How to Model and Enumerate Geographically Correlated Failure Events in Communication Networks

Vass, Balazs; Tapolcai, János; Hay, David; Oostenbrink, Jorik; Kuipers, Fernando

DOI
10.1007/978-3-030-44685-7_4

Publication date
2020

Document Version
Final published version

Published in
Guide to Disaster-Resilient Communication Networks

Chapter 4
How to Model and Enumerate Geographically Correlated Failure Events in Communication Networks

Balázs Vass, János Tapolcai, David Hay, Jorik Oostenbrink, and Fernando Kuipers

Abstract Several works shed light on the vulnerability of networks against regional failures, which are failures of multiple pieces of equipment in a geographical region as a result of a natural or human-made disaster. This chapter overviews how this information can be added to the existing network protocols through defining *shared risk link groups* (*SRLGs*) and *probabilistic SRLGs* (*PSRLGs*). The output of this chapter can be the input of later chapters to design and operate the networks to enhance the preparedness against disasters and regional failures in general. In particular, we are focusing on the state-of-the-art algorithmic approaches for generating lists of (P)SRLGs of the communication networks protecting different sets of disasters.

4.1 Introduction

The Internet is a critical infrastructure. Due to the importance of telecommunication services, improving the preparedness of networks to regional failures is becoming a key issue [4, 5, 8–10, 12, 13, 20–22, 27, 41]. The majority of severe network outages happen because of a disaster (such as an earthquake, hurricane, tsunami,

B. Vass (✉) · J. Tapolcai
High-Speed Networks Laboratory (HSNLab), MTA-BME Future Internet Research Group, Budapest University of Technology and Economics, Budapest, Hungary
e-mail: vb@tmit.bme.hu

J. Tapolcai
e-mail: tapolcai@tmit.bme.hu

D. Hay
School of Engineering and Computer Science, Hebrew University, Jerusalem, Israel
e-mail: dhay@cs.huji.ac.il

J. Oostenbrink · F. Kuipers
Delft University of Technology, Delft, The Netherlands
e-mail: J.Oostenbrink@tudelft.nl

F. Kuipers
e-mail: F.A.Kuipers@tudelft.nl

J. Rak and D. Hutchison (eds.), *Guide to Disaster-Resilient Communication Networks*, Computer Communications and Networks,
https://doi.org/10.1007/978-3-030-44685-7_4

tornado) taking down a lot of (or all) equipment in a given geographical area. Such failures are called *regional failures*. Many studies have touched on the problem of how to prepare networks to survive regional failures, where the first solutions have assumed that fibres in the same duct or within 50 km of every network node fail simultaneously (namely, in a single regional failure) [18, 43]. These solutions were further improved by examining the historical data of different type of disasters (e.g. seismic hazard maps for earthquakes) and identifying the hotspots of the disasters [5, 10, 12, 20, 21, 27]. The weak point of these approaches is that, during network equipment deployment, many of the risks are considered and compensated (e.g. an earthquake-proof infrastructure in areas with larger seismic intensity), implying that the historical data does not represent the current deployments and therefore, not the current risks. Thus, it may be more realistic to assume that any physically close-by equipment has a higher chance to fail simultaneously. More recent studies are purely devoted to this particular problem and adapt combinatorial geometric-based approaches to capture all of the regional failures and represent them in a compact way [2, 7, 22, 30, 31, 34, 37], where the major challenge is that regional failures can have arbitrary locations, shapes, sizes, effects, etc. This chapter is devoted to overview of the state of the art and suggests unified definitions, notions and terminology.

The output of the approaches discussed in this chapter can serve as the input of the network design and management tools. Currently, network recovery mechanisms are implemented to protect a small set of pre-defined failure scenarios. Each recovery plan corresponds to the failure of some equipment. Informally speaking, when a link fails, the network has a ready-to-use plan on how to recover itself. Technically, a set of so-called shared risk link groups (SRLGs) are defined by the network operators, where each SRLG is a set of links whose joint failure recovery mechanism should be prepared for. In this chapter, we are purely focusing on how to define SRLGs that cover all types of disasters, as recovery mechanisms for a specific SRLG are discussed in later chapters. We will also address refinements of the SRLG model defined in the next section.

4.2 Notions Related to Vulnerable Regions

When several network elements may fail together as a result of a single event, they are often characterized by *shared risk group*s (SRGs). Each SRG has a corresponding failure event (or events); when such an event occurs, all elements in the SRG fail together. Specifically, the communication network is modelled as a graph $G = \langle V, E \rangle$, whose vertices are routers, PoPs,[1] optical cross-connects (OXC) and users, while the edges are communication links (mostly optical fibres). SRGs are then defined as subgraphs $\langle V', E' \rangle$, where $V' \subseteq V$ and $E' \subseteq E'$.

[1]A point of presence (PoP) is an artificial demarcation point or interface point between communicating entities.

In many cases, it is sufficient to consider only *links* in SRGs, and in this case, these groups are called *shared risk link groups* (SRLGs). For example, an SRLG may contain one edge (to capture a single-link failure) or all edges that touch one vertex (to capture a single-node failure). SRLGs may be more complex and represent simultaneous failures of multiple network elements. In particular, in this chapter, we focus on geographically correlated failures in which links within a specific region fail together.

A set $\mathcal{S}$ of SRLGs can be used as an input to network design and network recovery/protection mechanisms to ensure these mechanisms withstand the failures corresponding to these SRLGs. For example, to ensure connectivity between a specific pair of nodes, protection mechanisms may construct two edge-disjoint paths when $\mathcal{S} = \{\{e\}|e \in E\}$, two node-disjoint paths when $\mathcal{S} = \{\{(u, v) \in E\}|v \in V\}$ or two paths that do not traverse the same geographical region when $\mathcal{S}$ corresponds to all sets of links that are physically close by.

The following definition captures the notion of SRLG introduced by regional failures, such as a natural disaster or an attack. For ease of presentation, we will call these failure events *disasters*, regardless of their cause.

Definition 4.1 *(SRLG)* A set of links $S \subseteq E$ is an SRLG if we may assume there will be a disaster that can cause all edges in S to fail together. If the disaster can be characterized by a bounded geographical area in the two-dimensional plane $D \subset \mathbb{R}^2$, and S is the set of edges that intersect with D, then S is called the *regional SRLG that represents* D and is denoted by $S = \text{SRLG}(D)$. If D is a circular disc, we call $\text{SRLG}(D)$ a *circular* SRLG.

Circular SRLGs, which are the most common in the literature, can also be characterized by the failure epicentre $p \in \mathbb{R}^2$ and the failure radius $r \in \mathbb{R}$. In this case, $S = \{e \in E|d(e, p) \leq r\}$, where $d(e, p)$ is the Euclidean distance between edge e and point p.

The likelihood of a disaster occurring is not the same at all points of the plane. For example, earthquakes are more likely to occur in rupture zones than in other places, and regions with lower altitude are more likely to suffer from floods. Thus, the probability of an event to occur is important. This probability is sometimes given in the form of an *epicentre distribution map*, which gives for each location $p \in \mathbb{R}^2$, the probability that a disaster happened with epicentre p. Moreover, the size (or radius) of the disaster can also be a random variable (e.g. earthquakes with a larger magnitude are less likely to happen than earthquakes with smaller magnitude, even if their epicentres are the same). Thus, it is customary to consider a set $\mathcal{D}$ of disasters $D \subseteq \mathbb{R}^2$ (that can be of infinite size) and attach a probabilistic measure to this set. For simplicity, let us assume that $\mathcal{D}$ is finite, and let $p_D = \Pr[\text{disaster } D \in \mathcal{D} \text{ occurs}]$.[2] We note that an SRLG S can represent more than one disaster in $\mathcal{D}$; thus, we denote by the $\text{support}(S) = \{D \in \mathcal{D}|S = \text{SRLG}(D)\}$.

[2]For infinite sets, one can use discretization and consider only finite number of sets, albeit with a small error.

Definitions 4.2–4.5 capture the probabilistic nature of disasters and their effect on SRLGs. An FP (Definition 4.2) tells the probability that the failed link set will be *exactly* S, while a CFP (Definition 4.3) tells the probability that *at least* S will fail:

Definition 4.2 *(FP)* Given a set $\mathcal{D}$ of disasters $D \subseteq \mathbb{R}^2$, a probability p_D for each disaster in $\mathcal{D}$ and a link set $S \subseteq E$, the *Link Failure State Probability (FP)* of S is $\text{FP}(S) = \sum_{D \in \text{support}(S)} p_D$. We note that if a disaster in support(S) actually occurs, then all links in S fail (with probability 1).

Definition 4.3 *(CFP)* Given a set $\mathcal{D}$ of disasters $D \subseteq \mathbb{R}^2$, a probability p_D for each disaster in $\mathcal{D}$ and a link set $S \subseteq E$, the *Cumulative Link Failure Probability (CFP)* of S is $\text{CFP}(S) = \sum_{T \supseteq S} \sum_{D \in \text{support}(T)} p_D$. We note that if a disaster in $\bigcup_{T \supseteq S}$ support(T) occurs, then all links in S fail (with probability 1).

In a sense, FPs are like probability density functions (PDFs), while CFPs are like their cumulative distribution functions (CDFs).

Sometimes it is imperative to investigate situations in which disasters do not necessarily cause the failure of the links even if they traverse the disaster area. These events are called *disasters with probabilistic outcome* (Definition 4.4). While these behaviours can be described with lists of FPs or CFPs, *two-stage PSRLGs* (Definition 4.5) offer an alternative way of encoding the effect of the disasters.

Definition 4.4 *(Disaster with probabilistic outcome)* Given a disaster D with probabilistic outcome, each $e \in \text{SRLG}(D)$ is attached a failure probability $p_{e,D}$ which is the probability that link e fails had disaster D occurred (for each $e \notin D$ has $p(e, D) = 0$).

Definition 4.5 *(Two-stage PSRLG)* Given a set $\mathcal{D}$ of disasters $D \subseteq \mathbb{R}^2$ with probabilistic outcome, a probability p_D for each disaster in $\mathcal{D}$ and a link set $S = \{e_1, e_2, \dots e_{|S|}\} \subseteq E$, the *two-stage probabilistic SRLG* of S is $(S, p_S; p_1, \dots, p_{|S|})$, where p_S is the probability that S will be hit by the next disaster D, and p_i is the probability that link e_i is hit by D if it hits S. They can be calculated as $p_S = \sum_{D \in \text{support}(S)} p_D$, and $p_i = \frac{1}{p_S} \sum_{D \in \text{support}(S)} p_D p_{e_i,D}$.

