

BAR-ILAN UNIVERSITY

Coping with Physical Attacks on Random Network Structures

Omer Gold (Goldgevicht)

Submitted in partial fulfillment of the requirements for the Master's
Degree in the Department of Computer Science, Bar-Ilan University

Ramat Gan, Israel

January 2014

This work was carried out under the supervision of
Prof. Reuven Cohen and Dr. Tali Kaufman
Department of Computer Science,
Bar-Ilan University.

Acknowledgments

I would like to thank several people, who helped me during this work and contributed to my experience throughout my Masters degree in general:

First, my sincere gratitude is to Prof. Reuven Cohen, for his great support and efforts, guiding me throughout my research, writing a paper, and finishing this thesis, with his patience, welcoming attitude and great knowledge, whilst allowing me the room to work in my own way.

I would like to express my deepest appreciation to Dr. Tali Kaufman for her support and care throughout the entire course of my degree, for paying attention and giving the confidence that made the collaboration with the department of Mathematics possible. I would like to thank Prof. Moshe Koppel, I had the honor to work with, teaching the “Automata and Formal Languages” course. Thank you for giving me the opportunity to influence, and being open to my ideas.

I would also like to thank the computer science department staff, who assist me many times and made things easier for me, especially Hila Atia.

Last but not the least, I would like to thank my brother Doron for encouraging me since the beginning of my academic track, and my beloved family for all their love and support.

Abstract

Communication networks are vulnerable to natural disasters, such as earthquakes or floods, as well as to physical attacks, such as an Electromagnetic Pulse (EMP) attack. Such real-world events happen at *specific geographical locations* and disrupt specific parts of the network. Therefore, the geographical layout of the network determines the impact of such events on the network's physical topology in terms of capacity, connectivity, and flow.

Recent works focused on assessing the vulnerability of a *deterministic* (geographical) network to such events. In this work, we focus on assessing the vulnerability of (geographical) *random networks* to such disasters. We consider stochastic graph models in which nodes and links are probabilistically distributed on a plane, and model the disaster event as a circular cut that destroys any node or link within or intersecting the circle.

We develop algorithms for assessing the damage of both targeted and non-targeted (random) attacks and determining which attack locations have the expected most disruptive impact on the network. Then, we provide experimental results for assessing the impact of circular disasters to communications networks in the USA, where the network's geographical layout was modeled probabilistically, relying on demographic information only. Our results demonstrates the applicability of our algorithms to real-world scenarios.

Our novel approach allows to examine how valuable is public information about the

network's geographical area (e.g., demography, topography, economy) to an attacker's destruction assessment capabilities in the case the network's physical topology is hidden, and examine the affect of hiding the actual physical location of the fibers on the attack strategy. Thereby, our schemes can be used as a tool for policy makers and engineers to design more robust networks by placing links along paths that avoid areas of high damage cuts, or identifying locations which require additional protection efforts (e.g., equipment shielding).

Overall, the work demonstrates that using stochastic modeling and geometry (geometric probability, computational geometry) with numerical analysis techniques can significantly contribute to our understanding of network survivability and resilience.

Contents

Acknowledgments

Abstract

1	Introduction	1
1.1	Background	1
1.2	Related Work	5
2	Our Stochastic Model and Problem Formulation	13
3	Damage Evaluation Scheme	16
3.1	General Idea	16
3.2	Evaluating the Damage of a Circular Cut Algorithm	18
3.3	Numerical Accuracy and Running Time Analysis	21
3.3.1	Geometric Preliminaries	21
3.3.2	Additive Approximation	26
3.3.3	Multiplicative Approximation	29
4	Find Sensitive Locations Scheme	33
4.1	Sensitivity Map for Circular Attacks and Maximum Impact	33
4.2	Random Attacks	36
4.3	Simulations and Numerical Results	38
5	Probabilistic Stations' Locations Model	42
5.1	Background	42
5.2	Probabilistic Stations' Locations Model and Problem Statement	44
5.3	FPTAS for the Expected Vertices-Worst-Cut	45
5.4	Links and Compound Components	49
6	Conclusions and Future Work	51
	Bibliography	54

List of Figures

1.1	Color map of the the USA population density in logarithmic scale. Data is taken from [20].	4
1.2	The fiber backbone operated by a major U.S. network provider [22] studied in [31, 32, 33].	8
1.3	Line segments cuts of length approximately 120 miles optimizing <i>TEC</i> - the red cuts maximize <i>TEC</i> and the black lines are nearly worst-case cuts. Taken from [33].	9
1.4	The impact of circular cuts of radius approximately 120 miles on the <i>MFST</i> between Los Angeles and NYC. Red circles represent cuts that result in <i>MFST</i> = 0 and black circles result in <i>MFST</i> = 1. Cuts which intersect the nodes representing Los Angeles or NYC are not shown. Taken from [33].	10
1.5	The fiber backbone operated by a major U.S. network provider [9] and an example of two attacks with probabilistic effects (the link colors represent their failure probabilities). Taken from [4].	11
2.1	Example of a circular cut D with TEC of $\sum_{i=1}^{11} c_i$. The green, blue and red links depicts the different intersected-link types: $\alpha - link$, $\beta - link$, and $\gamma - link$ respectively (as defined in section 3.1). The black links are not affected by D	15
3.1	$\alpha - link$ edge illustration	17
3.2	$\beta - link$ edge illustration	17
3.3	$\gamma - link$ edge illustration	17

3.4	The area K_u of the shadow by the circular cut from the point u	20
3.5	t_1 and t_2 are tangents to a circle of radius Δ centered at u and to the circular cut. The colored areas depicts the extremum of difference for possible γ -links endpoints emanating from a point within the circle centered at u at one side of the cut.	22
3.6	The area in grey depicts the extremum of difference for possible γ -links endpoints emanating from a point within the circle centered at u at one side of the cut.	23
3.7	The perpendicular tangents from a point u on the circular cut and from a point v at distance 2Δ from the circular cut. The area in grey depicts the extremum of difference for possible γ -links endpoints emanating from u and v at one side of the cut.	23
4.1	Color map of the the USA population density in logarithmic scale. Data is taken from [20].	39
4.2	Color map of the centers of circular cuts with radius $r = 5$ (approximately 130km). Red is most harmful.	40
4.3	Color map of the centers of circular cuts with radius $r = 8$ (approximately 208km). Red is most harmful.	40
4.4	Color map of the centers of circular cuts with radius $r = 10$ (approximately 260km). Red is most harmful.	41
5.1	Example of the region $C(u)$ of node u with 3 super-level-sets.	48

Chapter 1

Introduction

1.1 Background

In the last decades, telecommunication networks have been increasingly crucial for information distribution, control of infrastructure and technological services, as well as for economies in general. Large scale malfunctions and failures in these networks, due to natural disasters, operator errors or malicious attacks pose a considerable threat to the well being and health of individuals all over the industrialized world. It is therefore of considerable importance to investigate the robustness and vulnerabilities of such networks, and to find methods for improving their resilience and stability.

The global communications infrastructure relies heavily on physical infrastructure (such as optical fibers, amplifiers, routers, and switches), making them vulnerable to *physical attacks, such as Electromagnetic Pulse (EMP) attacks, as well as natural disasters, such as solar flares, earthquakes, hurricanes, and floods* [8], [16], [17], [40], [41]. During a crisis, telecommunication is essential to facilitate the control of physically remote agents, provide connections between emergency response personnel, and eventually enable reconstitution of societal functions. Such real-world disasters happen in specific geographic locations, therefore the geographical layout of the network has a

crucial factor on their impact.

