the problem which minimized the total number of fibers used in the two risk disjoint paths. Having shown that there is no polynomial algorithm to solve the problem exactly, he provides an ILP formulation for the problem which helps solve small instances of the problem.

There is a tremendous amount of literature relating to finding SRLG-disjoint paths, and many people have come up with different algorithms and heuristics to solve this problem. Xu *et. al.* investigated an important issue in the heuristics for SRLG paths that is avoiding failures in path determination caused by "traps" [8]. They have shown that the proposed heuristics for finding the SRLG paths run into traps in 30% of the time, i.e. cannot come up with a solution. On the other hand, as shown by Hu solving the exact problem using the ILP-based formulation is not possible for large instances of the network. Therefore, they proposed a modified heuristic algorithm for avoiding this problem.

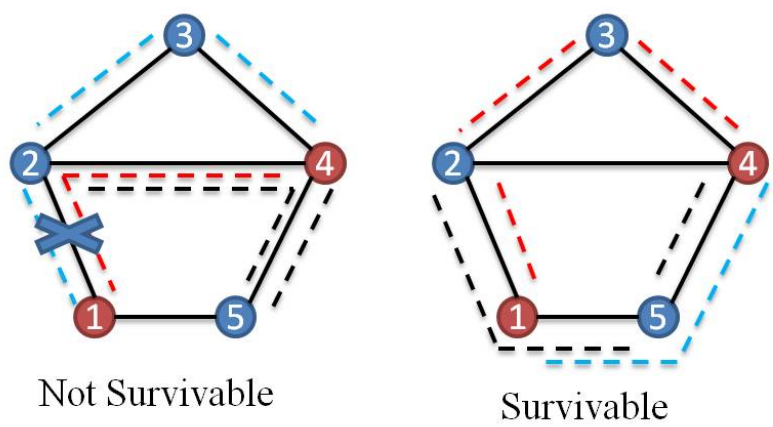
Later, Xu *et. al.* proposed a new algorithm which maximizes bandwidth sharing in the structure of SRLG, and at the same time avoids the traps [9]. In their proposed algorithm, they used a novel dynamic programming technique which achieves a higher bandwidth efficiency and lower request blocking probability.

Datta and Somani [10] proposed graph transformation techniques for protecting the multilayer network against single failures. In fact, they showed that although finding two diversely routed paths is NP-complete, there are certain restrictive failure sets which make the problem simpler. They showed that in such setting it is possible to restore the network against shared risk link failures, by using graph transformation techniques.

In all of these papers, the authors assumed that the physical and logical networks, and the routing of logical layer on the physical layer are given. Given these settings, the problem was to find the primary and backup paths. Another well-studied problem is network design, and the problem of finding a survivable routing of logical links on the physical topology. Modiano and Narula-Tam highlighted the fact that it is very important to route the lightpaths such that a single failure cannot disconnect the whole network [11]. Showing that the problem of survivable routing is NP-complete, they came up with a necessary and sufficient condition for survivability of light path routing that could be imposed in the ILP formulation. Moreover, since the problem was computationally hard to solve for large scale instances, they developed approximation algorithms for this problem. Figure 1-1 shows that changing the mapping of the logical layer on the physical layer can lead to a survivable routing. Here, the physical topology consists of 5 nodes and 6 fibers, and logical topology consists of the same 6 nodes and 6 lightpaths. Each lightpath can be routed on a set of fibers under a special mapping. In the logical topology, two paths can be found between nodes 1 and 4, one is exactly lightpath (1-4) and the other is the set of lightpaths (1-2),(2-3) and (3-4). In both mappings, lightpaths (1-2),(2-3),(3-4) and (4-5) are directly mapped on the underlying fiber. However, in the left mapping lightpaths (1-4) and (2-5) are mapped on the fibers (1,2)(2,4) and (2,4)(2,5); where in the right mapping lightpaths (1-4) and (2-5) are mapped on the fibers (1,5)(5,4) and (2,1)(1,5). Given this structure, in the left mapping both paths between nodes 1 and 4 are routed on fiber (1,2); thus the failure of this fiber disconnects both of paths. However, in the right mapping, no fiber is common between the two paths; therefore, these two paths are risk disjoint.



(a) Structure



(b) Mapping

Figure 1-1: In the left mapping, fiber 1-2 disconnects both paths between nodes 1 and 4; However, in the right mapping, no fiber can disconnect both paths. Therefore, the two paths are risk disjoint.

Later, Lee *et. al.* introduced the problem of maximizing the connectivity of layered networks [12]. They defined the Min-Cut and Max-Flow as an important connectivity metric, and showed that these metrics have a different meaning in the layered setting. In particular, they showed that the Min-Cut Max-Flow Theorem does not hold in a layered graph, and in fact computing each metric is NP-complete. They proposed Min Cross Layer Cut(MCLC) as a new metric for measuring connectivity of multilayer networks, and showed that a layered network with a large MCLC results in a more resilient network.

1.2 Outline and Contributions

Although the SRLG-disjoint paths problem has been well studied, there are several challenges to this approach. First, SRLG-disjoint paths may not always exist (Figure 1-2). Second, such a pair of paths could be very long and thus vulnerable (Figure 1-3). While by associating appropriate cost to a path, the SRLG-disjoint paths problem can be modified to find a path set avoiding long paths, the modified problem is known to be NP-hard [7] and there is no known algorithm with provable approximation guarantee.

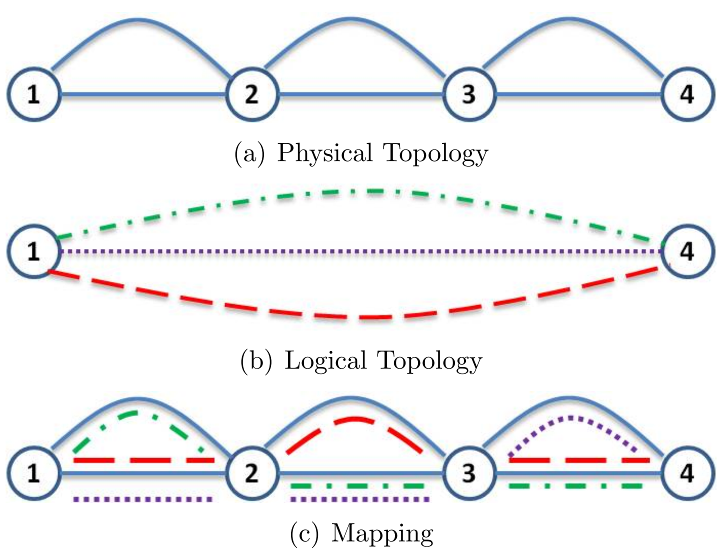
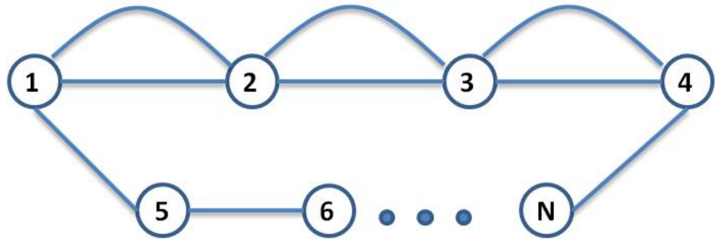
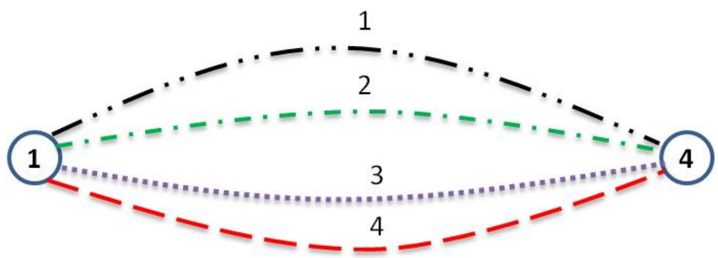


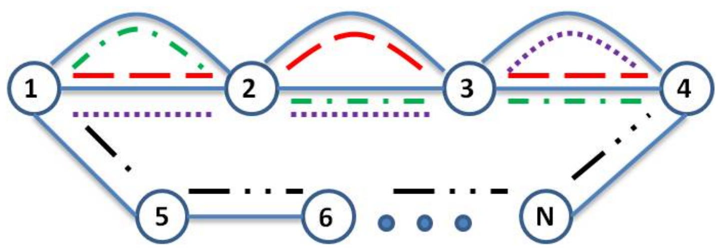
Figure 1-2: Topologies in Multilayer Networks



(a) Physical Topology



(b) Logical Topology



(c) Mapping

Figure 1-3: Long SRLG-disjoint Paths

In order to address these challenges, we take an alternative approach that is based on finding a set of paths that together will survive any single physical link failure. Thus, in the case that SRLG-disjoint paths do not exist, we may find three or more paths such that in the event of a fiber failure, at least one of the paths remain connected. This notion of *survivable path set* generalizes the traditional notion of SRLG-disjoint paths, and enables to provide protection for a broader range of scenarios. Our contributions can be summarized as follows:

- We introduce a new notion of survivable path set to provide protection even for the case where SRLG-disjoint paths do not exist;
- We prove the NP-hardness of the minimum survivable path set (MSP) problem;
- We show that under certain practical restrictions, the MSP problem is polyno-

mially solvable;

- We develop heuristics and approximation algorithms for the MSP problem.

Moreover, we look at other infrastructures with dependency between different layers. One of the most important infrastructures, which we discuss in this thesis, is the communication network (e.g. Internet) and power grid, and the strong dependency of communication network on the power network. We simplify the model of communication network and power grid, and develop a topology mapping between the two networks. This allows us to analyze the dependency of these two networks by formulating a set of similar reliability problems in this setting.

In the following section, we present the network model. In Chapter 2, we study the problem of finding a minimum set of paths that will survive any single fiber failure and develop several approximation algorithms. In Chapter 3, we design approximation algorithms for finding a survivable path set that uses the minimum number of fibers. In Chapter 4, we extend the layered network to model the dependency of communication networks on the power grid, and discuss the reliability problems in this new model. Finally, we provide simulation results, conclusions and future research directions in Chapter 5.

1.3 Network Model

We consider a layered network that consists of a logical topology $G_L = (V_L, E_L)$ built on top of a physical topology $G_P = (V_P, E_P)$ where V and E are the sets of nodes and links respectively. Each logical link (i, j) in E_L is mapped onto an $i - j$ path in the physical topology. This is called lightpath routing. Different lightpaths may use the same fiber (physical link), therefore when a fiber fails, all the lightpaths using that fiber will fail. Hence, a logical path survives the failure of any fiber that it does not use.

As mentioned above, we generalize the traditional notion of SRLG-disjoint paths to account for the case where there does not exist a pair of SRLG-disjoint paths. In a layered network, a set of logical paths is said to be *survivable* if at least one of the paths remain connected after any single physical link failure. Hence, a survivable set consisting of two paths is a pair of SRLG-disjoint paths. Note that there may exist a survivable path set while SRLG-disjoint paths do not exist. For example, consider the physical and logical topologies in Figure 1-2. Each dashed line in Figure 1-3(c) shows the lightpath routing of each logical link over the physical topology. Under this lightpath routing, each pair of logical paths between nodes 1 and 4 shares some fibers.

Suppose that we want to find a set of logical paths between nodes 1 and 4 in Figure 1-2 that can survive any single physical link failure. Clearly, there does not exist a pair of SRLG-disjoint paths as each pair of logical paths shares a fiber. However, it is straightforward to check that the set of 3 paths can survive any single fiber cut, although they are not SRLG-disjoint. This example shows that the traditional protection schemes based on SRLG-disjoint paths (such as the ones in [7]) may fail to provide protection against single physical link failures, while there exists a set of paths that can together provide protection. Our goal in this thesis is to address the problem of finding a set of survivable paths that together will survive any single fiber failure.

Chapter 2

Minimum Survivable Paths Set (MSP)

We start with the problem of finding a minimum survivable path set, i.e., the minimum cardinality set of paths between a pair of nodes s and t that survive any single physical link (fiber) failure. We first present a path-based Integer Linear Program (ILP) formulation for this problem, assuming that the entire set of $s-t$ paths with their routings over fibers is given. For each path j, let P_j be a binary variable which takes the value 1 if path j is selected, and 0 otherwise. The matrix $A \in R^{m \times n}$ refers to the mapping of all n paths over the m fibers such that $a_{ij} = 0$ if path j uses fiber i and $a_{ij} = 1$ otherwise. Let e be a $m \times 1$ vector of ones.

$$\text{minimize} \quad \sum_{j=1}^{n} P_j \tag{2.1}$$

$$\text{subject to} \quad A \times P \geq e \tag{2.2}$$

$$P_j \in \{0, 1\}, \quad j = 1, \cdots, n \tag{2.3}$$

In the above, the objective function is the number of selected paths. Each row $i \in \{1, \cdots, m\}$ in constraint (2.2) requires that at least one selected path survives the failure of fiber i, i.e., the selected path set should be survivable. Hence, the optimal solution to the above optimization problem gives a minimum survivable path set.

Although this formulation requires the knowledge of every path (which is possibly exponential in the number of fibers), the compact and clean expression of the path-based formulation enables us to analyze the useful properties of survivable path sets. Later, we will use this formulation to develop heuristics and approximation algorithms for finding a minimum survivable path set.

The MSP problem can also be formulated using a polynomial number of constraints and variables without enumerating all of the paths. Let P_{tot} denote the number of selected $s-t$ logical paths, E_L denote the set of logical links and E_L^k denote the set of remaining logical links after the failure of fiber k. Note that for survivability, each E_L^k should contain at least one of the selected paths. Let x_{ijk} be 1 if link (i,j) in E_L^k is selected to form an $s-t$ path over the remaining graph $G_L^k = (V_L, E_L^k)$, and 0 otherwise. Let y_{ij} be 1 if the selected path set uses logical link (i,j), and 0 otherwise. The following link-based formulation describes the MSP problem.

$$\text{minimize} \quad P_{tot} \tag{2.4}$$

$$\text{subject to} \quad \left.\begin{aligned} &\sum_{(s,j)\in E_L} y_{sj} = P_{tot} \\ &\sum_{(i,t)\in E_L} y_{it} = P_{tot} \\ &\sum_{(i,j)\in E_L} y_{ij} - \sum_{(j,i)\in E_L} y_{ji} = 0, \quad \forall i \neq s,t \end{aligned}\right\} \tag{2.5}$$

$$\left.\begin{aligned} &\sum_{(s,j)\in E_L^k} x_{sjk} = 1, && \forall k \\ &\sum_{(i,t)\in E_L^k} x_{itk} = 1, && \forall k \\ &\sum_{(i,j)\in E_L^k} x_{ijk} - \sum_{(j,i)\in E_L^k} x_{jik} = 0 && \forall k, \forall i \neq s,t \end{aligned}\right\} \tag{2.6}$$

$$y_{ij} \geq x_{ijk} \qquad \forall k,i,j \tag{2.7}$$

$$x_{ijk} \in \{0,1\} \qquad \forall k,i,j \tag{2.8}$$

The constraints in (2.6) require that each remaining logical graph E_L^k should contain an $s - t$ path, which guarantees the survivability against any single physical link failure. By the constraints in (2.7), logical link (i, j) is selected if it has been used in some remaining logical graph $G_L^k = (V_L, E_L^k)$. Hence, the constraints in (2.5) require that there should be total P_{tot} flows between nodes s and t over the selected logical links specified by y_{ij}'s. Consequently, the variable P_{tot} counts the total number of paths selected for survivability. In Section 5, we will use this formulation to verify the performance bound of our approximation algorithms.