If in case of each two-stage PSRLG S, links being part of S fail with the same probability, S is called a *homogeneous two-stage PSRLG*, or else it is a *heterogeneous two-stage PSRLG*.

Collectively, we call FPs, CFPs and two-stage PSRLGs as *probabilistic SRLGs (PSRLGs)*. Figure 4.1 depicts the connections between these notions.

We can convert a heterogeneous two-stage PSRLG into a list of FPs as follows. For an arbitrary heterogeneous PSRLG $(S, p_S; p_1, \dots, p_{|S|})$, the probability p_P of failing exactly a nonempty set $P \subseteq S$ is $p_S \prod_{e \in P} p_i$. This way one can store every non-empty set $P \subseteq S$ with probability p_P in a list of FPs. In case of a list of heterogeneous two-stage PSRGs, probabilities p_P add up. A list of homogeneous PSRLG can be transformed similarly.

Tables 4.1 and 4.2 give an overview of the works presented in this chapter. Papers offering lists of SRLGs and PSRLGs translate the composed geometric problem of

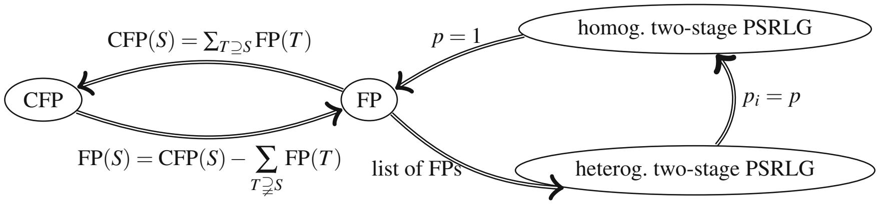


Fig. 4.1 Relation of probabilistic SRLGs (PSRLGs) (two-stage PSRLGs, FPs and CFPs): an FP is a homogeneous two-stage PSRLG with $p = 1$, which is a heterogeneous two-stage PSRLG with $p_i = p$. In addition, a heterogeneous two-stage PSRLG can be represented by a list of FPs, and lists CFPs and FPs are also easily interchangeable by definition

protecting telecommunication networks against regional failures to purely combinatorial and probabilistic problems, respectively.[3]

In the following, we present works showing that the composed geometric problem of protecting telecommunication networks against regional failures translates to combinatorial problems via generating (P)SRLGs. Then, one can use a variety of known tools to handle the translated combinatorial problems.

4.3 Calculating Lists of SRLGs

4.3.1 General Practices for SRLG Enumeration

Prior to presenting concrete algorithms for SRLG enumeration, we discuss the most important issues of the field.

As the size of SRLG list $\mathcal{S}$ determines the run-time and complexity of the mechanisms that use it, an important goal is to keep $\mathcal{S}$ as small as possible. For example, when two sets S_1, S_2 are in $\mathcal{S}$ and $S_1 \subseteq S_2$, it is sufficient to include only S_2 in $\mathcal{S}$; omitting S_1 from $\mathcal{S}$ usually does not affect the outcome of the underlying mechanisms.[4] This is due to the monotonicity (of network design/recovery mechanisms):

Definition 4.6 *(Monotonicity of mechanism)* A mechanism is monotone if for any S_1, S_2 such that $S_1 \subseteq S_2$, the actions the mechanism takes in response to S_1 is a subset of the actions it takes in response to S_2.

Moreover, some works use *over-approximation* to reduce the size of $\mathcal{S}$:

[3]Some papers like [6, 20, 40, 42, 43] are loosely related to the regional (P)SRLG generating problem, however, our goal is presenting the most relevant works in this field.
A list of (P)SRLGs can be used as a pre-computed input for various problems [3, 11, 14, 17].

[4]This is true for communication networks but not for networks in which there is no monotonicity in failures.
When attaching probability to the SRLGs, this no longer holds.

Table 4.1 Papers enumerating regional SRLGs

Paper	Current chapter	Goal	Geometric info.		Assumptions		Algorithms		
			Physical network	Planar/ spherical	Disaster shape	Single disaster	Precise/ approximate	Polynomial	Parametrized
–	Section 4.3.2.1	SRLG list	No	Plane	–	✓	Precise	✓	✗
Vass et al. [37, 38]	Section 4.3.2.2	SRLG list	Poor	Plane	Circular	✓	Precise	✓	✓
Tapolcai et al. [29, 30]	Section 4.3.2.3	SRLG list	Good	Plane	Circular	✓	Precise	✓	✓
–	Section 4.3.2.5	SRLG list	Good	Any	Set of known disasters	✓	Precise	✓	✗
Vass et al. [39]	Sections 4.3.2.4, 4.3.3	SRLG list	Any	Plane+sphere	Bounded by segments+arcs	✓	Precise+ approximate	✓	✓
Iqbal et al. [16]	Section 4.3.4.1	SRLG list	Good	Plane	–	✓	Precise	✗	✗
Neumayer et al. [22]	Section 4.3.4.2	Most vulnerable point	Good	Plane	Circular or line segment	✓	Precise	(✓)	✗
Pašić et al. [25, 26]	Sect. 4.5	SRLG list	Good	Plane	Any	✓	Approximate	✓	✗

While the rest of the papers consider deterministic disaster scenarios, in [25, 26] SRLGs are obtained from PSRLG lists

Table 4.2 Papers enumerating regional PSRLGs

Paper	Current chapter	Goal	Correlated link failures inside the disaster	Natural disaster/attack
Oostenbrink et al. [23]	Section 4.4.1.1	FP list	(✓)	–
Tapolcai et al. [31]	Section 4.4.1.2	CFP list	✓	Natural disaster
Valentini et al. [36]	Section 4.4.1.3	FP list + CFP list	✓	Natural disaster (earthquake)
Agarwal et al. [1, 2]	Section 4.4.2	Most vulnerable point	✗	Attack

Definition 4.7 *(Over-approximation)* $\mathcal{S}'$ over-approximates $\mathcal{S}$ is for every $S \in \mathcal{S}$ there exists $S' \in \mathcal{S}'$ such that $S \subseteq S'$. This relationship is denoted with $\mathcal{S}' \sqsupseteq \mathcal{S}$.

As an over-approximation, instead of including two sets S_1, S_2, one can include a single set $S_1 \cup S_2$ (this is especially appealing if $S_1 \cap S_2$ is of non-negligible size); such over-approximation, however, can degrade the outcome of the underlying mechanisms. For example, if a big over-approximation is plugged into network protection mechanism (e.g. one that computes secondary paths that are SRLG-disjoint from the primary paths), this will cause a performance degradation (namely, longer secondary paths). Thus, one need to keep the *degree of over-approximation* low:

Definition 4.8 *(SRLG of a disaster D with respect to over-approximate set $\mathcal{S}'$)* $SRLG_{\mathcal{S}'}(D)$ is a minimal size set $S' \in \mathcal{S}$ such that $SRLG(D) \subseteq S'$.[5]

Definition 4.9 *(Degree of over-approximation)* SRLG list $\mathcal{S}'$ (α, β)-over-approximates a set of disasters $\mathcal{D}$ if $|\text{SRLG}_{\mathcal{S}'}(D)| \leq \alpha|\text{SRLG}(D)| + \beta$ for all $D \in \mathcal{D}$.

Keeping low 1) the degree of over-approximation and 2) the number of listed SRLGs are two conflicting objectives. As the best practices for SRLG enumeration vary in function of the problem input, we will present a range of algorithms proposed for calculating SRLGs. For every geometric over-approximation of the disasters, one can give very badly behaving input networks (meaning arbitrarily high degrees of over-approximation), but many of these inputs are not realistic (e.g. if two links are very close, probably they share the same duct, etc.). The presented algorithms are conservative both with the number of listed SRLGs and with the degree of over-approximation in case of different classes of realistic inputs.

For regional SRLGs, over-approximation is achieved by taking a larger failure region. The most common practice is to take a simpler shape that completely contains the original failure region, e.g. *circular discs* (Sects. 4.3.2.2–4.3.2.3, 4.3.4.2), or *fixed shape* bounded by segments and arcs (Sects. 4.3.2.4, 4.3.3).

[5] If there are more than one equal-size sets that satisfy the condition, one is chosen arbitrarily.

Dealing with circular SRLGs are in fact over-approximations of regional SRLGs. Notice, for example, that one can over-approximate disasters contained by a circular disc with a certain radius r, with disasters of radius $r' > r$ (namely, assuming all disasters cause larger damage). If such an over-approximation is plugged into network protection mechanism (e.g. one that computes secondary paths that are SRLG-disjoint from the primary paths), this will cause a performance degradation.

Another very common practice is to assume that *in the investigated time period, there will be at most one disaster*. If one can enumerate the set $\mathcal{S}$ of SRLGs of single disasters, it is straightforward to compute SRLGs of multiple disaster events. For example, if two disaster can happen simultaneously, one might look at $\mathcal{S}' = \{S_1 \cup S_2 | S_1, S_2 \in \mathcal{S}\}$. Thus, we will concentrate on enumerating the SRLGs of single disaster events.

Lastly, we note that although many of the presented methods are designed to handle links which are considered as line segments (or geodesics) between their endpoints, these results can be extended to a more general setting, where the links are polygonal chains (or series of geodesics) at the price of a polynomial increase in run-time.[6]

4.3.2 *Precise Polynomial Algorithms Enumerating SRLGs*

4.3.2.1 SRLG Lists Induced by Hop Count

The current best practice to increase the resilience of the networks against disasters is to ensure that the primary and backup paths assigned to a connection are node disjoint. Compared to edge-disjointness, in this way, operators ensure that the distance between the nodes of the primary and backup paths (except at the terminal nodes) are in at least 1-hop distance from each other. The intuitive reasoning is that a link in a backbone network is typically a few hundred kilometres long, while natural disasters are never larger than a few hundred kilometres.

Let M_h denote the set of link sets ensuring a distance of h hops. For each node v, $M_{h=1}$ is containing the set of links incident to v. As an exception, let us list the single link failures in $M_{h=0}$, as this SRLG list ensures link-disjoint routing. For higher values of h, M_h can be defined as follows. To ensure an odd number of hops, for every node v, $M_{h=2k-1}$ contains the edges of a tree of shortest paths to v from the nodes not further from v than k hops. Similarly, for every link $e = \{u, v\}$, $M_{h=2k}$ contains the edges of a tree of shortest paths to e from the nodes not further from u or v than k hops. We can conclude that the number of SRLGs in M_h is low $|M_{h=2k}|$ being $|E|$ and $|M_{h=2k+1}|$ being $|V|$.

