UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DOCTORADO EN CIENCIAS DE LA INGENIERÍA CON MENCIÓN EN INGENIERÍA ELÉCTRICA

Maximizing reliability of data networks to face multiple failures via optimal route selection

---

Profesor supervisor: PhD. Jorge Pezoa Núñez
Departamento de Ingeniería Eléctrica
Facultad de Ingeniería
Universidad de Concepción

Tesis para optar al grado académico de Doctor en Ciencias de la Ingeniería con mención en Ingeniería Eléctrica

NICOLÁS BOETTCHER PALMA

CONCEPCIÓN - CHILE

2022

Prefacio

Esta tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias de la Ingeniería con mención en Ingeniería Eléctrica, de la Universidad de Concepción, Chile, y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Departamento de Ingeniería Eléctrica, durante el período comprendido entre el año 2015 y 2022, bajo la dirección del Doctor Jorge Pezoa Núñez.

Nicolás Boettcher Palma
nboettcher@udec.cl
Departamento de Ingeniería Eléctrica
Facultad de Ingeniería
Universidad de Concepción
Concepción, 2022

En la vida de un doctorante hay desafíos:
Nacimiento de un hijo,
Congelar por licencia médica,
Crisis social,
Pandemia mundial,
Trabajar y estudiar al mismo tiempo,
4 mudanzas,
Remodelación completa de casa en cuarentena,
pero hay que luchar hasta el final y ser ***resilientes****...*

Agradecimientos

El proceso de realizar esta tesis involucró a muchas personas y entidades entre los que se encuentran:

Universidad Diego Portales, que me apoyó financieramente y me dio la oportunidad y tiempo para cursar el doctorado.

Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), que me apoyó financieramente con la Beca Doctorado Nacional Folio: 2015-21150775.

Luciano Ahumada, por haberme ayudado a encontrar una casa de estudios que se alineara con mis temas de interés.

Jorge Pezoa, profesor supervisor de esta tesis, por haberme dado la oportunidad de trabajar en el área que me interesa y por su paciencia.

Yasmany Prieto, por ser un gran amigo y compañero con el que espero seguir compartiendo su amistad y vida académica.

Marcela Hernández e Inés Lillo por haberme ayudado a resolver trámites y mis dudas de estudiante y siempre haber estado dispuestas a ayudar.

Cora Diaz y a mis hijos, por haberme apoyado en todo el periodo de estudiante, por haberme comprendido y soportado por estar ausente mientras recueperaba el tiempo perdido en los estudios y por haberle pedido al viejito pascuero que termine el doctorado.

Contents

List of Figures

List of tables

Abstract

Natural disasters, depending on both their size and magnitude, can produce large-scale failures to the telecommunications network infrastructure. These failures can lead to service interruptions due to disconnection of network nodes. To avoid disconnection in the network, the proactive mechanism of having a route pair between the nodes has the fastest response times. The region disjoint geo-diverse routing algorithms are the most widely used to deal with large-scale geographic disasters, but restricts a minimum separation between routes and assumes failures can occur anywhere. In this work, we consider large-scale disasters are composed by multiple independent failures. To achieve a reliable network, we propose the following methodology. First, we consider a threat model based on a probability distribution of geo-located failures obtained from historical records. Second, we develop a risk model where the probability of link failure is proportional to the intersection area between the probability distribution and the vulnerable zone of the link. Finally, we compute a route pair between all couple of nodes in the network using routing algorithms that minimize the failure probability of the route pair. We apply our methodology in three real-world networks with fires detected by NASA satellites. We observe our routing algorithms select different route pair according to the hazard season. On average, algorithms proposed managed to establish routes in more than 99.83% of the cases, outperforming all the algorithms analyzed in the state of the art. Moreover, route pair solutions proposed, were able to obtain an average ATTR higher than 99.9% in the analyzed scenarios, despite the existence of multiple threat regions, taking advantage of the knowledge where failures have historically occurred.

Resumen

Las catástrofes naturales, dependiendo de su tamaño y magnitud, pueden producir fallos a gran escala en la infraestructura de la red de telecomunicaciones. Estos fallos pueden provocar interrupciones del servicio debido a la desconexión de los nodos de la red. Para evitar la desconexión de la red, el mecanismo proactivo de tener un par de rutas entre los nodos tiene los tiempos de respuesta más rápidos. Los algoritmos de enrutamiento geo-diverso por regiones son los más utilizados para hacer frente a los desastres geográficos a gran escala, pero restringen una separación mínima entre las rutas y suponen que los fallos pueden ocurrir en cualquier lugar. En este trabajo, consideramos que las catástrofes a gran escala están compuestas por múltiples fallos independientes. Para conseguir una red confiable, proponemos la siguiente metodología. En primer lugar, consideramos un modelo de amenaza basado en una distribución de probabilidad de fallos geolocalizados obtenida a partir de registros históricos. En segundo lugar, desarrollamos un modelo de riesgo en el que la probabilidad de fallo del enlace es proporcional al área de intersección entre la distribución de probabilidad y la zona vulnerable del enlace. Por último, calculamos un par de rutas entre todos los pares de nodos de la red utilizando algoritmos de enrutamiento que minimizan la probabilidad de fallo del par de rutas. Aplicamos nuestra metodología en tres redes del mundo real con incendios detectados por satélites de la NASA. Observamos que nuestros algoritmos de enrutamiento seleccionan diferentes pares de rutas en función de la época de peligro. En promedio, los algoritmos propuestos lograron establecer rutas en más del 99,83% de los casos, superando a todos los algoritmos analizados en el estado del arte. Además, las soluciones de pares de rutas propuestas, fueron capaces de obtener una media de ATTR superior al 99,9% en los escenarios analizados, a pesar de la existencia de múltiples regiones de amenaza, aprovechando el conocimiento de dónde se han producido históricamente los fallos.

1 Introduction

Telecommunication networks are a fundamental backbone for computer systems. Such networks are supported by an infrastructure deployed over large geographical areas. This infrastructure is vulnerable to natural disasters such as earthquakes, hurricanes, tsunamis, wildfire, etc., or to man-made disasters such as terrorist attacks, weapons of mass destruction, and even accidents; eventually affecting optical fibers which are responsible for providing connectivity between different points through established routes where data is transmitted [1]. Network connectivity, despite these infrastructure damages, can be maintained based on the provisioning mechanisms that the network operator can take into account when selecting routes to connect different locations [1]. For this purpose, several routing algorithms have been developed for proactive risk management. One mechanism is to compute disjoint alternative routes to maintain connectivity in case of failures [2, 3]. Thus, if one of the two routes is affected by a failure, the other is not. This mechanism allows IT managers to react faster to disasters with a reliable network.

To deal with large-scale failures in geographic network, region disjoint geo-diverse routing algorithms focus on the selection of both, short and geographically distant disjoint paths. Thus, the distance traveled by the routes and the distance between the nodes that compose them are the key to determine their efficiency [4]. Unlike routing algorithms typical of ISP telecommunications networks, which are based on network properties and network metrics such as delay, bandwidth, etc.; geo-diverse routing algorithms use information corresponding to another domain, the geographic information in which the network topology is located. To represent the coverage of failures, geometric representations of different sizes are used, where the disk is the most observed [3, 5]. Despite the existence of defined radius values for each type of failure [6], the classification of 'large' should be based on the distribution of failure sizes and where they occur, rather than a fixed value [7].

Even so, we have not identified techniques in the literature about how to solve the failure size selection problem that could affect a network topology. Therefore, better techniques are needed to help the network administrator to model threats to face real failures and be able to use them as input into routing algorithms.

Among the notified disaster areas we could find simulations have used radius of 50 km up to 1000 km [3, 8, 9, 10, 11, 12, 13, 14]. In these works, synthetic random threats models are used to generate the network failures, without considering historical empirical data about geo-located failures [15, 16, 17, 18, 19].

From the accuracy of the failure detection techniques, confidence about the area affected by the failure event can be determined. At higher resolution, it is possible to detect small areas affected by failures. Conversely, the lower the resolution, the larger the failure area detected, making invisible threat-free zones within it, such as non-seismic zones for earthquakes or hydrographic zones for fires. The above, coupled with risk models that do not consider multiple independent failures and require distance to the center of failure, contributes to the error in the failure probabilities of the network infrastructure.

1.1 Motivation

There are a large number of different hazards that can cause a network to lose connectivity [20, 21]. A taxonomy of the different types of hazards, including large-scale natural and man-made disasters, is established in [22]. These hazards are classified and categorized by characteristics such as the spatial region covered and the duration, which can last from seconds to hours, such as earthquakes or fires, respectively. Additionally, in [22], it is suggested the scope of large-scale catastrophes are both, non-local and non-repetitive catastrophic events. Nevertheless, there is evidence from large-scale natural catastrophes repeated in the same geographical areas. For example, hurricanes, formed mainly by a tropical phenomenon, in the United States have caused massive failures in power supplies and cut cables due to the force of the wind [23, 24, 25, 26, 27]. In addition, countries near the Pacific belt are the most affected by earthquakes. In Taiwan [28, 29], severe damage to multiple submarine cables has been reported causing service disruption in parts of Asia for weeks. Chile, despite being a highly seismic country and having recorded the largest earthquake in global history [30], is also susceptible to other threats. In its central zone, the most prone to fires, there have been fiber optic cuts in telecommunications networks due to multiple fires [31, 32].

Catastrophes caused by multiple independent sources, such as wildfires, have resulted in telephone, Internet and cable outages at several companies. Telecommunications outages have worsened in recent years as disasters have become more frequent and destructive. There are geographic areas with a higher probability of

dangerous fire season, such as California from United Stated, and it is necessary to prepare in advance to be able to fight the fire. The California Public Utilities Commission report [33] revealed that 85 000 wireless customers and 160 000 cable customers lost service during the 2017 fires in the Northern California Bay Area. This region is one of the most affected in United States by fires every year. Based on empirical data collected, the existence of Fire Return Periods has been determined [34]. As is well known, there are wildfires seasons, which are predicted every year in order to fight them effectively.

Due to the major problems caused by natural disasters in telecommunication networks, it is essential to be able to react in time to avoid loss of connectivity, without having to rely on infrastructure modification, which is not always economically viable. Determining where failures will occur and avoiding outages on active routes to maintain network reliability is our main motivation for this work.

1.2 State-of-the-art

This Chapter contains the state-of-the-art in threat and risk models, routing algorithms to face geographic failures and network performance metrics. A series of works that support the main ideas developed in the thesis are presented. Finally, a literature review matrix focusing on the main routing algorithms and performance metrics is presented.

1.2.1 Threat models

In the literature we can find that most of the works considering failures in the network, use natural failures generated by synthetic random failure models [3, 8, 9, 10, 11, 12, 13, 14]. Nevertheless, real failures may occur with different probabilities distribution according to its geography. Thus, the use of historical failure data can be used to predict the location of future failures. Within the available open datasets associated with natural catastrophes, there is a great diversity of records with a predominance of earthquakes, hurricanes and fires [15, 16, 17, 18, 19]. Due to not all datasets use the same coverage, methods, tools and detection techniques, interoperability between them is difficult.

A variety of failures detection techniques are presented in [35]. These include human observation, satellite systems, digital cameras, and wireless sensor networks. Although data collected by sensor networks are more accurate and real-time, satellite data have global coverage. For example, NASA provides global data collected since 2001 related to geo-located fire hotspots identified by satellite detection through Moderate Resolution Imaging Spectroradiometer (MODIS) sensors. MODIS can detect small-scale fires because they have a recognizable

thermal signature. Each time the satellite detects a fire source, it marks the location of the signal in the data set [36]. Each fire focus obtained from MODIS has a pixel resolution of 1000 m^2, and can even reach a resolution of 50 m^2 under ideal conditions. But considering that about 70 percent of the planet's surface is covered by clouds at any given time, these ideal circumstances are difficult to achieve.

On the other hand, the large-scale failures are mostly modeled using a circular area, represented by a disk shape, known as region failure [3, 5]. Authors usually designate simple geometric figures to represent the total coverage of large scale hazards, such as circles, lines, rectangles, among others [37, 38]. Other uses irregular regions to represent floods, tornadoes and stars to represent volcano eruptions [39, 40]. Most of these failure modeling approaches try to find the right balance between accuracy and the number of Shared Risk Link Groups (SRLG) [40]. The larger the size of a failure, the lower the accuracy of the SRLG when considering the areas actually affected by a failure.

The impact of real failures on telecommunication networks has become increasingly important, but few studies have considered modeling their historical records. Savas et al. [41] joins real earthquake, tornado, and hurricane hazard heat map for the territory of the United States, where it is observed that each hazard contributes with different risk zones. From these, it is possible to estimate the probability of a catastrophe occurring and the level of risk that a network device will be damaged by such a catastrophe. From the risk level, they filter out the areas of greatest hazard and assign to each one an equally sized disk, representing a threat. In [42], from the empirical data of Japan Seismic Hazards Information Station (J-SHIS) and International Best Track Archive for Climate Stewardship (IBTrACS), the failures representing the greatest damage are chosen, removing from the record areas where failures have occurred with less magnitude. Then, data is merged and from Monte Carlo simulations, randomly generate a fixed number of disaster events. In the same way, in [3], the Monte Carlo simulation is generated from empirical earthquake data. An uniform distribution of disasters over the deployment region is considered. This allows for some unevenly distributed regional failures, for example, tsunamis, which affect only those components of the network that are close to the sea.

To the best of our knowledge, there is no record of other work using fire to represent large-scale failures as a failure factor in telecommunication networks. In fact, we have also found no work that reflects the risk change in telecommunication networks according to hazard seasons.

1.2.2 Risk models

Different risk models are presented in [9, 43], deterministic models, Gaussian, probabilistic functions, and compound component, where the greater the distance between the link and the failure epicenter, the lower the probability of damage to the link. The composite component model, unlike the rest of the risk models, establishes a vulnerable zone around the link, represented by the hippodrome area containing the link. The area of the vulnerable zone is restricted to the length of the link and the radius of separation between the link and the hippodrome [3]. The value of the radius is a compromise between the cost and the robustness of the link to a given type of threat. In [44], the probability of link failure contained in the hippodrome is calculated as the area of the vulnerability zone divided into the area of the plane in which the network topology is located. In this way, it is considered that hazards can occur anywhere in the plane. Although the composite component model is widely used, we did not identify techniques in the literature on how to choose a value for this parameter.