2.1 MSP in general setting

In this section, we show that the MSP problem is NP-hard in general and discuss some algorithms that can be used to solve the problem. In Sections 2.2 and 2.3, we will study the MSP problem under practical constraints. Our first result pertains to the complexity of the MSP problem as stated in Theorem 2.1.1 below.

2.1.1 Complexity

Theorem 2.1.1. *Computing the minimum number of survivable paths in multilayer networks is NP-hard. In addition, this minimum value cannot be approximated within any constant factor, unless $P = NP$.*

The proof of Theorem 2.1.1 relies on a mapping between the survivable path set problem and the minimum set cover problem. Suppose that each path corresponds to a set of fibers that are not used by that path, i.e., survived. Then, finding a minimum survivable path set is equivalent to finding a minimum path set that survives (covers) all of the fibers.

Minimum Set Cover: Given a set of elements $E = \{e_1, e_2, ..., e_n\}$ and a family $F = \{C_1, C_2, ..., C_m\}$ of subsets of E, and the minimum value k such that there exist k subsets $\{C_{j1}, C_{j2}, ..., C_{jk}\} \subset F$ that cover E, i.e., $\cup C_{jl} \in C = E$ [13].

Proof. Given an instance of Minimum Set Cover Problem with ground set E and family of subsets R, we construct a physical topology $E = \{f_1, \cdots, f_m\}$ containing all m fibers and a logical topology $R = \{P_1, \cdots, P_n\}$, where each P_j corresponds to the set of fibers that survive in the failure of path j, i.e. all fibers that are *not* used by path j. It follows that the minimum number of logical paths that survives all the physical fibers is equal to the size of a minimum set cover. As the last step of proof, we need to show we can construct a physical topology with the given routing. Given the set of paths and the fibers used by each path (complement of fibers survived by each path), we can use the physical topology in [12]. The inapproximability result follows immediately from the inapproximabilities of the Minimum Set Cover problem. □

2.1.2 Approximation Algorithms

Greedy

Since the problem is computationally hard to solve, we consider heuristics and approximation algorithms that give a set of survivable paths in polynomial time. Owing to the similarity to the set cover problem, the heuristics that have been developed for set cover problems can be used here. In particular, a common approach to solve the set cover problem is the greedy algorithm. In order to apply the greedy algorithm to our setting, one needs to enumerate all of the paths with their routings on the fibers. In general, the number of paths in a multilayer network is exponential in the total number of fibers. Moreover, in each iteration, the greedy algorithm tries to find a path that survives the maximum number of fibers. This is equivalent to the Minimum Color Path problem, which is known to be NP-hard. [14]

Randomized Rounding

Another approach which can be used to approximate the set cover problem is randomized rounding. Randomized rounding gives an $O(\log m)$ approximation, where m is the number of fibers [15]. This is the best possible approximation for the MSP problem, which is due to the fact that the minimum set cover problem cannot be approximated within better than a $\log m$ factor [16].

Fortunately, practical systems impose certain physical constraints that make the survivable path-set problem easier to solve. For example, due to physical impairments and delay constraints, paths are typically limited in length. Furthermore, in WDM networks, the sharing of a fiber by the logical links is limited by the number of available wavelengths. In the following, we show that these physical limitations make the MSP problem tractable.

2.2 The Path Length Restricted Version

In this section, we assume that each logical path is restricted to use at most K fibers. Restricting the length of paths (i.e., number of fibers on each path) is a realistic assumption, because each logical link is typically constrained in the number of fibers that it may use, and each logical path is constrained in the number of logical links.

Lemma 2.2.1. *Under the path length restriction, the optimal number of survivable paths is at most* $K + 1$.

Proof. By the assumption, each path uses at most K fibers, and thus at least $m - K$ fibers are survived by a path. Suppose that we have selected an arbitrary path, and want to add other paths to form a survivable path set. In the worst case, each of the newly selected paths can survive only a single fiber which is not survived by the previously selected paths. Since there are at most K fibers that are not survived by the first path, we need at most K additional paths to survive the rest of the fibers. Therefore, the total number of paths will not exceed $K + 1$. □

Lemma 2.2.2. *In the path length restricted version of MSP, the total number of paths is polynomial in the number of fibers* m*, and can be enumerated in polynomial time.*

Proof. Under the assumption, a path can consist of up to K fibers, and thus at most K logical links. In a graph with n nodes there can be $O(n^K)$ paths of length up to K. Since the number of nodes is at most $2m$, the total number of logical paths of length up to K is $O(m^K)$. A simple exhaustive search can be used to enumerate the paths. □

Theorem 2.2.1. *The path length restricted version of the MSP problem can be solved in polynomial time.*

Proof. By Lemma 2.2.1, MSP needs at most $K+1$ paths to survive any single failure. Therefore, one can find the exact solution by searching through all subsets of paths with sizes $2, 3, ..., K+1$. This will take $O(P^{K+1})$ iterations where P is the total number of paths. On the other hand, by Lemma 2.2.2, the total number of paths is $O(m^K)$. Therefore, the total running time of exhaustive search is $O(m^{K(K+1)})$ which is polynomial in the total number of fibers. □

Although this exhaustive search returns an optimal solution, its running time can be prohibitive for large values of m and K. This motivates us to study heuristics and approximation algorithms with better running time. First, we consider a greedy algorithm, followed by a randomized algorithm based on ε-net which is a well-known technique in the area of computational geometry.

2.2.1 Greedy Algorithm

The first heuristic we consider is a greedy algorithm which is similar to the greedy algorithm for the minimum set cover problem. The input to the greedy algorithm is the set of paths with the set of fibers used by each path and the set of all fibers. The greedy algorithm is an iterative algorithm that works as follows. In the first iteration, it selects a path using the minimum number of fibers, and updates the set of fibers not survived by the selected path. This greedy path selection is repeated until the selected path set survives all of the fibers. Following the proof of Lemma 2.2.1, it can be shown that the greedy algorithm also finds a survivable path set with size at most $K+1$.

As discussed in Section 2.1, the greedy algorithm generally gives an $O(\log m)$ approximation to the minimum survivable path set. However, under the assumption of restricted path length, it provides a better approximation as stated in Theorem 2.2.2.

Theorem 2.2.2. *The greedy algorithm provides an $O(\log K)$ approximation in polynomial time for the path length restricted version of MSP.*

Proof. Let ξ be the size of minimum survivable path set. Let n_i be the number of fibers that are not survived after the i^{th} iteration of the greedy algorithm. Clearly, we have $n_1 \leq K$. Now, note that there is a path that survives at least $\frac{n_1}{\xi}$ of the remaining n_1 fibers, because otherwise the size of the optimal path set would be larger than ξ. Hence, in the second iteration, the greedy algorithm would select a path that survives at least $\frac{n_1}{\xi}$ of fibers. Thus,

$$n_2 \leq n_1 - \frac{n_1}{\xi} \leq K(1 - \frac{1}{\xi}). \tag{2.9}$$

Similarly,

$$n_3 \leq n_2 - \frac{n_2}{\xi} \leq K(1 - \frac{1}{\xi})^2, \tag{2.10}$$

and in general,

$$n_i \leq K(1 - \frac{1}{\xi})^i. \tag{2.11}$$

The greedy algorithm will terminate when $n_t < 1$, and this condition is satisfied when

$$K(1 - \frac{1}{\xi})^t < 1, \tag{2.12}$$

where t is the total number of iterations. Since $1 - x < e^{-x}$ for $x > 0$, inequality (2.12) is satisfied when

$$Ke^{-\frac{t}{\xi}} \leq 1 \Leftrightarrow t \leq \xi \times \log K. \tag{2.13}$$

Therefore, the greedy algorithm provides an $O(\log K)$ approximation.

To prove the polynomial time complexity, note that in each iteration of the greedy algorithm, the best path can be found in $O(m^K)$ by searching through all the paths (see the proof of Theorem 2.2.1). Furthermore, as mentioned above, the greedy algorithm terminates in at most $K + 1$ iterations. Therefore, the computational complexity of the greedy algorithm is $O(Km^K)$. □

Although the greedy algorithm runs significantly faster than the exhaustive search algorithm, its running time can still be prohibitive for large K and m. Hence, we develop a novel randomized algorithm which has a considerably better running time.

This algorithm builds upon solutions to the closely related Set Cover and Hitting Set problems [13]. In particular, the algorithm is based on ε-net, a concept in computational geometry, which provides an approximation algorithm for the Hitting Set problem.

2.2.2 ε-net Algorithm

Our ε-net algorithm is an iterative algorithm which selects each path with some probability. If all the fibers are survived by the selected path set in the first iteration, the algorithm terminates. Otherwise, it changes the probability of selecting each path and selects a new set of paths using the new probabilities, until all fibers are survived.

Let W_j be the weight of path j, initialized as $W_j = 1$. Define the weight of each fiber i to be the sum of the weights of paths surviving fiber i, i.e.,

$$W(f_i) = \sum_{j:a_{ij}=1} W_j. \tag{2.14}$$

Definition 2.2.1. A fiber is said to be ε-*Survivable* if

$$W(f_i) \geq \varepsilon \sum_{j=1}^{n} W_j \text{ for some } \varepsilon \in (0,1), \tag{2.15}$$

where n is the total number of paths.

Note that when all the paths have the same weight of 1, a fiber is ε-*Survivable* if it is survived by at least $\varepsilon \times n$ paths. Hence, if a fiber is ε-*Survivable* with large ε, then it is likely to be survived by randomly selected paths. This observation is exploited in our ε-net algorithm as discussed below.

By applying the randomized algorithm for the hitting set problem from [17] and [18], we can obtain a path-selection algorithm for selecting a random subset of paths that will survive all of the ε-*Survivable* fibers, with high probability. In particular, the algorithm finds a set of paths via s independent random draws (with replacement), such that in each draw, a path is selected from the entire path set according to the probability distribution $\mu(P_j) = \frac{W_j}{\sum_{j=1}^{n} W_j}, \forall j$.

Our ε-net algorithm iteratively applies this random path selection as follows. After each iteration, it checks the survivability of the selected path set. If not all fiber failures are survived, the algorithm doubles the weight of all paths that survive the failure of fibers in $\bar{S}$, where $\bar{S}$ is the set all the fibers that are not survived yet (so that such fibers are more likely to be survived by the new path set). The random path selection is repeated with the new probability distribution.

Let ξ be the optimal solution to the MSP problem. By applying the results in [19, 20], the following theorem can be proved.

Theorem 2.2.3. *Assume $s = c\frac{\log K}{\varepsilon} \log \frac{\log K}{\varepsilon}$, where c is a constant. The ε-net algorithm finds a set of survivable paths of size $O(\log K \log \xi)\xi$, with high probability.*

This theorem together with Lemma 2.2.1 implies that the ε-net algorithm finds a survivable path set of size $O(\log^2 K)\xi$. Moreover, it can be shown that the algorithm requires $O(K \log(\frac{m}{K}))$ iterations to achieve this performance bound. On the other hand, the path-selection algorithm needs to select $O(\frac{\log K}{\varepsilon} \log \frac{\log K}{\varepsilon})$ paths in each iteration. Therefore, the computational complexity of the ε-net algorithm is $O(K \log(K) \log(m) \log(\log(K)))$. Table 2.1 summarizes the performance of each algorithm under the path length restriction.

Method	Approximation	Running Time	T
ExS	Exact Solution	$O(m^{K(K+1)})$	D
Greedy	$O(\log K)$	$O(Km^K)$	D
ε-net	$O(\log K \log \xi)$	$O(K \log(K) \log(m) \log(\log(K)))$	P

Table 2.1: Performance bounds under path length restricted version: ExS-Exhaustive Search, T-Type, D-Deterministic, P-Probabilistic

2.3 Wavelength Restricted version

Another important practical constraint is that in WDM-based networks, the number of lightpaths using a fiber is limited to say W, which is the number of wavelengths supported over a fiber. In this section, we assume that a set of logically disjoint paths with their mapping on the physical topology is given, and the goal is to find a minimum survivable path set among those paths under the WDM restriction. Note that the set of logically disjoint paths can be abstract to a logical topology with two nodes and parallel links (e.g., the one in Fig. 3-1(a)). Clearly, in this setting, the WDM restriction implies that *each fiber can be used by at most W paths*. Using this property, it can be shown that the MSP problem under the WDM restriction can be solved in polynomial time. To prove this, we need the following lemma.

Lemma 2.3.1. *Under the wavelength restriction, the minimum number of survivable paths is at most $W+1$.*

Proof. Suppose that the minimum survivable path set contains more than $W+1$ paths. This implies that there exists a fiber whose failure disconnects at least $W+1$ paths (so that more than $W+1$ paths are needed for survivability), which contradicts to the fact that under the WDM restriction, each fiber can be used by at most W paths. □

Using the fact that the total number of paths that can use a fiber is restricted to W, we have show that the total number of logical paths in WDM setting is polynomial.

Lemma 2.3.2. *Under the WDM restriction, the number of given paths can be at most $W \cdot m$.*

Proof. By the Max-Flow Min-Cut Theorem, in the physical topology the number of disjoint paths between nodes s and t is equal to the minimum $s-t$ cut (MC). On the other hand, since a fiber can be used by at most W logical links, each physical path can carry at most W logical links. Therefore, the maximum possible number of logical paths is $W \cdot MC \leq W \cdot m$. □

Theorem 2.3.1. *Under the wavelength restriction, the MSP problem can be solved in polynomial time.*

Proof. By Lemma 2.3.2, the given path set contains $O(m)$ paths. By Lemma 2.3.1, we only need to enumerate path sets of size up to $W+1$ in order to find a minimum survivable path set. Clearly, this can be done in $O(m^{W+1})$ time. More details can be found in Appendix **??**. □

Although there exists a polynomial time optimal algorithm, it requires excessive computation for large values of W and m. As in the case of restricted path length, we have developed approximation algorithms with better running time. Table 2.2 shows the summary of our approximation algorithms under the wavelength restriction (See A.2 for details).