Figure 4.2 depicts the average number of links contained by SRLGs in M_h. Clearly, this average is 1 for $h = 0$ and is equal for the average nodal degree for $h = 1$.

[6] Polygonal chains can be dismantled to a set of line segments; the method can be applied, and then, the sets of line segments can be joined.

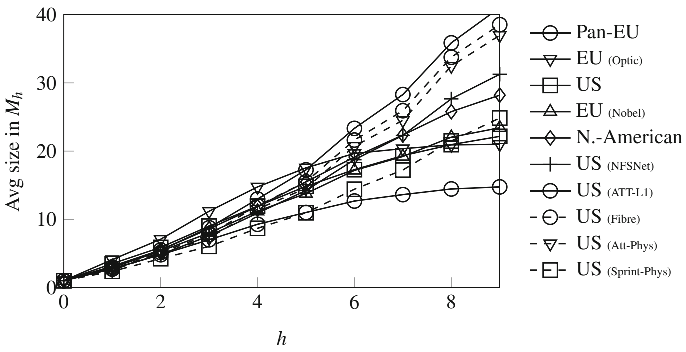


Fig. 4.2 Average number of links contained in the SRLGs of M_h in case of physical backbone topologies of [24]

According to simulation results, for bigger values of h, the average seems to grow slightly superlinearly before the growth slows down to plateau at $|V| - 1$.

Clearly, M_h can be computed in low-polynomial time of n. To generate Fig. 4.2, we generated M_h for $h \in \{0, \ldots, 12\}$ for all the networks on the figure in less than 17 seconds on a commodity laptop, using a code written in Python3, not optimized for speed.

4.3.2.2 SRLG Lists of Disasters Represented by Circular Discs Containing *k* Nodes

As mentioned before, the current best practice is to ensure that the primary and backup paths assigned to a connection are node disjoint. When ensuring a 1-hop distance, the root of the outages is usually because a disaster simultaneously hitting: (1) *close nodes*: when two nodes are placed close to each other; for example, in highly populated areas. (2) *parallel links*: when two links are placed close to each other because of some geographic reasons.

Unfortunately, handling the geometric information with the network topology is not part of the current best practice. Furthermore, the Internet service providers usually hire the links as a service from an independent company, called the physical infrastructure provider, and thus, operators have no information about the route of the links or the physical coordinates of the intermediate routing nodes.

In [38], a limited geometric information failure model is defined, which is based on the following assumptions:

1. The network is a geometric graph $G = \langle V, E \rangle$ embedded in a 2D plane.
2. The exact route of the conduits of the network links are not known but contained by a polygonal region.
3. The shape of the regional failure is assumed to be a circular disc with an arbitrary radius and centre position.

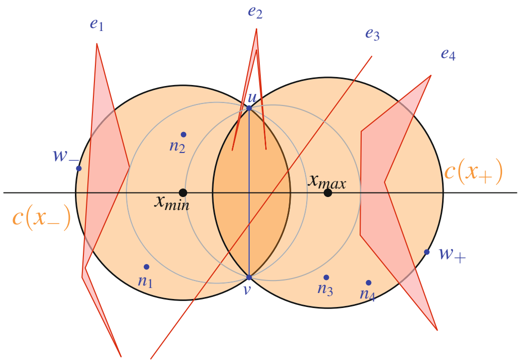


Fig. 4.3 Illustration of an apple with $k = 2$. Apple $A_k^{u,v}$ consists of specific ordered lists of links and nodes which can be hit by a disc from $\mathcal{C}_k^{u,v}$. For more details, please check [38]

4. It focuses on *regional link k-node failures*, which are caused by disasters that hit k nodes for $k \in \{0, |V| - 2\}$.

Paper [38] presents a low-polynomial algorithm for determining the set M_k of maximal regional link k-node failures. The proposed method is based on a set of (computational) geometric considerations. The key observation is that for any element of M_k, there exists a circular disc-shaped disaster having k nodes in the interior which has (1) two nodes on its boundary, or else (2) only one node u on its boundary and having an infinite radius. This allows us to enumerate all possible maximal failures using a sweep surface method as follows.

Let $\{u, v\} \subseteq V$ be two nodes for which the set $\mathcal{C}_k^{u,v}$ of circles which have k nodes in its interior and u and v on its boundary is not empty. These $\{u, v\}$ pairs are part of the set E_k of *k-Delaunay* edges, and their set can be determined in low-polynomial time [28, Thm. 2.4]. In [38], data structure *apple* $A_k^{u,v}$ is defined, which contains ordered lists of links and nodes which can be hit by a circle from $\mathcal{C}_k^{u,v}$. Suppose u and v are positioned as in Fig. 4.3. With the help of $A_k^{u,v}$, one can sweep through circles of $\mathcal{C}_k^{u,v}$ ordered by the abscissas of their centre points allowing to collect the set $M_k^{u,v}$ of maximal hit link sets by discs from $\mathcal{C}_k^{u,v}$. Then, the globally maximal elements of all lists $M_k^{u,v}$ are collected in M_k^2.

In the second case, the set of maximal failures M_1 from M_k for which exist a half-plane going through a node and hitting them can be calculated similarly via turning a half-plane around every node while checking the set of hit links and the number of hit nodes. Finally, M_k can be obtained by collecting the maximal elements of M_k^1 and M_k^2.

The process is sketched in Fig. 4.4. The complexity of the algorithm is low-polynomial and squared in the number of nodes n [38, Thm. 3, Cor. 25]. Besides theoretical upper bounds, simulation results show that the number of maximal failures is approximately $1.2n$ and $2.2n$ for $k = 0$ and $k = 1$, respectively (Fig. 4.5).

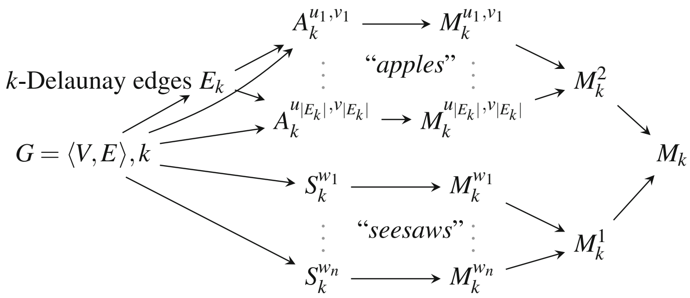
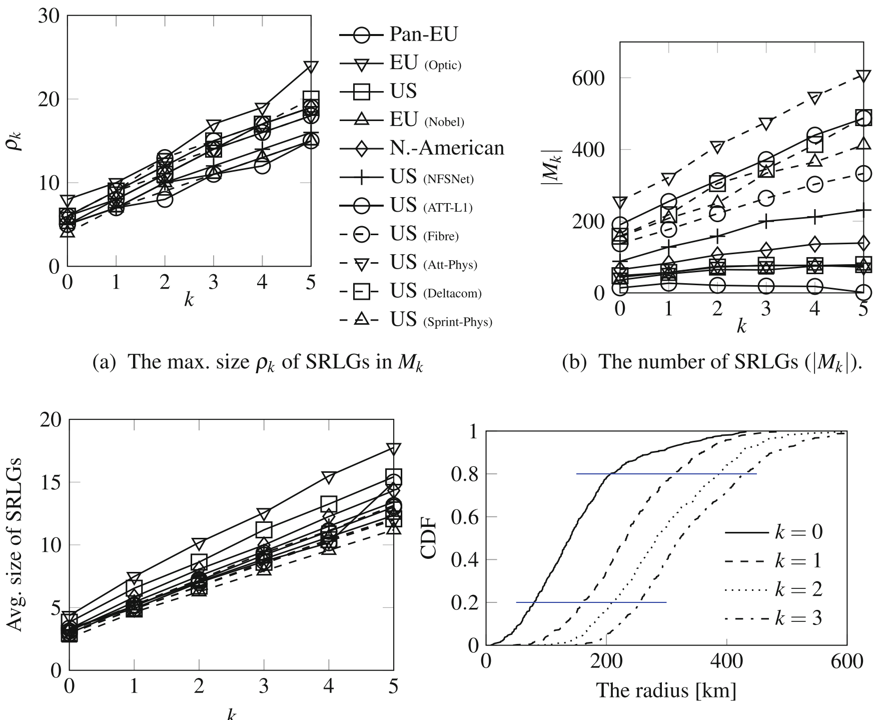


Fig. 4.4 Sketch of algorithm from [38] for enumerating set M_k of maximal link sets which can be hit by a circular disc hitting k nodes



(a) The max. size ρ_k of SRLGs in M_k

(b) The number of SRLGs ($|M_k|$).

(c) The average size of SRLGs in M_k

(d) The cumulative distribution function of the radius of the disk having a given k number of nodes interior.

Fig. 4.5 The edge density, number and size of SRLGs for each network and $k = \{0, \ldots, 5\}$ in case of polygonal chain links

4.3.2.3 Circular SRLG Lists of Disasters with Radius *r*

If the physical positions of the network elements are known, a fast systematic approach to generate the list M_r of maximal SRLGs that represent circular discs of a given radius r is clearly desired.[7]

Paper [30] presents a low-polynomial algorithm for computing M_r when links are considered as line segments (and the network is embedded in the plane). It shows that the number of elements of M_r is linear in the number of nodes in the network n, and its calculation can be done in a squared complexity of n (Theorem 6 of [30]). Simulations indicate that this list has a size of $\approx 1.2n$ in practice (see Fig. 4.7).

To be more precise with the theoretical results, Corollary 4 of [30] tells that the number of SRLGs in M_r is at most proportional to the product of 1) the number of nodes n plus the number of link intersections x and 2) in the cardinality ρ_r of the biggest link set contained. The computing time needed is $O\left((n+x)^2\rho_r^5\right)$ [30, Thm. 6]. We note that x is 0 or a small number, and according to simulation results, ρ_r increases linearly with r, suggesting an $O(n^2r^5)$ run-time for $r > 0$.