Although there has been significant research on network survivability, most previous works consider a small number of isolated failures or focus on shared risk groups (e.g., [6], [27], [34] and references therein). On the other hand, work on large-scale attacks focused mostly on cyber-attacks (viruses and worms) (e.g., [5], [18], [24]) and thereby, focus on the logical Internet topology. In contrast, we consider events causing a large number of failures in a specific geographical region, resulting in failures of network components (represented by nodes and links) which are geographically located within or intersecting the affected region.

This emerging field of *geographically correlated failures*, has started gaining attention only recently, e.g., [31, 32, 33, 30, 2, 4, 29, 28], in these works, algorithms were proposed for assessing the impact of such geographically correlated failures to a given *deterministic network*, and finding a location (or set of locations) where a disaster will cause maximum disruptive damage to the network, measured by either capacity, connectivity or flow terms. Various failure models were studied in these works, mainly in the form of a circular region failures or line-segment failures, such that any station or fiber within or intersecting the affected region is destroyed. Such failures were studied under the assumption of deterministic failures, as well as random failures (i.e. a circular region failure located randomly over the map), and probabilistic failures (e.g. gives the ability to study cases when a component's fails in proportion to its distance from the attacked point). While various failure models were studied in these works, in all of them, the network's layout is assumed to be deterministic, i.e. the geographical locations of nodes and links are known. To the best of our knowledge, spatial non-deterministic networks (e.g. spatial random networks) survivability were not studied under the assumptions of geographically correlated failures, and this work is the first to do so.

In this work we focus on the problem of geographically correlated failures in context

of *spatial random networks*, in particular, finding locations where a disaster or an attack on, will cause the most disruptive impact on the network. Before this work, that was one of the most significant unstudied problems in the field, as similar problems for deterministic networks was recently studied extensively.

We consider a stochastic model in which nodes and links are probabilistically distributed geographically on a plane. The motivation behind it is to examine the reliability of a network where we possess only partial (probabilistic) information about its geographical layout. For example, a geographically hidden network where the adversary possesses only partial information about the network topology or no knowledge at all. We show that valuable probabilistic knowledge about the network's geographical layout can be modeled from publicly available data, such as demographic maps, topographic maps, economy maps, etc. A simple example uses the fact that in densely populated areas the probability for stations (nodes) to exist is high compared to desolated areas, in which it is less likely to find many stations. Similarly, the probability for existence of a fiber (link) between two stations can be modeled as a function of the distance between the stations, the population density in the station's regions, and possibly other parameters relating to the endpoints and geography.

We study the impact of circular geographically correlated failures centered in a specific geographical location (referred as "circular cut"). The impact of such an event is on all the networks' component, i.e. stations, fibers (represented as nodes and links) intersecting the circle (including its interior). Each network component located within or intersecting the circular cut is considered to be destroyed. The damage is measured by the total expected capacity of the intersected links, or by the total number of failed network components.

It is relevant to note that in the case of a large-scale disaster or physical attack, where many links fail simultaneously, current technology will not be able to handle very large-scale re-provisioning (see for example, the CORONET project [11]). Therefore,

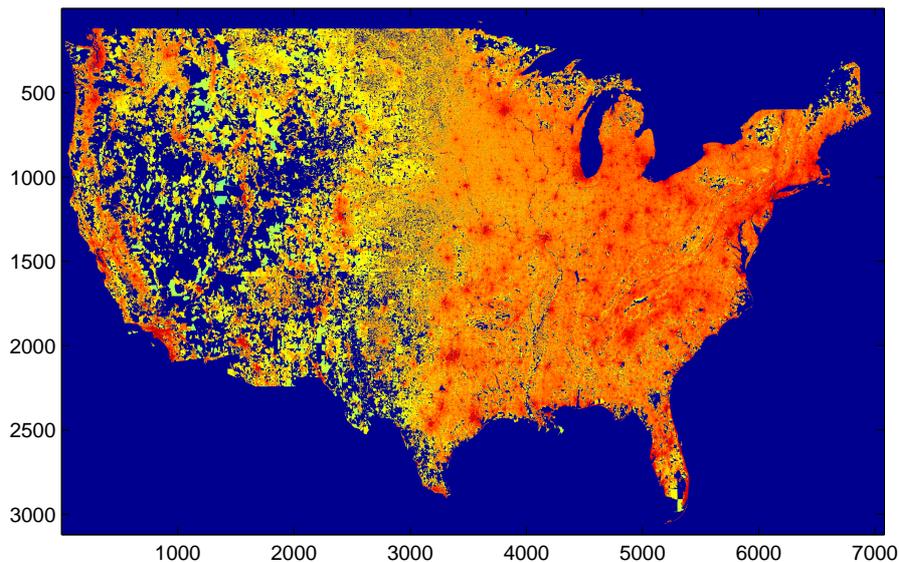


Figure 1.1: Color map of the the USA population density in logarithmic scale. Data is taken from [20].

we assume that lightpaths are static, implying that if a lightpath is destroyed, all the data that it carries is lost. However, the main data loss damage is caused by the inability to use those failed fibers, remaining the post-attack network (the network without the failed fibers) limited only to the components outside the attack's region. Thus, a scenario that the location and the size of the attack causes large-scale failures is fatal for the ability to transmit data not only from locations within the attack's region, but also from locations outside the region that are connected through fibers that go through the region.

The ability to probabilistically model a network using information such as demographic maps (see illustration in Fig. 1.1), terrain maps, economy maps, etc. can be used as an input to our algorithms for estimating the damage of attacks in different locations, and determining a location where an attack will cause maximum expected disruptive damage in terms of capacity or connectivity. This is important in assessing the expected damage from an attack by an adversary with limited knowledge. In order to design a more robust and well defended system one can consider the resilience of the

actual network topology compared to the appropriate random model, and also consider the effect of hiding the actual physical location of fibers on the attack strategy and expected damage by an adversary.

In section 4.3 we provide experimental results that demonstrate the applicability of our algorithms to estimate the expected impact of circular cuts in different locations on communication networks in the USA, and to find locations of (approximately) worst-case cuts for this model, where the network’s layout was modeled relying on demographic information only (see Fig. 1.1).

Overall in this work, we study the vulnerability of various *spatial random network* structures to geographically correlated failures. That is networks which their components’ locations (nodes and links) are distributed probabilistically on the plane. Using stochastic modeling, geometric probability and numerical analysis techniques, we demonstrate a novel approach to develop algorithms for finding locations in which circular disasters (of particular radius) cause the expected most significant destruction, allowing to identify locations which require additional protection efforts (e.g., equipment shielding). We also provide an algorithm to assess the impact of a ‘random’ circular disaster to the random network. To the best of our knowledge, our work is the first to study such geographically correlated failures in the context of *spatial random networks*. Before this work, that was one of the most significant unstudied problems in the field, as similar problems for deterministic networks was recently studied extensively.

1.2 Related Work

The issue of network survivability and resilience has been extensively studied in the past (e.g., [25, 6, 19, 42, 26, 27, 12, 5] and references therein). However, Most of these works concentrated on the logical network topology and did not consider the physical

location of nodes and links. When the logical (i.e., IP) topology is considered, widespread failures have been extensively studied [18], [24]. Most of these works consider the topology of the Internet as a random graph [5] and use percolation theory to study the effects of random link and node failures on these graphs. These studies are motivated by failures of routers due to attacks by viruses and worms rather than physical attacks.

Works that consider physical topology and fiber networks (e.g., [14], [27]), usually focused on a small number of fiber failures (e.g., simultaneous failures of links sharing a common physical resource, such as a cable, conduit, etc.). Such correlated link failures are often addressed systematically by the concept of shared risk link group (SRLG) [21] (see also section 1.1). Additional works explore dependent failures, but do not specifically make use of the causes of dependence [23], [37], [39].

In contrast with these works, we focus on failures within a specific geographical region, (e.g., failures caused by an EMP attack [17],[40]) implying that the failed components do not necessarily share the same physical resource.