A. Pašić et al. [45] proposes that high reliability and very high availability are determined by the underlying network infrastructure, the appropriate failure modeling and by the routing schemes used (i.e. the protection mechanism). In the proposal, the key to the failure modeling approach is the distance to the hazard epicenter, as in [46], where damage from large-scale hazards is considered to be associated with the distance between the link and the failure epicenter. B. Vass et al. [47] propose through historical data to determine the intensity of earthquakes from the distance of their epicenters. However, failures that span large territories originating from multiple, independent hazards, such as landslides caused by rainfalls [48] or fires [49, 50] caused by several outbreaks at the same time, do not follow a behavior consistent with the location where the failure originated. Thus, how the threat is modeled is critical to generate an accurate risk model.

Based on the analyzed works, it is obtained that a conditioning factor is to determine the distance to the epicenter of the large-scale hazard. This assumes that the fault originates at the center of mass of the figure and expands from that point. Frankel, A. et al. [51] manage to simulate large earthquakes from records of small earthquakes adjacent to the area of interest. This can also be achieved thanks to the EMSC [52], which collects real-time parametric data provided by 65 seismological networks in the Euro-Mediterranean region. These data are more accurate than large-scale failures, as they consider a smaller coverage. Based on this fact, we believe that multiple small-scale areas where damage has been evidenced following a deterministic model are more effective to represent large-scale failures compared to a single failure.

To consider that the risk of failure is equal for all points that are at the same

distance from the failure epicenter, we believe that geographic homogeneity among all such points must be fulfilled. As mentioned in [9], the waves in an earthquake depend on geographical characteristics and earth materials. When considering large-scale failures, it becomes more difficult to maintain geographical homogeneity, especially if we know that the geography is irregular for large extensions of land. In our work we do not follow these approaches, we consider that large-scale threats can be represented by multiple independent small-scale failure events that allow to highlight threat-free zones. Each point belonging to the area covered by small-scale events has the same risk, without differentiating the risk by proximity to an epicenter. This approach offers better advantages when modeling large-scale hazard risk, as the higher resolution allows highlighting threat-free zones among the threatened areas.

The dimensions of hazards, in addition to varying in size according to the geographical area in which they occur, can also be determined by climatic conditions. Carlson, A. R. et. al [53] shown, on the basis of empirical data, that fire ignition zones differ spatially and according to seasonal variations. In addition, the number of small fires is much larger than the number of large fires defined by a threshold of 400 ha. Depending on the diversity of biophysical features or ecoregions [54] that make up the geography, the spread of these disasters is not homogeneous [55]. In fact, in [53], it is obtained that the environmental factors driving the occurrence of a failure vary considerably between different regions. In the U.S., 18 distinct ecoregions have been categorized [56], which supports the fact that there is not a level playing field for hazards across geography. This indicates that considering large-scale failures could be misleading based only on the distance to the epicenter of the disaster. For this reason, we only consider records of small-scale failures, whose damage is homogeneous throughout their area of coverage.

1.2.3 Routing algorithms

To face natural catastrophes and maintain the communication routes, region disjoint geo-diverse routing algorithms have been developed. The origin of these algorithms seek a solution to the path geo-diverse problem based on the geo-diverse routing protocol (GeoDivRP) [4]. Considering that disasters may differ according their remoteness, Pašić, A. et. al. [40] suggested geo-diverse routing can be used to increase the network disaster survivability as long as disjoint paths are kept spatially separated according to failure regions. Therefore, accurate failure models that better understand the study area are needed. These kind of algorithms use the geographic diversity in which the network topology is located to choose paths separated by a minimum distance between them. This distance is calculated between the nodes that compose each path, also known as region

disjoint routes. The aim of these algorithms is to prevent different routes from sharing failure regions, determined by a diameter distance. Therefore, avoiding Shared Risk Groups (SRGs) or probabilistic SRGs (p-SRG) is essential to prevent network components from failing simultaneously because they are geographically close [57]. Izaddoost, A. et al. [9] propose to calculate the paths from the failure probabilities associated with each link as a cost to be minimized, as well as [3]. The k-shortest paths are calculated, using the Dijkstra's algorithm, and the resulting paths are ordered from lowest to highest cost. They consider that all links have positive failure probability and that the path is safe from failures when the failure probability of the path is below a threshold. Girão-Silva, R. et al. [13] obtained, despite using randomized failures with SRLG routing algorithms based with or without geodiversity constraints, similar results to those obtained for routes separated at small distances. Despite these proposals, we consider that there may be links with zero probability of failure, like shielded links [58], and that in case there are several routes with the same probability, a tie-breaker is necessary.

Since the region-disjoint paths problem is NP-hard, in [37] an optimization problem was developed. Through an ILP formulation, an exact solution for a geo-diverse path is found, and we denoted it as SRLGRA 1. This algorithm selects a route pair separated by a diameter D, denoted as the minimum distance among the nodes belonging to each route, and the sum of the distance traveled by both routes should be the minimum possible. The larger the chosen diameter, the higher the cost of the chosen routes, but the lower the probability that both links will be affected by a single failure [59]. Also, a heuristic was implemented to find a non-optimal solution, and we denoted it as SRLGRA 2. It also uses the euclidean link length as the cost as SRLGRA 1, but the path pair selection is calculated through two iterations of Dijkstra's shortest path algorithm [60]. The first iteration selects the shortest path between a couple of nodes. Next, nodes with distance equal to or less than D are removed from nodes on the first path that is not the source or destination. Finally, the second path is selected by running Dijkstra's algorithm on the pruned network.

By introducing more degrees of freedom in geo-diverse routing algorithms, more diversity is achieved. Iterative WayPoint Shortest Path (iWPSP) [61], is a heuristic for non-optimal solution which in addition to using a distance D between the nodes composing the multiple paths, uses an additional distance δ to choose central path nodes. The central nodes m, separated by $D + \delta$ among them, force the paths separation delivering greater diversity of paths as δ varies.

In the same paper, also Modified Link Weight (MLW) is presented. This heuristic statistically modifies the link weights and performs Dijkstra's algorithm to calculate the geo-diverse paths with the modified link weights in the network. From the results, iWPSP performs better than MLW when dealing with couple of

nodes near topology boundaries. A larger number of constraints can be found in [59], where the Minimum Cost Pair of D-Geodiverse Path (MCPD-GP) problem is presented. In addition to considering a minimum separation between the nodes of both paths, it also considers the separation of the links on both routes.

All of these region disjoint geo-diverse routing algorithms recommend using a larger separation radius than the failures, to prevent them from affecting both routes. In [44], Gour et al. discuss about how to choose the hazard radius value based on a real radius. If the network design radius is less than the real radius value, the network will be compromised. On the other hand, if the design radius is larger than the real radius value, longer paths will be obtained, in agreement with [13]. Despite obtaining longer paths, a reduction in the probability of simultaneous failure is achieved, especially when the network covers a larger geographical area [3]. The above contrasts with works such as [62] where the choice of a route is intended to reduce the route length. Thus, design radius selection offers a trade-off between cost and robustness, but depends on solving the failure size selection problem. In [59] the maximum distance problem D of geo-diverse paths is solved. Nevertheless, to the best of our knowledge, we could not find a solution to the failure size selection problem to deal with the real threats. Because of this, it is still a problem to determine a realistic threat size for region disjoint geo-diverse routing algorithms.

1.2.4 Performance metrics

To determine if the routing algorithms are successful in selecting routes between a couple of nodes, the success rate of an algorithm is defined in [37] as the quotient between the number of successful request of the algorithm, and the number of successful request of the exact solution. To the best of our knowledge, even exact solution algorithms can fail when the constraints cannot be met. Therefore, in our work, we improve this metric so to be based on the number of distinct couple of nodes in the network. Based on the EGPD metric, wich represent the Effective Geographic Path Diversity, in [4] the compensated Total Geographical Graph Diversity (cTGGD) is presented, which is useful as one global graph metric to characterize the graph resilience to area-based challenges. This metric measures the geographic route diversity normalized to the number of links in a network. Although routes may have high geographic diversity, the network may have low resilience to multiple failures.

Average Two-Terminal Reliability (ATTR) is used in several studies to measure the network reliability [9, 43, 58]. This metric represents the nodes probability to remain connected after random independent link failures. Modiano et al. [58] protect links to be resilient to failures of varying size. Using ATTR, the required connectivity to minimize the number of links to protect is calculated. In addition

to measuring the number of disrupted connections, in [9] also the network disruption time is calculated. In this work, we use ATTR to represent the number of node pairs remain connected by at least one path after threats and we not consider repairing the links.

Also, the routes chosen, depending on the type of routing algorithm used, may vary in the costs associated with the links. To represent the costs associated to the resulting routes, in [4], the number of links belonging to a path is counted, since it reflects the hops number through which it transits. In addition, in [59] the geographical distance between the nodes of each **s-d** node pair, is defined as SD length, also known as path length in [13]. Considering that the shortest route is generally the best route, metrics such as path stretch [3] are used to measure the ratio of the route length to the shortest route. In this thesis we consider that short routes are not necessarily the most resilient. This is why we limit ourselves to comparing the average of both, length and links, to compare the performance of the routing algorithms.

1.2.5 Literature Review Matrix

The literature review matrix is presented in Table 1.1, where from a chronological organization the most relevant works in routing algorithms and performance metrics are discussed in this thesis.

Table 1.1: Literature review matrix

Ref	year	Purpose	Threats	Routing algorithms	Performance metrics
[43]	2013	Modeling network vulnerability when the event has a probabilistic nature, defined by an arbitrary probability density function	Monte Carlo algorithm to generate a max expected damage location	Predefined routes	ATTR
[9]	2014	A probabilistic failure model is proposed based on wave energy behaviour.	Probability of failure of each link according to the distance to the epicenter of the failure.	Dijkstra's algorithm using failure probabilities	ATTR, restoration time
[41]	2014	Re-assigning resources among connections by leveraging their degraded-service tolerance	Probability of occurrence of a disaster through risk maps	SRG-disjoint paths	BW, Network load
[37]	2015	Finding critical network regions based on polygons and finding two region-disjoint paths.	Finding a critical region that contains a set of predefined nodes number	SRLG-disjoint routing algorithm	Number of disconnected pairs, average shortest path length, success rate
[61]	2015	Proposing two heuristics for solving the path geodiverse problem (PGD)	Finding a critical region that contains a set of predefined nodes number	iWPSP and MLW algorithms	cTGGD

Ref	year	Purpose	Threats	Routing algorithms	Performance metrics
[4]	2015	Considering delay-skew requirement in geographically diversepaths	Artificial threats to generate SRLG-disjointness paths	iWPSP with delay-skew constraint	Link counts
[58]	2017	Increasing network connectivity by shielding links	A disk shaped failure that can occur anywhere in the network.	Physical connectivity	ATTR
[59]	2017	Determining the minimum cost pair of D-geodiverse paths	Artificial threats to generate SRLG-disjointness paths	Minimum Cost Pair of D-Geodiverse Path (MCPD-GP)	SD length
[3]	2019	Finding the minimal risk path between end node pairs to tolerate random regional failures	Monte Carlo algorithm to generate a specific sized area from hazard maps	Local Search Algorithm for nodes with minimal vulnerable area	Path stretch
[42]	2019	Proposing a network and disaster model capable of modeling a sequence of disasters in time	Monte Carlo algorithm to generate a max expected damage location	Physical connectivity	ATTR
[13]	2020	Estimating the increase of the path lengths compared to simple link-disjointness	Artificial threats to generate SRLG-disjointness paths	SRLG-disjoint path pairs with geodiversity constraints	Path length

1.3 Hypothesis and research questions

The hypothesis is:

If empirical and geolocated data are used to model link hazards and link failure risks, the optimal selection of geolocated route pairs between network nodes based on minimizing the probability of multiple and independent link failures caused by natural disasters will allow selection of a route pair in the totality of cases and improve network reliability compared to the most widely used SRLG-based optimal geo-diverse route selection algorithm in the literature.

The research questions driving this thesis are:

i) Will the history of geo-located threats in a geography help to increase network reliability?

ii) How to model multiple independent threats?

iii) How to model the risk of network links to deal with multiple threats?

iv) Is it possible for threat-free areas to exist where the network is located?

v) Does the probability distribution of failure vary according to the hazard season?

vi) Is it feasible to increase network reliability through route pair selection based on threat risk?

vii) Which metrics are the most accurate for measuring network reliability based on route pair selection?

1.4 Objectives

General objective:

Propose a threat model and a risk model based on empirical data that allows, through routing algorithms based on geo-located factors, external to the network, to increase the network reliability to face multiple independent failures caused by natural disasters.

Specific objectives:

i) To develop a threat model to determine the representation and occurrence of threats in an area of geographic interest from geo-located historical records.

ii) To develop a risk model to determine the probability of failure of all links belonging to the network.

iii) To develop optimization problems and heuristics to allow selecting a pair of disjoint and maximally disjoint paths between a couple of nodes with the minimum probability of failure.

iv) To implement, and improve if necessary, solutions observed in the literature to face geographic threats, through the selection of a route pair using geo-diversity, to compare with our algorithms.

v) To use or improve metrics to determine the network reliability based on a route pair selection between nodes.

1.5 Contributions

The main author contributions to the state-of-the-art of network reliability through an optimal pair route selection are: The development of a threat model based on a probability distribution of geo-located failures calculated from historical failures in an interest area. The development of a risk model where the probability of link failure is proportional to the intersection area between the probability distribution and the vulnerability zone of the link. The development of optimization problems and heuristics that allow the pair selection of disjoint and maximally disjoint paths with the minimal probability of failure. Finally, the improvement of performance metrics.

The following works contain results served as a basis to achieve the contributions mentioned above and part of them are described in the Chapter 2 and 3 of the manuscript.

1.5.1 Network reliability improved through physical selection

A conference and a journal [63, 64] are the result of research on the concepts of correlated failures and reliability assessment through ATTR. These works addressed the problem of having failures associated with the low diversity in hardware and software of nodes used in the network, which can be concurrent, independent of their remoteness or closeness in the physical network. The result was that the optimal selection of different cultures of nodes location within the network, allows reducing the impact of failures associated with the culture of the nodes. From these works, which point to the moment of creation, relocation or updating of

nodes in the network, which are infrequent processes, the need to solve the problem of improving reliability through more frequent processes, such as the optimal selection of routes between nodes, was triggered.