Algorithm 4.1 Sketch of algorithm proposed in [30]

```
Require: Graph G = ⟨V, E⟩ embedded in plane, radius r
Ensure: List M_r of maximal SRLGs of disasters being circular discs with radius r
M'_r := ∅
Calculate X := {points of edge crossings}
for w ∈ V ∪ X do
    Determine E_w := {edges not further from w than 3r}
    for e_1, e_2 ∈ E_w do
        Calculate circles c_i described in Fig. 4.6a
    end for
    for e ∈ E_w do
        Calculate circles c_j described in Fig. 4.6b with w as point
        Calculate circles c_k described in Fig. 4.6c
    end for
    Refresh[8] M'_r with link sets hit by circles c_i, c_j, c_k (1 circle at a time)
end for
return M'_r as M_r
```

In the followings, we give an overview of the proposed algorithm (Algorithm 4.1),[8] which relies on a series of geometric considerations. The most important one is Theorem 1 of [30], which leverages that the link sets possibly hit by any of

[7]Reference [18] offers a mistaken heuristic for computing M_r. It claims the disc failures having nodes of the network as their centre point represent the worst case of failures of radius r, which is clearly not the case. Consider, e.g. a network being an equilateral triangle with side length 3, and $r = 1$; here, M_r consists of a single SRLG containing all the 3 links instead of the 3 link-pairs claimed by [18].

[8]Refreshing in Algorithm 4.1 means that M' is the set of maximal failures among which are already checked, and if f is maximal amongst them, it is added to M' and all f's subsets are eliminated from M'; or if f is not maximal inM', nothing happens.

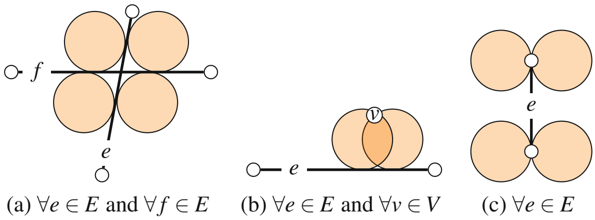


Fig. 4.6 Disc failures examined

the infinite number of possible disaster locations can be determined via checking the effect of a quadratic number of discs on the network edges. In particular, for a positive real r and a non-empty set of edges H which is hit by a circular disc of radius r, there exists a disc c of radius r which hits the edges of H such that at least one of the following holds (see Fig. 4.6 for illustrations): (a) there are two non-parallel links in H such that c intersects both of them in a single point. These two points are different. (b) There are two links in H such that c intersects both of them in a single point. These two points are different, and one of them is an endpoint of its interval. (c) Disc c touches the line of a link $e \in H$ at an endpoint of e.

Intuitively, there is no reason for checking for the circles described in Fig. 4.6 in case of two network elements which are much further apart than the disaster radius r. Indeed, one can build up the solution of the global problem based on some local calculations, as follows. Let X be the set of link intersection points. After determining X, one has to collect edges not further from w than $3r$ into a set E_w, for all $w \in V \cup X$, then determine the maximal failures of sets E_w and finally, get the result by collecting the maximal elements of the resulting lists.

With the help of some additional computational geometric ideas, for determining M_r, one could achieve a near-linear computing complexity in the number of nodes n [29].[9]

4.3.2.4 Circular SRLG Lists of Disasters with Radius *r* on a Sphere

The Earth is not flat, as its shape (geoid) is much more like a sphere. With this in mind, we can deduce that when studying spread-out networks (e.g. the optic fibre network of the USA), in order to reach a higher precision, one should consider that networks are embedded on a spherical surface instead of the much more common planar embedding. Note that [39] found that the spherical counterpart of M_r denoted M_r^s and $M_r^p = M_r$ can be different even in the case of a network having a geographical extension of 100 km.

More precisely, [39] took network `AboveNet` [32] and its shrunk instances, where `AboveNet/c` means that `AboveNet` was rotated such that the average lat and lon coordinates to be both 0; then, each coordinate was divided by c.

[9]Under certain conditions, the complexity of the algorithm presented in [29] is optimal.

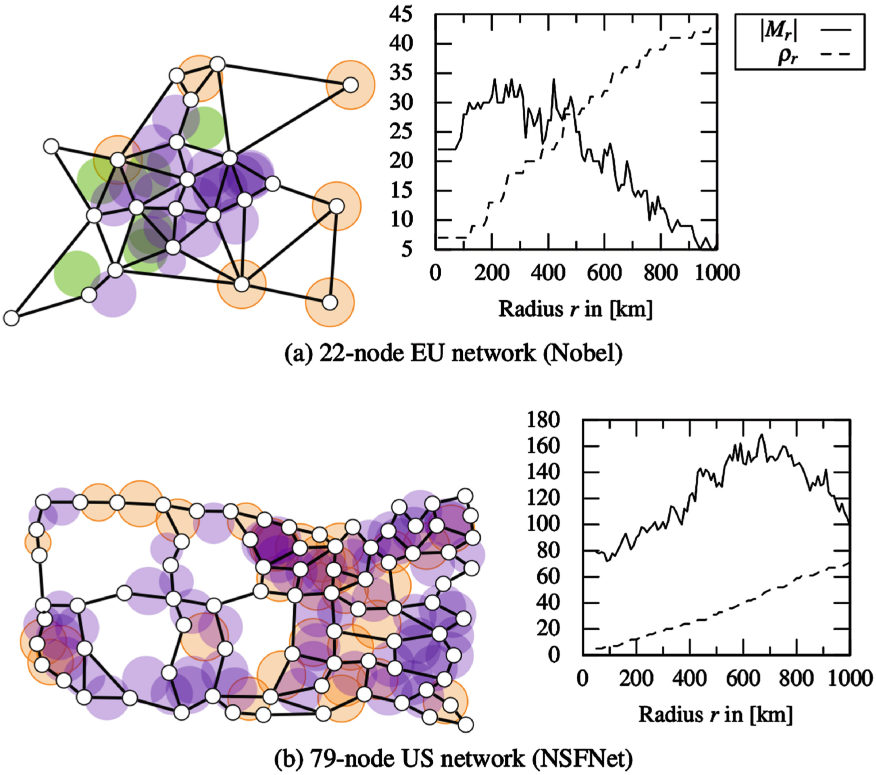


Fig. 4.7 Simulation results for determining list M_r of maximal SRLGs of disasters being circular discs with radius r

As a similarity measure, $\mathcal{M}(r) := |M_r^p \triangle M_r^s| / (|M_r^p| + |M_r^s|) \in [0, 1]$ (the ratio of SRLGs, which are present in only one of M_r^p and M_r^s) was used : if $\mathcal{M}(r)$ is close to 1, it means the two lists are very different, while if it is close to 0, it means there are few differences. Radius $r = 8$ was set to be a bit larger than the half of the diameter of the current network; $r = 0$ was set to be a small radius; the rest of the r values were linearly interpolated.

Figure 4.8 shows that while in case of `AboveNet`, M_r^p and M_r^s are almost entirely different for many values of r, the tendency is that $\mathcal{M}(r)$ decreases as the physical size of the network decreases, which nicely fits the intuition. Surprisingly, $\mathcal{M}(r)$ is not 0 for every range r even for `AboveNet/300`, which equals to the case when the approximative network diameter is 104 km, `AboveNet/400` (having a diameter of approx. 74 km) being the most spread out instance where M_r^p and M_r^s are the same for all investigated r ranges.

As a rule of thumb, it can be said that the difference between the planar and spherical representation of the network can result in different SRLG lists even in case of networks having a geographic extension as small as 100 km.

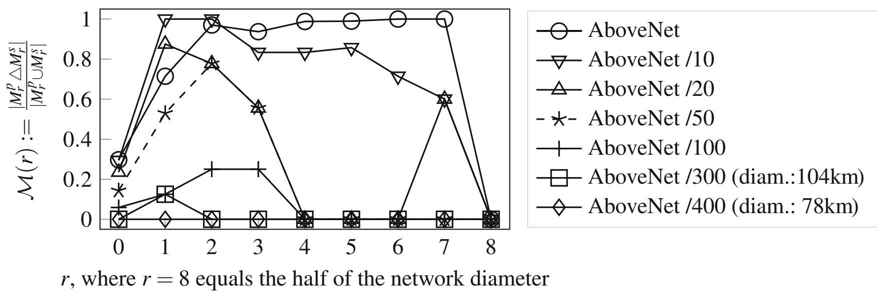


Fig. 4.8 Ratio of those SRLGs which are different in M_r^p and M_r^s, i.e. $|M_r^p \triangle M_r^s|/|M_r^p \cup M_r^s|$

Regarding to calculating list M_r^s of maximal SRLGs of disasters represented as circular discs with radius r on a sphere, basically, the same ideas could be repeated on the sphere as we saw earlier. While considering a more general model, where links of networks are represented as polygonal chains consisting of at most γ line segments between their endpoints (where γ is a parameter), paper [39] presents an approach similar to the one seen in Sect. 4.3.2.3 for determining M_r^s. However, as it only aims to present that 'planar' approaches can be repeated on the sphere, it provides a moderately sophisticated algorithm and complexity bound on determining M_r^s. For the sake of complexity analysis, an additional parameter λ is defined, which is the maximal length of the list of suspected maximal failures M while collecting the maximal failures.

According to Cor. 9 of [39], if both x and λ is $O(n)$, and γ is bounded by a constant, the list M_r^s of maximal link sets which can be hit by a circular disc on the sphere can be computed in $O(n^4\rho_r)$. Simulation results show that ρ_r is proportional to $\frac{2r}{\text{diam}}$ in the interval (0, diam/2], where *diam* is the geometric diameter of the network. This means an $O(n^4\frac{r}{\text{diam}})$ total running time in practice.

4.3.2.5 SRLG Lists of Disaster Sets

In Sects. 4.3.2.1–4.3.2.4, we investigated the possibilities of SLRG list enumeration when the concrete disaster zones are not known, and thus, they are over-approximated with circular discs. In reality, the affected region greatly depends on the properties of the disaster, as well as those of the surrounding area. For example, the region affected by an earthquake depends on the earthquake's magnitude, as well as the properties of the rocks and sediments that the earthquake waves travel through. Thus, it makes sense to base the SLRGs on a variety of possible failure shapes.

If we do know the set $\mathcal{D}$ of representative disasters (along with the physical embedding of the network G), the SRLG enumerating process becomes trivial:

1. $\forall D \in \mathcal{D}$, compute SRLG($D$) and collect these link sets in list $F_{\mathcal{D}}$
2. return the maximal elements of $F_{\mathcal{D}}$ as a list $M_{\mathcal{D}}$.