Method	Approximation	Running Time	T
ExS	Exact Solution	$O(W^{W+1}m^{W+1})$	D
Greedy	$O(\log m)$	$O(W^2m)$	D
ε-net	$O(\log W \log \xi)$	$O(W \log(W) \log(m) \log(\log(W)))$	P

Table 2.2: Approximation bounds under wavelength restricted version: ExS-Exhaustive Search, T-Type, D-Deterministic, P-Probabilistic

Chapter 3

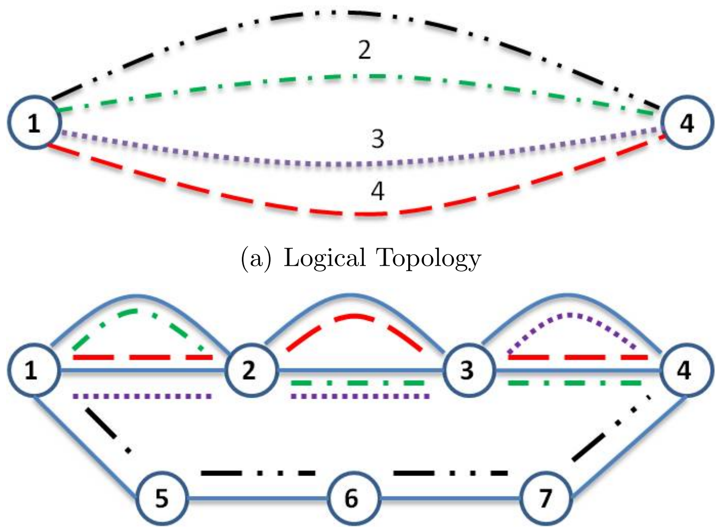
Minimum number of physical fibers in Survivable Paths (MFSP)

Our focus so far has been on providing protection using the minimum number of paths. In the this section, our goal is to find a survivable path set that uses the minimum number of fibers. This problem seems to have a direct connection to the minimum cost survivable path set problem where the cost of a path is the number of fibers used by that path. However, this is not true owing to the fact that costs of paths are not additive, i.e., a fiber that is used by multiple paths only adds one unit of cost. In order to make this point clear, consider Fig. 3-1. A minimum cost survivable path set problem will find paths 1 and 2 as the set of survivable paths with total cost 7, while the MFSP problem will find paths 2, 3 and 4 as the optimal survivable path which has the total cost 6. In the next section we will develop ILP formulations, and analyze the complexity of MFSP.

3.1 ILP Formulation

3.1.1 Path-Based Formulation

We start with an ILP formulation of the problem. Similar to the MSP problem, the MFSP problem can be formulated in several different ways, but here we only



(a) Logical Topology

(b) Routing

Figure 3-1: Routing in Multilayer Network

present the path-based formulation which will be used for developing heuristics and approximation algorithms. Given the set of paths and associated fibers, for each path j, assign a binary variable P_j which takes the value 1 if path j is selected and 0 otherwise. Similarly, for each fiber i, assign a binary variable f_i which takes the value 1 if fiber i is selected and 0 otherwise. The matrix A and vector e are defined in the same way as in the MSP formulation (2.1)-(2.3).

$$MFSP: \qquad \text{minimize} \quad \sum_{i=1}^{m} f_i \tag{3.1}$$

$$\text{subject to} \quad A \times P \geq e \tag{3.2}$$

$$f_i \geq P_j \qquad \forall f_i \in P_j \tag{3.3}$$

$$P_j \in \{0,1\} \quad \forall P_j \tag{3.4}$$

In the above, the objective function is the number of fibers used by the selected paths. Again, the constraints in (3.2) require the selected path to be survivable. The constraints in (3.3) relate the selected paths and fibers, such that a fiber is selected if at least one of the paths using the fiber is selected. Clearly, the optimal solution to the above optimization problem gives a set of survivable paths that use the minimum number of fibers.

3.1.2 Link-Based Formulation

The idea of Link-Based formulation for MFSP problem is the same as formulation for MSP. The only difference is that we do not need to find the paths in the main logical topology using flow constraints and minimize the number of paths.

For each fiber r, let f_r be a binary variable which takes the value 1 if fiber r is selected, and 0 otherwise. Similar to MSP link-based formulation, variables x_{ijk} refers to the logical links (i, j) and constraint (3.6) refers to the flow constraints in the remaining logical topology E_L^k correspondent to the failure of fiber k.

$$\text{minimize} \qquad \sum_{r=1}^{m} f_r \tag{3.5}$$

$$\text{subject to} \left.\begin{aligned} &\sum_{(s,j)\in E_L^k} x_{sjk} = 1 && \forall \text{ Fiber } k \\ &\sum_{(i,t)\in E_L^k} x_{itk} = 1 && \forall \text{ Fiber } k \\ &\sum_{(i,j)\in E_L^k} x_{ijk} - \sum_{(j,i)\in E_k} x_{jik} = 0 && \forall k, \forall i \neq s, t \end{aligned}\right\} \tag{3.6}$$

$$f_r \geq x_{ijk} \qquad \forall i, j, k, \forall f_r \in x_{ijk} \tag{3.7}$$

$$x_{ijk} \in \{0, 1\} \qquad \forall i, j, k \tag{3.8}$$

Constraint (3.7) shows the relation between the selected logical links and fibers, such that fiber i is selected if at least a logical linkusing f_i is selected. The objective function is the sum of selected fibers. Hence, solving this ILP formulation will find a set of survivable paths that uses minimum number of fibers.

3.1.3 Cut-Based Formulation

A set of paths will survive any single failure, if in the occurence of any fiber failure there exist at least one path from s to t.

Let f_r be a binary variable for each fiber r, and y_{ij} be a binary variable for each logical link ij. Let N be the set of all nodes in the logical topology. The objective function is minimizing the total number of selected fibers.

$$\text{minimize} \sum_{r=1}^{n} f_r \tag{3.9}$$

$$\text{subject to} \sum_{(i,j)\in E_L^k : i\in S, j\in \bar{S}} y_{ij} \geq 1 \quad \forall k, S \subset N, S \neq N, \emptyset$$

$$s \in S, t \in \bar{S} \tag{3.10}$$

$$f_r \geq y_{ij} \qquad \forall i, j, \forall f_r \in y_{ij} \tag{3.11}$$

$$y_{ij} \in \{0, 1\} \tag{3.12}$$

Define "$s-t$ cut" as a cut $[S, \bar{S}]$ such that $s \in S$ and $t \in \bar{S}$. Constraint (3.10) shows that for every "$s-t$ cut" in remaining graph E_L^k, there exist at least one logical link from S to $\bar{S}$. This will gaurantee the survivability of network in the failure of fiber k. Constraint (3.11) builds the relation between selected logical links and fibers, such that a fiber i will be selected if at least one logical link using f_i is selected. Constratints (3.10) and (3.11) select the fibers used by a set of survivable paths and the objective function will find the solution to the MFSP problem.

3.2 MFSP Complexity

The MFSP problem can be shown to be NP-hard.

Theorem 3.2.1. *Computing the set of survivable paths using the minimum number of physical fibers is NP-hard. In addition, this minimum value cannot be approximated within any constant factor, unless $P = NP$.*

Proof. We provide a mapping from the Minimum 3-Set Cover problem, which is a special version of the Set Cover problem where each set has exactly 3 elements, to the MFSP problem. The Minimum 3-Set Cover problem is NP-hard, and holds all the inapproximability properties of the Minimum Set Cover problem.

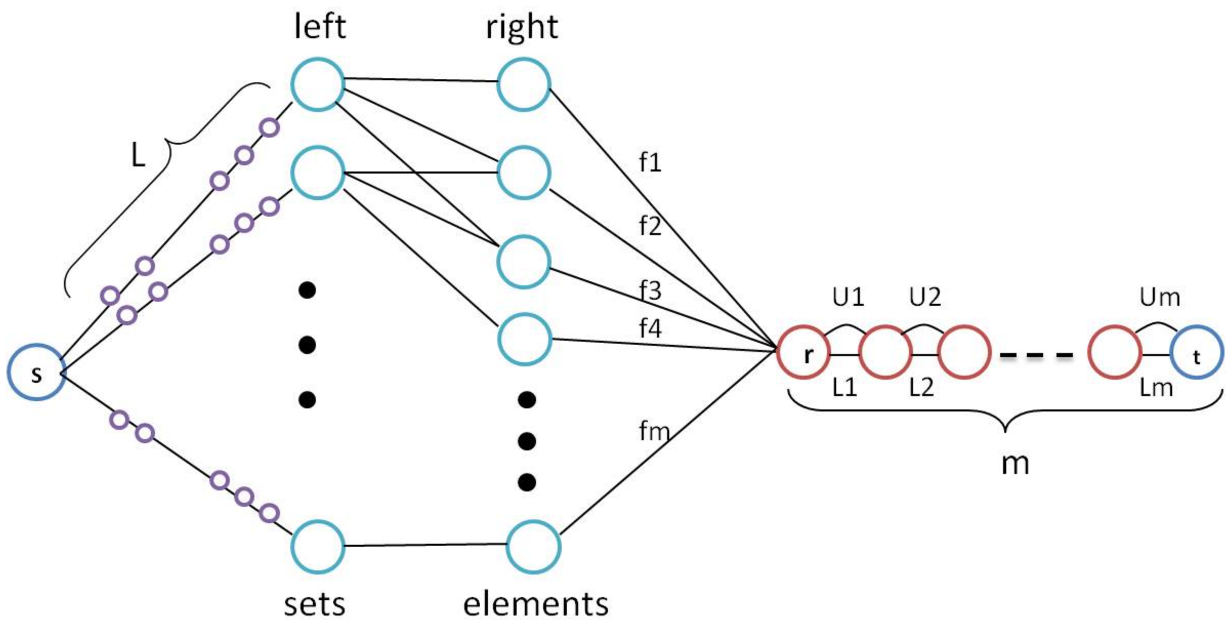


Figure 3-2: Physical Topology

Consider an instance of the Minimum Set Cover problem with the ground set E and a family of subsets F. Suppose that each subset in F contains only 3 elements. To show a mapping, we construct a physical topology as shown in Fig. 3-2, such that each node on the left corresponds to a subset in $F = \{C_1, ..., C_{|F|}\}$ and the nodes on the right are the elements of $E = \{e_1, ..., e_m\}$. Node j on the left is connected to node i on the right if and only if $e_i \in C_j$. Note that a node on the left is connected to only three nodes on the right (i.e., each set contains only three elements).

We can construct a logical topology and its lightpath routing over the physical topology; such that for protection, we need to have m paths from s to t that pass through all the nodes on the right. Moreover, since each path between s and the nodes on the left uses a large number of fibers, we should select a survivable path set that uses the minimum number of nodes on the left. Consequently, the minimum fiber survivable path set for the aforementioned layered network gives a minimum set cover for the given instance of E and F, which shows the NP-hardness of the MFSP problem.

In the physical topology shown in Figure 3-2, nodes s and t are the starting and ending nodes. Each node on the left side (n nodes) is connected to 3 nodes on the right such that all the nodes on right are covered by the nodes on left. There are L nodes between s and each node on the left where L is a large number (say $L \geq 3m + 3n$ so that the left hand side should be the first priority when minimizing the used number of fibers) and there are $m \geq 3$ nodes on the tail of the graph such that every node on right connects to the tail through the first node r.

The logical topology is shown in Figure 3-3.

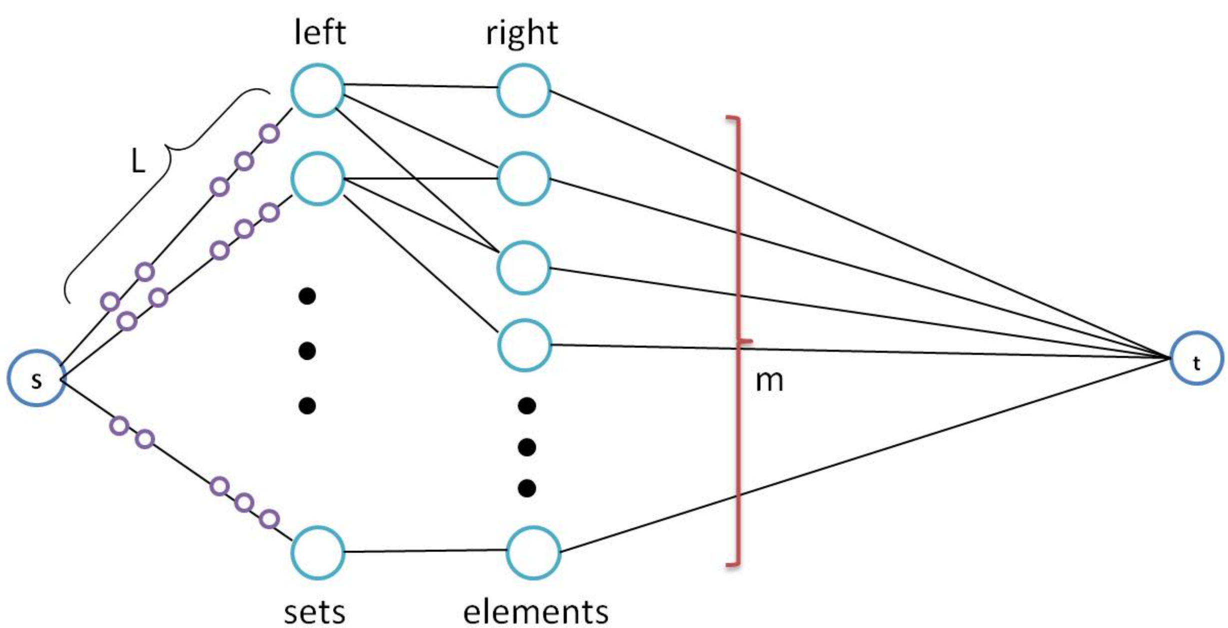


Figure 3-3: Logical Topology

In the logical topology, from s to the nodes on right, each fiber is also a lightpath, while from nodes on right side to t, there are m parallel lightpaths with a specific routing. The first lightpath will be routed on fibers f_1, U_1 and L_2 to L_m, the second lightpath will be routed on fibers f_2, L_1, U_2 and L_3 to L_m and so on. Therefore, lightpath i will use fibers f_i, U_i and all the other L_js ($j \neq i$).

To survive any single failure in the firbers from right nodes to node t, we need to have at least m paths, each going through one of the parallel logical links. These m paths will not share any fiber from nodes on left to nodes on right, thus any single failure on the fibers between left and right nodes will be survived. Finally, to survive any fiber failure from node s to nodes on the left, it is enoough that at least two of paths use disjoint logical links from node s to left nodes.

Consequently, it is enough just to have m paths covering all nodes on the right hand side. On the other hand, paths between s and left nodes use a large number of fibers. To have a set of paths using the minimum number of fibers, we need to pick the minimum number of nodes from left, to cover all the nodes on right which is a mapping from minimum 3-set cover problem to our problem. The remaining of the proof is explained in the main text. □

Since the MFSP problem is a reduction from the minimum 3-set cover problem, it is unlikely that the MFSP problem has an efficient optimal algorithm. For this reason, we develop new heuristics and approximation algorithms. In particular, as in the previous section, we focus on the practical scenario where the number of paths on a fiber is at most W, i.e., the wavelength restricted setting. We first present a greedy algorithm, and then a randomized rounding algorithm based on the path-based formulation for MFSP.