1.5.2 Network reliability improved through the logical selection

A conference and a journal [65, 66] contain the main ideas of this research. Threats to the region in which the network is located and the risk associated with links resulting from multiple independent threats external to the network are modeled. Then, we seek to improve network reliability through the optimal paths selection to maintain connectivity between nodes in the presence of failures. This new approach complements in parallel the one discussed in Sec. 1.5.1. The main result was that based on empirical results of multiple small-scale failures, the proposed solutions for route pair selection achieve higher reliability than geo-diversity algorithms focused on large-scale threats, widely used in the literature to face geographic threats.

1.5.3 Journal papers, conference and seminar presentations.

The following journal papers, conference and seminar presentations were obtained as a result of this thesis, and they allowed the dissemination of this research work:

- Boettcher, N. A., Prieto, Y., & Pezoa, J. E. (2018, September). Micro Failure Region Models Inducing Massive Correlated Failures on Networks Topologies. In International Conference on Information Technology in Disaster Risk Reduction (pp. 130-141). Springer, Cham.
- Boettcher, N. (2021, December). Maximizando la resiliencia en redes de telecomunicaciones. ED740 PhD Seminar, Doctorado en Ingeniería Eléctrica, Universidad de Chile.
- Boettcher, Nicolás A., Yasmany Prieto, and Jorge E. Pezoa. Maximizing the telecommunication network reliability through a pair of routes selection by exploiting geo-located failure records. IEEE Transactions on Network and Service Management, Manuscript submitted for publication.

2 Materials and methods

This chapter presents the main ideas that support the development of this thesis. For the selection of a route pair, a method capable of choosing the route pair with the lowest probability of failure, based on historical geo-located failures records is presented. The threat model and risk model that allows calculating the probability of failure associated to each link is presented in Sec. 2.3. The mathematical modeling of the proposed solutions for the selection of a route pair between a couple of nodes is presented in Sec. 2.4. The real historical failures and the real-world networks used to implement the proposed methodologies are introduced in Sec. 2.5 and Sec. 2.6 respectively. Finally, the metrics to evaluate the performance of the approach proposed are described in Sec. 2.7.

2.1 Problem statement

The reliability of a communication network is defined as its ability to maintain connectivity through paths among all its nodes without interruptions in the presence of failures. A geo-located communication network is mathematically represented by the undirected graph $G = (V, E)$, where $V = \{1, 2, \ldots, n\}$ is the set of geo-located communication nodes and $E = \{e_1, e_2, \ldots, e_\epsilon\}$ is the set of straight communication links between nodes. The length, in kilometers, of each link, is represented by the euclidean distance between the nodes compose it, by $l(e_i)$. In addition, the geography in which G is located is susceptible to different types of hazards, which generate network failures.

We denote a failure as a disk associated with the geographic coordinates of the center of the disk that forms it and its diameter. The set of failures occurring in an area of interest A, is denoted by $f = \{f_1, f_2, \ldots, f_\varsigma\}$, where ς corresponds to the number of failures for a specific time window of measurement.

We consider a threatened area to be any geographic area in which historical

failures have been detected. If no failure events have ever been recorded in an area, we consider it to be a threat-free area. The longer the period of recorded failures, the greater the reliability provided by the data with respect to the risk of the threatened area.

In our methodology, we first consider a threat model based on a probability distribution of geo-located failures obtained from historical failures in an area. Second, we develop a risk model where the probability of link failure is proportional to the intersection area between the probability distribution of geo-located failures and the vulnerable zone of the link. Finally, we compute a route pair between all couple of nodes belonging to the network through routing algorithms that minimize the probability of the route pair failure.

2.2 Rationale

Through a simple case study, we explain the beneficial impact of determining a route pair based on knowledge of geo-located failures records occurred in the geography where the network topology is located. A telecommunication network immersed in Italy, its infrastructure is at risk of failure, as there are records of failures that are repeated every season in the same areas where they have been historically recorded.

Figure 2.1: Real-world network topology with geo-located failures records.

A representation of the geo-located network topology and their threats is shown in Fig. 2.1, where each name corresponds to the city where the node is located. The red circles correspond to the historical records where failures have

occurred. Additionally, information corresponding to the length of the largest set of failures detected, represented by D, is provided.

The network operator wants to establish a route pair, primary and backup paths, to establish a connection between two nodes and maintain the connectivity in case a failure interrupts the connection of one route. Thus, by keeping at least one route active after the failure, it avoids incurring costs associated with compensating customers for SLA non-compliance. The correct selection of a route pair from among all available options requires a methodical process to maximize the network reliability.

To select the routes, the operator must first choose a couple of nodes. For the example, he uses PESC and PARM as source and destination nodes respectively. Now, he must choose a route pair selection algorithm. For this, he uses two different routing algorithms paradigms. The first one consists in using a region-disjoint geo-diverse routing algorithm, which is widely used in the literature to deal with geographical threats. The second one corresponds to the routing algorithm proposed, based on the minimal probability of failure.

The region-disjoint geo-diverse routing algorithm is based on the selection of the shortest route pair that are separated by a minimum distance D. For the example, the length D is used, which corresponds to the largest failure recorded in the network. Based on that distance, two routes will be chosen together, using an exhaustive method until the route pair is found, whose nodes meet the separation D and the sum of the length of both routes is the minimum possible. As can be seen in Fig. 2.2a, the two selected routes (blue and green paths) meet the requirements described above.

The following routing algorithm to be used corresponds to our proposal. From the total number of geo-located failures recorded, a threat model is obtained to represent the probability distribution of failure in the geographic plane. Additionally, the vulnerable zone associated to each link is represented by a hippodrome. Then, the sum of the areas intersected by the hippodromes with the failure probability distribution areas is divided by the total area of recorded threats. The result is presented as the probability of link failure with respect to the total network failures. Based on the above, the route pair having the minimum failure probability between the couple on nodes is calculated. As can be seen in Fig. 2.2b, the two selected routes have the minimal probability of being affected by any threat, since they are not intersected by any failure record.

Finally, let us assume that in each season the failures occur again in the same areas, so that the links represented by the red lines fail. For the first case, we would obtain that neither route remains active, since both fail, despite being geographically separated by D. This occurs because this algorithm is limited to considering a single large-scale failure and considers that failures can occur anywhere. From the constraint of defining a threat size, using any of the algorithms

discussed in the state of the art [37, 61], the same problem would have been reached, since they are all based on the same principle. In the second case, both routes will continue to be active, since the links are located over a geography that has never reported this type of failures, being considered to be safe links. This allows that, despite being close to threatened areas, they can have priority to be selected over the rest of the links. This means that, despite the existence of threat areas where the network is located, it is possible to select a route pair with the minimal probability of failure. While the disjoint region model attempts to separate the route pair based on D without knowing if failures are likely to occur in the area, the proposed model selects the routes with the minimal probability of failure based on the probability distribution generated from the historical data.

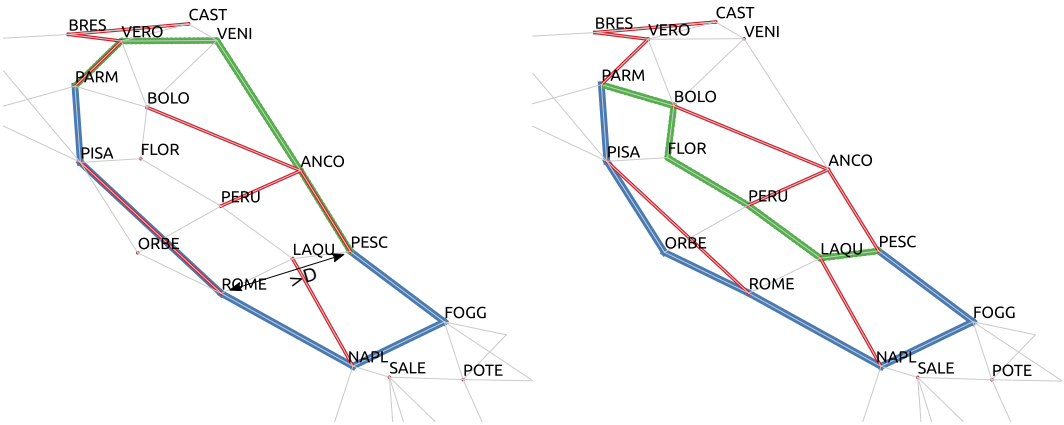


Figure 2.2: A route pair selection for different paradigms: (a) region-disjoint geo-diverse routing, and (b) minimal risk-based routing.

Minimal probability of failure route selection requires historical geo-located failure information to be effective. Despite this, it offers the network operator clear advantages over regional-disjoint geo-diverse route selection algorithms by eliminating the need to determine a failure size and avoids forcing route separation. Based on databases containing extensive historical data and accurate failure detections, better network resilience will be achieved by using maximally disjoint routes that share links with zero probability of failure on both routes.

The optimal design of a routing algorithm that avoids routes that may be threatened involves prior knowledge of historical failures in the areas where the links are located. In addition, it can be complemented by prior knowledge of the safe geographical areas in which the network topology is located. This problem represents a huge challenge in terms of modeling the main threats according to their history and modeling the risk of the links belonging to the network.

2.3 Threat model and Risk model

Major natural disasters are generally associated with adverse climatic conditions, depending on the season of the year, determining the level of devastation they achieve, and their occurrence. To better analyze the threats, we differentiate them according to their seasonality within a year. We define a season as a time window equal to a month, in order to be able to evaluate twelve different seasons per year. The shape representing the failure area and the refresh failure detection time may vary depending on the nature of the instrumentation used.

Considering the geo-located failure data and their frequency of occurrence, year by year, in an area of interest A, will allow defining the likelihood that a threat will induce a failure in the infrastructure. For this purpose, it will be considered whether these zones are close to the infrastructure or not. For simplicity and because of the shape of the data networks, we will use hippodromes to determine zones under threat of failure.

To define the threat model, we first obtain the areas affected by failures during a season. It is possible that there are failure areas intersecting with each others in the same season, such as a hours-long failure. In these cases, we interpret it, as the expansion of the failure in the geography during this season. Thus, as shown Fig. 2.3, each season is composed by f and represented by $\nu = \bigcup_{i=1}^{\varsigma} f_i$.

(a) (b)

Figure 2.3: A season of $\varsigma = 5$ is (a) composed by f, and (b) represented by ν.

We consider the threat areas as the union of the recorded failures areas, not as a single disk representing the totality of the records. Thus, we avoid considering hazards in areas where they have never occurred, i.e. tsunamis above 1000 m sea level, earthquakes in non-seismic areas or fires in hydrographic areas. Also, we define each geographic coordinate $(x, y) \in \nu$ has the same probability of failure occurring in it. Now, in order to represent the frequency of occurrence of hazards, we calculate ν for the same month in a set of years defined by ψ. Each intersection among ν represents that a failure has occurred again at the same location. We determine for each time a geographic coordinate failure is detected at the same location, its incidence factor increase. The greater the number of annual intersections among ν, the greater the probability of failure in that area.

Given an area of interest A, associated to the total network coverage, the region R in which the failures occur are recorded every m-th month. The geographic coordinate cumulatively record the existence of failures in a given m-th month as shown in Eq. 2.1. This process considers all failures, in a month, with the same probability at the same location, regardless of how long ago they occurred.

$$R^m_{x,y} = \sum_{i=1}^{|\psi|} F^{m,i}_{x,y} \tag{2.1}$$

$$F^{m,i}_{x,y} = \begin{cases} 1 & \text{if } (x,y) \in \nu \\ 0 & \text{otherwise} \end{cases}$$

, where $F^{m,i}_{x,y}$ represent the failure existence at the geographic coordinates (x, y) in the m-th month of i-th year belonging to the set ψ of historical recorded years. The totality of the failures occurring in the area of interest A are represented by the sum of the failures present in this area, denoted by Γ^m_A, is shown Eq. 2.2.

$$\Gamma^m_A = \sum_{(x,y)\in \mathrm{A}} R^m_{x,y} \tag{2.2}$$

From Γ^m_A, a probability function g can be constructed to calculate the probability for a given coordinates presents a failure in a m-th month, according to the incidence rate in the cumulative history, as shown Eq. 2.3. The more times a failure has occurred at a coordinates, the higher the probability to occur again at the same location, with respect to the area of interest.

$$g^m(x,y) = \frac{R^m_{x,y}}{\Gamma^m_A} \tag{2.3}$$

From now on, it is possible to determine the probability of failure occurring in a geographical location belonging to an area of interest where the network is located. In this work, as there are several failures that compose ν, it is not possible to represent where the failure originates. Furthermore, the probability of failure within each ν does not follow the behavior of a given distribution, since it only depends on which failures areas intersect among them and not on an analytical function.

Figure 2.4: Vulnerable area represented by a hippodrome.

Each link e_i has a vulnerable zone, as shown Fig. 2.4, which corresponds to an area delimited by a hippodrome with distance h from any point of the link around it [43, 67], denoted as e_i^h. If $e_i^h \not\cap \Gamma_A$, then e_i is safe from threat, otherwise $P(e_i) > 0$. The larger the data set from which the threat model is constructed, the less error there will be in assuming a link cannot be affected by a threat.

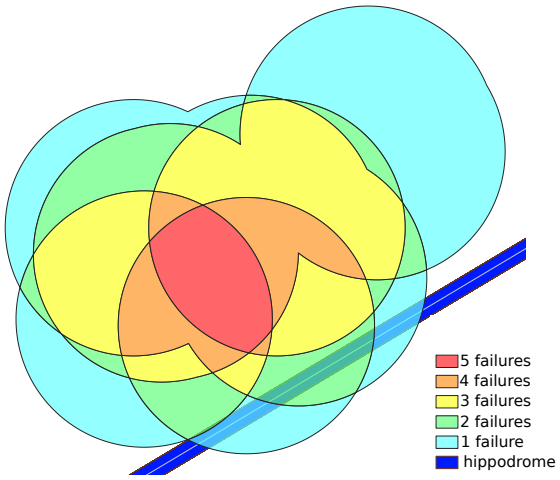


Figure 2.5: Annual threat model close to a hippodrome.

An example, as can be seen in Fig. 2.5, where the hippodrome, represented by the blue color, is intersected by multiple ν corresponding to the same month in $|\psi| = 10$ years. Each color represents the number of intersected ν per geographic coordinate within the analyzed period. In ten years, only the red area managed to record five failures in the same zone. Despite the link being close to threats areas where failures have occurred more than two times, it is only affected by areas that comprise one or two occurrences. The green area intersecting the hippodrome registers two failure occurrences unlike the cyan color that only registers one failure occurrence within the same period. The rest of the link is not affected by failures in the analyzed period. The larger the h value from the hippodrome, the greater the sensibility of the link to distant threats, can further increase its probability of failure.