A closely related theoretical problem is the *network inhibition problem* [35], [36]. Under that problem, each edge in the network has a destruction cost, and a fixed budget is given to attack the network. A feasible attack removes a subset of the edges, whose total destruction cost is no greater than the budget. The objective is to find an attack that minimizes the value of a maximum flow in the graph after the attack. However, previous works dealing with this setting and its variants (e.g., [10], [36]) did not study the removal of (geographically) neighboring links. Until the recent papers [31, 32, 33] by *Neumayer et al.*, perhaps the closest to this concept was the problem formulated in [7].

Similarly to deterministic networks, the subject of *random networks* and *survivability* was well studied in the past (e.g. [5], [12], [24], [13]), most of these works model the network as a random graph without considering the physical location of nodes and links, and focus on the robustness of the graph's structure, ignoring a physical

embedding in the *plane*. Some works considered geometry into the random model by considering distances to the graph (e.g. [15]). However, previous works dealing with such models did not study the removal of (geographically) neighboring links.

Recently, the subject of *geographically correlated failures* was proposed (e.g., [31, 32, 33]), where failures happens within a *specific* geographical region and span an extensive geographic area, disrupting all physical network equipment within the affected region. Novel works have been made recently to study the impact of various types of geographically correlated failures to a given *deterministic network* [33], [29], [30], [1], [1], [4], [3]. In these works various failure models were studied, mainly in the form of a circular region failures or line-segment failures, such that any station or fiber within or intersecting the affected region is destroyed. Such failures were studied under the assumptions of deterministic failures, as well as random failures (i.e. a circular region failure located randomly over the map), and probabilistic failures (e.g. gives the ability to study cases when a component's failure probability is proportional to its distance from the attacked point). While various failure models were studied in these works, in all of them, the network's layout is assumed to be deterministic, i.e. the geographical locations of nodes and links are known. To the best of our knowledge, our work is the first to study such geographically correlated failures in the context of *spatial random networks*.

We give a brief survey on works which are the most closest and relevant to our work. Similarly to the failure model we study in our work, most of these works consider a geographically correlated failure by a circular region of a specific radius (referred as “circular cut”), or by a line-segment of specific length (referred as “line-segment cut”). Such a cut destroys any station or fiber within or intersecting it. However, we note again, in all these works, the subject network is deterministic and given.

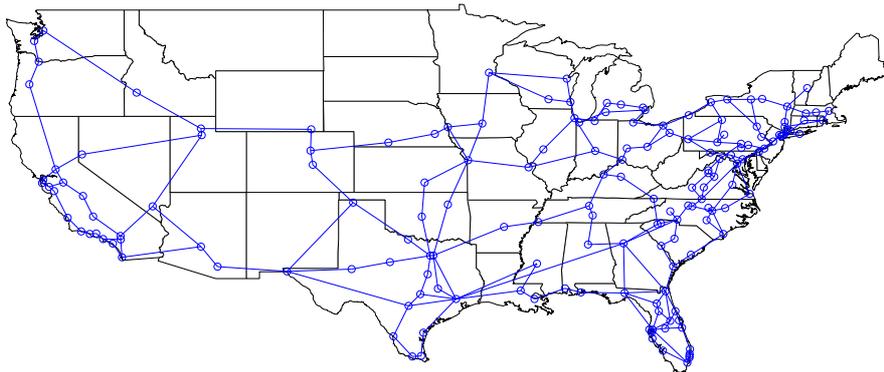


Figure 1.2: The fiber backbone operated by a major U.S. network provider [22] studied in [31, 32, 33].

Deterministic Failures In [31, 32, 33] *Neumayer et al.* formulated the problem of finding a location where a *physical* disaster or an attack on, will cause the most disruptive impact on the network in terms of capacity, connectivity and flow criterions (termed by them as the *geographical network inhibition problem*). They designed algorithms for solving the problem, and demonstrated simulation results on a U.S infrastructure (see Fig. 1.2). To the best of our knowledge, [31, 32, 33] were the first to study this problem. In these works they studied the affect of circular cuts and line segments cuts under different *performance measures*, in particular, provided polynomial time algorithms to find a *worst-case cut*, that is a cut which maximizes/minimizes the value of the performance measure. The performance measures of a cut, as studied in these works are:

- *TEC* - The total expected capacity of the intersected links with the cut.
- *ATTR* - The fraction of pairs of nodes that remain connected (also known as the average two-terminal reliability of the network)
- *MFST* - The maximum flow between a given pair of nodes s and t .
- *AMF* - The average value of maximum flow between all pairs of nodes.

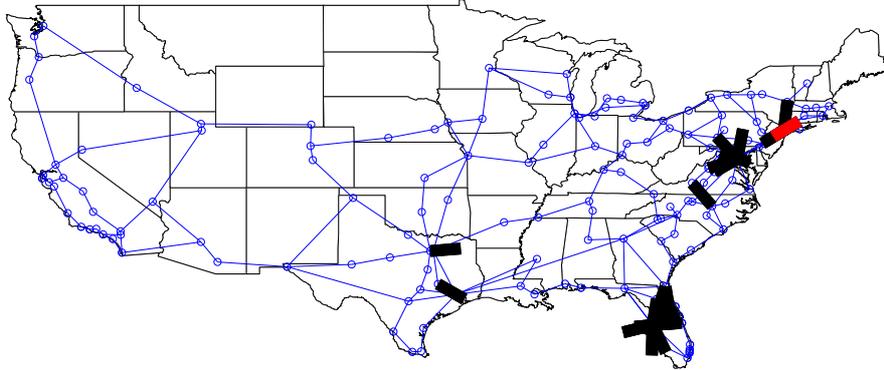


Figure 1.3: Line segments cuts of length approximately 120 miles optimizing TEC - the red cuts maximize TEC and the black lines are nearly worst-case cuts. Taken from [33].

For performance measure TEC , the worst-case cut obtains a maximum value, while for the rest, it obtains a minimum value.

We give illustrations from some of their simulation results on a U.S infrastructure which are relevant to our simulation results further in this work (in section 4.3). Fig. 1.3 shows that TEC value is large in areas of high link density, such as areas in Florida, New York, and around Dallas. Fig. 1.4 shows that cuts that minimize the MFST performance measure between Los Angeles and NYC were found close to both to Los Angeles and NYC, and the southwest area also appeared to be vulnerable. These results are relevant to the simulation results for our model as shown in section 4.3

In [1] *Agarwal et al.* presented improved runtime algorithms to the problems presented in [31, 32, 33] by the extensive use of tools from the field of *computational geometry*. In addition, they study some extensions of these problems, such as allowing *multiple* disasters to happen simultaneously and provided an approximation algorithm for the problem of finding k points in which circular disasters (of particular radius) cause the most significant destruction.

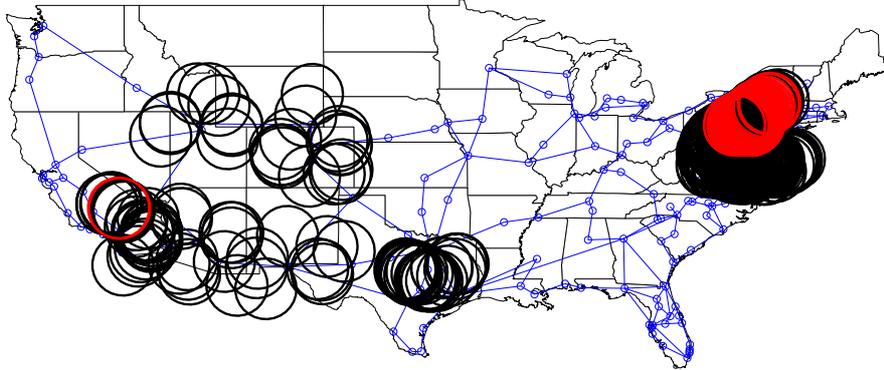


Figure 1.4: The impact of circular cuts of radius approximately 120 miles on the $MFST$ between Los Angeles and NYC. Red circles represent cuts that result in $MFST = 0$ and black circles result in $MFST = 1$. Cuts which intersect the nodes representing Los Angeles or NYC are not shown. Taken from [33].