Note that a detailed knowledge on the nature of possible events taking down some elements of the network in a region might allow us to refine the definition of a disaster (in Definition 4.1) to leverage our additional insights on the failure schemes (e.g. in case of a flood, optical cables traversing the river on the body of a bridge might fail, while aerial cables may remain operational).[10] With all this, if, for each disaster D, computing SRLG(D) can be done in polynomial time, the above-mentioned method returns $M_{\mathcal{D}}$ in polynomial time.

Solutions presented in Sects. 4.3.2.1 and 4.3.2.5 can be viewed as the no and full information versions of the (non-probabilistic) SRLG enumerating problem, respectively. We could see that these cases can be handled algorithmically easily. In Sects. 4.3.2.2–4.3.2.4, we tackled the same problem in case of different qualities of knowledge on the physical embedding of the network.

4.3.3 Approximate Polynomial Algorithms Listing SRLGs

We could see in Sect. 4.3.2.3 a part of a sophisticated theory and relatively complex algorithms which have to be built to be able to provide an algorithm for determining just a single kind of regional SRLG list. This raises the question if one could approach the problem better or at least more general. As we will see in this section, the answer is yes. In a sense, one of the aims of paper [39] is to show that while there is a struggle for fast algorithms determining basically any kind of SRLG list *precisely*, with relatively low effort, one can design discretized approaches which can make small mistakes but which might be permissible given the uncertainty in the failure modelling and the network data. We note that in [38], links are represented as polygonal chains (or chains of geodesics) between their endpoints, allowing to represent real topologies accurately.

For a point P (in the plane or on the sphere) and node $v \in V$, let the node-distance couple be $[v, d(v, P)]$, where $d(v, P)$ is the distance of v to P. Let $v(P)$ be the list of node-distance pairs of all nodes $v \in V$. We define $e(P)$ to be the list of edge-distance pairs defined similarly. It can be proved that for a given point P, $v(P)$ can be computed in $O(n)$ and $e(P)$ in $O(n + x)$ (where x is the number of edge crossings).

The plan is to determine these lists for enough points which are also placed well enough to be able to determine the maximal SRLG lists based on these node-distance and edge-distance lists. Let $\mathcal{P}$ denote the set of points P for which we want to construct $v(P)$ and $e(P)$.

Let us restrict ourselves to planar geometry for a moment. Intuitively, we can calculate M_r by including the grid points of a sufficiently fine grid (let us say containing $1\,\text{km} \times 1\,\text{km}^2$) in $\mathcal{P}$. On a sphere, we should choose a similar nice covering.

[10]Probabilistic refinements are presented in Sect. 4.3.4.1.

Algorithm 4.2 Approximate algorithm for determining the maximal r-range SRLG lists

Require: $G = \langle V, E \rangle, r, \mathcal{P}$, geometry type g, coordinates of nodes and polylines of edges
Ensure: M_r^g
$M_r'^g := \emptyset$
for $P \in \mathcal{P}$ **do**
    determine $e(P)_{\text{hit}}$
    if $e(P)_{\text{hit}} \neq \emptyset$ **then**
        refresh[11] $M_r'^g$ with $e(P)_{\text{hit}}$
    end if
end for
return $M_r'^g$ as M_r^g

Algoirthm 4.2[11] is an example discretized algorithm for determining M_r^g (where 'g' stands for *geometry type*, M_r^g being M_r^s or $M_r^p = M_r$, if the geometrical representation is planar or spherical, respectively) for circular discs, which has a complexity of $O(|\mathcal{P}|n\frac{r}{\text{diam}})$ under some practical assumptions, and it has a low-polynomial complexity indifferent of the nature of the problem input [39, Thm. 11, Cor. 13]. We can see that although Algoirthm 4.2 is much simpler to implement, it is competitive with much more complex precise Algorithm 4.1 in terms of asymptotic run-time.

Regarding to its accuracy, (1) let $d_{\mathcal{P}}$ be the maximal distance of any geometric location from the (closed) convex hull of the geometric embedding of graph G to the closest point of set $\mathcal{P}$, i.e. $d_{\mathcal{P}} := \max_{t \in \text{conv}(G)} \min_{p \in \mathcal{P}} \text{dist}(p, t)$, and (2) let us denote the relationship of two (link) sets E_1 and E_2 by $E_1 \sqsupseteq E_2$ (E_1 over-approximates E_2) if and only if for all $e_2 \in E_2$, there exists an $e_1 \in E_1$, such that $e_1 \supseteq e_2$ (e_1 over-approximates e_2), conform with Definition 4.7. Using these notations, $M_r^g \sqsupseteq H_r^g \sqsupseteq M_{r-d_P}^g$, where H_r^g is the output of Algorithm 4.2 [38, Thm. 11]. Based on this, if one wants to protect disasters caused by discs with radius r, it is only needed to run Algorithm 4.2 initializing the radius as $r + d_{\mathcal{P}}$. Furthermore, by choosing $\mathcal{P}$ such that $d_{\mathcal{P}}$ to be small, one can avoid enumerating overprotective SRLGs, more precisely, $\lim_{d_{\mathcal{P}} \to 0} H_r^g = M_r^g$, for any fixed network.[12]

Considering non-circular SRLGs, engineering fast precise algorithms for determining SRLG lists for arbitrary disaster shape instead of a circle is not trivial,[13] but approximate algorithms similar to the one described for determining M_r^g can be easily designed. In short, while the disc is invariant to rotation, the only additional hardness here is that the different orientations of the fixed shape should be also considered. In other words, one should check the links hit by the shape at every centre point *and every orientation*. Discretizing the possible orientations of the shape can be handled just as discretizing the places of centre points. Based on this idea,

[11] Similarly to the precise algorithms, refreshing in Algorithm 4.2 means that M_r^g is the set of maximal failures among which are already checked, and if $e(P)_{\text{hit}}$ is maximal amongst them, it is added to M_r^g and all $e(P)_{\text{hit}}$'s subsets are eliminated from M_r^g; or if $e(P)_{\text{hit}}$ is not maximal in M', nothing happens.

[12] In other words, the degree of over-approximation (Definition 4.9) tends to $(1, 0)$ as $d_{\mathcal{P}} \to 0$.

[13] Reference [35] tackles a similar problem.

[39, Alg. 4.] approximately calculates the list $M_{\texttt{shape}}$ of maximal failures which can be caused by a disaster shape in $O(a|\mathcal{P}|n\frac{r}{\text{diam}})$ under some practical assumptions, where a is the number of orientations of the shape which are considered. Its complexity is low-polynomial indifferent of the problem input, and, in limit, the output is precisely $M_{\texttt{shape}}$, both in the plane and on the sphere [39, Thm. 15, Cor. 17].

4.3.4 More SRLG Enumerating Approaches

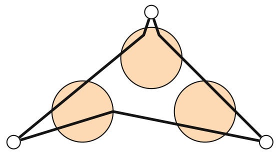
4.3.4.1 SRLGs of Spatially-Close Fibres

Paper [16] proposed to call a pair of fibres *spatially-close* if their distance is at most r', i.e. they can be covered with a circular disc of radius $\frac{r'}{2}$. They propose to define SRLGs as sets of fibres where any pair of fibres are spatially-close; in other words, any pair of fibres can be covered with a circular disc of radius $\frac{r'}{2}$. See Fig. 4.9 highlighting the difference between this model and the one shown in Sect. 4.3.4.1. The intuition is that r' is a small parameter representing those fibres that are close together having a higher probability of failing simultaneously due to regional failures. The high-level idea is to provide an approach that generates SRLGs, not considering failure shapes, but simply considers a threshold r': any fibre pairs with a separation distance smaller or equal to r' are considered spatially-close.

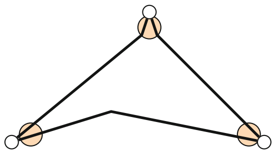
In [16], three spatially-close fibre problems are considered: (1) finding all pairs of spatially-close fibre segments, (2) finding all spatially-close intervals of fibre to a set of other fibres and (3) grouping spatially-close fibres into SRLGs.

Fibres are modelled as non-straight concatenations of fibre segments of irregular lengths. Each of these fibre segments is a straight line connecting two fibre points of known geodetic coordinates (latitude and longitude). For easier calculations, the coordinates are projected to two-dimensional Cartesian coordinates, embedding the fibres into the 2-D plane.

Problem 4.1 (*Detection of Spatially-Close Fibre Segments (DSCFS)* [16]) Given a set E of fibres and a distance r'. Each fibre $e \in E$ is associated with a set $\mathcal{T}_e$ of T_e fibre segments. Each fibre segment $t \in \mathcal{T}_e$ is associated with two fibre points (u_{t1}, v_{t1}) and (u_{t2}, v_{t2}). Find all fibre segment pairs of different fibres that have a minimum separation distance of at most r'.



(a) There are 3 SRLGs, each is a pair of links



(b) There is a single SRLG with three links

Fig. 4.9 Example of SRLG according to **a** Definition 4.1 and **b** Sect. 4.3.4.1

Clearly, the DSCFS problem is solvable in polynomial time, as the naive approach (computing the separation distance of all fibre segments) has a time complexity of $O(|E|^2 T^2)$, where T is the maximum number of fibre segments per fibre. In practice, the run-time can be reduced significantly by storing each segment in an R-tree[14] [16].

The probability that two spatially-close fibres fail simultaneously depends on the length of the interval(s) of the fibres that are close together.

Problem 4.2 (*Intervals to a Set of Spatially-Close Fibres (ISSCF) Problem* [16]) Given a fibre e_i, a set $\mathcal{Y}$ of Y fibres and a distance r'. Each fibre e_i or $e_j \in \mathcal{Y}$ is associated with a set $\mathcal{T}_i/\mathcal{T}_j$ of T_i/T_j fibre segments. Each fibre segment $t \in \mathcal{T}_i/\mathcal{T}_j$ is associated with two fibre points (u_{t1}, v_{t1}) and (u_{t2}, v_{t2}). Find the intervals of fibre e_i that have a minimum separation distance of at most r' to any fibre $e_j \in \mathcal{Y}$.

This problem can be solved in $O(YT^2)$ time, where T is the maximum number of fibre segments per fibre, by first finding all fibre segments of Y that are spatially-close to e_i and then computing the spatially-close intervals to these segments by solving sets of quadratic equations (see Alg. 3 in [16]).

Finally, if a set of fibres are grouped into an SRLG if every pair of fibres is spatially-close to each other:

Problem 4.3 (*Grouping of Spatially-Close Fibres (GSCF) Problem* [16]) Given a set $\mathcal{F}$ of F spatially-close fibre pairs. Group all fibres that are spatially-close to each other, *such that the number of distinct SRLGs is minimized*.