To calculate the link failure probability, we use Eq. 2.4, where the probability of failure in the hippodrome e_i^h is equal to the sum of all probability of failure in the region intersecting the hippodrome of i-th link. For simplicity, starting from the probability of failure at the hippodrome, we consider $P(e_i) = P(e_i^h)$.

$$P(e_i^h) = \sum_{(x,y)\in e_i^h} g(x,y) \tag{2.4}$$

2.4 Route pair algorithms

In this section we present two optimization problems to select the route pair between a couple of nodes with the minimal probability of joint failure. Also, two heuristics to select the route pair between a couple of nodes with the minimal probability of non-joint failure are presented. Finally, we improve a region-disjoint geo-diverse routing heuristic to compare the results of our proposal. We defined the couple of nodes s and d as source and destination nodes respectively, also called terminal nodes, where $(s, d) \in V$. For each link $e_i \in E$, two variables $x_i, y_i \in \{0, 1\}$ are defined. If e_i is on path $\mathcal{P}_1$, then $x_i = 1$, otherwise $x_i = 0$; if e_i is on path $\mathcal{P}_2$, then $y_i = 1$, otherwise $y_i = 0$;

2.4.1 Optimization problems

Two optimization problems are defined to solve the Minimal Probability of Joint Failure route pair selection. The first, for disjoint route pair selection and the second, for maximal disjoint route pair selection. For both, a route pair $\mathbf{p}$ is selected from a set of all possible routes $\mathcal{P}_{s,d}$ between nodes $(s, d) \in V$.

Since the failure events are independent and the failure probabilities are very small, we consider that selecting the paths with minimum failure probability with $-\sum_{(i)} \log(1 - p_i)$ is equivalent to $-\prod_{(i)} (1 - p_i)$, without affecting the optimal solution, as posed in [68]. This property is feasible to use for both, disjoint and maximally disjoint route pair selection, where only links with probability of failure equal to zero can share both routes. A weight $\omega_i = -\log(1 - P(e_i)) + \frac{l(e_i)}{K}$ is defined for each link $e_i \in$ E and can be calculated beforehand. The value $\frac{l(e_i)}{K}$ is a negligible value compared to the smallest failure probability in the network, where we used $K = 10^{10}\Gamma_A^i$. With this value we avoid that a link with $P(e_i) = 0$ can be reused in the same path. Also, it is feasible to untie a choice of routes, based on $l(e_i)$, when both routes have the same probability of failure.

The first proposed optimization problem, presented in Eq. 2.5, is defined as Minimal Probability of Joint Failure for Disjoint Route Pair (MPJF-DRP).

$$\mathbf{p}^* = \underset{\mathbf{p} \in \mathcal{P}_{s,d}}{\operatorname{argmin}} \sum_{i=1}^{\epsilon} \omega_i (x_i + y_i), \tag{2.5}$$

subject to:

$$(x_i + y_i) \leq 1, \forall i \in 1, 2, \ldots, \epsilon \tag{2.6}$$

The objective function is to minimize the probability of joint failure between primary and backup routes. Constraint (2.6) states that e_i is used by at most one path.

Moreover, we present a second optimization problem, depicted in Eq. 2.7, defined as Minimal Probability of Joint Failure for Maximal Disjoint Route Pair (MPJF-MDRP), where an additional variable is added to the constraints. For each $e_i \in E$, a variable $b_i \in \{0, 1\}$ is defined. If e_i is on both paths, then $b_i = 1$, otherwise $b_i = 0$. The difference with the first optimization problem is that it allows sharing links between both routes, in case they are threat-free links, allowing to further decrease the minimal probability of joint failure of the route pair.

$$\mathbf{p}^* = \underset{\mathbf{p} \in \mathcal{P}_{s,d}}{\operatorname{argmin}} \sum_{i=1}^{\epsilon} \omega_i (x_i + y_i), \tag{2.7}$$

subject to:

$$P(e_i) b_i = 0, \forall i \in 1, 2, \ldots, \epsilon \tag{2.8}$$

$$(x_i + y_i) \leq 1 + b_i, \forall i \in 1, 2, \ldots, \epsilon \tag{2.9}$$

The objective function is to minimize the probability of joint failure between primary and backup routes. Constraint (2.8) prevents a link with a non-zero failure probability from being chosen for use on both routes. Constraint (2.9) states that link will be used by at most one path, unless it is enabled to be on both routes. The proposed optimization problems were resolved using the Gurobi optimizer solver 9.1.1 [69].

2.4.2 Minimal risk routing heuristics

In addition to the optimization problems, we present two heuristics. These use the weights ω_i, computed for the proposed optimization problems, to obtain the shortest routes serially. Instead of selecting the route pair jointly as optimization problems do, it selects them sequentially, as SRLGRA 2 [60] does. In this way, we determine what advantages there are to using a solution generated from the selection of the route pair with the minimal joint failure probability.

On Alg. 1, the pseudo code for the Minimal Probability of non-Joint Failure for Disjoint Route Pair (MPnJF-DRP) is presented. From the network and the terminal nodes, the primary path $\mathcal{P}_1$ is calculated running Dijkstra's algorithm to obtain the least weight route. Then, the links belonging to $\mathcal{P}_1$ are pruned from the network and the backup path $\mathcal{P}_2$ between the same couple of nodes is calculated using Dijkstra's algorithm again. Both calculated routes, one after the other, are the resulting pair of routes.

Unlike the disjoint process, the maximally disjoint route pair heuristic, only prunes links whose probability of failure is greater than zero, thus allowing the reuse of links that have never had a previous failure record. The pseudo code

for the Minimal Probability of non-Joint Failure - Maximal Disjoint Route Pair (MPnJF-MDRP) is presented in Alg. 2.

Algorithm 1 MPnJF-DRP

Input:
$G :=$ graph
$s, d :=$ source and destination node
$\omega_i :=$ the weight associated to e_i

Output
$\mathcal{P}_1$,$\mathcal{P}_2$

1: **function** DIJKSTRA(G, m, n)
2: **return** path$_{m,n}$
3: **function** PRUNE(G, path)
4: Prune from G every links $\in$ to path
5: **return** G'
6: **procedure** MPNJF-DRP
7: Primary $\leftarrow$ Dijkstra$\{G, s, d\}$
8: G$'$ $\leftarrow$ Prune$\{G$,Primary$\}$
9: Backup $\leftarrow$ Dijkstra(G', s, d)
10: **return** Primary,Backup

2.4.3 SRLG algorithm

In this work we present the Double WayPoint Shortest Path (DWPSP) heuristic. It is a simplified and improved version of iWPSP heuristic [61], since it considers only two routes ($k = 2$), but it also solves nodes re-utilization problem in the same path.

The iWPSP, as can be seen in Fig. 2.6, consists in the nodes choice S_i, m_i and T_i, source neighbor, middle and destination neighbor respectively. Then, finds the paths that link them using the Dijkstra's algorithm. It uses a distance d to establish a minimum separation between the neighbouring nodes of both, the source and destination. Also, adds a δ to separate in $d + \delta$ the middle nodes m_i between them.

As can be seen in Fig. 2.7, iWPSP can present problems when choosing the routes due to the mechanism of how the routes are built between both ends, using only the information of the m_i node, without previous knowledge of the nodes used to build the route. Reused links may exist between $S_i - m_i$ and $m_i - T_i$ as can be seen in Fig. 2.7a, where the link Barcelona-Lyon is used twice. In addition, a loop can be created within a route, unnecessarily lengthening the link and bypassing the desired effect of moving the link away by $(d+\delta)$ as can be seen in Fig. 2.7b. In both cases a node reuse problem in the same path is detected, on Lyon and Berlin respectively.

Algorithm 2 MPnJF-MDRP

Input:
$G :=$ graph
$s, d :=$ source and destination node
$\omega_i :=$ the weight associated to e_i

Output
$\mathcal{P}_1$,$\mathcal{P}_2$

function DIJKSTRA(G, m, n)
  return path$_{m,n}$
function PRUNE(G, path)
  Prune from G every links with $P(e_i) > 0 \in$ to path
  return G'
procedure MPNJF-MDRP
  Primary $\leftarrow$ Dijkstra$\{G, s, d\}$
  G$' \leftarrow$ Prune$\{G$,Primary$\}$
  Backup $\leftarrow$ Dijkstra(G', s, d)
  return Primary,Backup

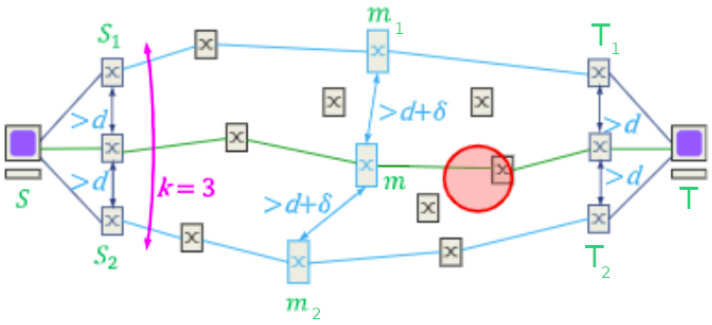


Figure 2.6: iterative WayPoint Shortest Path (iWPSP) heuristic.

To solve this, a node memory buffer has been implemented in Alg. 3. Before calculating Dijkstra's algorithm between a couple of nodes, it prunes from the network all links intersecting the nodes of the path already calculated, except for the links intersecting the last node. In this way, it is impossible to reuse the same nodes on the same path. In case it is not feasible to find a solution, the path is computed in reverse order. If in both cases it is not possible to obtain a route, a new m_i node is searched for both paths that is ϕ farther away from $(d + \delta)$, thus fulfilling the minimum distance requirements, and the feasibility of finding a route pair is checked again.

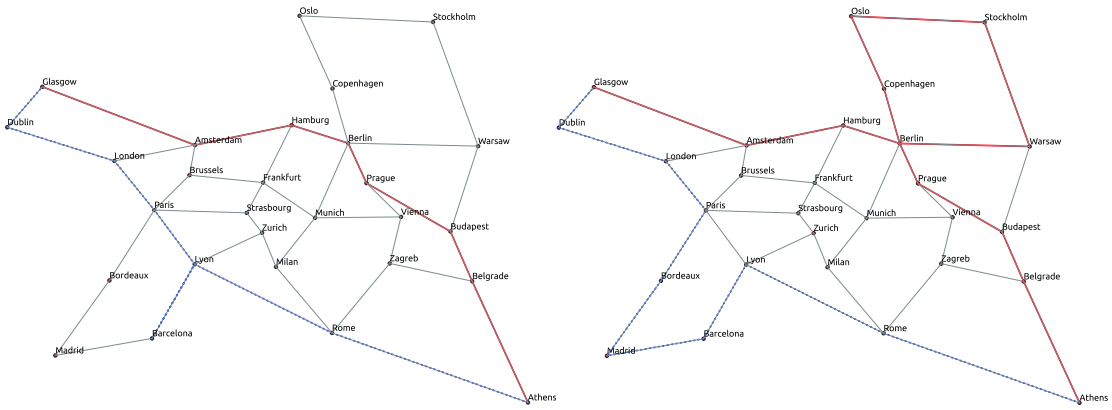
SRLG algorithms require an input D, corresponding to the diameter of the largest expected failure. A solution to the failure size selection problem, proposed

Algorithm 3 Double WayPoint Shortest Path heuristic.

Input:
$G :=$ graph
$S, T :=$ source and target node
$d :=$ separation distance between paths
$\delta :=$ delta distance when selecting waypoint node
$\phi :=$ incremental distance

Output
$\mathcal{P}_1$,$\mathcal{P}_2$

function PRUNE(G,path,a)
    N $\leftarrow$ Select all nodes $\in$ path less a
    G' $\leftarrow$ Prune from G all $e_i \cap N$
    return G'
function DIJKSTRA($path, a, b$)
    G' $\leftarrow$ Prune{G,path,a}
    return shortest_path between a,b from G'
function CHECKNODETWICE(path)
    if exist a repeted node in the path **then**
        return 1
    return -1
function GETPAIR(δ)
    $[m_1, m_2] \leftarrow$ GetMiddlePoint{δ}
    Primary $\leftarrow$ GetRoute{S, S_1, m_1, T_1, T}
    Backup $\leftarrow$ GetRoute{S, S_2, m_2, T_2, T}
    if (Primary OR Backup) == -1 **then**
        GetPair{$\delta + \phi$}
    return path
function GETROUTE(S, S_n, m, T_n, T)
    path $\leftarrow e_i \cap \{S, S_n\}$
    path $\leftarrow$ path + Dijkstra(path, S_n, m)
    path $\leftarrow$ path + Dijkstra(path, m, T_n)
    path $\leftarrow$ path + Dijkstra(path, T_n, T)
    if CheckNodeTwice{path} **then**
        path $\leftarrow e_i \cap \{T, T_n\}$
        path $\leftarrow$ path + Dijkstra(path, T_n, m)
        path $\leftarrow$ path + Dijkstra(path, m, S_n)
        path $\leftarrow$ path + Dijkstra(path, S_n, S)
    if CheckNodeTwice{path} **then**
        return -1
    return path
function GETMIDDLEPOINTS(δ)
    Create a paralel segment separated by $d + \delta$ above and below to L (L_1, L_2) respectively.
    Choose m' above L_1 with a minimum distance greater than $d + \delta$ to m.
    Choose m'' below L_2 with a minimum distance greater than $d + \delta$ to m.
    return m', m''
procedure DWPSP
    Create an imaginary straight line L intersecting S and T
    $m \leftarrow$ the middle point of L.
    Create two paralels segments above and below to L, with a distance d.
    Choose neighbour node S_1, T_1 above L with a minimum distance $> d$ to L.
    Choose neighbour node S_2, T_2 below L with a minimum distance $> d$ to L.
    Primary, Backup $\leftarrow$ GetPair{δ}



(a) d=0 km and δ=400 km. (b) d=150 km and δ=600 km.

Figure 2.7: Issues detected in iWPSP heuristic in Nobel-EU, solved in DWPSP.

by us, from the empirical data approach, is described below. From the records obtained by the data set, we compute the smallest circle [70] enclosing intersected failures in ν to find the largest diameter to represent D for each hazard season, as shown in Fig. 2.8.