Random Failures In [29], [30] by *Neumayer and Modiano*, they develop tools to assess the impact of a ‘random’ geographic disaster to a given network. They studied the impact of both random circular cut (i.e. a circular region failure of particular radius located randomly on the plane) and random line-segment cut.

The random location of the disaster can model failure resulting from a natural disaster such as a hurricane or collateral non-targeted damage in an EMP attack.

Intuitively, the probability of a link to fail under a random cut is proportional to the length of the link. They also compared *independent failures* versus *failures correlated to a random cut* under the $ATTR$ performance measure, assuming independent link failures such that links fail with the same probability as in the random cut case. Thus the probability a link fails is still a function of its length, however links fail independently. Their results shows that correlated failures (e.g., from a random circular cut) are fundamentally different from independent failures.

In addition they showed simple insights about network design problems in the context of random cuts. In the proposed problems the location of every node is fixed and the goal is to find a set of links most robust to some metric under some constraints.

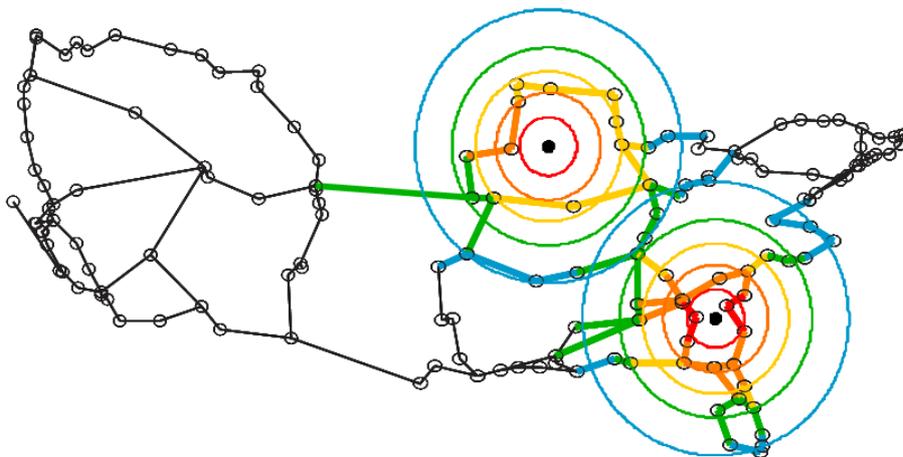


Figure 1.5: The fiber backbone operated by a major U.S. network provider [9] and an example of two attacks with probabilistic effects (the link colors represent their failure probabilities). Taken from [4].

Probabilistic Failures An extensive work on the field of geographically correlated failures, generalizing previous work and providing interesting extensions was made recently in [2, 4] by *Agarwal et al.* In these works they proposed a probabilistic failure model in the context of geographically correlated failures. In this model an attack induces a spatial probability distribution on the plane, specifying the damage probability at each location (see Fig. 1.5). They consider probability functions which are non-increasing functions of the distance between the epicenter and the component, assuming these functions have a constant description complexity. Then, they develop fully polynomial time approximation schemes to obtain the expected vulnerability of the network in terms similar to the *TEC* performance measure discussed earlier. Their algorithms also allow the assessment of the effects of several simultaneous events. Another notable extension provided in these works is the study of the vulnerability of *compound components*, that is a component consists by a finite number of network components, allowing *lightpaths* investigation, by noting a lightpath as a compound component comprised by a finite number of links.

Recently, *Agarwal, Kaplan, and Sharir* presented in [3] an improved runtime algorithms to the algorithms presented in these works, as an outcome of their result in

their paper about the complexity of the union of ‘random Minkowski sums’ of s -gons and disks (where the disks’ radius is a random non-negative number). We note that this is one of the most recent works in this field (by the time of finalizing this thesis).

The novel techniques in these works are comprised with extensive use of tools from the field of *computational geometry*, and in particular, *theory of arrangements*, *randomized algorithms*, and state of the art results in this field. In chapter 5 we provide an overview for one of the algorithms presented in these works, as it is relevant to the problems and algorithm we provide in that chapter.

Summary Summarizing the previous work in this emerging field of *geographically correlated failures*, we saw that when considering the network’s geographic layout to be *deterministic*, various failure models were deeply studied in these works. In our work we study the vulnerability of various *spatial random network* structures. That is networks which their components’ locations (nodes and links) are probabilistically distributed on the plane. Using stochastic modeling, geometric probability and numerical analysis techniques, we develop a novel approach to develop an algorithm for finding locations in which circular disasters (of particular radius) cause the expected most significant destruction. We also provide an algorithm to assess the impact of a random circular disaster to the random network. Before this work, that was one of the most significant unstudied problems in the field, as similar problems for deterministic networks was recently studied extensively. To the best of our knowledge, our work is the first to study such geographically correlated failures in the context of *spatial random networks*.

Chapter 2

Our Stochastic Model and Problem Formulation

We study the model consisting of a random network immersed within a bounded convex set $B \subseteq \mathbb{R}^2$. We consider nodes as stations and links as cables that connect stations. Stations are represented through their coordinates in the plane \mathbb{R}^2 , and links are represented by straight line segments defined by their end-points. The network is formed by a stochastic process in which the location of stations is determined by a stochastic point process. The distribution of the Poisson process is determined by the intensity function $f(u)$ which represents the mean density of nodes in the neighborhood of u . The number of nodes in a Borel set B follows a Poisson distribution with the parameter $f(B)$, i.e., the integral over the intensity of all points in the Borel set. Furthermore, the number of nodes in disjoint Borel sets are independent. An introduction to Poisson Point Processes can be found in [38].

In our model we consider a network $N = ((x', y') \sim PPP(f()), link \sim y(), c \sim H(), Rec)$ where nodes are distributed in the rectangle $Rec = [a, b] \times [c, d]$ through a *Spatial Non-Homogeneous Poisson Point Process* $PPP(f())$ where $f()$ is the *intensity function* of the *PPP*. Let $y(p_i, p_j)$ be the probability for the event of existence of

a link between two nodes located at p_i and p_j in Rec . $H(c, p_i, p_j)$ is the cumulative distribution function of the link capacity between two connected nodes, i.e. $P(C_{ij} < c) = H(c, p_i, p_j)$ where p_i and p_j are the locations of nodes i and j , respectively. It is reasonable to assume that $y(p_i, p_j)$ and $H(c, p_i, p_j)$ can be computed easily as a function of the distance from p_i to p_j and that the possible capacity between them is bounded (denoted by $\max\{capacity\}$). We assume the following: the intensity function f of the PPP , y , and the probability density function h (the derivative of H), are functions of constant description complexity. They are continuously differentiable and Riemann-integrable over Rec , which also implies that our probability functions are of bounded variation over Rec , as their derivatives receive a maximum over the compact set $Rec \subseteq \mathbb{R}^2$.

We note that Poisson process is *memoryless* and *independent*, in particular, if we look on a specific region with high intensity values then it is likely to have multiple stations in this region. However, by our assumptions, it is reasonable that if a disaster destroys some nodes in a region, then it is highly likely that it destroys other nodes located nearby in this region. There are various point process models that can be used to model a random network, for example, Matérn hard-core process and Gibbs point processes class [38], which can be used where it is wanted that a random point is less likely to occur at the location u if there are many points in the neighborhood of u or for the hard-core process where a point u is either “permitted” or “not permitted” depending whether it satisfies the hard-core requirement (e.g. far enough from all other points). Methods similar to those presented here for the Poisson process can be applied to these models, as well.