3.3 Approximation Algorithms

3.3.1 Additive Cost Greedy Algorithm (ACG)

Recall that the goal is to find a survivable path set that uses the minimum number of fibers. Hence, it is desired to select a path that uses a small number of fibers while surviving many new fibers (i.e., fibers not survived by already selected paths) as possible. Note that this is clearly different from the MSP problem where the number of fibers does not matter. The Additive Cost Greedy algorithm requires the set of paths and associated fibers as input. We define a new cost metric in order to take into account the two factors simultaneously. Let C_j be the number of fibers used by path j. The "amortized cost" AC_j of path j, which is updated for every iteration, is defined as follows:

$$AC_j = \frac{C_j}{\#\text{newly survived fibers by } P_j},$$

where the denominator is the number of fibers survived by path j and not survived by the previously selected paths. Our greedy algorithm selects a path with minimum amortized cost, updates the amortized costs of the remaining paths, and continue until all the fibers are survived. This greedy algorithm, which we call the Additive Cost Greedy algorithm, gives an approximate solution.

Theorem 3.3.1. *The Additive Cost Greedy algorithm provides an $O(W \log m)$ approximation to the MFSP problem.*

Before proving Theorem 3.3.1, we need to prove two other Lemmas. Consider a set S of survivable paths and let F be the total number of fibers used by all paths in S.

Lemma 3.3.1. *Under the wavelength restricted assumption, the following inequality holds:* $\frac{1}{W}\sum_{j=1}^{|S|} C_j P_j \leq F$

Proof. Figure 3-4 shows the relation between paths and the fibers used by them. There exist an edge between node j on left (P_j) and node i on right (f_i) if $f_i \in P_j$. Since each path j is using C_j fibers, the total number of edges is $\sum_{j=1}^{|S|} C_j P_j$. On the other hand, by assumption, each fiber can be used by at most W paths, therefore each right node can be incident to at most W edges. Thus, we have $\sum_{j=1}^{|S|} C_j P_j \leq WF$, which completes the proof. □

Lemma 3.3.2 in below is a direct consequence of Lemma 3.3.1.

Lemma 3.3.2. *Let S be a set of survivable paths and R be the set of all possible survivable path sets ($S \in R$). Let ξ be the minimum number of fibers used by a survivable path set. Then, we have:* $\frac{1}{W}\min_{S\in R}\sum_{j=1}^{|S|} C_j P_j \leq \xi \leq \min_{S\in R}\sum_{j=1}^{|S|} C_j P_j$.

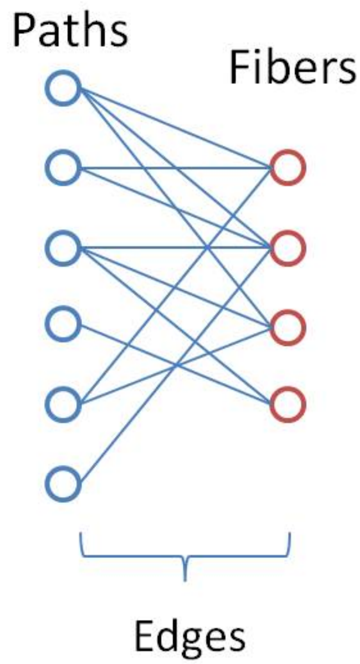


Figure 3-4: relation between paths and fibers

Proof. Lemma 3.3.1 gives the left inequality. In the right inequality, $\min_{S \in R} \sum_{j=1}^{|S|} C_j P_j$ outputs a set S of survivable paths, therefore it is feasible and gives an upperbound for the optimal solution. □

By Lemma 3.3.2, the optimal solution to the problem $\min_{S \in R} \sum_{j=1}^{|S|} C_j P_j$ provides a W approximation to the MFSP problem. Note that this problem seeks to find a set of survivable paths with minimum cost where the cost of a path is the number of fibers used by that path, and these costs are assumed to be additive. Clearly, this problem is a reduction from minimum cost set cover problem. Since the minimum cost set cover problem is NP-hard, finding a set of survivable paths with minimum additive costs is also NP-hard. Therefore, we use the explained additive cost greedy algorithm to approximate the additive cost survivable path set problem. Now we can prove the $O(W \log m)$ bound stated in Theorem 3.3.1.

Proof. Let ξ be optimal value of MFSP problem. By the argument in [21], the additive cost of paths selected by ACG is not larger than $O(\log m)\xi$, i.e.,

$$\frac{Greedy}{\log m} \leq \min_{P \in S} \sum_{j=1}^{n} C_j P_j \tag{3.13}$$

where *Greedy* denotes the additive cost of ACG. Combining equation (3.13) and Theorem 3.3.2 gives the following inequality:

$$\frac{Greedy}{W \log m} \leq \xi \tag{3.14}$$

□

Note that the number C_j in the additive cost of path j does not change over iterations. That is, the additive cost implicitly assumes that selecting path j will add C_j fibers to the total cost, while only the number of new fibers is added to the total cost. Therefore, one can better take into account the actual change to the cost by updating C_j as the number of fibers that are used by path j and not used by the previously selected paths. In Section 5, we will show that this Non-additive Cost Greedy (NACG) algorithm works better than the ACG, by finding survivable path sets with fewer fibers.

3.3.2 Randomized Rounding Algorithm

Randomized rounding is a widely used technique to solve difficult integer optimization problems. In general, randomized rounding scheme solves the Linear Program (LP) relaxation of the original ILP formulation, and rounds the solution randomly. In our case, the LP relaxation of the MFSP problem is given as

$$\text{LP relaxation:} \quad \text{minimize} \quad \sum_{i=1}^{m} f_i \tag{3.15}$$

$$\text{subject to} \quad A \times P \geq e \tag{3.16}$$

$$f_i \geq P_j \qquad \forall f_i \in P_j \tag{3.17}$$

$$0 \leq P_j \leq 1. \tag{3.18}$$

Let P_j^* and f_i^* be the optimal values of path j and fiber i. Note that the above path-based LP uses the set of paths and associated fibers as input. Our randomized rounding algorithm to solve the MFSP problem works as follows:

1. Initialize $S = \emptyset$. Solve the relaxed problem.
2. Select each path j with probability P_j^*, and add it to S if selected.
3. Repeat step 2 for T times.

Since paths are selected randomly, some fibers may not be survived in one iteration. Clearly, as the number of iterations T increases, the probability of surviving all of the fibers increases. On the other hand, it may increase the number of selected paths and thus fibers. Therefore, the parameter T determines the survivability probability and the approximation quality of the solution. The following theorem characterizes this relationship.

Theorem 3.3.2. *With $T = O(\log \frac{m}{1-q})$ iterations, the randomized rounding algorithm gives an $O(W \log \frac{m}{1-q})$ approximation with probability at least q.*

Proof. We first find an upper bound on the expected number of fibers selected in each iteration (which gives the approximation quality of the solution), and then, the probability of survivability is derived.

Expected Number of Selected Fibers

Note that fiber i is selected if any of the paths using the fiber is added to the path set S. Moreover, in each iteration, each path j is added with probability P_j^*. To count the number of selected fibers, define a random variable F_i for each fiber i such that $F_i = 1$ if fiber i is selected and 0 otherwise. The expected number of fibers selected in each iteration can be written as

$$E[\sum_{i=1}^{m} F_i] = \sum_{i=1}^{m} \mathbf{Pr}(F_i = 1) = \sum_{i=1}^{m}(1 - \mathbf{Pr}(F_i = 0)) \tag{3.19}$$

Therefore, we need to compute $\mathbf{Pr}(F_i = 0)$, which is the probability of a fiber not being selected. Note again that a fiber is not selected if none of the paths using the fiber are selected. It follows that

$$\mathbf{Pr}(F_i = 0) = \prod_{j:a_{ij}=0} (1 - P_j^*) \tag{3.20}$$

$$\geq \prod_{j:a_{ij}=0} (1 - f_i^*) \quad \text{(by constraint (3.17))} \tag{3.21}$$

where the equality is due to the independence of path selections. Let w_i be the number of paths that use fiber f_i, i.e., $w_i = |\{j : a_{ij} = 0\}|$. Then, we can obtain

$$\mathbf{Pr}(F_i = 0) \geq \prod_{j:a_{ij}=0} (1 - f_i^*) = (1 - f_i^*)^{w_i}, \tag{3.22}$$

$$\mathbf{Pr}(F_i = 1) \leq 1 - (1 - f_i^*)^{w_i}. \tag{3.23}$$

Finally, by using the fact that $1 - (1 - x)^n \leq nx$, the probability of selecting a fiber can be upper-bounded as

$$\mathbf{Pr}(F_i = 1) \leq w_i f_i^*. \tag{3.24}$$

Combining (3.19) and (3.24) yields the following bound on the expected number of fibers selected in each iteration:

$$
\begin{aligned}
E[\sum_{i=1}^{m} F_i] &\leq \sum_{i=1}^{m} w_i f_i^* \\
&\leq W \times \sum_{i=1}^{m} f_i^* \quad \text{(by wavelength restriction)} \\
&= W \times LP^*, \qquad (3.25)
\end{aligned}
$$

where LP^* is the optimal value of the LP relaxation.

Probability of Survivability

Next, we derive an upper bound on the probability that the selected path set is *not* survivable, by applying the idea of the feasibility argument in [15]. First, for each fiber i, the probability that the selected path set cannot survive the failure of fiber i can be written as follows:

$$\mathbf{Pr}(\text{fiber } i \text{ not survived in one iteration}) \qquad (3.26)$$

$$= \mathbf{Pr}(\text{none of paths surviving fiber } i \text{ are picked}) \qquad (3.27)$$

$$= \prod_{j:a_{ij}=1} (1 - P_j^*) \qquad (3.28)$$

$$\leq \prod_{j:a_{ij}=1} e^{-P_j^*} \qquad \text{using}(1 - x \leq e^{-x}) \qquad (3.29)$$

$$\leq e^{-\sum_{j:a_{ij}=1} P_j^*} \leq \frac{1}{e}. \qquad \text{(using constraint 3.16)} \qquad (3.30)$$

Since the randomized rounding runs for T iterations with $T = \log \frac{m}{1-q}$, we can obtain

$$\mathbf{Pr}(f_i \text{ not covered in all iterations}) \leq \frac{1}{e^{\log \frac{m}{1-q}}} = \frac{1-q}{m}. \qquad (3.31)$$

Thus, by the union bound,

$$\mathbf{Pr}(\text{there exist an unsurvived fiber}) \leq m \times \frac{1-q}{m} = 1 - q. \qquad (3.32)$$

Approximation Result

By (3.25), the total expected number of fibers after T iterations is bounded as

$$E[\text{Total \# fibers}] \leq W \log \frac{m}{1-q} LP^* \tag{3.33}$$

Since the solution is in integer form, it is an upperbound for the ILP solution. Thus, with probability at least q,

$$\frac{E[\text{Total \# fibers}]}{W \log \frac{m}{1-q}} \leq ILP \leq E[\text{Total \# fibers}]. \tag{3.34}$$

□

3.3.3 Random-Sweep Greedy Algorithm (RSG)

Next, we present a new Greedy algorithm for the MFSP problem, which is called the Random-Sweep greedy. Unlike the Greedy algorithm discussed in Section 2, the RSG removes a path (from the selected path set) which survives the fibers covered by other paths; so that the size of the path set can be further reduced while maintaining the survivability. Although we could not quantify the performance of this algorithm, it performs near optimally in some scenarios as will be shown in Section 5.

The RSG algorithm also requires the knowledge of the set of paths and associated fibers. Let S_j be the set of fibers that are survived by path j. Moreover, let the cost C_j of path j in each iteration be the number of fibers that are used by path j, and not used by the previously selected paths. Using the cost function C_j, define the amortized cost AC_j as the ratio of C_j to the number of newly survived fibers by path j. The first two iterations of RSG are the same as the Non-Additive Cost greedy algorithm. That is, in each iteration, it selects a path with minimum amortized cost. If the first two paths survive all of the fibers, the algorithm terminates. Otherwise, it continues as follows.

Suppose the RSG algorithm is in the i^{th} iteration. First, find a path, say i, with minimum amortized cost among the remaining paths. Then, pick a path, say j,

Method	Approximation	Running Time	T
ExS	Exact	$O(W^{W+1}m^{W+1})$	D
ACG	$W \log m$	$O(W^2 m)$	D
ε-net	$W \log W \log \xi$	$O(W \log(W) \log(m) \log(\log(W)))$	P
RR	$W \log \frac{m}{1-p}$	$\log \frac{m}{1-p}$	P
RSG	nearly Opt.	$O(W^2 m)$	D

Table 3.1: Approximation bounds under wavelength restricted version of MFSP: ExS-Exhaustive Search, RSG-Random Sweep Greedy, ACG-Additive Cost Greedy, RR-Randomized Rounding, T-Type, D-Deterministic, P-Probabilistic

randomly from the previously selected paths and find $S_j \cup S_i$, which is the set of fibers that are survived by either path i or path j. If there exists a path k among the previously selected paths such that $S_k \subset S_i \cup S_j$, remove path k from the selected paths. Note that removing such a path does not affect the survivability of the selected path set, i.e., the same set of fibers are still survived after the removal. More importantly, we can possibly decrease the number of fibers used by the selected paths.

Table 3.1 shows the summary of our algorithms for the MFSP problem. Note that we have also developed an ε-net algorithm and its details can be found in A.3.

Chapter 4

Communication Networks Dependency on the Power Grid

Optical Communication Networks are merely one example of a multilayer network. In fact, there is a strong dependency between almost every network and utility [22]. Figure 4-1 shows such connections between critical infrastructures. There is an extensive literature on the reliability of the power grid or communication networks, as separate networks, but there is very little research on modeling the effects of these networks on each other, especially from the perspective of reliability. In this chapter, we will focus on the dependency of communication networks on the power grid, and show that a single failure in the power grid may lead to a large scale failure in the communication network.