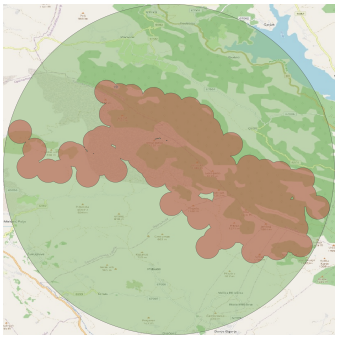


Figure 2.8: Smallest sized circle from recorded data representing D.

To compare the performance of the proposed algorithms, we choose from the literature SRLGRA 1 and SRLGRA 2, for which we will use the same mechanism as for DWPSP to compute D.

The optimization problem describing the operation of SRLGRA 1 is explained in Appendix A. When the constraint for the pair of disjoint routes with a specific D value is not met, then it is not possible to establish communication between both. To increase the DWPSP chance of obtaining routes, we set $\delta + \phi = 0$ to relax the separation constraint on intermediate nodes.

2.5 Empirical failures

In order to use the proposed threat model, records of geotemporal failures are required to calculate the effects they produce in different networks of the globe. Any type of failure can be used with the proposed model, but for this work, we choose fire as failure. Fire is a threat in most parts of the world, generally reported as large-scale areas, representing regions containing multiple independent small-scale failures. In addition, we have access to geolocated global fire events, which have been recorded over time via NASA satellites. We considered that all fire events recorded by MODIS may be of high risk to damage the network infrastructure, so we did not filter out any events. We obtained the totally complete annual records available from MODIS, comprising the range between the year 2001 and the year 2019. Records between the years 2001 and 2018 were used to import into the threat model. While those for the year 2019, for each season we used each day's failures as a different scenario. The failures occurring each day will be considered as unique failures to that day, i.e., each day starts with all its links active. To represent the coverage of each recorded failure geographic coordinate, we use a disk of diameter 1000 meter, since 500 meter is the radius of each fire focus derived from the MODIS database [71].

MODIS instrumentation onboard two NASA satellites ensures at least 4 daily measurements [72]. During an exploratory data analysis, we observed that in no case did the fire outbreaks remain in the same place in a month, thus indicating fire expansion or extinguishment. With the above, we eliminate the probability of a fire reigniting at the same point of origin in the same month. For the risk model, we consider hippodromes with a $h = 40$ m [73]. This value represents the case where flammable material, such as wood, is found near the link, which could cause the link to be damaged. The h value assigned to the hippodrome depend on how close the threat must be to influence the performance of the link. We kept the h independent of the geographical area and season of the threat. To select the failures to be taken into account in the threat model proposed, we filter by the geographic area in which the networks are located reported below. For each network area, an additional 1 km buffer was established to ensure that the edge links could also be affected by nearby failures of 500 m radius.

2.6 Real-world networks selection

In this work, we looked for topologies with different both, degrees of connectivity and geographical coverage. Three real-world telecommunication networks used in previous works on region disjoint geo-diverse route selection in networks were chosen [61, 74]. AT&T, Nobel-EU and Italian main backbone, were obtained

from InternetZoo [75], SDNLib [76], and based on what was published in [37], respectively. In Fig. 2.9 the geo-located topologies can be observed, where the associated name to each node corresponds to the name of the city where it is located. The main graph properties of the real world telecommunication networks to be used in this work are presented in the Table 2.1, where diversity in both degree and diameter is demonstrated.

Table 2.1: Network properties of each topology chosen.

Topology	Nodes	Links	Degree	Diameter
Italy	32	60	3.75	7
Nobel-EU	28	42	3	8
AT&T	25	57	4.48	5

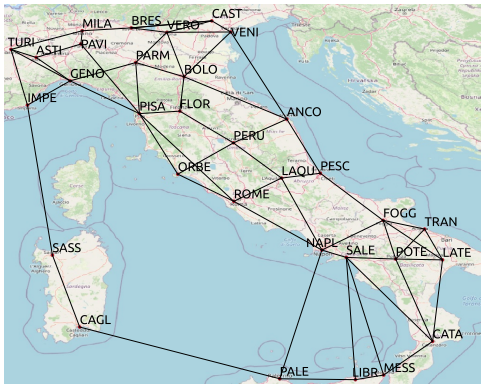
Otherwise, there are also geographical characteristics, less explored in the literature, that help us to better understand the sensibility of the network to different types of hazards such as climate, difficulty of access to the terrain, altitude, etc. The topologies analyzed in this work are located in the northern hemisphere at similar latitudes, so they maintain the same hazard seasons characteristics in the year [77]. The geographical properties of the topologies are shown in Table 2.2, where the perimeter is defined as the sum of the links length belonging to the edge network, coverage is the geographic area enclosed by the perimeter, and the average cost per link is the average euclidean length from all links belonging to the network. Also, the average annually fires detected by MODIS for each network area is presented, calculated from the number of failures obtained between 2001 and 2018.

Table 2.2: Geographical network properties of each topology.

Topology	Perimeter [km]	Coverage [km^2]	Average cost per link [km]	Average annually fires detected
Italy	2613	420 996	76	5444
Nobel-EU	9985	3 958 080	184	25 293
AT&T	11 141	6 775 562	685	121 586

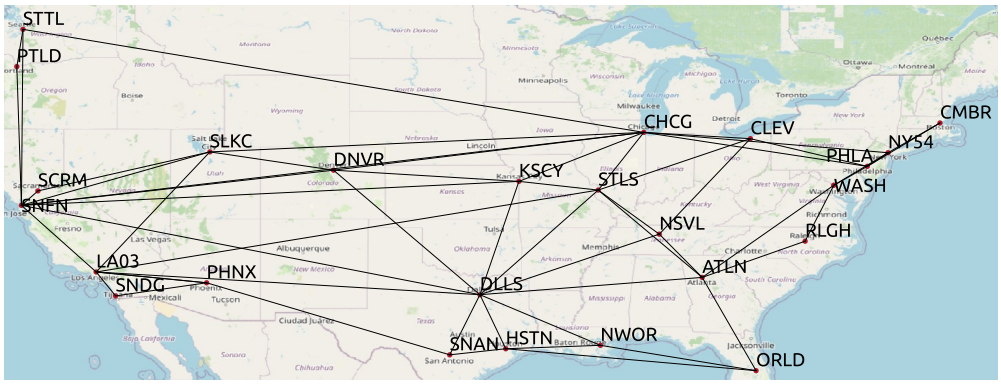
Infrastructures with different sized coverage areas were chosen to have geographic diversity and to increase the probability of finding safe links along the records. As expected, for similar latitudes, the greater the geographic coverage, the greater the number of fires recorded. Although the Italy network has more nodes and links than the other topologies analyzed, Table 2.2 shows that it covers a smaller geographical area. Indicators such as this show that, even if a network has a high degree of connectivity, it can be highly affected by small-scale failures

when the network has a small coverage area. This kind of information is relevant when calculating routes to avoid failures of a specific size. The smaller the geographic coverage, the smaller the failures size to disable the connection between a couple of nodes in the network, because the nodes are located closer to each other.



(a) Italy backbone

(b) Nobel-EU



(c) AT&T

Figure 2.9: Real-world geo-located telecommunication network topologies.

2.7 Performance Metrics

The ability of a network to remain active after threats to a couple of nodes depends initially on its ability to establish routes between them and whether at least one route remains active after failures.

The ability of the network to provide a route pair between a couple of nodes, given a particular routing algorithm, is measured with the success rate (SR). Next to it, the ability to maintain at least one route after network failures occurred is

calculated by Average Two Terminal Reliability (ATTR). In addition, as characteristics of the selected route pairs, the connectivity between them are calculated from the threat probability according to the historical record through Average Connectivity (AC). Also, to represent the cost of resource consumption by each routing algorithm, we calculated the path pair length (PPL).

2.7.1 Reliability and Feasibility

Average Two-Terminal Reliability quantifies how well the network is connected after the occurrence of failure events. This metric represent the probability of a randomly chosen couple of nodes remain connected. Unlike works seen in the literature where they measure connectivity through active links, we will measure connectivity based on pre-established working routes. Thus, the Two Terminal Reliability between two nodes (s, d), is defined by $Z_{s,d} \in \{0, 1\}$. If after failures at least one route is available between (s, d), then $Z_{s,d} = 1$, otherwise $Z_{s,d} = 0$. Therefore, the ATTR of a network after failures occur, calculated from previous route pairs selected between two nodes (s, d) is given by:

$$\text{ATTR} = \binom{n}{2}^{-1} \sum_{s \neq d} Z_{s,d} \tag{2.10}$$

, where $\binom{n}{2}$ is the binomial coefficient and n is the amount of nodes connected to the topology. To calculate the ATTR, we use as active routes all the route pairs selected by each routing algorithm.

However, achieving an ATTR = 1 is not always possible, since it will depend on whether connectivity is achieved between the nodes prior to the failure. In our case, it will depend on the feasibility of selecting a route pair by the chosen routing algorithm, especially when there are constraints that prevent finding a solution. To calculate the success rate of an algorithm to generate routes between all couple of nodes in the network, we improve the SR presented in [37]. We measure the reachability between two nodes (s, d) through $\mathcal{R}_{s,d}$ and a routing algorithm. If the routing algorithm is able to deliver a solution for the node pair, then $\mathcal{R}_{s,d} = 1$, otherwise $\mathcal{R}_{s,d} = 0$. The above is closely tied to the routing algorithm used, since even if infrastructure exists to connect the two nodes, the algorithm may not be able to deliver a solution. Instead of calculating SR as the reachability between all nodes belonging to the network divided by the number of node pairs connected by routes generated by a single solution routing algorithm, we propose to divide it by the total number of node pairs in the network, as Eq. 2.11 shown.

$$\text{SR} = \binom{n}{2}^{-1} \sum_{s \neq d} \mathcal{R}_{s,d} \tag{2.11}$$

, where $\binom{n}{2}$ is the binomial coefficient and n is the amount of nodes connected to the topology. Because prior to failures there will always be higher connectivity, the ATTR $\leq$ SR condition must always be satisfied.

Finally, from the previously established routes and the detected failures, we calculate the number of failed single routes and the number of both, the primary and backup routes that failed. Single failed routes correspond to all routes containing a failure that crosses the hippodrome belonging to a network. As a subset of single failure are dual failed routes, which represent pairs of failed routes, where connectivity between a couple of nodes ceased to exist. Failed dual routes are also known as number of disconnected peers, a metric used in [37]. In both cases, single and dual failures, the total number of failed routes are normalized with respect to the total number of routes established by the routing algorithm. From these metrics it is possible to determine how reliable each algorithm is in keeping its routes available after a failure.

2.7.2 Average Connectivity

AC represents how well connected a couple of nodes $(s, d) \in E$ remain after the failure of links connecting them, due to failure events. To represent mathematically the average connectivity we were based on [68]. It is computed from the failure probability $P_{i,j}$ associated to the j-th link belonging to the i-th path, provided by the risk model discussed in Sec. 2.3. Since the failure events are independent, the failure probability of the i-th path, denoted by P_i, corresponds to the multiplication of the failure probabilities of the k links that make up the path. To allow compatibility with scenarios where there are safe links with $P_{i,j}$ equal to zero, we calculate P_i based on the complement of $P_{i,j}$. Then, the average connectivity between (s, d) is calculated as the complement of the failure probability of the two paths between these nodes as denoted by Eq. 2.12.

$$P_i = 1 - \prod_{j=1}^{k}(1 - P_{i,j}), \qquad , \qquad \mathrm{AC}_{s,d} = 1 - \prod_{i=1}^{2} P_i \tag{2.12}$$

Therefore, the average connectivity for a network, defined in Eq. 2.13, is the average $\mathrm{AC}_{s,d}$ among all couples of nodes belonging to the network.

$$\mathrm{AC} = \binom{n}{2}^{-1} \sum_{s \neq d} \mathrm{AC}_{s,d} \tag{2.13}$$

, where $\binom{n}{2}$ is the binomial coefficient and n is the amount of nodes connected to the topology. This metric allows to know what percentage of connectivity a route has according to the historical records of its links. Thus, before a failure occurs,

we can predict how connected the network will remain after the failure event. The lower the failure probability of the links belonging to a route, the higher the average connectivity value. Thus, with failure probabilities close to zero, it is feasible to obtain average connectivity values higher than 99.999%. Although this is a good indicator, there is still a probability that links may fail. To compare high AC values, we propose the Average Annual Disconnectivity (AAD) metric. In Eq. 2.14, AAD is defined as the complement to the annual average of the average connectivity for each month of the year, where AC_i is the AC value for the i-th month. The smaller the value of AAD, the better the annual average connectivity.

$$AAD = 1 - \frac{1}{12}\sum_{i=1}^{12} AC_i \tag{2.14}$$

2.7.3 Path Pair Length

PPL is a metric based on a single path length [78, 79], since the number of hops and the distance covered by the routes are usually factors associated with economic costs. Path Pair Length allow us to calculate the average of: number of **links** used by both routes, and the **distance**, in kilometers, traveled by them for PPL_l and PPL_d respectively. We defined them as show in Eq. 2.15 for each node pair belonging to the network, represented as the k-th node pair in the m-th month.

$$\mathrm{PPL}^k_{l,m} = \sum_{i=1}^{\epsilon} x_i + y_i, \mathrm{PPL}^k_{d,m} = \sum_{i=1}^{\epsilon} l(x_i)x_i + l(y_i)y_i \tag{2.15}$$

Generalizing, for a whole topology, we can obtain the monthly PPL as shown in Eq. 2.16, representing the monthly PPL among each pair of distinct nodes belonging the topology in the m-th month.

$$\mathrm{PPL}_{l,m} = \binom{n}{2}^{-1} \sum_{k=1}^{\binom{n}{2}} \mathrm{PPL}^{\mathrm{k}}_{\mathrm{l,m}}, \mathrm{PPL}_{d,m} = \binom{n}{2}^{-1} \sum_{k=1}^{\binom{n}{2}} \mathrm{PPL}^{\mathrm{k}}_{\mathrm{d,m}} \tag{2.16}$$

where n is the number of nodes belonging to the network. Finally, the APPL value corresponds to the average among every m-th PPL values as Eq. 2.17 shows.

$$\mathrm{APPL}_l = \frac{1}{12}\sum_{m=1}^{12} \mathrm{PPL}_{\mathrm{l,m}}, \mathrm{APPL}_d = \frac{1}{12}\sum_{m=1}^{12} \mathrm{PPL}_{\mathrm{d,m}} \tag{2.17}$$

To achieve a fair comparison, the average PPL values among different routing algorithms for the same network, must be obtained from the same couple of nodes set. Since not all algorithms are able to select a route pair for all couple of nodes,

to calculate PPL, we only consider the couple of nodes in the network where all the routing algorithms were able to establish a route pair.