Definition 2.1 (Circular Cut). *A circular cut D , is a circle within Rec determined by its center point $p = (x_k, y_k)$ and by its radius r , where $p \in Rec_r = [a + r, b - r] \times [c + r, d - r]$.*¹ *We sometimes denote the cut as $\text{cut}(p, r)$ and Rec_r as Rec (depending*

¹For simplicity we assume that D can only appear in whole within Rec .

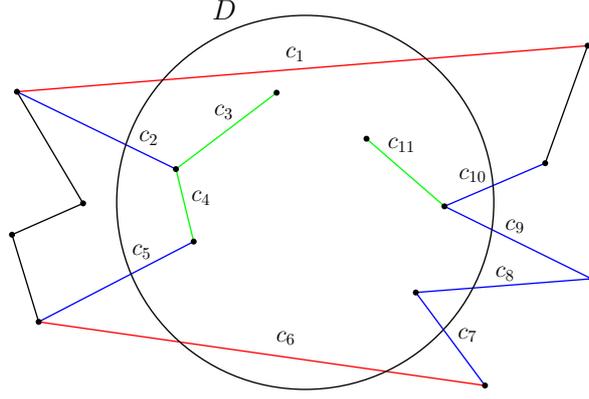


Figure 2.1: Example of a circular cut D with TEC of $\sum_{i=1}^{11} c_i$. The green, blue and red links depicts the different intersected-link types: α - link, β - link, and γ - link respectively (as defined in section 3.1). The black links are not affected by D .

on the context). Such a cut destroys all fibers (links) that intersect it (including the interior of the circle).

Our goal is to assess the vulnerability of the network to circular attacks (cuts). We consider a fiber to be destroyed (failed fiber) if it intersects the cut (including the interior of the circle), namely, the attack's influence region. The impact is measured by the *total expected capacity of the intersected links* (TEC), or by the total expected number of intersected links, which is equivalent to the previous measure when all fibers have capacity 1 (see illustration in Fig. 2.1). This will be done in three manners:

1. Provide an algorithm for evaluating the total expected capacity of the intersected links (TEC) of the network with a circular attack in a *specific location*.
2. Provide an algorithm for finding an attack location (or a set of locations) which has the highest expected impact on the network, that is, a *worst case attack* (one with the highest TEC value).
3. Provide an algorithm to assess the expected impact on the network from a *random circular attack* (such that the attack's location is probabilistically distributed).

The first is done in chapter 3, the second and third in chapter 4.

Chapter 3

Damage Evaluation Scheme

3.1 General Idea

We develop a scheme to evaluate the total expected capacity of the intersected links (TEC) of the network with a circular attack in a *specific location*. This, will be useful for the developing an algorithm that finds the attacks that have the highest expected impact on the network, as described in chapter 4. First, we present the general idea behind it. We divide the intersection of a cut D (denoted also by $\text{cut}(p, r)$) with a graph's edges into 3 independent types:

- α – *link*, is the case where the entire edge is inside $\text{cut}(p, r)$, which means both endpoints of the edge are inside $\text{cut}(p, r)$ (see illustration in Fig. 3.1).
- β – *link*, is the case where one endpoint of the edge is inside $\text{cut}(p, r)$ and the other endpoint is outside of $\text{cut}(p, r)$ (see illustration in Fig. 3.2).
- γ – *link*, is the case where both endpoints are outside of $\text{cut}(p, r)$ and the edge which connects the endpoints intersects $\text{cut}(p, r)$ (see illustration in Fig. 3.3).

Note that any intersected link with cut $D = \text{cut}(p, r)$ belongs to exactly one of the above types. Fig. 2.1 depicts a circular cut with the different types of intersected

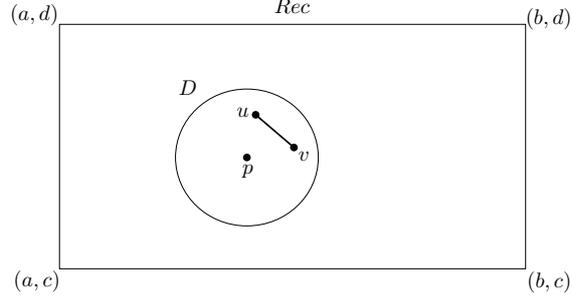


Figure 3.1: α – link edge illustration

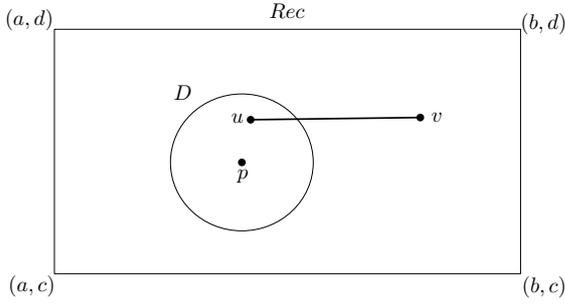


Figure 3.2: β – link edge illustration

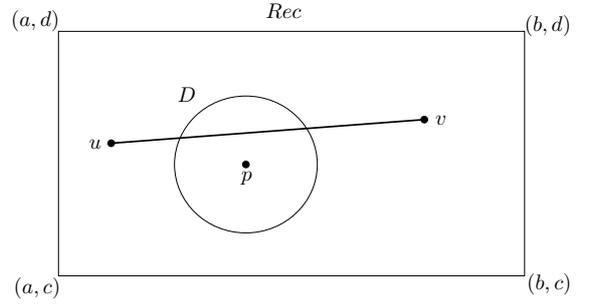


Figure 3.3: γ – link edge illustration

links. For $\sigma \in \{\alpha, \beta, \gamma\}$, let X_σ be the total capacity of all the intersected σ – link type edges with cut D , namely the damage determined by σ – links. Thus, it holds that the expected capacity of the intersected links of types α , β and γ is determined by:

$$E[X_\alpha] = \frac{1}{2} \iint_D \iint_D f(u)f(v)g(u,v) dudv \quad (3.1)$$

$$E[X_\beta] = \iint_{Rec-D} \iint_D f(u)f(v)g(u,v) dudv \quad (3.2)$$

$$E[X_\gamma] = \frac{1}{2} \iint_{Rec-D} \iint_{Rec-D} f(u)f(v)g(u,v)\mathcal{I}(u,v,D) dudv \quad (3.3)$$

where $g(u,v) = y(u,v) \int_0^{\max\{capacity\}} h(c,u,v)c dc$ is the expected capacity between two nodes at points u and v (determined by the probability of having a link between them,

times the expected capacity of this link). $\mathcal{I}(u, v, D)$ is the indicator function, giving one if the segment (u, v) intersects the circle D and zero otherwise.

Denote by $X = X_\alpha + X_\beta + X_\gamma$, the total damage determined by all the intersected links with cut D . Due to the linearity of expectation, we get that the total expected capacity of the intersected links (TEC) is $E[X] = E[X_\alpha] + E[X_\beta] + E[X_\gamma]$. Hence, it is sufficient to evaluate the expected damage caused by each of the 3 types of the intersected links separately. Summing them all together is the total expected damage caused by $D = cut(p, r)$.