4.1 Power grid Model

In the literature of power grid's reliability, people mostly have used the linearized power flow model as the base model. This model is also known as DC model. Here, we briefly explain the DC model [23]. Suppose we have a directed network $G_P = (V_P, E_P)$ with nodes V_P and edges E_P. Given this, we have the following properties for an operating power grid:

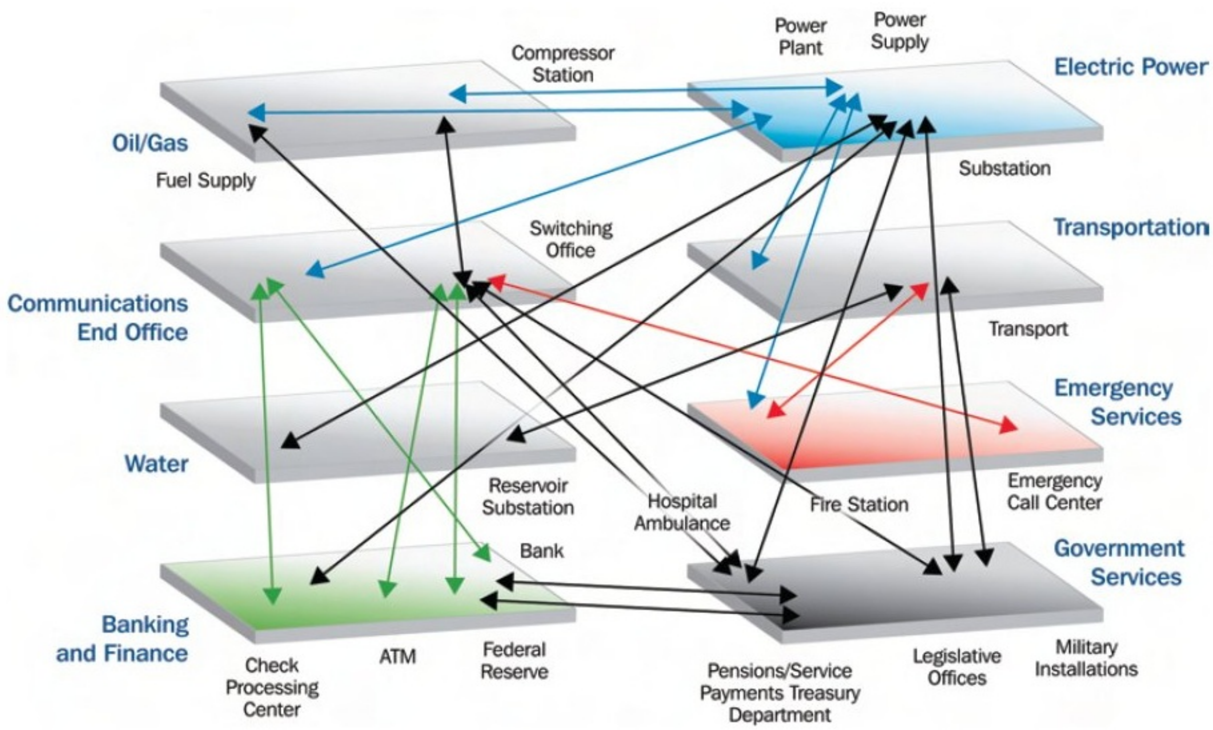


Figure 4-1: Interdependency between Infrastructures in Today's World (Source: [1])

- There are three types of nodes in the power grid: 1- Supply Nodes or Generator denoted by set G; 2- Demand Nodes or Load denoted by set D; 3- Substations which are neither generator nor load. In fact, these are the nodes that form the network between the source and destination.

- The power of the generator node i is bounded below by zero and above by P_{max}.

- For each demand node i there exist a nominal demand D^i_{nom}.

- The arcs E_P in the network corresponds to the power lines, where each power line has positive resistance x_{ij} and maximum power capacity u_{ij}.

Given these parameters of the power grid $G_P = (V_P, E_P)$, we can find the flow on each arc. However, unlike in communication networks, the power flow equations should satisfy some physical constraints. These physical constraints are due to the fact that electricity flows over the links based on Kirchhoff's law. Therefore, unlike common networks we cannot control the flow of electricity, unless we change some of the physical constraints of the network such as line resistance or capacity.

In the linearize model of the power flow in network $G_P = (V_P, E_P)$, the following constraints should be satisfied [23]:

$$\sum_{j:(i,j)\in E_P} f_{ij} - \sum_{j:(j,i)\in E_P} f_{ji} = \begin{cases} p_i & i \in C \\ -D_i & i \in D \\ 0 & \text{Otherwise} \end{cases} \tag{4.1}$$

$$\theta_i - \theta_j - x_{ij} f_{ij} = 0 \quad \forall (i,j) \tag{4.2}$$

$$|f_{ij}| \leq u_{ij} \quad \forall (i,j) \tag{4.3}$$

$$P_i^{min} \leq P_i \leq P_i^{max} \quad \forall i \in C \tag{4.4}$$

$$0 \leq D_j \leq D_j^{nom} \quad j \in D \tag{4.5}$$

In this formulation, the flow in the line ij is denoted by f_{ij} and the phase angel of the voltage at each node i is denoted by θ_i. Here, equation 4.1 is the typical network flow constraint, and equations 4.3,4.4 and 4.5 are capacity constraints for power lines, sources (generator), and destinations (demands). The only unusual constraint is equation 4.2 which relates the power and resistance of line ij to the phase angel of its end nodes θ_i and θ_j.

Now suppose that one of the nodes or links fail. To satisfy the constraints in the new setting, the electricity itself reroutes in the network based on equation 4.2. Now, if the power in a line exceeds its capacity u_{ij}, the overloaded line heats up and eventually fails. This is referred to as a *cascading failure.* In many cases, the cascading failure can cause very major blackouts in the power grid. Blackouts in northeast America in 2003 [24] or in Italy in 2003 [25] are examples of widespread outage due to the cascading failures in the power grid.

Despite all of these complexities in the power grid, our goal is to analyze the effect of failures in the power grid on communication networks. Therefore, we only focus on the connectivity of the power grid, rather than the amount of flow in the lines, or the satisfied demand; this can be done by ignoring the power line capacities, i.e.

cascading failures due to overloading are ignored. Under this situation, a substation operates, i.e. has power, if and only if it is connected to a generator (whether directly or through a path of other substations.

To simplify the model, we assume that generators are not prone to failure. We can also assume that, without loss of generality, there is only one generator G in the power grid, and all of the substations should be connected to G in order to get power. This can be done by connecting all of the generators to a super generator. Figure 4-2 shows the process of transforming a network with multiple generators to a network with a single super generator.

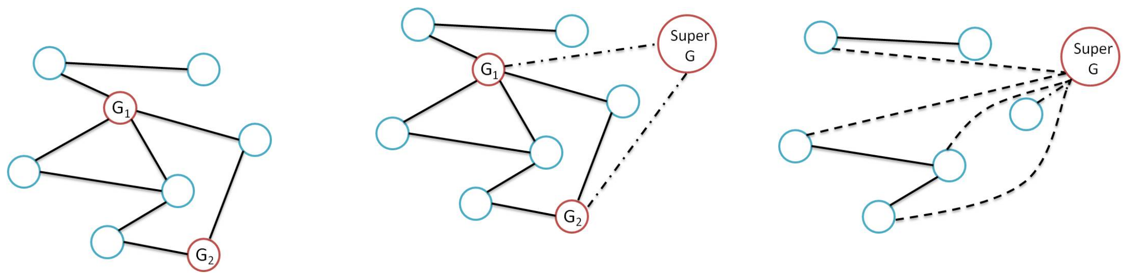


Figure 4-2: Process of Transformation

In our model, the communication network is the logical layer, and the underlying power grid is the physical layer such that any failure in the physical layer may lead to failures in the logical layer. As explained above, the power grid can be modeled with a graph $G_P = \{V_P, E_P\}$ where V_P's are the substations, and the E_P's are the power lines connecting the substations to each other and to the generator. Moreover, the topology of communication network is denoted by graph $G_L = \{V_L, E_L\}$ where the nodes V_L are the routers, and the links E_L are the links connecting them.

In this model, we analyze the network with respect to single failures in the power grid. However, unlike in optical communication networks, the failures in the physical topology are not independent. In fact, if a substation fails, it can disconnect other substations from the generator and results in their failure.

Moreover, the dependency between the logical and physical layer is through the nodes. Here, each router in the communication network needs electric power for operation, and it receives this power from the power substations in the physical layer. Therefore, if a substation fails, all of the routers that get their power from this substation will also fail. Note that if a router is connected to more than one substation, all of the substations should fail in order to cause the failure of that router. This is in contrast to optical communication networks where if a lightpath is routed on multiple fibers, the failure of each fiber could result in the failure of that lightpath. This demonstrates another key difference between the power grid and optical communication networks. To simplify the model, we assume that every router is getting power from only one substation. Figure 4-3 shows the model of the communication network dependency on the power grid.

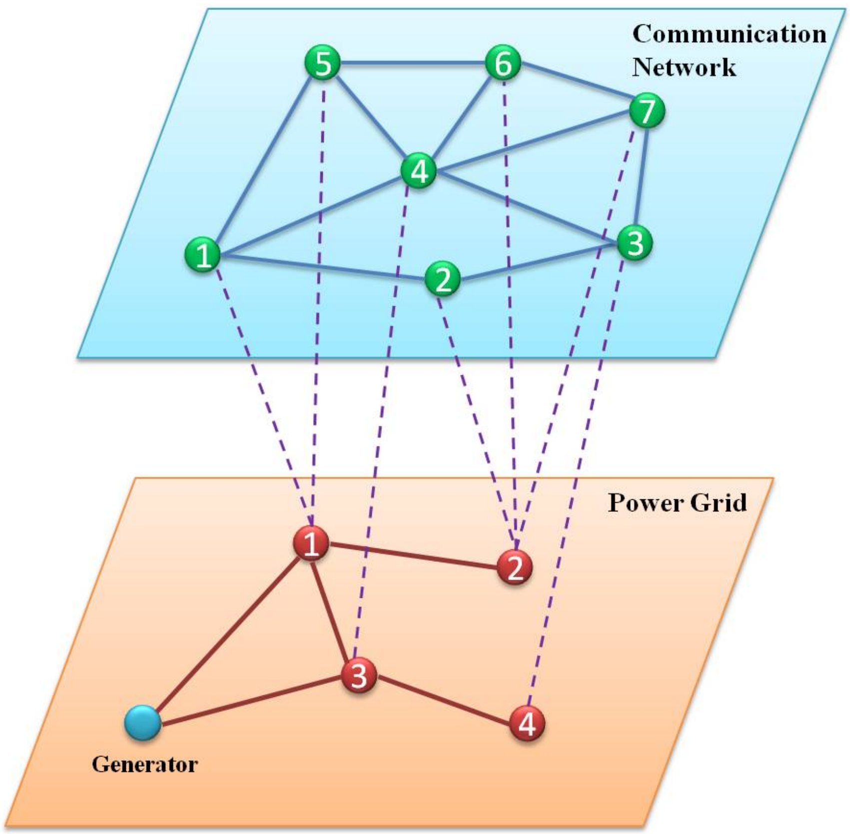


Figure 4-3: Communication Network and Power Grid Dependency Model

Figure 4-4 shows how a single failure in the power grid can cause multiple failures in the communication network. As shown, substation 1 fails at the beginning and causes the failure of routers 1 and 5. Moreover, the failure of substation 1 disconnects substation 2 from the generator. Subsequently, the failure of substation 2 causes the failure of routers 2, 6 and 7.

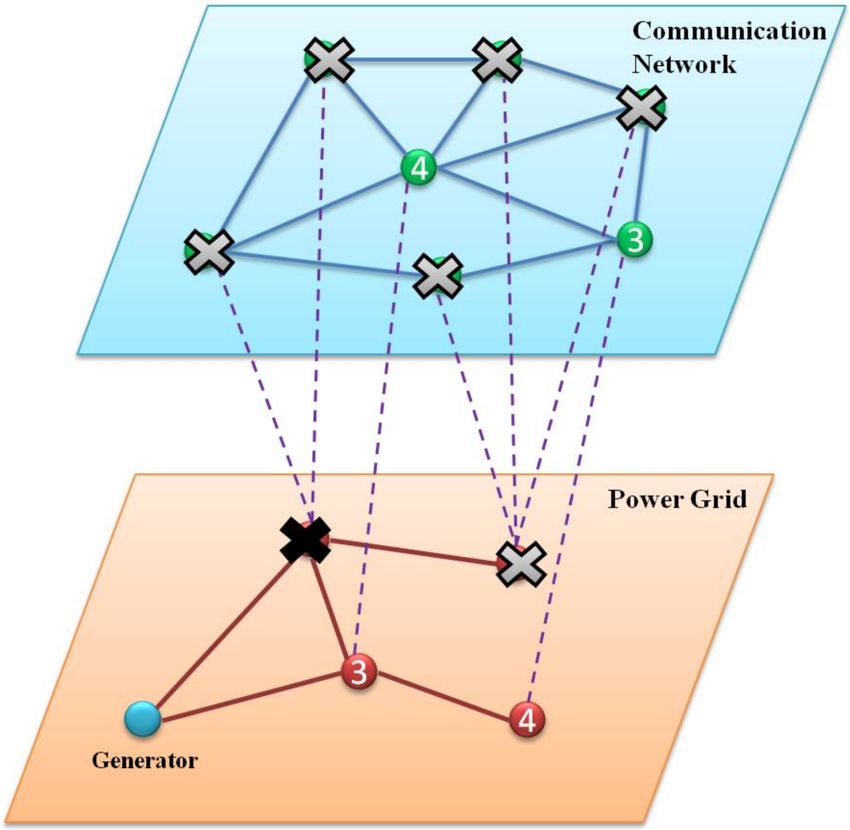


Figure 4-4: Failure of one substation cascading in both power grid and communication network

In the following sections, we will try to answer some of the fundamental questions regarding reliability of communication networks in this setting. In order to increase the reliability of communication network, we tend to use redundant paths: one as primary path, and the other as back-up path which is disjoint from the primary one. However, in this multilayer setting, disjointness of paths in communication layer is not enough. This is due to the fact that two logically disjoint paths may use the same substation and fail simultaneously. In the next section, we focus on the problem of finding two primary and back up paths that are both physically and logically disjoint.

4.2 SRLG-Disjoint Paths in Communication-Power Network

Consider two nodes s and t in the communication network layer. Suppose that we want to find one primary path, and one back up path that is physically disjoint from the primary path, in the sense that failure of one substation cannot cause the failure of both paths. Here, the risks are the failure of substations, and the problem reduces to finding two SRLG-disjoint paths between source s and destination t. Hu proved that the problem of finding two SRGL disjoint paths in a general setting is NP-complete [7]. Using a similar structure, we prove that finding two SRLG-disjoint paths in our special setting is also NP-complete.

Theorem 4.2.1. *In two layer setting of communication network and power grid, the problem of finding two SRLG-disjoint paths between two nodes in the communication network is NP-complete.*

Proof. This proof is based on the reduction of the NP-complete problem of set splitting to our problem of finding two SRLG-disjoint paths between two nodes in the communication network. To show this reduction, we need to construct a special power grid and communication network which fits in the structure of set splitting problem.

Set Splitting Problem: Given a collection of subsets C_i $(i = 1, \cdots, K)$ of a finite set C, the problem is finding two disjoint subsets of C such that each subset contains at least one element in C_i for $i = 1, \cdots, K$ [13].

First, we construct a special communication network as our logical layer $G_L = \{V_L, E_L\}$, using a similar notation as in [7] (See Figure 4-5). As shown in Figure 4-5, graph G_L consists of K subgraphs $G_{L,1}, \cdots, G_{L,K}$. Each subgraph contains a set of nodes, and every node in subgraph i is connected to every node in subgraph $i+1$. Moreover, node s is connected to all of the nodes in subgraph $G_{L,1}$, and node t is connected to all of the nodes in subgraph $G_{L,K}$.