3 Results

This chapter presents a series of experiments that were carried out to evaluate the performance of the proposed route pair selection algorithms. The results presented below reflect the advantages of using empirical data as a basis for establishing the risk of the links belonging to the network.

3.0.1 Route pair selection

To calculate the path pair between the nodes in each network, we first generate the threat and risk model from the geo-located failure records as discussed in Sec. 2.3. Twelve scenarios were analyzed, representing each month of the year, in which the number of fire failures among the three topologies was recorded. In Fig. 3.1 percentage distribution of fires by month can be seen between 2001 and 2018, including both years. The month with the lowest and the highest amount of fire reported corresponds to December and August respectively. Also, we can observe between May and November there are more fires detected, which belongs to the fire season in the northern hemisphere. In addition, the results are in agreement with [7], where the months of March and April, show an increase in fires caused by humans. On the other side, between the months of June and September, a strong component caused by lightning strikes. For each month, we calculate the threat and the risk model, for every network topology in the range of years between 2001 and 2018, as detailed in Sec. 2.3. From the risk model, we obtained the number of safe links per month. Fig. 3.2 shows the percentage of safe links, normalized by the number of links by each network. Here, we can obtain the lowest safe links value within the year, between the months of July and October, which is consistent with the high fire season. On the other hand, between December and January, the highest number of safe links is achieved. Despite this, there is no clear correlation between the number of fires and the number of secure links that would allow this information to be generalized to any network topology.

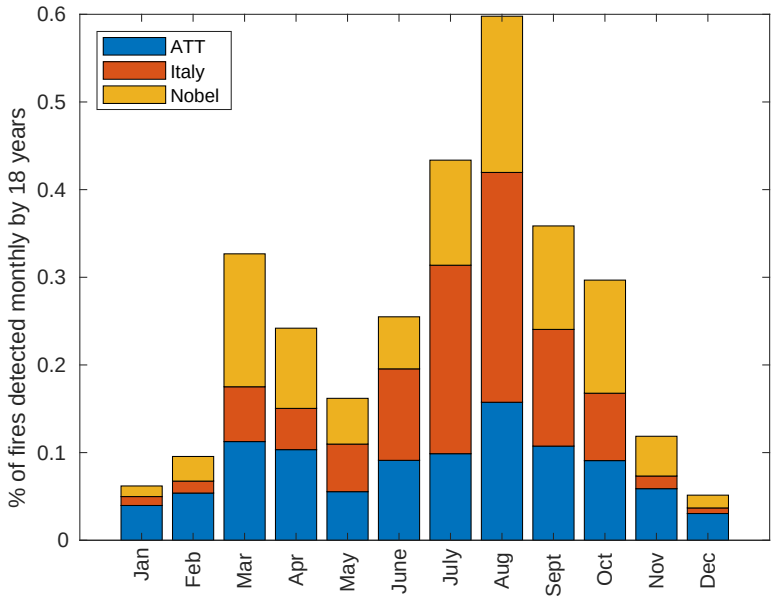


Figure 3.1: Percentage of totally fire sources monthly registered between years 2001 and 2018 by MODIS.

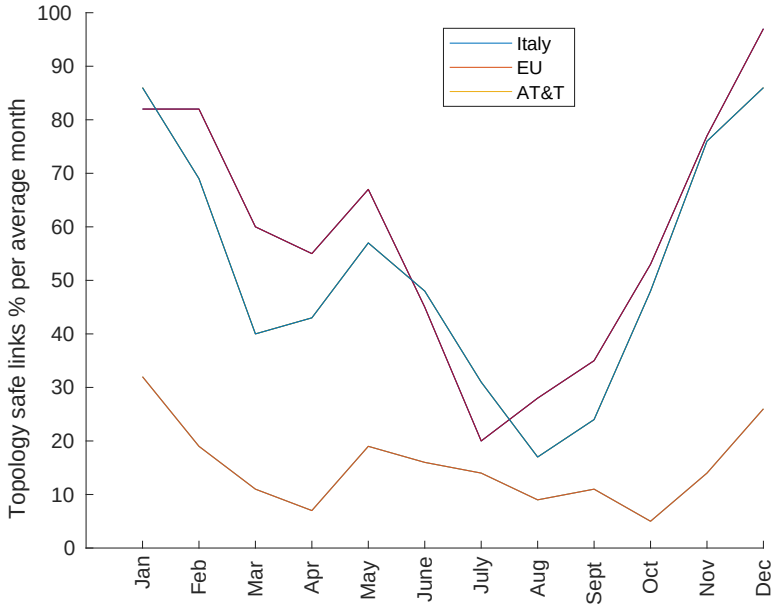
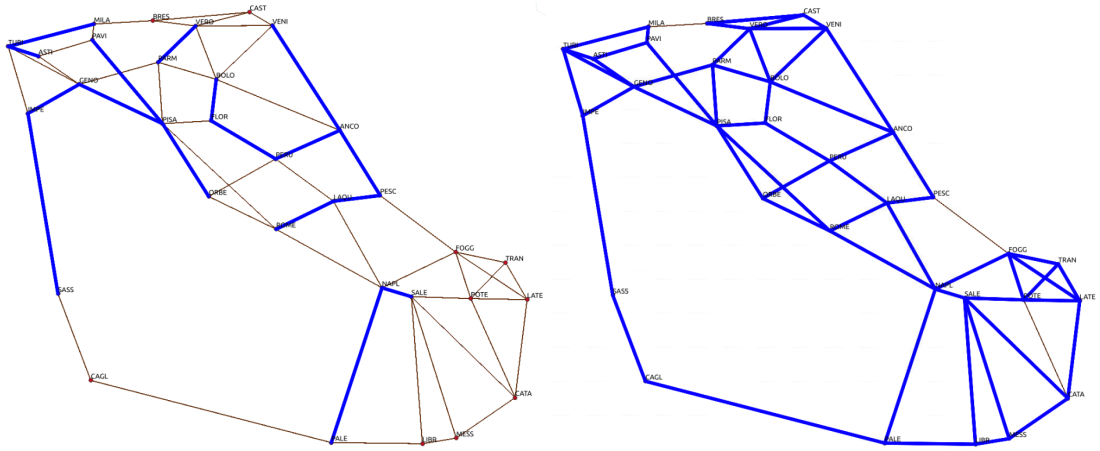


Figure 3.2: Network topology safe links % per month between 2001 and 2018.

Although Italy topology has smaller links than Nobel topology, there is no clear difference between the safe links detected. This shows that links are affected by threats according to their geographical location and not according to their

length, even though their length may lead to an increased probability of failure.

Fig. 3.3 shows the safe (blue) links in the Italy topology, where 28% and 97% of safe links are obtained for the month of August and December, respectively. Although December is the month with the highest number of safe links, it is not free from failures. On the other hand, in August, there is also a presence of safe links. The above answers our research questions, where it is obtained that the probability distribution of failure varies according to the season and it is feasible that exist threat-free areas where safe links are located. Although many links are not safe year-round, it is still feasible to use them seasonally.



(a) (b)

Figure 3.3: Safe links for Italy detected between the years 2001-2018 for the months with the most and least threats, (a) August and (b) December respectively.

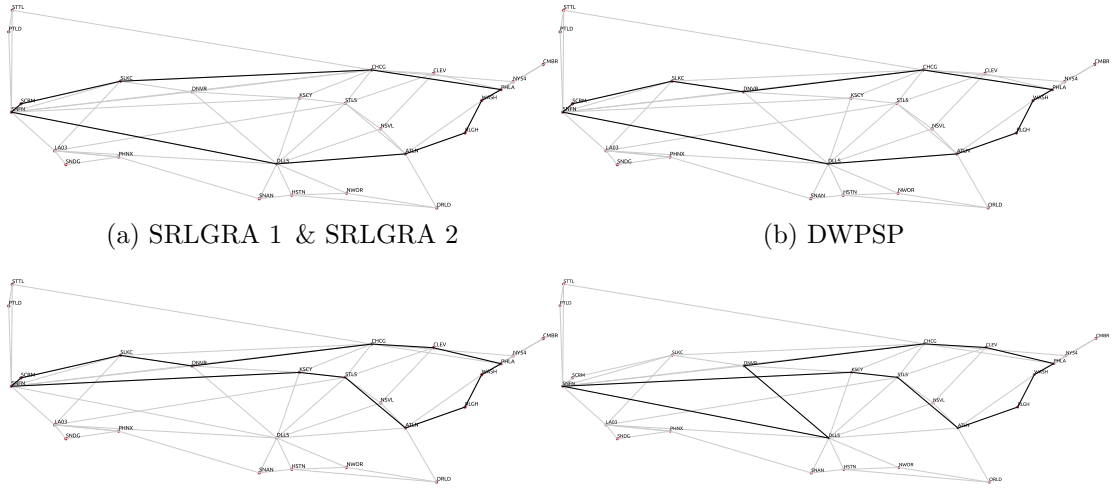
To calculate the largest failure size for each hazard station, we use the technique proposed in Sec. 2.4.3. Table 3.1 represents the largest diameter calculated from each ν. Here, we can obtain that although the largest diameters are obtained from the month with the highest frequency of fires (August), this does not allow us to conclude that the diameter is directly proportional to the number of fires detected. In cases such as the AT&T network, it is observed that in May, the smallest circle diameter is greater than those obtained in months with fewer fires, such as March.

Once we have the probabilities of failure for each link and the failure sizes for each month, we proceeded to calculate the route pair between all couple of nodes for each network topology. As in [44], we ignored the connectivity between two points if a backup path between the two points was not feasible. This triggers the SR value of the network to decrease and an ATTR equal to one cannot be achieved. Despite using different routing algorithms, it is possible to find, in some

Table 3.1: Maximum monthly fire cluster diameter registered in Km between 2001 and 2018.

Topology	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Italy	5	11	9	7	7	9	15	15	9	11	7	5
Nobel	9	11	11	9	11	23	25	37	13	11	9	7
AT&T	15	15	21	31	55	111	121	185	65	45	35	45

cases, equal solutions between them. In order to better compare the algorithms, Fig. 3.4 shows the routes obtained with our algorithms are totally different from those computed with the region disjoint geo-diverse routing algorithms.



(a) SRLGRA 1 & SRLGRA 2

(b) DWPSP

(c) MPJF-DRP

(d) MPnJF-DRP & MPnJF-MDRP & MPJF-MDRP

Figure 3.4: Pair routes between Raleigh and Sacramento for each algorithm.

3.0.2 Average Connectivity, Reliability & Feasibility

Because the AC has very high values in all cases, exceeding 99.999%, it is difficult to compare the values. For this reason, we calculate the average annual disconnection rate, previously defined by AAD in Eq. 2.13.

Table 3.2 shows the AAD values obtained from the route pairs selected by each algorithm associated to every network topology. The algorithm with the smallest AAD for each network topology is MPJF-MDRP, since it minimizes the routes probability of joint failure, being able to reuse safe links. In addition, it is observed that the disconnection of the three SRLG-based algorithms in all cases at least doubles the proposed algorithms. Even, for the smallest coverage network

Table 3.2: AAD values for each network topology x 10^{-10}

Algorithm	Italy	Nobel	ATT	Average
SRLGRA 1	620.68	21.971	7.36	216.67
SRLGRA 2	620.68	21.971	7.36	216.67
DWPSP	647.658	28.779	14.743	230.39
MPnJF-MDRP	36.33	10.096	3.375	16.60
MPnJF-DRP	45.292	9.815	3.319	19.48
MPJF-MDRP	36.31	9.019	3.295	16.21
MPJF-DRP	48.636	11.935	3.431	21.33

topology, the reached ADD values are up to 12 times higher than those obtained by our algorithms.

In order to compare the ATTR results between each algorithm, it is first necessary to know if all of them were able to choose a route pair between a couple of nodes, prior to the existence of failures in the network. In this way, it can be evaluated whether a small ATTR value is the product of a high link failure or the product of an inability to select a route pair between a couple of nodes.

The following Tables 3.3, 3.4, 3.5 show the results obtained using the seven routing algorithms used in this work, composed by three region disjoint geo-diverse routing algorithms and four risk minimization routing algorithms. As can be seen, in the three topologies, our proposed algorithms based on risk minimization are the best prepared to avoid single failures in active routes. Moreover, given that each topology is located in a different geographical area, it is possible to find scenarios where all our algorithms obtain a better performance, as occurs in AT&T.

Table 3.3: Reliability for Italy topology

Algorithm	SR	Single failure	Dual failure	ATTR
SRLGRA 1	1.000 00	0.020 26	0.000 51	0.998 85
SRLGRA 2	0.995 97	0.022 88	0.002 12	0.994 65
DWPSP	0.880 54	0.112 64	0.047 72	0.878 24
MPnJF-MDRP	1.000 00	0.009 68	0.002 21	0.992 20
MPnJF-DRP	0.999 50	0.011 45	0.000 29	0.999 41
MPJF-MDRP	1.000 00	0.008 54	0.001 74	0.993 64
MPJF-DRP	1.000 00	0.010 17	0.000 01	0.999 98

Table 3.4: Reliability for Nobel topology

Algorithm	SR	Single failure	Dual failure	ATTR
SRLGRA 1	1.000 000	0.018 881	0.000 086	0.999 829
SRLGRA 2	0.920 635	0.095 883	0.039 548	0.920 471
DWPSP	0.709 877	0.304 055	0.144 618	0.709 210
MPnJF-MDRP	0.999 780	0.011 300	0.001 036	0.997 912
MPnJF-DRP	0.981 041	0.033 467	0.009 534	0.981 041
MPJF-MDRP	1.000 000	0.010 934	0.000 982	0.998 014
MPJF-DRP	1.000 000	0.013 641	0.000 003	0.999 993

Table 3.5: Reliability for AT&T topology

Algorithm	SR	Single failure	Dual failure	ATTR
SRLGRA 1	0.994 722	0.056 014	0.005 151	0.989 821
SRLGRA 2	0.987 222	0.063 064	0.008 763	0.982 606
DWPSP	0.666 667	0.379 247	0.169 667	0.661 258
MPnJF-MDRP	1.000 000	0.026 279	0.001 215	0.997 562
MPnJF-DRP	1.000 000	0.027 315	0.001 178	0.997 639
MPJF-MDRP	1.000 000	0.025 890	0.001 251	0.997 491
MPJF-DRP	1.000 000	0.026 247	0.001 096	0.997 805

Table 3.6: Average Reliability for geo-routing algorithms.