In other words, we want to find all maximal SRLGs, where a maximal SRLG is a set of fibres that are spatially close to every other fibre in the set and which is not a subset of any other SRLG.

In [16], a heuristic algorithm was given that first transforms it to the Maximal Clique Enumeration (MCE) problem. Second, a variant of the Bron–Kerbosch algorithm [33] to find all maximal cliques is used to find all maximal SRLGs. Note that MCE is an $\mathcal{NP}$-hard problem in general.

4.3.4.2 A Single Worst SRLG in Case of a Fixed Metric

Reference [22] presents three flavours of problems for finding a most vulnerable place of the network in case of multiple network vulnerability measures.[15] The first problem assumes that the network is bipartite in the topological and geographic sense and that the cuts are vertical line segments. In the latter two problems network, links can be in almost arbitrary locations on the plane. In one of the problems, the disaster shapes are line segments in any direction. In the other, the disasters are circular discs

[14] An R-tree is an efficient tree data structure for storing spatial objects. Objects are grouped based on their minimum bounding rectangle.

[15] The investigated measures are: (1) the total expected capacity of the intersected links, (2) the fraction of pairs of nodes that remain connected, (3) the maximum flow between a given pair of nodes, (4) the average value of maximum flow between all pairs of nodes.

with a given range. To solve the problem instances, in [22], both MILP formulations and polynomial algorithms are given.

We say that a vulnerability metric μ is *monotone*, if, according to μ, for any link set $E_1 \subseteq E_2$, the failure of E_2 is *worse* than the failure of E_1. We note that in the natural condition when the vulnerability metric or a protection mechanism is monotone (cf. Definition 4.6), the worst SRLG will be part of the set of exclusion-wise maximal SRLGs fulfilling a given criteria c (e.g. SRLGs that can be hit by line segments in any direction, circular discs with a given range or by a disaster from a disaster set $\mathcal{D}$), which can be determined using the techniques depicted in Sects. 4.3.2 and 4.3.3 and 4.5 (which uses PSRLGs as an intermediate step). Thus, a worst SRLG can be found according to Algorithm 4.3[16]:

Algorithm 4.3 Worst SRLG of a vulnerability metric or protection mechanism

Require: graph $G = \langle V, E \rangle$, criteria c, monotone vulnerability metric or protection mechanism μ
Ensure: worst SRLG S according to μ fulfilling criteria c.
1: Calculate SRLG list M_c of maximal SRLGs fulfilling criteria c (e.g. as in Sects. 4.3.2, 4.3.3 or 4.5)
2: Compute the value $\mu(S)$ for each $S \in M_c$
3: **return** an $S \in M_c$ for which $\mu(S)$ is worst

4.4 Calculating Lists of PSRLGs

4.4.1 Computing Lists of FPs and CFPs

4.4.1.1 Computing FPs From Disaster Sets

Most algorithms for analysing the vulnerability of networks to disasters, or for creating regional (P)SRLG lists, assume the regional failure takes a fixed shape everywhere in the network area. In reality, the affected region greatly depends on the properties of the disaster, as well as those of the surrounding area.

Oostenbrink and Kuipers [23] proposes computing the vulnerability of a network to a set of representative disasters $\mathcal{D}$ (each of any shape), instead of to a fixed disaster shape. Each disaster $D \in \mathcal{D}$ is assigned a disaster area $D \subseteq \mathbb{R}^2$, and an occurrence probability p_D.[17] As the probability of simultaneous disasters is low (ignoring strongly correlated events such as aftershocks, which can be combined into a single composite disaster), it is assumed exactly one disaster will occur, i.e. $\sum_{D \in \mathcal{D}} p_D = 1$. Furthermore, it is assumed that if a disaster D occurs, all links

[16] Algorithm 4.3 is polynomial assuming the SRLG enumeration and calculating the metric value runs in polynomial time.

[17] In contrast, Sect. 4.3.2.5 depicts the non-probabilistic version of this regional failure model.

intersecting its disaster area will fail. A disaster D can take any shape and does not have to be connected, as long as it is possible to compute if a line segment intersects it.

A representative disaster set $\mathcal{D}$ can be obtained in a variety of ways, preferably in collaboration with experts (e.g. seismologists). For example, one can use a tool to randomly generate sets of possible disasters, use the last N historical disasters or construct a set of custom disaster scenarios. The concept of a set of representative disasters is similar to that of the stochastic event set used in Cat modelling, for which a large number of models exist. In all cases, it is necessary to convert hazard intensity values such as ground motions to a binary area D using some threshold function. [23] gives an example of converting earthquake scenarios to a disaster set $\mathcal{D}$.

Note that if the disaster locations and shapes are both finite discrete random values (e.g. a division of the plane into grid points), we can generate a finite disaster set $\mathcal{D}$ of all possible disasters by simply adding each possible combination of disaster location and shape to $\mathcal{D}$.

Let a failure state s be defined as a set $s \subseteq E$, where $e \in s$ if and only if e has failed. Now, the failure set $S(D)$ of links that are affected by disaster D is the set of links that intersect D. Note that this definition of $S(D)$ is equivalent to that of a regional SLRG, $SRLG(D)$.

Let $\mathcal{S}$ be the random value indicating the failure state after the disaster. Given a disaster set $\mathcal{D}$, we can obtain the distribution of $\mathcal{S}$ as follows [23]:

1. $\forall D \in \mathcal{D}$, compute $S(D)$
2. $\forall s \in S[\mathcal{D}]$ (the image of S), store $S^{-1}(s) = \{D \in \mathcal{D} | S(D) = s\}$
3. $\forall s \in S$, $\text{FP}(s) = P(\mathcal{S} = s) = \sum_{D \in S^{-1}(s)} p_D$.

We now have each possible failures state, as well at its occurrence probability. That is, we have the complete list of FPs, based on the disaster set $\mathcal{D}$.

4.4.1.2 Calculating Lists of CFPs Based on Correlated Link Failures

A study dealing with the probabilities of correlated link failures is [31], which models the regional failures as having a random epicentre and a random size (described with a size parameter s in [0, 1]). Two assumptions are made: (1) in the investigated time period, there is at most one disaster and (2) for every possible failure epicentre and failure sizes $s_1 < s_2$, the region destroyed by disaster with size s_1 is totally contained by the region hit by the one with size s_2.

Figure 4.10 briefly depicts the model. It shows an example network and the corresponding failure probabilities. Suppose we need to establish a high-availability connection from the top node through the working path of link b and protection path $a - e$. The unavailability of the working path can be computed as $\text{CFP}(b) = 0.0055$, and for the protection path, it is $\text{CFP}(a) + \text{CFP}(e) - \text{CFP}(a, e) = 0.00986$. In the traditional approach, the two paths are assumed to fail independently; thus, the total

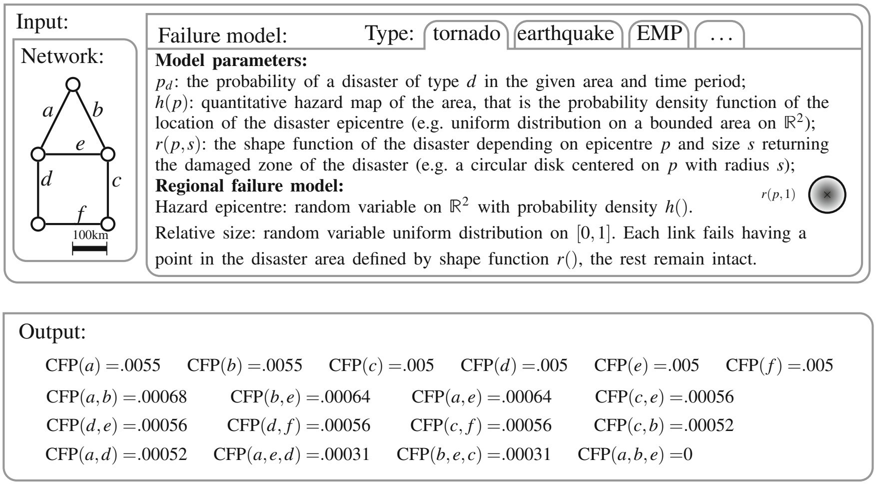


Fig. 4.10 An illustration of the CFP problem inputs and outputs

connection availability is estimated as $1 - 0.0055 \cdot 0.00986 = 0.999945$, i.e. four nines. However, considering the joint failure probabilities of the links (provided in the example), the total connection availability should be $1 - \text{CFP}(a, b) - \text{CFP}(b, e) + \text{CFP}(a, b, e) = 0.9987$, i.e. not even three nines, which is a significant difference.

Now, an implementation strategy follows which uses discrete functions instead of continuous ones. The problem is discretized by defining a sufficiently fine resolution, say 1 km and placing a grid of 1 km × 1 km squares over the plane to assume that the disaster regions $r(p, s)$ and hit link sets $R(p, s)$ are 'almost identical'[18] for every size s and disaster centre point p inside each grid cell c. This way the whole integration problem translates to a summation. The inputs are defined over the grid, and the Euclidean plane is considered as a Cartesian coordinate system.

Let r denote the absolute maximum range of a disaster in km. Let $(x_{\min}, y_{\min})$ be the bottom left corner and $(x_{\max}, y_{\max})$ the top right corner of a rectangular area in which the network lies. It is sufficient to process each c in the rectangle of bottom left corner $(x_{\min} - r, y_{\min} - r)$ and top right corner $(x_{\max} + r, y_{\max} + r)$, and we denote by $c_{i,j}$ the grid cell in the i-th column and jth row. In this range, for each $c_{i,j}$, we will consider the probability $h_{i,j}$ of the next disaster having epicentre p in the cell $c_{i,j}$. (i.e. $h_{i,j} = \int_{p \in c_{i,j}} h(p) \mathrm{d}p$, where $h(p)$ is the probability density function of the disaster epicentre).

To build up the list of CFPs, an associative array CFP[ ] is used, which can be addressed by a set of links $\{e_1, e_2, \ldots, e_k\}$ and returns its cumulative failure probability. For this, in the pre-computation process, we have to extract the contribution of $c_{i,j}$ to the cumulative failure probability of every subset S of links. We do this by

[18] In particular, we may assume that the probability $f(e, p)$ that link e fails if a disaster with epicentre p happens is independent of p as long as it is in c. We denote this common value by $f(e, c)$.

working with the list $\mathcal{S}_{i,j} = (e_1, e_2, \ldots, e_k)$ (where e_i is the ith closest link to cell $c_{i,j}$) and increments the CFP values accordingly, i.e. $\text{CFP}[\{e_1\}] += h_{i,j} \cdot f(e_1, c_{i,j})$, $\text{CFP}[\{e_2\}] += h_{i,j} \cdot f(e_2, c_{i,j})$, $\text{CFP}[\{e_1, e_2\}] += h_{i,j} \cdot f(e_2, c_{i,j})$, etc.