We say a logical node is correspondent to a set of risks if the failure of each element

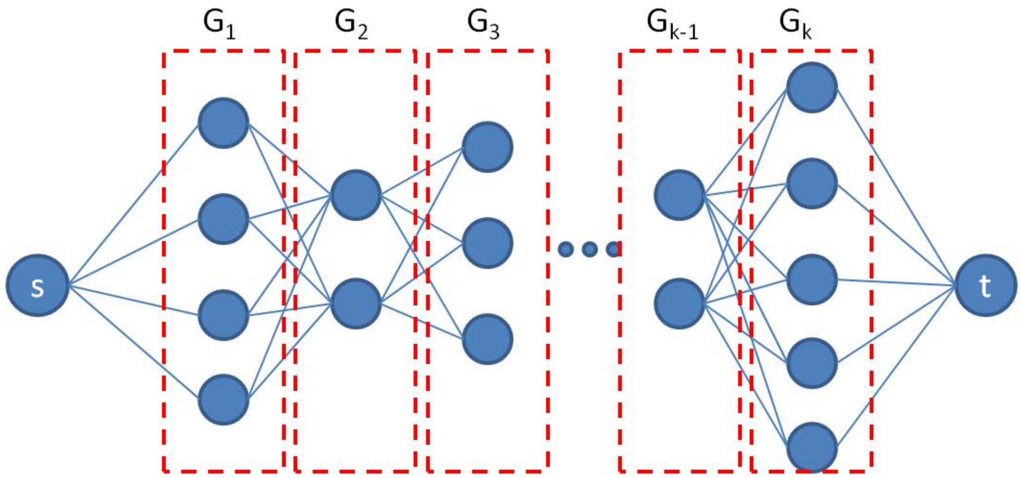


Figure 4-5: Logical Layer Special Structure

in the set leads to the failure of that logical node. In our setting, the risks are the substations, and the elements in risk set are the nodes in the physical topology V_P. We assume that in our special structure, every logical node is correspondent to a single risk; this implies that the failure of the substations should be independent. Otherwise, if $S1$ causes $S2$ to fail and $S2$ causes logical node i to fail, both $S1$ and $S2$ will be in the risk set of node i. Under the assumption of disjoint substation failures, it is easy to construct a physical layer correspondent for the logical layer, and a mapping between the logical and physical layer in Figure 4-5. We construct the physical topology as a star graph with one generator G in the center, and $|V_P|$ nodes as the substations directly connected to the generator G. To show the mapping between two layers, we connect every logical node to its risk in the physical layer (see Figure 4-6).

We also define a set of risks R_i for every subgraph $G_{L,i}$, where R_i is the set of substations that lead to the failure of nodes in $G_{L,i}$. Clearly, this is equivalent to the union of risks of all of the nodes in $G_{L,i}$.

Now, the mapping from the set splitting problem to our problem is similar to [7]. Based on the definition of Set Splitting problem, if we set C as the set of all risks, and $R_i = C_i \quad \forall i = 1, \cdots, K$ as the set of risks in subgraph $G_{L,i}$, then the reduction of set splitting problem to the problem of finding two SRLG-disjoint paths in the

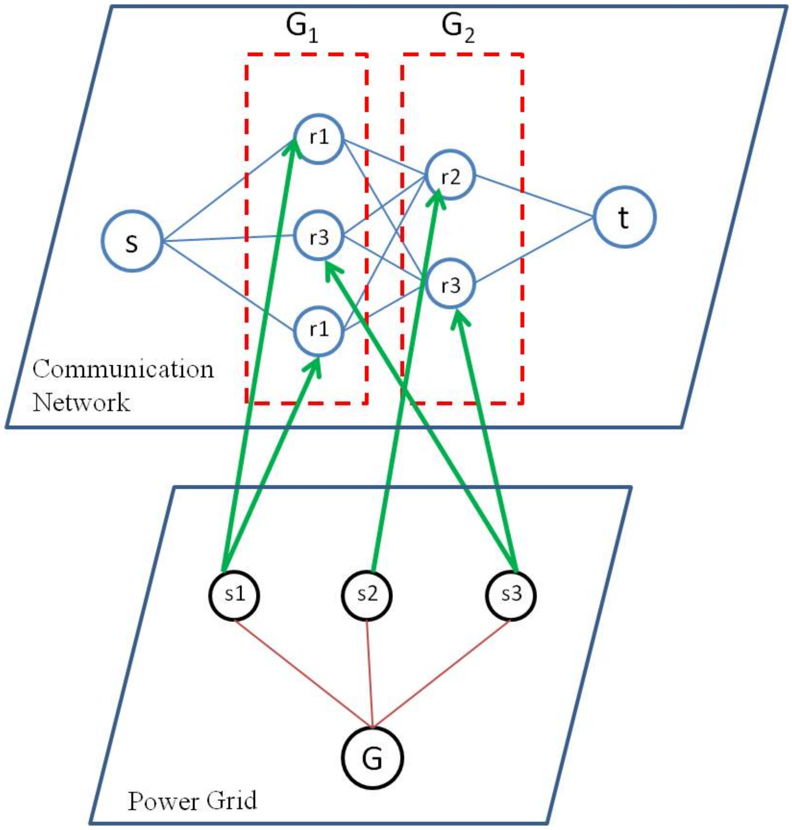


Figure 4-6: Special Structure of Logical and Physical Layers with Mapping

communication-power network is straightforward. Since the set splitting problem is NP-complete [13], finding the SRLG-disjoint paths is also NP-complete.

□

Similar to the optical communication network, SRLG-disjoint paths may not exist in this model; thus, we look at the problem of finding the minimum number of paths in the communication network that are needed to survive any single failure in the power grid.

4.3 MSP in Communication-Power Network

Suppose that the set of SRLG-disjoint paths do not exist. In order to have a reliable connectivity between two nodes s and t, we should find the minimum number of paths in the communication network that are needed to survive any single failure in the power grid, i.e. substation failure.

The formulation of MSP problem is similar to the MSP problem in chapter 2. Suppose, we have m substations in the physical layer, and n paths between nodes s and t in the logical layer. A path P_j fails if at least one of its logical nodes fails due to the failure of a single substation. Define matrix $A \in R^{m \times n}$ such that $a_{ij} = 1$ if substation i will not cause the failure of path j, and $a_{ij} = 0$ otherwise. Under this definition, the MSP problem can be formulated similar to the formulation in chapter 2 as follows:

$$\text{minimize} \quad \sum_{j=1}^{n} P_j \tag{4.6}$$

$$\text{subject to} \quad A \times P \geq e \tag{4.7}$$

$$P_j \in \{0, 1\}, \quad j = 1, \cdots, n \tag{4.8}$$

Given the set of paths, we show that constructing matrix A can be done in polynomial time. If the substations fail independently, then a logical node fails only due to the failure of the substation to which it is directly connected. Therefore, it is straightforward to obtain the matrix A. Now, if the substations can be dependent, the failure of one substation will cascade in the power grid. Moreover, each failed substation will lead to the failure of a set of logical nodes directly connected to it. Therefore, if we can find the failures cascaded in the grid, we can find the failed logical nodes immediately. On the other hand, in our power grid model, a substation operates if and only if it is connected to a generator. Therefore, we can easily find the cascading failures by removing the failed node and checking the connectivity of each substation to the generator. Repeating this procedure for all of the nodes, it will take $O(m^2)$ to obtain the matrix A in general dependent case.

Theorem 4.3.1. *Computing the minimum number of paths between nodes s and t in the communication network that are needed to survive any single substation failure is NP-complete. In addition, this minimum value cannot be approximated within any* *constant factor, unless P = NP.*

Proof. Here, we show that we can construct a special power grid and communication network on top of that in a way that the MSP problem in the communication network can be reduced from the NP-complete problem of Minimum Set Cover.

Minimum Set Cover Problem: Given a set of elements $E = \{e_1, e_2, \cdots, e_n\}$ and a family $F = \{C_1, C_2, \cdots, C_m\}$ of subsets of E, the problem is finding the minimum value k such that there exist k subsets $\{C_{j1}, C_{j2}, \cdots, C_{jk}\} \subset F$ that cover E, i.e., $\cup C_{j_l} \in C = E$ [13].

Similar to the proof of Theorem 2.1.1, given an instance of Minimum Set Cover Problem with ground set E and family of subsets R, we construct a physical topology $E = \{s_1, \cdots, s_m\}$ such that each s_i corresponds to a substation. We also construct a logical topology of parallel paths $\{P_1, \cdots, P_n\}$, where each path corresponds to the subset of substations such that their failure will not affect that path. It follows that the minimum number of logical paths that survives all of the physical fibers is equal to the size of a minimum set cover.

As the last step of proof, we need to show that we can construct such logical and physical topology with the given mapping. Let the physical topology be a star graph where a single generator is in the center, and all of the m substations are directly connected to that generator. Moreover, let the logical topology be the set of parallel paths $p_1, \cdots, p_n$ that are logically disjoint, i.e. do not share any logical node. For each path p_j, we are given the set of substations P_j such that their failure will not affect path p_j. Therefore, we set the size of each path p_j to exactly $m - |P_j|$ logical nodes and connect each logical node to one of the substations not in P_j. This way, each logical path survives the failure of P_j substations. Figure 4-7 shows an example of this mapping.

The inapproximability result follows immediately from the inapproximability of the Minimum Set Cover problem.

□

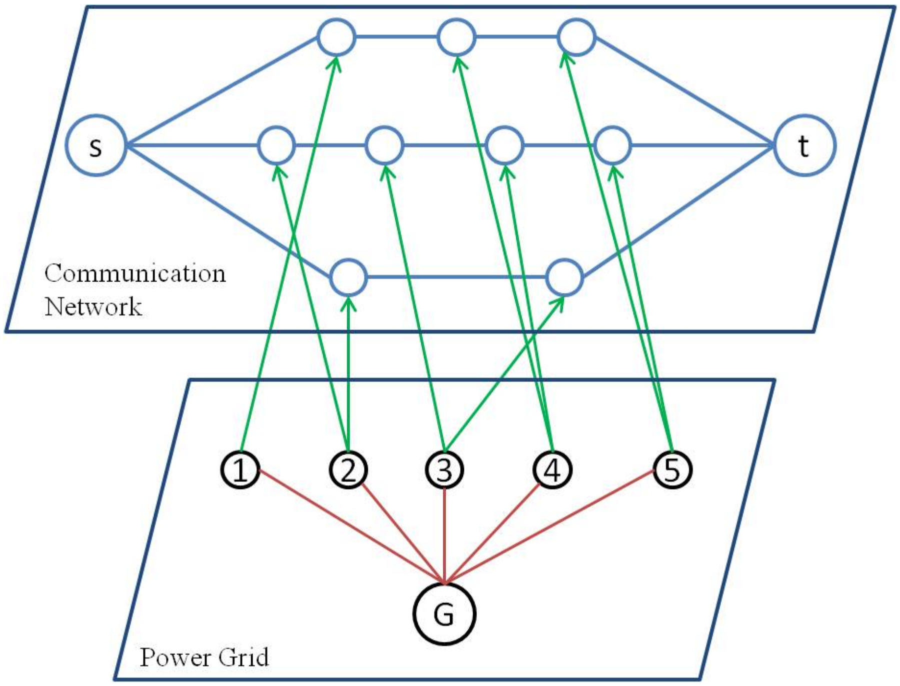


Figure 4-7: Mapping of the communication network on the power grid for $E = \{s_1, \cdots, s_5\}$ and $R = \{\{s_2, s_3\}, \{s_1\}, \{s_1, s_4, s_5\}\}$.

4.4 Maximum Number of Risk-Disjoint Paths

In this section we focus on the problem of finding the maximum number of paths that are risk disjoint, i.e. failure of any substation can affect one and only one of these paths.

Let $P_1, \cdots, P_n$ be the set of all logical paths between nodes s and t. Define matrix $B \in R^{m \times n}$ as the complement of matrix A in the MSP formulation, i.e. $b_{ij} = 1$ if $a_{ij} = 0$, and $b_{ij} = 0$ otherwise. This means that if the failure of substation i results in the failure of path j, then we set $b_{ij} = 1$. Here, the problem would be finding the maximum number of paths such that each substation is used by at most one path. This problem can be formulated as follows:

$$\text{maximize} \quad \sum_{j=1}^{n} P_j \tag{4.9}$$

$$\text{subject to} \quad B \times P \leq e \tag{4.10}$$

$$P_j \in \{0, 1\}, \quad j = 1, \cdots, n \tag{4.11}$$

In the above formulation, the objective function in 4.9 is to maximize the total number of selected paths. Moreover, each row i in constraint 4.10 limits the paths that can use substation i to at most 1. Similar to the MSP problem, this is a path based formulation which requires the knowledge of all logical paths between nodes s and t, which is exponential in the general case.

In the following, we prove that the problem of finding the maximum number of risk-disjoint paths is NP-complete by a reduction from the well-known maximum set packing problem.

Theorem 4.4.1. *Computing the maximum number of risk disjoint paths between nodes s and t is NP-complete. In addition, this maximum value cannot be approximated within any constant factor, unless $P = NP$.*

Proof. The proof is based on the reduction of well-known Maximum Set Packing problem to a special structure of our problem. First, we will construct the special power grid and communication network, and then we will show the reduction.

Set Splitting Problem: Given a set of elements $E = \{e_1, e_2, \cdots, e_n\}$ and a family $F = \{C_1, C_2, \cdots, C_m\}$ of subsets of E, the problem is finding the maximum value k such that subsets $\{C_{j1}, C_{j2}, \cdots, C_{jk}\} \subset F$ are mutually disjoint [13].

Given an instance of Maximum Set Packing Problem with ground set E and family of subsets F, we construct a physical topology $E = \{s_1, \cdots, s_m\}$ which contains m substations. We also construct a logical topology of parallel paths $P_1, \cdots, P_n$, and each path corresponds to the subset of substations such that their failure leads to the failure of that path. Therefore, finding the maximum risk disjoint paths is equal to finding the maximum number of sets P_j's that do not share any risk, i.e. mutually disjoint. This is exactly equivalent to solving the Maximum Set Packing problem. Since this problem is a well-known NP-complete problem, the problem of finding maximum risk disjoint paths is also NP-complete.

The inapproximability result follows immediately from the inapproximability of the Maximum Set Packing problem. □

Chapter 5

Simulation Results and Conclusion

5.1 Simulation Results

We compare the performance of our algorithms using both large-scale random network topologies, as well as the US backbone network topology. In particular, we compare the following algorithms:

- ILP-based optimal algorithm computed by CPLEX; denoted by ILP
- Simple Greedy algorithm from Section 2.2.1; denoted by MSPG
- Additive Cost Greedy algorithm from Section 3.3.1; denoted by ACG
- Non-additive Cost Greedy algorithm from Section 3.3.1; denoted by NACG
- Random-Sweep Greedy algorithm from Section 3.3.3; denoted by RSG
- Randomized rounding algorithm from Section 3.3.2; denoted by RR
- ε-net algorithm from Section 2.2.2; denoted by EPS

5.1.1 Performance in Large-scale Random Topologies with Path Length Restriction

We first consider a random layered network where the logical topology consists of 10 paths between nodes s and t. This layer is mapped onto the physical topology containing 100 fibers, using the mapping structure shown in [12]. In the K restricted version of the problem, each path consists of at most K fibers. For each value of K, we generate 1000 random topologies each with 10 paths routed on the physical topology in a way that each path can select up to K fibers at random, uniformly and independently. We then apply our algorithms to each network in order to find a minimum survivable path set (i.e., to solve the MSP problem). Note that the performances of Randomized Rounding and ε-net algorithms depend on the survivability guarantee of the algorithms, which are 99.9% and 100% respectively for the results shown below.