Algorithm	SR	Single failure	Dual failure	ATTR
SRLGRA 1	0.998 241	0.031 721	0.001 917	0.996 166
SRLGRA 2	0.967 942	0.060 610	0.016 810	0.965 908
DWPSP	0.752 363	0.265 316	0.120 668	0.749 568
MPnJF-MDRP	0.999 927	0.015 755	0.001 489	0.995 890
MPnJF-DRP	0.993 512	0.024 079	0.003 669	0.992 696
MPJF-MDRP	1.000 000	0.015 123	0.001 323	0.996 380
MPJF-DRP	1.000 000	0.016 686	0.000 370	0.999 258

Table 3.6 shows the average results among the three network topologies. As can be seen from the above results, all SRLG-based algorithms were not able to deliver an SR equal to one in all scenarios. Therefore, it is important to have an effective mechanism to solve the failure size selection problem or to use techniques such as ours that do not rely on path separation. This is a relevant point when comparing different route selection algorithms, since without the SR value, an ATTR of less than one could indicate failures in both routes, since they were never established. Furthermore, from the tables it can be extracted that of all the route pairs generated, there is a clear tendency for maximally disjoint routes to have fewer single failed links, despite having a lower ATTR. In cases in which the time to fix each link is very long and the costs associated to repairing are high, an excellent alternative would be to consider maximally disjoint route selection, to minimize the costs associated with repairing active links.

Although there are fewer single failure links in MPJF-MDRP, the number of dual failure is higher than that obtained in MPJF-DRP. This occurs because despite having safe links according to our records, some of them were affected by failures in 2019. If we compare the percentage of single failure obtained by the best region disjoint geo-diverse routing algorithm, we observe that our MPJF-DRP algorithm was able to reduce by 47.4% the link failures belonging to the selected routes. Finally, it is obtained that our MPJF-DRP problem is the one that achieves a higher ATTR in all analyzed topologies. Also reports a lower number of single and dual failures in the links belonging to the network topology, compared to the region disjoint geo-diverse routing algorithms.

3.0.3 Route costs

The comparison of *A*PPL values are exposed in the Table 3.7. Since SRLGRA 1 will always choose the combination of routes with the shortest distance traveled it always get the smallest values. SRLGRA 2 being a non-optimal heuristic based on SRLGRA 1, always obtains the same or a higher value than SRLGRA 1, being the most similar to it. On the other side, DWPSP is the geo-diverse routing algorithm analyzed that obtains highest PPL values, due to the fact that its mechanism of obtaining the shortest route between intermediate points requires that both routes have a node near the center of the imaginary line joining the two nodes (s, d) and two neighbors. This also forces there to be at least five nodes on each route pair, so there can be no direct links between origin and destination. Within the disjoint algorithms analyzed, our proposals obtain higher APPL values. Additionally, the results delivered by the maximally disjoint algorithms obtain on average less costs than the disjoint proposals, reducing up to 12% the APPL value. This is largely because a connection between two neighboring nodes can be achieved using the same link for both routes (in case there are no threats in the hippodrome). For

disjoint routes, it must necessarily have a minimum of three links, which increases the APPL value (one for the primary path and two for backup or vice versa). Also,

Table 3.7: $A\text{PPL}_l$ and $A\text{PPL}_d$ values for each algorithm per topology.

Algorithm	Italy		Nobel		AT&T	
	APPL_l	APPL_d	APPL_l	APPL_d	APPL_l	APPL_d
SRLGRA 1	8.50	1 184.84	8.45	3 148.36	5.86	5 089.93
SRLGRA 2	8.68	1 218.42	8.48	3 175.06	5.86	5 340.86
DWPSP	10.88	1 528.59	10.32	3 973.66	8.33	6 743.22
MPnJF-MDRP	10.00	1 401.02	9.63	3 612.61	8.37	6 919.84
MPnJF-DRP	11.44	1 672.78	10.36	4 034.06	8.53	7 086.47
MPJF-MDRP	9.74	1 439.43	9.39	3 679.48	8.25	6 831.34
MPJF-DRP	10.96	1 640.63	10.08	3 986.44	8.36	6 939.05

the distance traveled is not directly correlated with the number of links used. It is observed for different algorithms where the distance traveled increases, but the number of links used decreases. This is because not all links are the same length. Comparing the monthly PPL_d values within a year serves to highlight the diversity

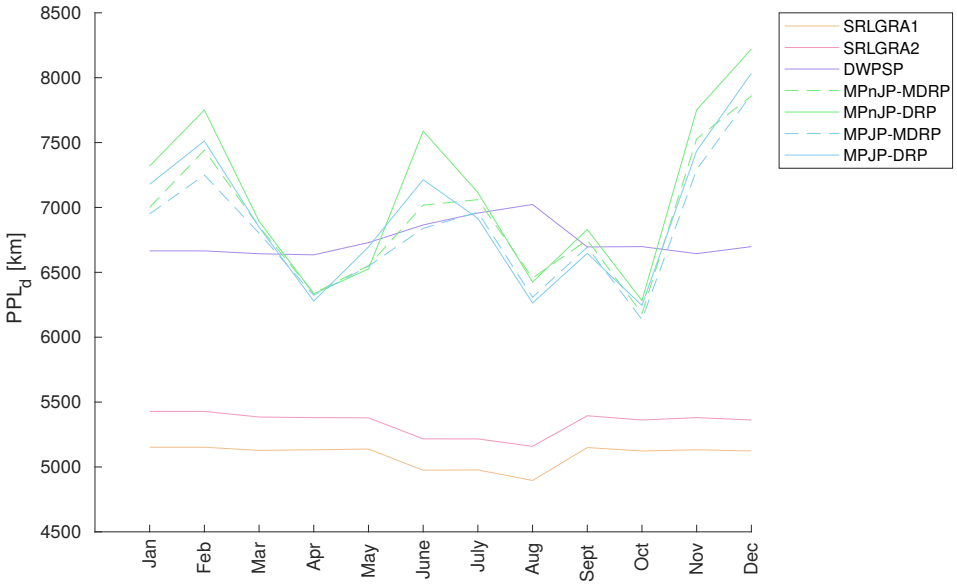


Figure 3.5: PPL_d for every month for AT&T network.

of routes that may exist within a network topology. If the value remains constant, we consider that the routes selected from one month to another are the same. Fig. 3.5 shows the PPL_d calculated monthly for each routing algorithm in the

biggest coveraged network, AT&T. The proposed algorithms, on average, obtain a standard deviation 5 times higher than that obtained by the region disjoint geo-diverse routing algorithms, which evidences a more accurate adjustment to the changes occurring monthly in the network as a result of threats.

4 Conclusions and Future Work

In this work, it is proposed to improve network reliability through the idea of selecting a route pair minimizing the risk on both routes, based on real geo-located failure records. The methodology presented comprises a series of steps to bring geotemporal knowledge of threats to the route pair selection method: a threat model using the geographic and temporal properties of failures to represent the probability of failure distribution in an interest area; a risk model that assing a probability of failure to each link, and the algorithms for a route pair selection between a couple of nodes based on minimal risk. This work is not only a contribution to the theory of network reliability through threats. It also allows the network operator, in addition to increase the availability of both routes, to achieve greater dynamism in the routes according to the hazard season. In addition, it avoids having to solve the failure size selection problem that delivers realistic failure diameters for a zone of interest, for which we do not evidence techniques to facilitate its selection.

All these advantages greatly facilitate the work of the network operator, since they allow the construction of a reliable network, independent of the network operator expertise to face threats. The threat model presented allows to use empirical data associated to a geographical and temporal dimension. Moreover, by considering only the detected threat area, the proposed model does not modify the probability of failure distribution, which occurs when clustering failures over huge terrains as a simple geometric figure. The main idea exploited by the method presented, is that it facilitates the identification of the regions where the greatest number of failures have been located and at the same time detect the threat-free areas. The solution of the optimization problem allows selecting a route pair with the lowest probability of joint failure between a pair of nodes, as demonstrated in the Sec. 2.4. In addition, by having knowledge of all the failures in the geographical environment where the network is located, it is able to be reliable to more

than one failure region, adapting better to massive disaster scenarios. From the results obtained in the three topologies, the reliability was increased with respect to that achieved by the region disjoint geo-diverse routing algorithms. An average ATTR of more than 99.9% was achieved in the face of real threats. These results allow us to validate that our hypothesis was correct. In addition, with our optimization problems, 100% connectivity by paths was achieved in network topologies tested, while no region disjoint geo-diverse routing algorithms achieved the same. Although there are areas with no previous records of failure events, this does not mean that it is a threat-free region. Climate change and human intervention in the geography make the model much more difficult to predict, as conditions where the network is located are not the same for all the time. The network operator must determine the validity of the empirical data used to calculate failure probabilities. Thus, the greater the certainty of the existence of safe links, the more the ATTR can be increased.

4.1 Future Work

A limitation of this work is the number of empirical threats used to generate the threat model. It is desirable that the threat model can continue to be powered and that its performance can be measured with the most recent years uploaded to the databases. Applying maximum entropy models to hazards would make it possible to avoid relying on the fact that a threat-free zone only corresponds to regions with no failure records. By applying these models, maximally disjoint routes would be expected to improve their performance. It is also hoped to obtain more global threats databases to compare with the observed fire performance.

The threat model can be improved as a multi-layer model, which considers several types of threats at once. Apart from considering threats extracted from databases, it could consider features of the geography that help to reinforce the hypothesis of threat-free zones, like eco-regions. An example of this would be to complement a fire safe area with a hydro graphic zone.

Test with a greater diversity of real-world network topologies, to determine if there are geographical characteristics that help to minimize risks from different threats.

A

ILP definition

The optimization problem defined as ILP is proposed to solve a geo-route pair problem. For each link $(i,j) \in \mathcal{L}$, two variables $x_{ij}, y_{ij} \in \{0,1\}$ are defined. If a link (i,j) is on path $\mathcal{P}_1$, then $x_{ij} = 1$, otherwise $x_{ij} = 0$; if a link (i,j) is on path $\mathcal{P}_2$, then $y_{ij} = 1$, otherwise $y_{ij} = 0$. The distances $d(i,j)$ between the nodes can be calculated beforehand in polynomial time. The 0–1 ILP formulation is as follows:

$$\min \sum_{(i,j)\in\mathcal{L}} \omega(i,j)(x_{i,j} + y_{i,j})$$

subject to

$$\sum_{j\in\mathcal{N}}(x_{ij} - x_{ji}) = \begin{cases} 1, & \text{if } i \equiv s \\ -1, & \text{if } i \equiv d \\ 0, & \text{otherwise} \end{cases} \tag{A.1}$$

$$\sum_{j\in\mathcal{N}}(y_{ij} - y_{ji}) = \begin{cases} 1, & \text{if } i \equiv s \\ -1, & \text{if } i \equiv d \\ 0, & \text{otherwise} \end{cases} \tag{A.2}$$

$$x_{ij} + y_{kl} \leq 1, \text{if}(d(i,k) \leq D, i \notin \mathcal{M}, k \notin \mathcal{M}) or \tag{A.3}$$
$$(d(i,l) \leq D, i \notin \mathcal{M}, l \notin \mathcal{M}) \text{or}$$
$$(d(j,k) \leq D, j \notin \mathcal{M}, k \notin \mathcal{M}) \text{or}$$
$$(d(j,l) \leq D, j \notin \mathcal{M}, l \notin \mathcal{M}), \text{where} \mathcal{M} = \{s,d\}$$

$$x_{sd} + y_{sd} \leq 1. \tag{A.4}$$

The objective function represents the total weight of the region-disjoint paths. The equality conditions (1) and (2) are "conservation rules" and ensure that for all the nodes (different from s and d) in both $\mathcal{P}_1$ and $\mathcal{P}_2$, the number of incoming and outgoing links is the same. For the source node s, there is exactly one outgoing

link for both $\mathcal{P}_1$ and $\mathcal{P}_2$, while for the destination node d there is exactly one incoming link for both $\mathcal{P}_1$ and $\mathcal{P}_2$. Condition (3) gives the region-disjointness constraint, preserving two nodes different from s and d, one in link $(i, j) \in P_1$ and one in link $(k, l) \in \mathcal{P}_2$ to be on a distance at most D. If there is a direct link from s to d, condition (4) states that link will be used by at most one path. Condition (4) is not a sub-case of (3).

Bibliography

[1] Ashraf, M. W., Idrus, S. M., Iqbal, F., Butt, R. A., & Faheem, M. (2018, December). Disaster-resilient optical network survivability: a comprehensive survey. In Photonics (Vol. 5, No. 4, p. 35). Multidisciplinary Digital Publishing Institute.

[2] Kuipers, F. A. (2012). An overview of algorithms for network survivability. International Scholarly Research Notices, 2012.

[3] Xie, A., Wang, X., & Lu, S. (2019). Risk minimization routing against geographically correlated failures. IEEE Access, 7, 62920-62929.

[4] Cheng, Y., Medhi, D., & Sterbenz, J. P. (2015). Geodiverse routing with path delay and skew requirement under area-based challenges. Networks, 66(4), 335-346.

[5] Wang, X. (2011, December). Network recovery and augmentation under geographically correlated region failures. In 2011 IEEE Global Telecommunications Conference-GLOBECOM 2011 (pp. 1-5). IEEE.

[6] Netstandar (2017, September 18). How Do I Determine the Appropriate Distance between My Site and the Recovery Site. [Blog Post]. retrieved from http://www.netstandard.com/far-far-enough-disaster-recovery-site/. Last accessed 18 Feb 2020.

[7] Nagy, R., Fusco, E., Bradley, B., Abatzoglou, J. T., & Balch, J. (2018). Human-related ignitions increase the number of large wildfires across US ecoregions. Fire, 1(1), 4.

[8] Pašić, A., Girão-Silva, R., Vass, B., Gomes, T., & Babarczi, P. FRADIR: A Novel Framework for Disaster Resilience. In 2018 10th International Workshop on Resilient Networks Design and Modeling (RNDM), pp. 1–7. IEEE, Norway (2018).

[9] Izaddoost, A., & Heydari, S. S. (2014). Enhancing network service survivability in large-scale failure scenarios. Journal of Communications and Networks, 16(5), 534-547.

[10] Izaddoost, A., & Heydari, S. S. (2017). Risk-adaptive strategic network protection in disaster scenarios. Journal of Communications and Networks, 19(5), 509-520.