3.2 Evaluating the Damage of a Circular Cut Algorithm

We present an approximation algorithm *EDCC* (see pseudo-code in Algorithm 1) for evaluating the total expected capacity of the intersected links (TEC) of the network with a circular attack (cut) D in a *specific location*. Later, we use this algorithm to find attack locations with the (approximately) highest expected impact on the network. We give two different approximation analyses for algorithm *EDCC* output. One is an additive approximation, and one is multiplicative. Although additive approximations in general are better than multiplicative, our analysis of the additive approximation depends on the maximum value of the functions $f(\cdot)$ and $g(\cdot, \cdot)$ over Rec , where $g(\cdot, \cdot)$ stands for the expected capacity between two points $u, v \in Rec$. While the maximum of $f(\cdot)$ and $g(\cdot, \cdot)$ over Rec can be high and affect the running time of the algorithm, for practical uses on "real-life" network it is usually low enough to make the running time reasonable. The multiplicative analysis which does not depend on the maximum of $f(\cdot)$ and $g(\cdot, \cdot)$ over Rec , depends on the variation bound of $f(\cdot)f(\cdot)g(\cdot, \cdot)$, namely a constant M which is an upper bound for the derivative of $f(\cdot)f(\cdot)g(\cdot, \cdot)$ over Rec . Define an

Algorithm 1: EvaluateDamageCircularCut (EDCC): Approximation algorithm for evaluating the damage of a circular cut.

```

1: capacity procedure EDCC( $N, cut(p, r), \epsilon$ )
2:  $D \leftarrow Circle(cut(p, r))$  //  $D$  is a representation of the circular cut.
3:  $Grid \leftarrow ComputeGrid(Rec, r, \epsilon)$ 
4: for every  $u, v \in Grid$  do
5:   // Compute the expected capacity between two points  $u$  and  $v$ :
    $g(u, v) \leftarrow y(u, v) \int_0^{\max\{capacity\}} h(c, u, v) c dc$ 
6: The following steps are calculated over  $Grid$ :
7:  $E[X_\alpha] \leftarrow \frac{1}{2} \iint_D \iint_D f(u) f(v) g(u, v) dudv$ 
8:  $E[X_\beta] \leftarrow \iint_{Rec-D} \iint_D f(u) f(v) g(u, v) dudv$ 
9:  $E[X_\gamma] \leftarrow \frac{1}{2} \iint_{Rec-D} f(u) \mathbf{evaluateGamma}(u, D, Grid) du$ 
10: return  $E[X_\alpha] + E[X_\beta] + E[X_\gamma]$ 
Procedure evaluateGamma( $u, D, Grid$ )
11: Create two tangents  $t_1, t_2$  to  $D$  going out from  $u$ .
12: Denote by  $K_u$  the set of points which is bounded by  $t_1, t_2, D$  and the boundary
    of the rectangle  $Rec$  (see Fig. 3.4).
13: return  $\iint_{K_u} f(v) g(u, v) dv$ 

```

additive ϵ -approximation to the TEC as a quantity \tilde{C} satisfying $C - \epsilon \leq \tilde{C} \leq C + \epsilon$ for $\epsilon > 0$, where C is the actual expected capacity intersecting the cut. Similarly, define a multiplicative ϵ, ε -approximation to the cut capacity as a quantity \tilde{C} satisfying $(1 - \varepsilon)C - \epsilon \leq \tilde{C} \leq (1 + \varepsilon)C + \epsilon$.

The algorithm uses numerical integration based on the division of Rec into squares of edge length Δ (we refer to Δ as the “grid constant”). The different approximations are pronounced in the $ComputeGrid(Rec, r, \epsilon)$ function in the algorithm, which determines the grid of constant Δ . Let $Grid$ be the set of these squares center-points. The algorithm evaluates the integrals numerically, using the points in $Grid$. Intuitively: the denser is the grid, the more accurate the results, at the price of requiring additional time to complete. In section 3.3 we examine the relation between the accuracy parameter ϵ and the grid constant Δ , this relation determines the implementation of

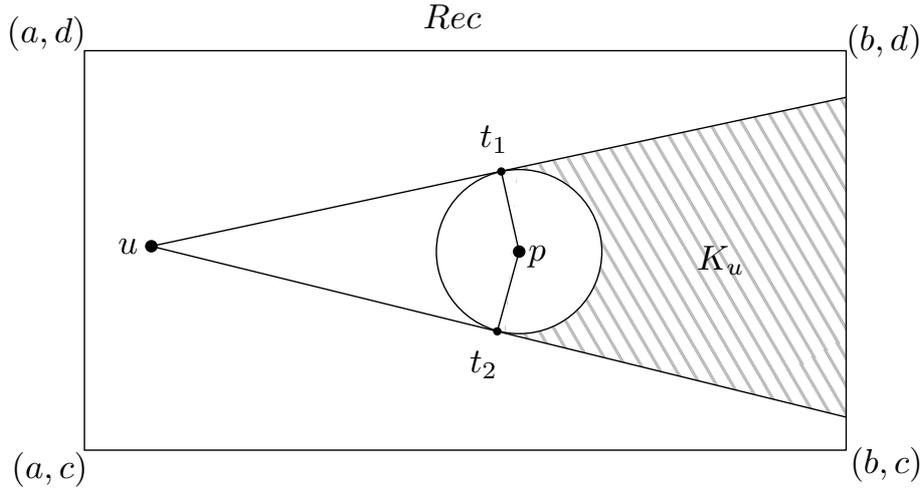


Figure 3.4: The area K_u of the shadow by the circular cut from the point u .

$ComputeGrid(Rec, r, \epsilon)$ and the running time of our algorithm.

When computing the expectation of the γ -links damage caused by a cut D , we use the following lemma:

Lemma 3.2.1. *For every node $u \in Rec \setminus D$, the set K_u (see lines 11-12 in Algorithm 1 and illustration in Fig. 3.4) satisfies: for every $v \in K_u$, (u, v) intersects D , and for every $w \in Rec \setminus (K_u \cup D)$, (u, w) does not intersect with D .*

Proof. The set, A_u bounded by both tangents and the boundary of the rectangle, containing the circle, is convex, as it is the intersection of convex sets (a triangle and a rectangle). This set, minus the union of radii from the center of the circle to both tangents is a disconnected set. The point u belongs to the connected set containing u , whereas the point v belongs to the connected set K_u , which is separated from u by the union of radii. That is, any path between u and v contained in the convex set A_u , must intersect with the union of radii, and thus with the circle.

For any point $p \in Rec$ which is not on the tangents' rays, the line segment (u, p) is either completely inside A_u or completely in $Rec \setminus A_u$ (except for point u). For any point $w_1 \in Rec \setminus A_u$, since A_u is bounded by the rays from u which are tangent to the circle, it follows that the segment (u, w_1) is completely outside A_u (except u), and thus

does not intersect with the circle. The set $S_u = A_u \setminus (K_u \cup D)$ is a star domain, such that for any point $w_2 \in S_u$ the segment (u, w_2) is completely contained in S_u , and thus does not intersect with the circle.

□

Thus, we run over every point u in *Grid* which is outside the cut, and compute all possible γ -link damage emanating from u , using procedure *evaluateGamma*($u, D, Grid$). Summing them all together and dividing by 2, due to double-counting, we get the total expected γ -link damage.

3.3 Numerical Accuracy and Running Time Analysis

3.3.1 Geometric Preliminaries

In this section we give theorems regarding the accuracy and the running time of Algorithm 1 (*EDCC*). The first theorem is in subsection 3.3.2, gives an additive bound on the error in the calculation of the damage caused by a cut of radius r using numerical integration with grid constant Δ . This gives us the relation between the accuracy parameter ϵ and the grid constant Δ . This relation gives the implementation of *ComputeGrid*(Rec, r, ϵ) in $O(1)$ by choosing Δ small enough such that the error will be not more than ϵ . Then, in Theorem 3.3.4 we give a bound on the running time of the algorithm for any accuracy parameter $\epsilon > 0$.