For the probability CFP(S) of failing at least the set S of links, we need to look up S in CFP. If not found, then CFP(S) $= 0$. The query time of sets can be reduced to a constant with very high probability (with the help of hashing). Using self-balancing binary trees, the worst-case query time is always $O(\rho \log((n + x)\rho))$, where ρ is the maximum number of links hit by a disaster.

The drawback of the CFP list is that it has an $\Omega(2^\rho)$ space complexity, which makes it inefficient for bigger network densities. With this in mind, one can build up also a list of FPs representing the same disasters, which will be significantly shorter, but some pre-computations will be needed to determine CFP(S).

4.4.1.3 CFPs and FPs from Historical Earthquake Data

Intuitively, the models presented in Sects. 4.4.1.1 and 4.4.1.2 are somewhat related. In fact, both models could be used for computing lists of FPs and CFPs. In [36], an approach for determining the list of CFPs and FPs based on the available historical earthquake data is presented. This approach can be viewed as a special case of both models presented in Sects. 4.4.1.1 and 4.4.1.2.

Namely, the next possible earthquake has a random epicentre taken from a set of grid points over the evaluated area, and the disaster area has also a random range (which is a function of the earthquake moment magnitude) taken from a discretized scale. This results in a set of disaster scenarios with some probabilities as in Sect. 4.4.1.1 and also can be viewed as a discrete version of the model in Sect. 4.4.1.2.

Simulation results of [36] show that the graph of $f(x) =$ '*the probability of the x^{th} most probable* (C)FP' follows an exponential distribution in case of FPs and fits the power law in case of CFPs. In other words, if one stores only (C)FPs having a probability higher than a threshold T, lowering T by several orders of magnitude does not cause a severe increase in the number of listed FPs, but the size of the CFP list explodes.

4.4.2 Probabilistic Modelling of the Worst Place of a Disaster

Similarly to [22] (in Sect. 4.3.4.2), [2] aims to find the single worst place of a disaster under a certain metric. While [22] models the disaster effect to be deterministic (every network element which has an intersection with the disaster area fails with probability 1), in [2], every link has a probability $\in [0, 1]$ of failing in case of each disaster place. However, an incompleteness of the paper is that, in case of a fixed disaster, it considers that the affected links fail independently of each other.

To be more precise, the model of [2] is the following. They define a failure probability distribution function $f : Q \times \mathbb{R}^2 \rightarrow R \geq 0$. Given a disaster location

$P \in \mathbb{R}^2$ and $q \in Q$, $f(q, p) = f_q(p)$ is the probability that q is affected by the disaster at p. For a compound component π composed of a sequence of simple components $q_1, \ldots, q_r$, the probability of being damaged by a disaster at a location p is denoted as $f_\pi(p)$ and being defined to be the probability that at least one of its simple component is damaged, i.e. $f_\pi(p) = 1 - \prod_{q \in \pi}(1 - f_q(p))$.

For finding the most vulnerable point according to various metrics (expected component damage, average two-terminal availability and expected maximum post-attack flow), [2] presents Las Vegas and Monte Carlo algorithms. It also offers approximate solutions to the problem of finding the worst arrangement of k simultaneous disasters (attacks), which is a generalization of the $\mathcal{NP}$-hard maximum set cover problem [15].

4.4.3 On Two-Stage PSRLGs and Denomination Issues

The first paper considering probabilistic SRLGs was [19]. There, the structure which in this chapter is called 'two-stage PSRLG' is named simply as 'probabilistic SRLG' (PSRLG). Since we felt that FPs and CFPs deserve to be called probabilistic SRLGs at least as much as the structure defined in [19], we call these structures collectively as PSRLGs and name the [19]-PSRLG as 'two-stage PSRLG'.

Due to this historical reason, we believe it worth presenting its model even though [19] does not tackle the question of calculating PSRLGs (it only uses PSRLGs as inputs for a diverse routing problem). Lee et al. [19] defines the two-stage PSRLGs as follows. There is a set R of SRLG events that can incur link failures. Each SRLG event $r \in R$ occurs with probability π_r, and once an SRLG event r occurs, link (i, j) will fail independently of the other links with probability $p^r_{i,j} \in [0, 1]$. Link (i, j) is part of the resulting (two-stage) PSRLG if $p^r_{i,j} > 0$. This definition from [19] is slightly generalized in Definition 4.5 while keeping the form of the data structure.

As Fig. 4.1 also suggests, using lists of two-stage PSRLGs, one could store the same information more compactly as in lists of FPs or CFPs. However, there are numerous open questions related to this field; as to the best of our knowledge, no paper investigated how to enumerate in an efficient way lists of two-stage PSRLGs.

We believe that lists of FPs or CFPs are the right standard structures describing the probabilistic effects of the disasters, and any other versions of PSRLGs may be viewed as model-specific compact representations of these lists.

4.5 Advanced: SRLG Lists Obtained from PSRLG Lists

It is a natural idea to list the (maximal) link sets which have a probability of failing together higher than a given threshold T (like in [26]).[19] Obviously, for this, as an intermediate step, one has to generate a set of probabilistic SRLGs. More precisely, CFP is the most useful structure in this context, since, by definition, for a link set S, the cumulative failure probability CFP(S) is the probability that at least the links of S will fail. The advantage of this approach is that SRLG lists can be generated based on sophisticated objectives. Algorithm 4.4 sketches this framework.

Algorithm 4.4 SRLG list obtained from CFP list

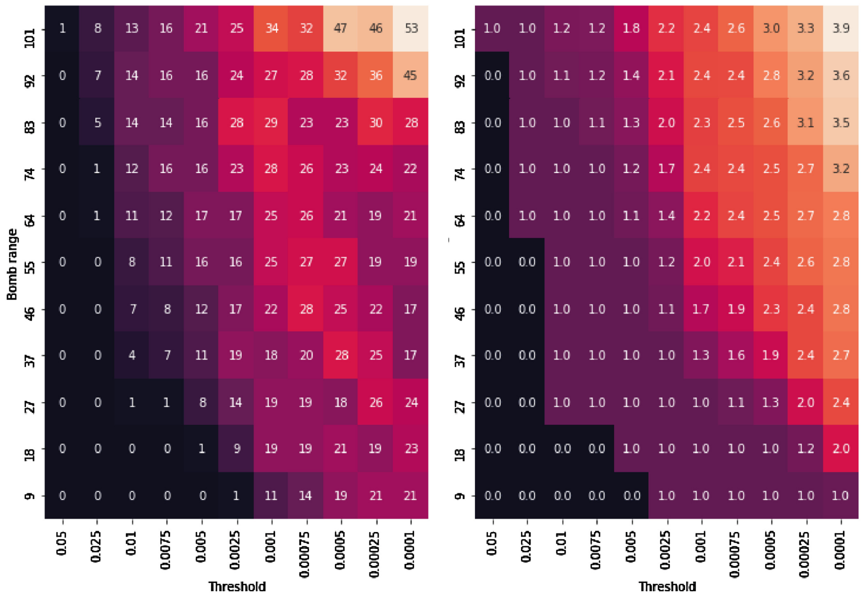
Require: graph $G = \langle V, E \rangle$, threshold $T \in [0, 1]$, CFP model C (e.g. as in one of [23, 25, 26, 31]), additional parameters needed for C
Ensure: list M_T of maximal link sets having a CFP at least T
1: Calculate list L of CFPs according to C
2: Collect CFPs of L with probability $\geq T$ in list F_T
3: **return** list M_T of maximal elements of F_T

As a concrete example, [26] modifies the CFP enumerating model presented in [31] (and Sect. 4.4.1.2) in order to take in count also the availabilities of the links. Compared to [31], a link e with low availability makes CFPs it is involved in to have higher probabilities, while reliable links decrease these probabilities. Figure 4.11 shows the cardinality of the output M_T and average length of SRLGs in M_T in function of threshold T and maximal disaster area R in case of backbone topology `16_optic_pan_eu` [24]. Note that the unit of R is not a `km`, as the Euclidean distances are altered during the CFP enumerating process, based on the availabilities.

Figure 4.11 shows that a radius $R \geq 80$ (which roughly corresponds to the 20% of the network diameter) or larger combined with a threshold $T \leq 0.001$ yields a high number of maximal probable failures. This translates to the fact that a bigger disaster possibly hits a larger number of edges, and the failures above the small threshold cannot be dominated by only a few sets from M_T.[20] Further observations of [26] are that: (i) if $R \in [0, 80]$, M_T is likely to contain only a handful of most probable SRLGs; (ii) similar $R \cdot T$ value indicates similar cardinality of $\mathcal{M}_T$. For reasonable disaster sizes, M_T has a manageable size, with its cardinality being comparable with the number of network elements. In addition, one can observe that the average size of the SRLGs scales with the radius.

[19] In Sect. 4.4.1.2, we could see how CFPs (and thus all kind of PSRLGs) can be used for calculating the availability of services. To leverage the probabilistic information stored in PSRLGs in case of resilient routing, one needs to calculate a list of SRLGs based on a PSRLG input.

[20] Of course, in a non-practical extreme case of R being greater than half of the network diameter, it is possible that $M_T = \{E\}$, meaning $|M_T| = 1$.



(a) $|M_T|$ (b) Average SRLG length in M_T

Fig. 4.11 Cardinality of $|M_T|$ and avg. length of SRLGs in M_T in function of probability threshold T and maximum disaster radius R in case of backbone topology `16_optic_pan_eu` [24]

4.6 A Mind Map of the Chapter

See Fig 4.12.