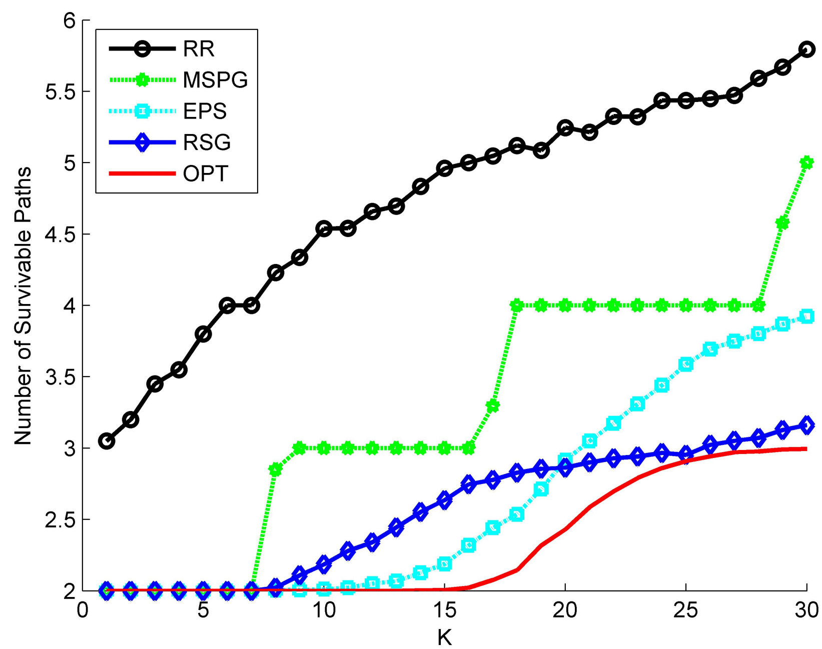


Figure 5-1: Comparison of algorithms for MSP under path length restriction

Figure 5-1 compares the average number of survivable paths found by each algorithm. It can be seen that as the value of K increases, the number of paths increases. This is due to the fact that when K is large, logical paths consist of more fibers; therefore, more logical paths are needed since they can share more fibers. Figure 5-2 compares the logarithm of the running time of the algorithms. It can be seen that the Randomized Rounding algorithm is the fastest, while the RSG algorithm and the ε-net algorithm have larger running times. Note also that the running times are

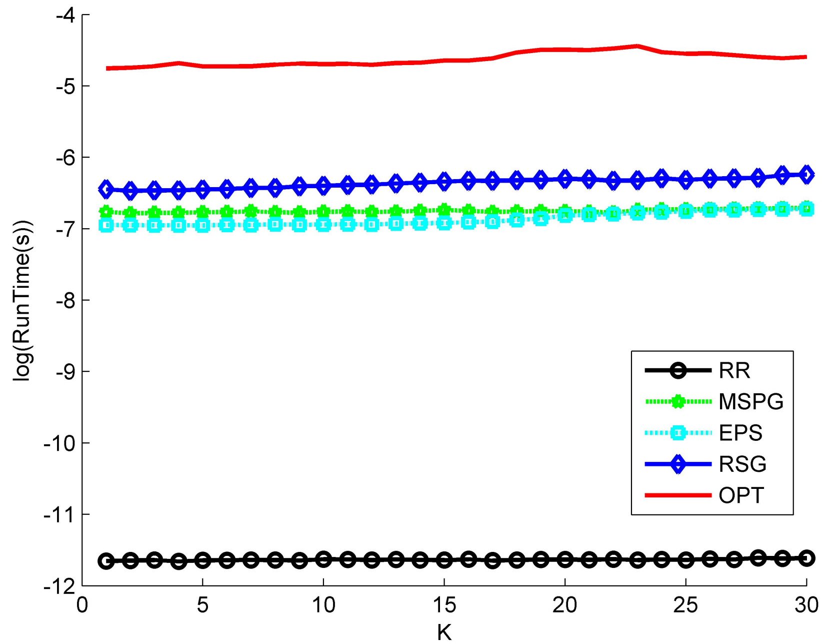


Figure 5-2: Run Time comparison of different heuristics with respect to optimal

nearly independent of K.

5.1.2 Performance in Large-scale Random Topologies with WDM Restriction

Similar to the previous section, we consider a random layered network with 20 paths between nodes s and t in the logical layer. For each W, we generate 1000 random topologies under the the wavelength restriction where at most W paths can be assigned to each fiber. In order to solve the MSP problem, we have applied our algorithms to each network. The survivability guarantees of the Randomized Rounding and ε-net algorithms are 99.9% and 100% respectively for the results shown below.

Figure 5-3 compares the average number of survivable paths found by each algorithm. It can be seen that as the value of W increases, the number of paths increases. This is due to the fact that when W is large, more logical paths can share a fiber, and therefore, more logical paths are needed since a single physical link failure can lead to a large number of logical path failures. Note that the Random-Sweep Greedy (RSG) algorithm is closest to the optimal. Figure 5-4 compares the running time of each

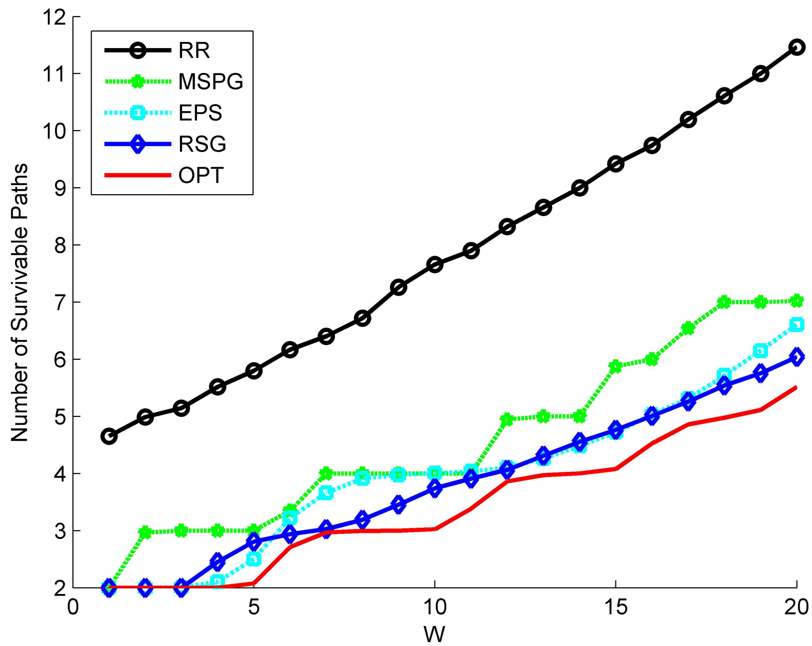


Figure 5-3: Comparison of algorithms for MSP under WDM restriction

algorithm. Similar to the K-restricted version, we observe that randomized rounding is the fastest.

5.1.3 Performance of MFSP in Large-scale Random Topologies with WDM Restriction

We first consider a random layered network where the logical topology consists of 50 paths between nodes s and t. This layer is mapped onto the physical topology containing 100 fibers, using the mapping structure shown in [12]. In the wavelength restricted version of the problem, at most W paths can be assigned to each fiber. For each value of W, we generate 1000 random topologies each with 50 paths that are randomly routed on the physical topology. We then apply our algorithms to each network in order to find a survivable path set using the minimum number of fibers (i.e., to solve the MFSP problem). Note that for Randomized Rounding the performance depends on the survivability guarantee of the algorithm, which is 99.9% for the results shown below.

Figure 5-5 compares the average number of fibers in the survivable path set found

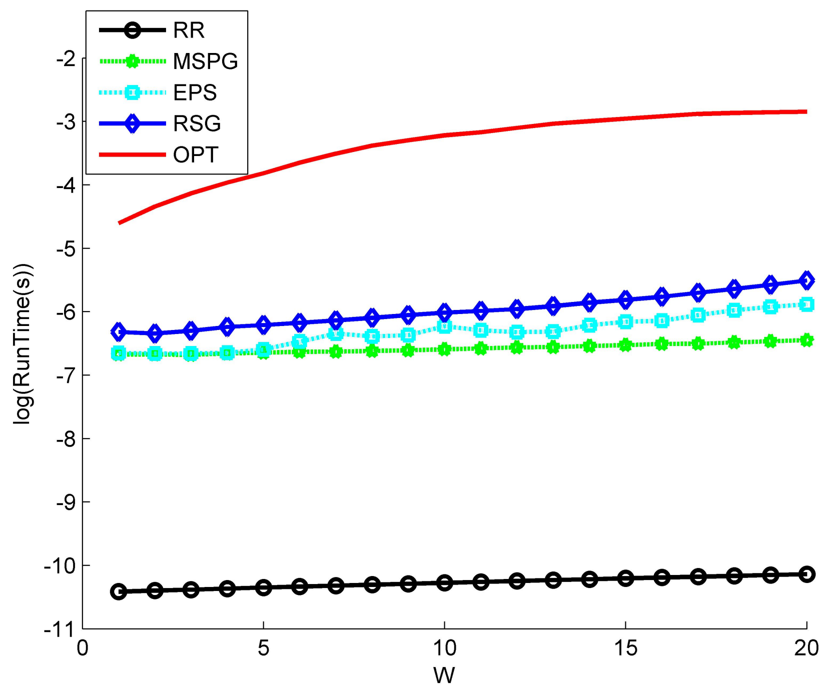


Figure 5-4: Run Time Comparison of Different heuristics with respect to Optimal

by each algorithm. It can be seen that as the value of W increases, the number of used fibers increases. This is due to the fact that when W is large, more logical paths can share a fiber, and therefore more logical paths are needed since a single physical link failure can lead to a large number of logical path failures. Note that the Random-Sweep Greedy (RSG) algorithm is nearly optimal, and the performance of ε-net algorithm is better than RSG for large values of W. Figure 5-6 compares the logarithm of the running time of the algorithms. It can be seen that the Randomized Rounding algorithm is the fastest, while the RSG algorithm which gives the closest to optimal solution, and the ε-net algorithm which performs nearly optimal for networks with large values of W, have larger running times. Note also that the running times are nearly independent of W for all of the proposed algorithms. In contrast, obtaining the exact optimal solution using the ILP formulation becomes quickly impractical as W increases.

Next, we consider larger networks where there are 1000 fibers in the physical topology and 500 paths in the logical topology, with W ranging from 1 to 40. Figure 5-7 shows the performance of the various algorithms as a function of W. The

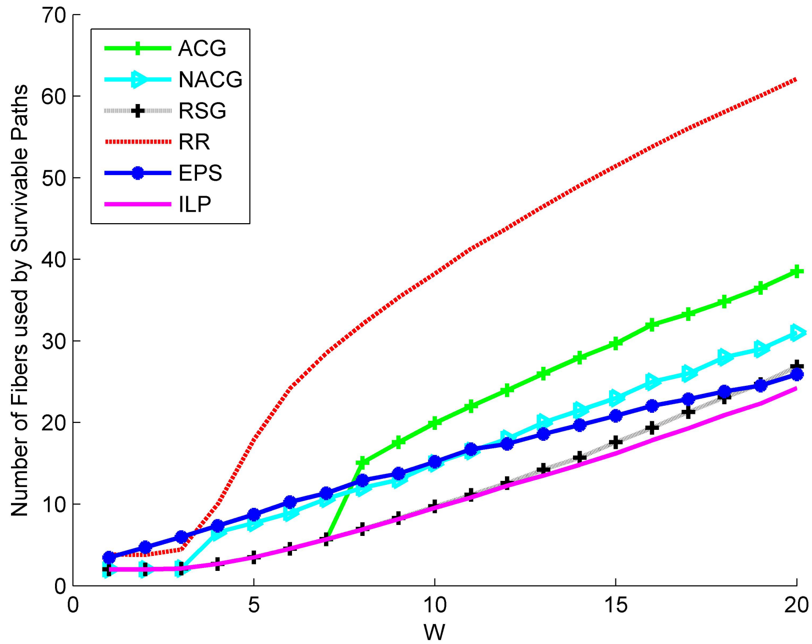


Figure 5-5: Comparison of algorithms for MFSP problem: Approximation quality in random networks

performance of the ILP-based algorithm is omitted since CPLEX often fails to find a solution within a reasonable amount of time. Again we see that the RSG algorithm considerably outperforms the rest of algorithms.

5.1.4 Performance in Real Networks

Next, we examine the performance of the approximation algorithms over the US backbone topology shown in Figure 5-8, with the objective of finding a minimum survivable path set between nodes 4 and 22 [26]. For the logical topology, we generated random graphs with eight nodes (including nodes 4 and 22) each of degree 4. We use shortest path lightpath routing for the logical links.

Table 5.1 shows the average number of paths and average running time of each algorithm. It can be seen that the RSG and randomized rounding algorithms are nearly optimal; furthermore, the randomized rounding gives a solution almost instantly. We also note that the survivability guarantees of the Randomized Rounding and ε-net algorithms are 99.9% and 100% respectively for the results shown in the table.