In subsection 3.3.3, we give a combined multiplicative and additive bound on the error for a grid constant Δ , which is different from the additive bound by being independent on the maxima of the functions f and g , thus, for a given accuracy parameters ε, ϵ , one can choose Δ which is independent on the maxima of the functions f and g , small enough such that the error will be not more than $(1 + \varepsilon)C + \epsilon$, where C is the

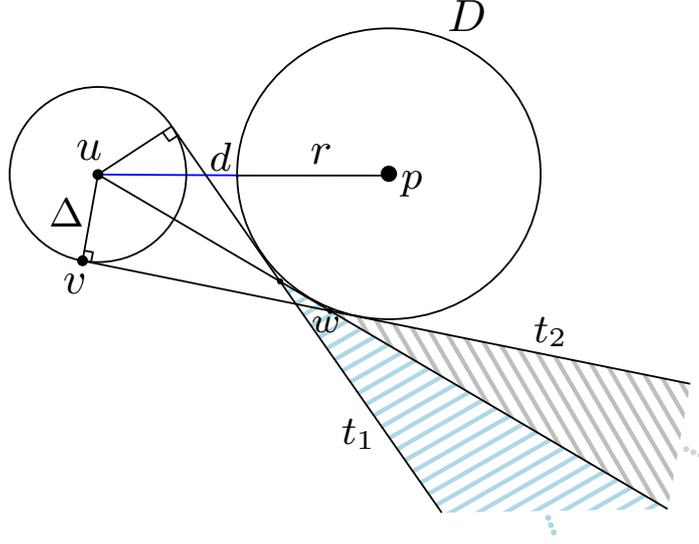


Figure 3.5: t_1 and t_2 are tangents to a circle of radius Δ centered at u and to the circular cut. The colored areas depicts the extremum of difference for possible γ -links endpoints emanating from a point within the circle centered at u at one side of the cut.

actual TEC of the cut. With this multiplicative-additive approximation we can bound the running time independently on the maxima of the functions f and g , useful for the case these maxima are high (usually in "real-life" networks it is not the case and the additive approximation will have a reasonable running time, as described in section 3.2).

We restrict our results to the case where $\Delta < r/2$, as otherwise the approximation is too crude to consider.

Some technical results are needed for proving the theorems in this section. We first notice that, given a point u at a distance d from a circular cut, and a point v at a distance Δ from u , the difference between the area of endpoints of γ -links starting at point u to those starting at point v is bounded in the area between (a) the segment from u to the boundary of Rec tangent to the cut (b) the segment from v to the boundary of Rec tangent to the cut (c) the boundary of Rec . Plus the area bounded by (a), (b) and the boundary of the circular cut. Furthermore, the extrema of the area of difference

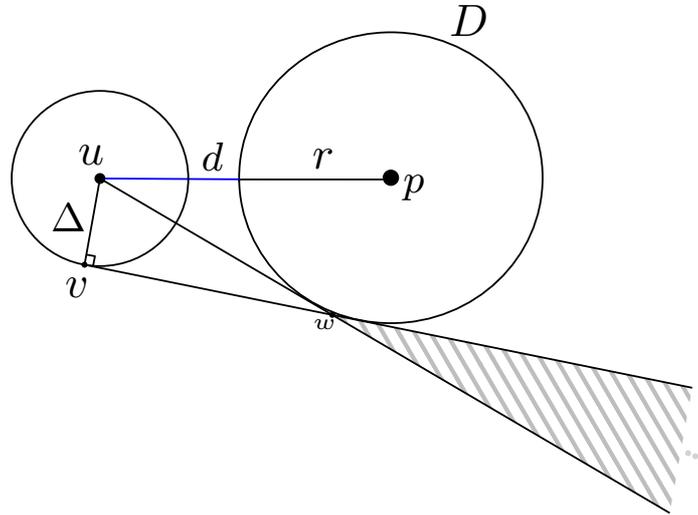


Figure 3.6: The area in grey depicts the extremum of difference for possible γ -links endpoints emanating from a point within the circle centered at u at one side of the cut.

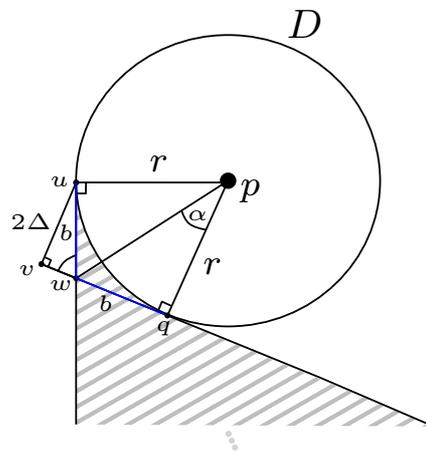


Figure 3.7: The perpendicular tangents from a point u on the circular cut and from a point v at distance 2Δ from the circular cut. The area in grey depicts the extremum of difference for possible γ -links endpoints emanating from u and v at one side of the cut.

are obtained in the cases where the segment from v to the boundary of Rec tangent to the cut is also tangent to the circle centered at u with radius Δ . See Fig. 3.5 and 3.6.

For an accurate numeric bound of the error in the theorems of this section, one should add a factor of 2 to the results of the following lemmas (to bound also the area of possible γ -links endpoints obtained by the additional tangents emanating from u and v to the other side of the circular cut).

Lemma 3.3.1. *The area bounded by (a) the tangent at a point u on the circumference of the circular cut (b) the tangent to the cut from any point v at distance 2Δ from u (c) the circumference of the cut and (d) the boundary of Rec , is bounded by $c\sqrt{\Delta}$ for some constant c . See Fig. 3.7.*

Proof. The extremum of the difference between the two tangents is when the tangent to the circle is also a tangent to the circle of radius 2Δ centered at u , i.e., when the segment of length 2Δ starting at point u is perpendicular to the tangent at its other endpoint, v . See Fig. 3.7.

Let q be the point of intersection of the second tangent and the circular cut, and let w be the point of intersection between the two tangents. Let p be the center of the circle. We have $\angle upw = \angle wpq = \alpha$ and angle $\angle uuv = 2\alpha$.

Let $b = d(u, w) = d(w, q)$ and $d(w, v) = a = \sqrt{b^2 - 4\Delta^2}$. We have $\tan \alpha = b/r$, where r is the radius of the circle, and $\tan 2\alpha = \frac{2\Delta}{\sqrt{b^2 - 4\Delta^2}}$. Using $\tan 2\alpha = 2 \tan \alpha / (1 - \tan^2 \alpha)$ one obtains $b^2 = (\Delta^2 r^2 + \Delta r^3) / (r^2 - \Delta^2)$. Thus, $r\Delta \leq b^2 \leq 2r\Delta$. Assume that both tangents hit the same side of Rec , the area of the triangle bounded by both tangents and the boundary of Rec is bounded by

$$\frac{1}{2} \mathcal{D}^2 \sin 2\alpha \leq \mathcal{D}^2 \sin \alpha \leq \mathcal{D}^2 \tan \alpha \leq \mathcal{D}^2 \sqrt{\frac{2\Delta}{r}}, \quad (3.4)$$

where \mathcal{D} is the diagonal of Rec . Now, if the tangents hit different sides of Rec , there are two cases:

If they hit perpendicular sides, then the area bounded by them and the sides is a quadrangle, taking a line segment from the corner point of Rec which is the intersection of the perpendicular sides to w we triangulate this area, obtaining two triangles. The angle, near point w of each triangle is bounded by 2α which is bounded by $\pi/2$, since we assume $\Delta < r/2$. Thus from (3.4) the area is bounded by $2\mathcal{D}^2\sqrt{\frac{2\Delta}{r}}$.

If the tangents hit parallel sides we can triangulate the area bounded by them and the sides, by taking two line segments from the corners of Rec within this area to the point w , obtaining three triangles. Using similar arguments as the previous case, we obtain that the area is bounded by $3\mathcal{D}^2\sqrt{\frac{2\Delta}{r}}$.