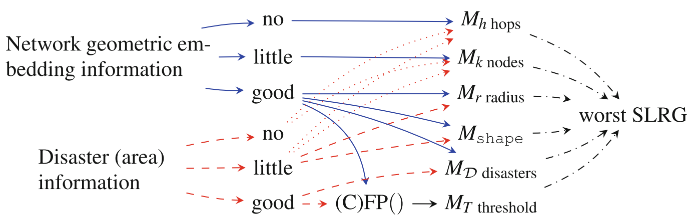


Fig. 4.12 A mind map of the chapter. (C)FP() stores PSRLGs, while lists M_* consist of SRLGs. SRLGs are used, e.g. by resilient routing, while PSRLGs, e.g. in determining service availability

4.7 Conclusions

In this chapter, we overviewed the state-of-the-art algorithms for enumerating *regional shared risk link groups (SRLGs)* and *regional probabilistic SRLGs (PSRLGs)*; these structures are key in translating the composed geometric problem of protecting telecommunication networks against regional failures to purely combinatorial and probabilistic problems, respectively. We showed that the best technique to choose for enumerating the vulnerable regions varies on (1) the available geometric information on the network topology, (2) (probabilistic) information on the effects of possible disasters in the network area and (3) the desired output structure (SRLG/PSRLG). In the chapter, first, we presented a range of deterministic approaches for enumerating maximal regional SRLGs under various conditions. Then, for regional PSRLG enumeration, we visited some models, which are easily tunable to the available knowledge on the network topology and the disasters. Finally, as an advanced technique, we described an SRLG enumerating approach, which uses an arbitrary probabilistic model in an intermediate step.

Acknowledgements This chapter is based on work from COST Action CA15127 ("Resilient communication services protecting end-user applications from disaster-based failures—RECODIS") supported by European Cooperation in Science and Technology (COST). Part of this work is supported by the Hungarian Scientific Research Fund (grant No. OTKA K124171, K128062), by the BME-Artificial Intelligence FIKP grant of EMMI (BME FIKP-MI/SC), and the HUJI Cyber Security Center together with the Israel National Cyber Directorate in the Prime Minister's Office.

References

1. Agarwal PK, Efrat A, Ganjugunte SK, Hay D, Sankararaman S, Zussman G (2010) Network vulnerability to single, multiple, and probabilistic physical attacks. In: Military Communications Conference (MILCOM). IEEE, pp 1824–1829
2. Agarwal PK, Efrat A, Ganjugunte SK, Hay D, Sankararaman S, Zussman G (2013) The resilience of WDM networks to probabilistic geographical failures. IEEE/ACM Trans Netw (TON) 21(5):1525–1538
3. Colman-Meixner C, Dikbiyik F, Habib F, Tornatore M, Chuah CN, Mukherjee B (2014) Disaster-survivable cloud-network mapping. Photonic Netw Commun 27(3):141–153
4. de Sousa, A., Santos, D.: The minimum cost *D*-geodiverse anycast routing with optimal selection of anycast nodes. In: 2019 15th International Conference on the Design of Reliable Communication Networks (DRCN), pp 21–28. https://doi.org/10.1109/DRCN.2019.8713729
5. Dikbiyik F, Tornatore M, Mukherjee B (2014) Minimizing the risk from disaster failures in optical backbone networks. J Lightwave Technol 32(18):3175–3183
6. Dinh TN, Thai MT (2015) Network under joint node and link attacks: vulnerability assessment methods and analysis. IEEE/ACM Trans Netw 23(3):1001–1011
7. Gardner MT, Beard C (2011) Evaluating geographic vulnerabilities in networks. In: IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR), pp 1–6
8. Gerstel O, Jinno M, Lord A, Yoo SB (2012) Elastic optical networking: a new dawn for the optical layer? IEEE Commun Mag 50(2):s12–s20

9. Gomes T, Tapolcai J, Esposito C, Hutchison D, Kuipers F, Rak J, de Sousa A, Iossifides A, Travanca R, Andre J, Jorge L, Martins L, Ugalde PO, Pašić A, Pezaros D, Jouet S, Secci S, Tornatore M (2016) A survey of strategies for communication networks to protect against large-scale natural disasters. In: 8th International Workshop on Resilient Networks Design and Modeling (RNDM), pp 11–22. https://doi.org/10.1109/RNDM.2016.7608263
10. Habib MF, Tornatore M, De Leenheer M, Dikbiyik F, Mukherjee B (2012) Design of disaster-resilient optical datacenter networks. J Lightwave Technol 30(16):2563–2573
11. Habib MF, Tornatore M, Mukherjee B (2013) Fault-tolerant virtual network mapping to provide content connectivity in optical networks. In: Optical Fiber Communication Conference. OSA, pp OTh3E–4
12. Harter IBB, Schupke D, Hoffmann M, Carle G et al (2014) Network virtualization for disaster resilience of cloud services. IEEE Commun Mag 52(12):88–95
13. Heidemann J, Quan L, Pradkin Y (2012) A preliminary analysis of network outages during Hurricane Sandy. University of Southern California, Information Sciences Institute
14. Hmaity A, Musumeci F, Tornatore M (2016) Survivable virtual network mapping to provide content connectivity against double-link failures. In: Design of Reliable Communication Networks (DRCN). IEEE, pp 160–166
15. Hochbaum D, Pathria A (1998) Analysis of the greedy approach in problems of maximum k-coverage. In: Naval research logistics (NRL)
16. Iqbal F, Trajanovski S, Kuipers F (2016) Detection of spatially-close fiber segments in optical networks. In: 2016 12th International Conference on the Design of Reliable Communication Networks (DRCN). IEEE, pp 95–102. https://doi.org/10.1109/DRCN.2016.7470840. http://ieeexplore.ieee.org/document/7470840/
17. Jaumard B, Hoang HA (2013) Design and dimensioning of logical survivable topologies against multiple failures. J Opt Commun Netw 5(1):23–36
18. Kushwaha A, Kapadia D, Gumaste A, Somani A (2018) Designing multi-layer provider networks for circular disc failures. In: International Conference on Optical Network Design and Modeling (ONDM), Dublin, Ireland
19. Lee HW, Modiano E, Lee K (2010) Diverse routing in networks with probabilistic failures. IEEE/ACM Trans Netw 18(6):1895–1907
20. Long X, Tipper D, Gomes T (2014) Measuring the survivability of networks to geographic correlated failures. Opt Switching Netw 14:117–133
21. Mukherjee B, Habib M, Dikbiyik F (2014) Network adaptability from disaster disruptions and cascading failures. IEEE Commun Mag 52(5):230–238
22. Neumayer S, Zussman G, Cohen R, Modiano E (2011) Assessing the vulnerability of the fiber infrastructure to disasters. IEEE/ACM Trans Netw (TON) 19(6):1610–1623
23. Oostenbrink J, Kuipers F (2017) Computing the impact of disasters on networks. ACM SIGMETRICS Perform Eval Rev 45(2):107–110
24. Orlowski S, Wessäly R, Pióro M, Tomaszewski A (2010) SNDlib 1.0: survivable network design library. Networks 55(3):276–286
25. Pašić A, Girao-Silva R, Vass B, Gomes T, Babarczi P (2018) FRADIR: a novel framework for disaster resilience. In: International Workshop on Resilient Networks Design and Modeling (RNDM), Longyearbyen, Svalbard, Norway
26. Pašić A, Girao-Silva R, Vass B, Gomes T, Mogyorósi F, Babarczi P, Tapolcai J (2019) FRADIR-II: an improved framework for disaster resilience. In: International Workshop on Resilient Networks Design and Modeling (RNDM), Nicosia, Cyprus
27. Souza Couto R, Secci S, Mitre Campista M, Costa K, Maciel L (2014) Network design requirements for disaster resilience in IaaS clouds. IEEE Commun Mag 52(10):52–58
28. Su TH, Chang RC (1990) The k-Gabriel graphs and their applications. In: Algorithms. Lecture Notes in Computer Science, vol 450. Springer, Berlin (1990)
29. Tapolcai J, Rónyai L, Vass B, Gyimóthi L. Fast enumeration of regional link failures caused by disasters with limited size. Technical report. http://lendulet.tmit.bme.hu/lendulet_website/wp-content/papercite-data/pdf/tapolcai2019fast.pdf

30. Tapolcai J, Rónyai L, Vass B, Gyimóthi L (2017) List of shared risk link groups representing regional failures with limited size. In: Proceedings of IEEE INFOCOM, Atlanta, USA
31. Tapolcai J, Vass B, Heszberger Z, Biró J, Hay D, Kuipers FA, Rónyai L (2018) A tractable stochastic model of correlated link failures caused by disasters. In: Proceedings of IEEE INFOCOM, Honolulu, USA
32. The internet topology zoo. http://www.topology-zoo.org/dataset.html
33. Tomita E, Tanaka A, Takahashi H (2006) The worst-case time complexity for generating all maximal cliques and computational experiments. Theor Comput Sci 363(1):28–42
34. Trajanovski S, Kuipers F, Van Mieghem P et al (2013) Finding critical regions in a network. In: IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, pp 223–228
35. Trajanovski S, Kuipers FA, Ilić A, Crowcroft J, Van Mieghem P (2015) Finding critical regions and region-disjoint paths in a network. IEEE/ACM Trans Netw (TON) 23(3):908–921
36. Valentini A, Vass B, Oostenbrink J, Csák L, Kuipers FA, Pace B, Hay D, Tapolcai J (2019) Network resiliency against earthquakes. In: International Workshop on Resilient Networks Design and Modeling (RNDM), Nicosia, Cyprus
37. Vass B, Bérczi-Kovács E, Tapolcai J (2017) Enumerating shared risk link groups of circular disk failures hitting k nodes. In: Procedings of International Workshop on Design of Reliable Communication Networks (DRCN), Munich, Germany
38. Vass B, Kovács E, Tapolcai J. Enumerating maximal shared risk link groups of circular disk failures hitting k nodes. Technical report. http://lendulet.tmit.bme.hu/lendulet_website/wp-content/papercite-data/pdf/vass2019enumerating.pdf
39. Vass B, Németh L, Tapolcai J (2019) The Earth is nearly flat: precise and approximate algorithms for detecting vulnerable regions of networks in plane and on sphere. In: Networks. Wiley (2019)
40. Veremyev A, Boginski V, Pasiliao EL (2014) Exact identification of critical nodes in sparse networks via new compact formulations. Optim Lett 8(4):1245–1259
41. Xie A, Wang X, Lu S (2019) Risk minimization routing against geographically correlated failures. IEEE Access 7:62920–62929. https://doi.org/10.1109/ACCESS.2019.2916834
42. Yallouz J, Rottenstreich O, Orda A (2014) Tunable survivable spanning trees. ACM SIGMETRICS Perform Eval Rev 42(1):315–327
43. Zang H, Ou C, Mukherjee B (2003) Path-protection routing and wavelength assignment (RWA) in WDM mesh networks under duct-layer constraints. IEEE/ACM Trans Netw (TON) 11(2):248–258