[11] Rahnamay-Naeini, M., Pezoa, J. E., Azar, G., Ghani, N., & Hayat, M. M. Modeling stochastic correlated failures and their effects on network reliability. In Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN), pp. 1–6. IEEE, Hawaii (2011).

[12] J. Tapolcai, L. Rónyai, B. Vass and L. Gyimóthi, "List of shared risk link groups representing regional failures with limited size," IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, 2017, pp. 1-9, doi: 10.1109/INFOCOM.2017.8057040.

[13] Girão-Silva, R., Nedic, B., Gunkel, M., & Gomes, T. (2020). Shared Risk Link Group disjointness and geodiverse routing: A trade-off between benefit and practical effort. Networks, 75(4), 374-391.

[14] A. de Sousa and D. Santos, "The Minimum Cost D-Geodiverse Anycast Routing with Optimal Selection of Anycast Nodes," 2019 15th International Conference on the Design of Reliable Communication Networks (DRCN), 2019, pp. 21-28, doi: 10.1109/DRCN.2019.8713729.

[15] Corrected Earthquake Data of Turkey in 1910-2017. https://www.kaggle.com/caganseval/earthquake. Last accessed 30 Mar 2021.

[16] Mueller, C. S. (2019). Earthquake catalogs for the USGS national seismic hazard maps. Seismological Research Letters, 90(1), 251-261.

[17] Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... & Townshend, J. (2013). High-resolution global maps of 21st-century forest cover change. science, 342(6160), 850-853. http://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.7.html. Last accessed 30 Jul 2021.

[18] National Centers for Environmental Information. https://www.ncei.noaa.gov/pub/. Last accessed 30 Mar 2021.

[19] Short, K. C. (2021). Spatial wildfire occurrence data for the United States, 1992-2018 [FPA_FOD_20210617].

[20] Anakhov, P., Zhebka, V., Grynkevych, G., & Makarenko, A. (2020). Protection of Telecommunication Network From Natural Hazards of Global Warming. Eastern-European Journal of Enterprise Technologies, 3(10), 105.

[21] AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., ... & Sadegh, M. (2020). Climate extremes and compound hazards in a warming world. Annual Review of Earth and Planetary Sciences, 48, 519-548.

[22] Cetinkaya, E. K., & Sterbenz, J. P. (2013, March). A taxonomy of network challenges. In 2013 9th International Conference on the Design of Reliable Communication Networks (DRCN) (pp. 322-330). IEEE.

[23] Comfort, L. K., & Haase, T. W. (2006). Communication, coherence, and collective action: The impact of Hurricane Katrina on communications infrastructure. Public Works management & policy, 10(4), 328-343.

[24] Kwasinski, A. (2013, February). Lessons from field damage assessments about communication networks power supply and infrastructure performance during natural disasters with a focus on Hurricane Sandy. In FCC Workshop Network Resiliency. Brooklyn, New York, NY, USA.

[25] Heidemann, J., Quan, L., & Pradkin, Y. (2012). A preliminary analysis of network outages during hurricane sandy. University of Southern California, Information Sciences Institute.

[26] Kwasinski, A., Weaver, W. W., Chapman, P. L., & Krein, P. T. (2009). Telecommunications power plant damage assessment for hurricane katrina–site survey and follow-up results. IEEE Systems Journal, 3(3), 277-287.

[27] O'Reilly, G., Jrad, A., Nagarajan, R., Brown, T., & Conrad, S. (2006, November). Critical infrastructure analysis of telecom for natural disasters. In Networks 2006. 12th International Telecommunications Network Strategy and Planning Symposium (pp. 1-6). IEEE.

[28] Kitamura, Y., Lee, Y., Sakiyama, R., & Okamura, K. (2007). Experience with restoration of asia pacific network failures from taiwan earthquake. IEICE transactions on communications, 90(11), 3095-3103.

[29] Urushidani, S., Aoki, M., Fukuda, K., Abe, S., Nakamura, M., Koibuchi, M., ... & Yamada, S. (2014). Highly available network design and resource management of SINET4. Telecommunication Systems, 56(1), 33-47.

[30] Lomnitz, C. (2004). Major earthquakes of Chile: a historical survey, 1535-1960. Seismological Research Letters, 75(3), 368-378.

[31] Onemi. 14 incendios activos y corte de fibra. 10 de Febrero 2020. https://www.onemi.gov.cl/alerta/resumen-nacional-de-incendios-forestales-8/#collapse36. Last accessed 30 Mar 2021.

[32] Onemi. 45 incendios activos y corte de doble fibra. 11 de Marzo 2015. https://www.onemi.gov.cl/alerta/resumen-nacional-de-incendios-forestales/#collapse78. Last accessed 30 Mar 2021.

[33] Broadband Consortium. Telecommunications Outage Report: Northern California Firestorm 2017. Report Published: April 2018. http://www.mendocinobroadband.org/wp-content/uploads/1.-NBNCBC-Telecommunications-Outage-Report-2017-Firestorm.pdf. Last accessed 30 Mar 2021.

[34] Tiribelli, F., Morales, J. M., Gowda, J. H., Mermoz, M., & Kitzberger, T. (2019). Non-additive effects of alternative stable states on landscape flammability in NW Patagonia: fire history and simulation modelling evidence. International journal of wildland fire, 28(2), 149-159.

[35] Alkhatib, A. A. (2014). A review on forest fire detection techniques. International Journal of Distributed Sensor Networks, 10(3), 597368.

[36] Justice, C. O., Giglio, L., Korontzi, S., Owens, J., Morisette, J. T., Roy, D., ... & Kaufman, Y. (2002). The MODIS fire products. Remote sensing of Environment, 83(1-2), 244-262.

[37] Trajanovski, S., Kuipers, F. A., Ilić, A., Crowcroft, J., & Van Mieghem, P. (2014). Finding critical regions and region-disjoint paths in a network. IEEE/ACM Transactions on Networking, 23(3), 908-921.

[38] Neumayer, S., Zussman, G., Cohen, R., & Modiano, E. (2011). Assessing the vulnerability of the fiber infrastructure to disasters. IEEE/ACM Transactions on Networking, 19(6), 1610-1623.

[39] Mukherjee, B., Habib, M. F., & Dikbiyik, F. (2014). Network adaptability from disaster disruptions and cascading failures. IEEE Communications Magazine, 52(5), 230-238.

[40] Pašić, A., Girão-Silva, R., Mogyorósi, F., Vass, B., Gomes, T., Babarczi, P., ... & Rak, J. (2021). eFRADIR: An enhanced framework for disaster resilience. IEEE Access, 9, 13125-13148.

[41] Savas, S. S., Habib, M. F., Tornatore, M., Dikbiyik, F., & Mukherjee, B. (2014). Network adaptability to disaster disruptions by exploiting degraded-service tolerance. IEEE Communications Magazine, 52(12), 58-65.

[42] Oostenbrink, J., & Kuipers, F. (2019, May). The risk of successive disasters: A blow-by-blow network vulnerability analysis. In 2019 IFIP Networking Conference (IFIP Networking) (pp. 1-9). IEEE.

[43] Agarwal, P. K., Efrat, A., Ganjugunte, S. K., Hay, D., Sankararaman, S., & Zussman, G. The resilience of WDM networks to probabilistic geographical failures. IEEE/ACM Transactions on Networking, **21**(5), 1525–1538 (2013)

[44] Gour, R., Kong, J., Ishigaki, G., Yousefpour, A., Hong, S., & Jue, J. P. (2018). Finding survivable routes in multi-domain optical networks with geographically correlated failures. Journal of Optical Communications and Networking, 10(8), C39-C49.

[45] A. Pašić et al., "FRADIR-II: An Improved Framework for Disaster Resilience," 2019 11th International Workshop on Resilient Networks Design and Modeling (RNDM), 2019, pp. 1-7, doi: 10.1109/RNDM48015.2019.8949142.

[46] Kabir, M. N., Rahman, M. A., Azad, S., Azim, M. M. A., & Bhuiyan, M. Z. A. (2018). A connection probability model for communications networks under regional failures. International Journal of Critical Infrastructure Protection, 20, 16-25.

[47] B. Vass et al., "Probabilistic Shared Risk Link Groups Modeling Correlated Resource Failures Caused by Disasters," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 9, pp. 2672-2687, Sept. 2021, doi: 10.1109/JSAC.2021.3064652.

[48] Sharma, V. K., & Rawat, P. V. S. (2013). Post-disaster slope stability evaluation of catastrophic events in Uttarakhand. Indian J Geosci, 67(3-4), 337-346.

[49] California's Biggest Wildfire. https://www.nytimes.com/2019/06/11/us/california-ranch-wildfire-wasp-nest.html. Last accessed 30 Mar 2021.

[50] Europe wildfires out of control after continental europe is affected with 45.9 degree celsius record. https://bit.ly/3mHDvBX. Last accessed 30 Mar 2021.

[51] Frankel, A. (1995). Simulating strong motions of large earthquakes using recordings of small earthquakes: the Loma Prieta mainshock as a test case. Bulletin of the Seismological Society of America, 85(4), 1144-1160.

[52] European-Mediterranean Seismological Centre; editing status 2021-09-03; re3data.org - Registry of Research Data Repositories. http://doi.org/10.17616/R3N93X last accessed: 2021-09-26.

[53] Carlson, A. R., Sebasky, M. E., Peters, M. P., & Radeloff, V. C. (2021). The importance of small fires for wildfire hazard in urbanised landscapes of the northeastern US. International Journal of Wildland Fire, 30(5), 307-321.

[54] Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., ... & Saleem, M. (2017). An ecoregion-based approach to protecting half the terrestrial realm. BioScience, 67(6), 534-545.

[55] Stanković, I., Žeželj, M., Smiljanić, J., & Belić, A. (2014). Modelling of Disaster Spreading Dynamics. In High-Performance Computing Infrastructure for South East Europe's Research Communities (pp. 31-42). Springer, Cham.

[56] Malamud, B. D., Millington, J. D., & Perry, G. L. (2005). Characterizing wildfire regimes in the United States. Proceedings of the National Academy of Sciences, 102(13), 4694-4699.

[57] Dikbiyik, F., Tornatore, M., & Mukherjee, B. (2014). Minimizing the risk from disaster failures in optical backbone networks. Journal of Lightwave Technology, 32(18), 3175-3183.

[58] Zhang, J., Modiano, E., & Hay, D. (2017). Enhancing network robustness via shielding. IEEE/ACM Transactions on Networking, 25(4), 2209-2222.

[59] de Sousa, A., Santos, D., & Monteiro, P. (2017, March). Determination of the minimum cost pair of D-geodiverse paths. In DRCN 2017-Design of Reliable Communication Networks; 13th International Conference (pp. 1-8). VDE.

[60] Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. Numerische mathematik, 1(1), 269-271.

[61] Cheng, Y., Gardner, M. T., Li, J., May, R., Medhi, D., & Sterbenz, J. P. (2015). Analysing geopath diversity and improving routing performance in optical networks. Computer Networks, 82, 50-67.

[62] K. Foerster, A. Kamisiński, Y. Pignolet, S. Schmid and G. Tredan, "Improved Fast Rerouting Using Postprocessing," 2019 38th Symposium on Reliable Distributed Systems (SRDS), 2019, pp. 173-17309, doi: 10.1109/SRDS47363.2019.00028.

[63] Prieto, Y., Pezoa, J. E., Boettcher, N., & Sobarzo, S. K. (2017, October). Increasing network reliability to correlated failures through optimal multiculture design. In 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) (pp. 1-6). IEEE.

[64] Boettcher, N., Prieto, Y., Restrepo, S., & Pezoa, J. (2018). Resilient multiculture network design in the presence of exploit-triggered correlated failures. IEEE Latin America Transactions, 16(9), 2336-2344.

[65] Boettcher, N. A., Prieto, Y., & Pezoa, J. E. (2018, September). Micro Failure Region Models Inducing Massive Correlated Failures on Networks Topologies. In International Conference on Information Technology in Disaster Risk Reduction (pp. 130-141). Springer, Cham.

[66] Boettcher, Nicolás A., Yasmany Prieto, and Jorge E. Pezoa. Maximizing reliability of data networks to face multiple failures via optimal route selection. IEEE Transactions on Network and Service Management, Manuscript ready for submit for publication.

[67] Neumayer, S., & Modiano, E. (2016). Network reliability under geographically correlated line and disk failure models. Computer Networks, 94, 14-28.

[68] Lee, H. W., Modiano, E., & Lee, K. (2010). Diverse routing in networks with probabilistic failures. IEEE/ACM Transactions on networking, 18(6), 1895-1907.

[69] Gurobi Optimizer version 8.1.0. Gurobi Optimization, Inc., October 2018.

[70] Drager, L. D., Lee, J. M., & Martin, C. F. (2007). On the geometry of the smallest circle enclosing a finite set of points. Journal of the Franklin Institute, 344(7), 929-940.

[71] Earth Data NASA. FIRMS File Download. https://firms.modaps.eosdis.nasa.gov/download/. Last accessed 30 Mar 2021.

[72] Earth Data NASA. FIRMS Frequently Asked Questions. https://earthdata.nasa.gov/faq/firms-faq. Last accessed 30 Mar 2021.

[73] Cohen, J. D. (2000). What is the wildland fire threat to homes?. Thompson Memorial Lecture, School of Forestry, Northern Arizona University, Flagstaff, AZ, 10 April 2000.

[74] Ashraf, M. W., Idrus, S. M., Iqbal, F., & Butt, R. A. (2018). On spatially disjoint lightpaths in optical networks. Photonic Network Communications, 36(1), 11-25.

[75] The Internet Topology Zoo Homepage. http://www.topology-zoo.org. Last accessed 29 Aug 2020.

[76] Orlowski, S., Wessäly, R., Pióro, M., & Tomaszewski, A. (2010). SNDlib 1.0—Survivable network design library. Networks: An International Journal, 55(3), 276-286.

[77] Csiszar, I., Denis, L., Giglio, L., Justice, C. O., & Hewson, J. (2005). Global fire activity from two years of MODIS data. International Journal of Wildland Fire, 14(2), 117-130.

[78] Gomes, T., Santos, D., Girão-Silva, R., Martins, L., Nedic, B., Gunkel, M., Vass, B., Tapolcai, J., & Rak, J. (2020). Disaster-Resilient Routing Schemes for Regional Failures. 483-506.

[79] Hayashi, Y., & Matsukubo, J. (2006). Geographical effects on the path length and the robustness in complex networks. Physical Review E, 73(6), 066113.