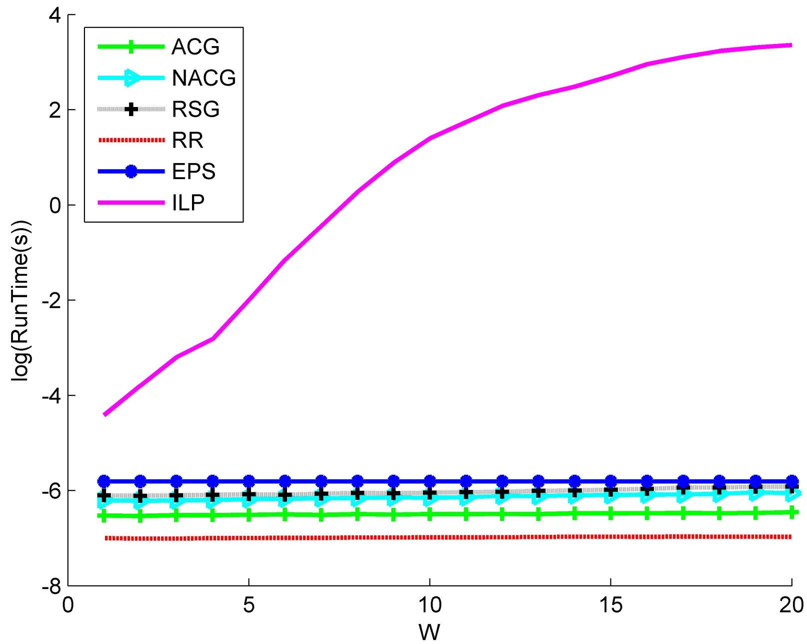


Figure 5-6: Run Time Comparison of Different heuristics with respect to Optimal

Method	Number of Paths	Running Time (ms)
ILP	2.0069	7.2133
RSG	2.0160	2.0167
RR	2.0482	0.0272
MSPG	2.2241	0.1911
EPS	2.551	1.6000

Table 5.1: Comparison of Algorithms for MSP in Real Networks

5.2 Summary

We considered the problem of finding survivable paths in layered networks. The traditional disjoint paths approach for protection cannot be directly applied to layered networks, since physically disjoint paths may not always exist in such networks. To address this issue, we introduced the new notion of *survivable path set*. We showed that in general the problem of finding the minimum size survivable path set (MSP) and the problem of finding the minimum fiber survivable path set (MFSP) are NP-hard and inapproximable. However, under practical constraints, we are able to develop both optimal and approximation algorithms for the MSP and MFSP problems. We also modeled the dependency of data communication network on the power grid as a

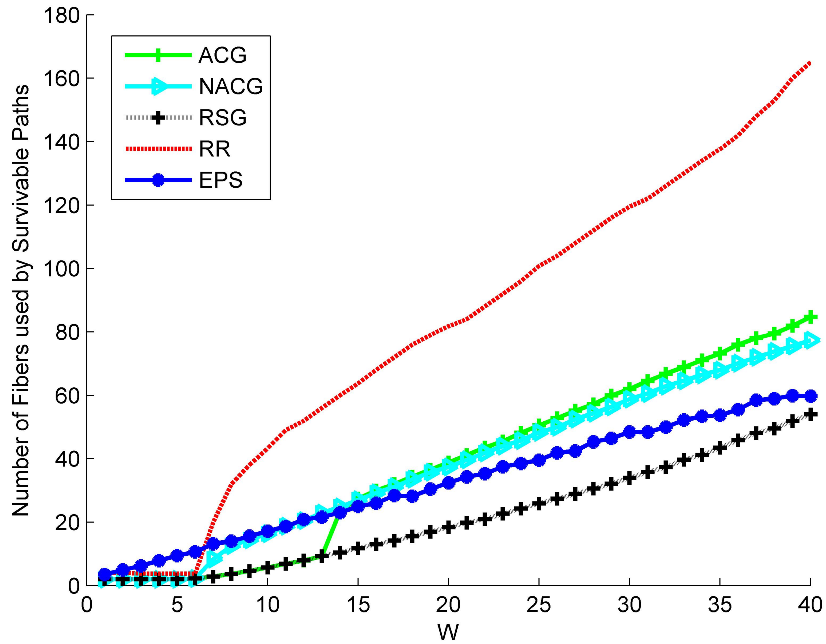


Figure 5-7: Comparison of Approximation Algorithms in Large Networks

layered network, and we showed how failures in the power grid lead to failures in the communication network. We proved that the problems of finding two SRLG-disjoint paths, minimum survivable path set and the maximum number of risk disjoint paths between a source and destination in the communication network are NP-complete.

5.3 Future Work

As explained in Chapter 1, modern networks are coupled and should be modeled as multilayered networks capturing the interdependency between the networks. OWDM-based network is the only network technology that has been well studied as a multi-layer network. Power grid and its influence on other networked infrastructures is a very important subject which has gained interest in recent years. Many researchers are working on the issue of power grid reliability. On the other hand, as explained the communication network strongly depends on the power grid, such that a single failure in power grid can lead to multiple failures in the communication network. In the future, we hope to study the layered network architecture in more details; with focus on the aspect of network design, i.e. how to design a communication network

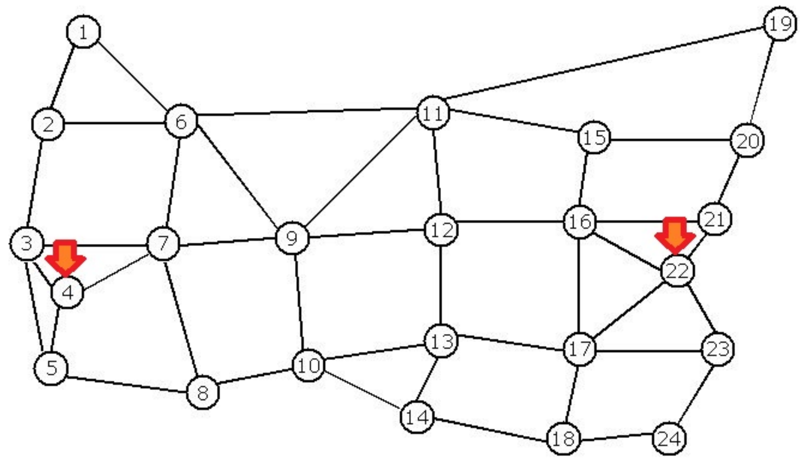


Figure 5-8: Physical Topology

which is reliable to the failures in power grid.

Recently, Rosato *et. al.* investigated the 2003 blackout in Italy [27]. They showed that such a large scale blackout was due to the interdependency between the communication network and the power grid, in a way that failures in the power grid resulted in failures in the communication network, and failures in the communication network resulted in failures in the power grid. Based on this observation, one can model the behavior of the two networks as a layered network with two way dependency, such that failure in one network leads to failure in the other network. We plan to define new metrics for survivability in such networks, and develop algorithms for robust network design on this setting.

Appendix A

Proofs

A.1 Proof of Theorem 2.2.3

In the procedure of ε-net algorithm, the “path-selection” algorithm will be applied iteratively, and checks the survivability of the selected path set after each iteration. If not all fibers are survived, the algorithm doubles the weight of all paths that survive the failure of fibers in $\bar{S}$, where $\bar{S}$ is the set all the fibers that are not survived yet, and repeat the random path selection.

Let ξ be the optimal solution of MSP. Based on the results in [19, 20], it can be shown that if in each iteration the selected subset of paths survive a “good” subset of fibers, in $O(\xi \log(\frac{m}{\xi}))$ iterations, the algorithm will return a set of survivable paths, with high probability. A subset is “good” if it is an ε-net.

Definition A.1.1. Consider a set system $F = (X, R)$, where X is the set of elements and R is the set of subsets of X. A set $H \subset X$ is an “*ε-net* ” of F if $S \cap H \neq \emptyset$, for every subset $S \in R$ for which $|S| \geq \varepsilon|X|$.

Lemma A.1.1 claims that it is guaranteed that in each iteration the selected paths survive a “good” subset of fibers

Lemma A.1.1. *For all $\varepsilon \in (0, \frac{1}{2})$, if $s = c\frac{\log K}{\varepsilon} \log \frac{\log K}{\varepsilon}$, where c is a constant, the path-selection algorithm selects a subset of paths that survives all of the ε-Survivable fibers with high probability.*

For the proof of Lemma A.1.1, we use the new techniques found by Haussler and Welzl in [28]. Theorem A.1.1 is an improvement on their work [18, 19].

Before presenting Theorem A.1.1, we need to define VC-dimension.

Definition A.1.2. Let R be a set system on a set X. Let us say that a subset $A \subset X$ is shattered by R if each of the subsets of A can be obtained as the intersection of some $S \in R$ with A, i.e. if $R|_A = 2^A$.

Define the VC-dimension of R, denoted by $dim(R)$, as the supremum of the sizes of all finite shattered subsets of X. If arbitrarily large subsets can be shattered, the VC-dimension is ∞.

Theorem A.1.1. *Let $F = (X, R)$ denote a set system with weights $w(u)$. For every $\varepsilon \in (0, \frac{1}{2})$, a random sample of X according to the probability distribution $w(u) = w(X)$ is likely to to be an ε-net with respect to $w(u)$, if the sample contains $O(\frac{d}{\varepsilon}\log(\frac{d}{\varepsilon}))$ elements, where d is the VC-dimension of the set system.*

To prove Lemma A.1.1, it is enough to show that $O(\frac{\log K}{\varepsilon}\log(\frac{\log K}{\varepsilon}))$ paths are needed to cover all ε-*Survivable* fibers.

Let $X = \{P_1, \cdots, P_n\}$ be the set of paths in our problem, and $R = \{f_1, \cdots, f_m\}$ be the set of subsets of X, where each fiber-set f_i corresponds to the set of paths that survives the failure of fiber i. Therefore, an ε-*net* will cover all ε-*Survivable* fibers.

In this setting, a subset A of paths is shatterd if intersection of A with every fiber-set produces all the subsets of A. For instance, $A = \{P_1, P_2, P_3\}$ is shattered by R if there exist 8 fibers $f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8$ such that $\{A \cap f_1 = P_1, A \cap f_2 = P_2, A \cap f_3 = P_3, A \cap f_4 = \{P_1, P_2\}, A \cap f_5 = \{P_1, P_3\}, A \cap f_6 = \{P_2, P_3\}, A \cap f_7 = A, A \cap f_8 = \emptyset\}$.

Lemma A.1.2. *In path length restricted version of MSP, VC-dimension d is less than* $\log K$.

Proof. Suppose VC-dimension is d. Then, by definition, there exist a subset of paths A of size d which intersection of A with all fiber-sets generates all subsets of A. In particular, for every $P_j \in A$, half of the subsets created by $A \cap R$ should contain P_j and the other half should not contain it.

Under the path length restricted assumption, each path uses at most K fibers, and survives at least $m-K$ remaining fibers. Therefore, at least $m-K$ fibers contain a particular path j. Thus,

$$2^{d-1} \leq m-K \text{and} d \leq 1+\log(m-K). \tag{A.1}$$

On the other hand, for each P_j at most K fibers do not contain it. Hence,

$$2^{d-1} \leq K \text{and} d \leq 1+\log K. \tag{A.2}$$

By combining both equations A.1 and A.2, we will have the following result:

$$d \leq 1+\log K. \tag{A.3}$$

□

A.2 Epsilon-Net in WDM

Using the same techniques discussed in the Section A.1, we have the following Theorem:

Theorem A.2.1. *The ε-net algorithm finds a set of survivable paths of size $O(\log W \log \xi)\xi$, with high probability and terminates in $O(\xi \log(\frac{m}{\xi}))$ iterations.*

To prove Theorem A.2.1, we need Lemma A.2.1. Then, by an argument similar to the one to prove Thm 2.2.3, the ε-net algorithm will find a set of survivable paths with a $\log W \log \xi$ approximation bound.

Lemma A.2.1. $\forall\varepsilon \in (0,1)$, *if $s = c\frac{\log W}{\varepsilon} \log \frac{\log W}{\varepsilon}$ where c is a constant, the path-selection algorithm selects a subset of paths that survives all of the "-Survivable fibers with high probability.*

Proof. Similar to the argument for the proof of Theorem A.1.1, it is enough to prove Lemma A.2.2. □

Lemma A.2.2. *In wavelength restricted version of MSP, VC-dimension d is less than W.*

Proof. Let $X = \{P_1, ..., P_n\}$ be the set of all paths, and $R = \{f_1, ..., f_m\}$ be the set of all fiber sets where a fiber is associated i.e., $P_j \in f_i$ if and only if path j survives fiber i's failure.

Let VC-dimension be d. Then, by the definition of VC-dimension, there exist a subset A of paths such that $|A| = d$ and intersection of A with all fiber sets generates all the subsets of A. In particular, there exist a fiber set i such that $f_i \cap A = \emptyset$, which means $A \subset X - f_i$. Thus,

$$d = |A| \leq |X - f_i|. \tag{A.4}$$

On the other hand, under the wavelength restricted assumption, each fiber can be used by at most W paths. Therefore,

$$n - W \leq |f_i|, \quad \forall f_i. \tag{A.5}$$

Combining inequalities (A.4) and (A.5) results in $d \leq W$.

□

A.3 Epsilon-Net in MFSP

Combining the results of Randomized algorithm described in subsection 2.2 and Theorem 3.3.2 results in the following corollary.

Corollary 1. *Using the ε-net algirithm, one can find an $O(W \log W \log \xi)$ approximation.*

Proof. Similar to the proof of Theorem 3.3.1. □

Bibliography

[1] *Severe Space Weather Events - Understanding societal and economic impacts: A Workshop report*. The National Academies Press, Washington, D.C., 2008.

[2] Biswanath Mukherjee. *Optical WDM Networks*. Springer, NY, 2006.

[3] L.M. Gambardella F. Ducatelle. Survivable routing in ip-over-wdm networks: An efficient and scalable local search algorithm. *Optical Switching and Networking, Elsevier*, 2(2):86–99, Sep. 2005.

[4] Biswanath Mukherjee. *Optical Communication Networks*. McGraw-Hill, NY, 1997.

[5] R. Bhandari. Optimal diverse routing in telecommunication fiber networks. In *IEEE INFOCOM*, June 1994.

[6] I.P. Kaminow and T.L. Koch. *Optical Fiber Telecommunications*, volume IIIA.

[7] J. Q. Hu. Diverse routing in optical mesh networks. *IEEE Transactions on Communications*, 51(3):489–494, March 2003.

[8] D. Xu et al. Trap avoidance and protection schemes in networks with shared risk link groups. *Lightwave Technology*, 21(11):2683–2693, Nov. 2003.

[9] D. Xu, Y. Xiong, and C. Qiao. A new promise algorithm in networks with shared risk link groups. In *IEEE Globecom*, Dec. 2003.

[10] P. Datta and A. K. Somani. Diverse routing for shared risk resource group (srrg) failures in wdm optical networks. In *First Annual International Conference on Broadband Networking*, pages 120–129, Oct. 2004.

[11] E. Modiano and A. Narula-Tam. Survivable lightpath routing: a new approach to the design of wdm-based networks. *IEEE Journal on Selected Areas in Communications*, 2002.

[12] K. Lee, E. Modiano, and H. Lee. Cross layer survivability in wdm-based networks. *IEEE/ACM Transactions on Networking*, 2011.

[13] M. Garey and D. Johnson. *Computers and Intractibility, a Guide to the Theory of NP-Completeness.* W.H. Freeman and Co., 1979.

[14] S. Yuan, S. Varma, and J. P. Jue. Minimum color path problems for reliability in mesh networks. In *IEEE INFOCOM*, 2005.

[15] P. Raghavan and C. D. Thompson. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. In *Combinatorica*, volume 7, pages 365–374, 1987.

[16] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *ACM*, 41(5):960–1071, Sep. 1994.

[17] K. Varadarajan. Epsilon nets and union complexity. In *25th annual symposium on Computational geometry*, Jun. 2009.

[18] J. Matousek. *Lectures on Discrete Geometry.* Springer-Verlag, NJ, 2002.

[19] G. Even, D. Rawitz, and S. Shahar. Hitting set when the vc-dimension is small. *Information Processing Letters*, 95(2):358–362, Jul. 2005.

[20] H. Bronnimann and M. Goodrich. Almost optimal set covers in finite vc-dimension. *Discrete Computational Geometry*, 14(1):463–479, Oct. 1995.

[21] V. Chvatal. A greedy heuristic for the set covering problem. *Mathmatics of Operations Research*, 4(3):233–235, Aug. 1979.

[22] `http://infrascapedesign.wordpress.com/2011/10/31/infrastructural-interdependency/`.

[23] D. Bienstock and A. Verma. The n-k problem in power grids: new models, formulations and computation. *SIAM Journal on Optimization*, 20:2352 2380, 2010.

[24] Technical report.

[25] Technical report.

[26] `http://www.optical-network.com/topology.php`.

[27] F. Tiriticco S. Meloni S. De Porcellinis R. Setola V. Rosato, L. Issacharoff. Modelling interdependent infrastructures using interacting dynamical models. *International Journal of Critical Infrastructures*, 4(1):63–79, 2008.

[28] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. In *Second Annual Symposium on Computational Geometry*, 1986.