Finally, the area bounded by the segments uw and wq and the circular arc uq is bounded by the area of triangle Δuwq , which, in turn is bounded by $b^2 \leq 2r\Delta \leq \mathcal{D}^2\sqrt{\frac{2\Delta}{r}}$. \square

Lemma 3.3.2. *For a point u at a distance d from a circular cut $\text{cut}(p, r)$ and a given $\Delta > 0$, the area between the circumference of the cut, the boundary of Rec , the tangent to the cut from u and the tangent to the cut from a point v located at a distance Δ from u is bounded by $c\sqrt{\Delta}$ for some constant c . See Fig. 3.6.*

Proof. Let w be the point of intersection between the two tangents (from point u to the cut and from point v to the cut). We have $\sin(\angle u w v) = \Delta / (\sqrt{(r+d)^2 - r^2} + b)$, where b is the distance between the point of intersection of each of the tangents with the cut and the point w . we have $\sin(\angle u w v) = \Delta / (\sqrt{2rd + d^2} + b)$.

Now for $d < \Delta$ the point v is located within a distance of $d + \Delta < 2\Delta$ from the circumference of the cut, and the Lemma follows from Lemma 3.3.1.

If $d > \Delta$ we have $\sin(\angle u w v) < \Delta / (\sqrt{2rd}) < \Delta / (\sqrt{2r\Delta})$. Thus, if the tangents hit the same side of Rec the triangular area between the tangents and the side is bounded by $\mathcal{D}^2\sqrt{\Delta/r}$ where \mathcal{D} is the diagonal of Rec .

If the tangents hit different sides, then we can triangulate the area bounded by them

and the sides similarly as in Lemma 3.3.1. Denote by $\alpha = \angle uvv$ the angle between the two tangents. For each triangle, the angle near point w is bounded by α which is bounded by $\pi/2$ (since we assume $\Delta < r/2$). Thus, the area bounded by the tangents and Rec (which is not containing the circle) is bounded by $3\mathcal{D}^2\sqrt{\Delta/r}$.

Similarly to Lemma 3.3.1 the area between the two tangents and the circle is also bounded by $b^2 \sin(\pi - \angle uvv) = b^2 \sin(\angle uvv) < \sqrt{2r^3\Delta}$, as $b < r$ since both tangents intersect the same quadrant of the cut. \square

3.3.2 Additive Approximation

Theorem 3.3.3. *For a grid of constant Δ , a point $p \in Rec$, and the result \tilde{C} for $cut(p, r)$ obtained by Algorithm 1, it holds that $C - \epsilon < \tilde{C} < C + \epsilon$, where C is the actual TEC value for $cut(p, r)$, and $\epsilon = c_0 \cdot \sqrt{\Delta}$ for some constant $c_0 > 0$ that depends on the maximum values of $f(\cdot)$ and $g(\cdot, \cdot)$, their variation bound, the sides of Rec and the radius r .*

Proof. By standard arguments on numerical integration the error in calculating the integral over any region is bounded by $M\Delta$ times the area of integration (that is bounded by the area of Rec , $|Rec|$), where M is a bound on the variation rate for $f(\cdot)$, $g(\cdot)$, and the product $f(\cdot)f(\cdot)g(\cdot, \cdot)$ over Rec . Additionally, the cumulative error value $|C - \tilde{C}|$ consists of the following:

Any point in a grid square is within a distance of $\Delta/\sqrt{2} < \Delta$ of the grid point (square center). The additional difference in the integral over α and β links is bounded by the integral over the area of inaccuracy around the circular cut (grid squares which are partially in the cut and partially outside). This is bounded by an annulus of radii $[r - \Delta/\sqrt{2}, r + \Delta/\sqrt{2}]$ around the center of the circular cut of area $2\sqrt{2}\pi r\Delta$. Thus, we obtain an error which is bounded by $2\sqrt{2}\pi r\Delta T|Rec|$, where T is a bound on the maximum value of $f(\cdot)$, $g(\cdot)$, and the product $f(\cdot)f(\cdot)g(\cdot, \cdot)$ over Rec .

The additional error is in the calculation of γ -links, and obtained in three terms, one term is determined in the procedure $evaluateGamma(u, D, Grid)$ where the area of inaccuracy is around K_u (grid squares which are partially in K_u and partially outside). This area is bounded by $2\sqrt{2}\mathcal{D}\Delta$ where \mathcal{D} is the diagonal length of Rec . This gives an error term which is bounded by $2\sqrt{2}\mathcal{D}\Delta T|Rec|$.

The second error term in the γ -links calculations is obtained by considering the change in the functions $f(u)$ w.r.t $f(w)$, and $g(u, \cdot)$ w.r.t $g(w, \cdot)$ in $K_u \cap K_w$, for a point w within distance Δ from u . For convenience, for a point u , denote $\tilde{f}_u(v) = f(u)f(v)g(u, v)$. Taking into account the change in the integrated function $\tilde{f}_u(v)$ and $f_w(v)$ over $K_u \cap K_w$, for a point w within distance Δ from u , we obtain an error, bounded by $\iint_{K_u \cap K_w} (\tilde{f}_u(v) + M\Delta) dv - \iint_{K_u \cap K_w} \tilde{f}_u(v) dv \leq \iint_{K_u \cap K_w} M\Delta dv \leq M\Delta|Rec|$. Thus, obtaining an error bounded by $\Delta M|Rec|^2$.

The third, and the most significant error term, is obtained using Lemma 3.3.2 which implies that for any point $u \in Rec \setminus D$ in a grid square (except grid squares which are partially in D and partially outside), the area of symmetric difference between K_u and K_w for the grid point (square center) w nearest to u (such that the euclidean distance $d(u, w) < \Delta$) is bounded by $a\sqrt{\Delta}$, where a is some constant (see subsection 3.3.1), depending on the radius of the cut r and the sides of the rectangle Rec . Thus, we obtain an error bounded by $a|Rec|T\sqrt{\Delta}$.

Taking into account the errors in this numerical integration from all terms above, one obtains that the leading term in the error, as $\Delta \rightarrow 0$, is $|C - \tilde{C}| \leq const\mathcal{D}^2|Rec|T\sqrt{\frac{\Delta}{r}}$, where \mathcal{D} is the diagonal length of Rec (see subsection 3.3.1). Thus, the accuracy depends on $\sqrt{\Delta}$, as well as on the sides of the rectangle, the radius of the cut, the maxima of $f(\cdot)$, $g(\cdot, \cdot)$ and their variation bound in Rec . \square

Using Theorem 3.3.3 and the given in subsection 3.3.1, $ComputeGrid(Rec, r, \epsilon)$ function in the algorithm can be implemented by selecting the value of Δ guaranteeing

that the additive error will be at most ϵ . We now give a bound on the running time of the *EDCC* algorithm for any additive accuracy parameter $\epsilon > 0$.

Theorem 3.3.4. *For a Rec of area A with diagonal length \mathcal{D} , attack of radius r , and an additive accuracy parameter $\epsilon > 0$, the total running time of Algorithm 1 (*EDCC*) is $O\left(\frac{A^{10}\mathcal{D}^{16}T^8}{\epsilon^8 r^4} + \frac{A^{10}T^4M^4}{\epsilon^4}\right)$, where T is a bound on the maximum value of $f(\cdot)$, $g(\cdot)$, and the product $f(\cdot)f(\cdot)g(\cdot, \cdot)$ over *Rec*, and M is a bound on the variation rate for these functions over *Rec*.*

Proof. The algorithm is based on performing numerical integration over pairs of grid points (square-center points). The denser is the grid, the more pairs of points we have in the grid, thus the running time is determined by the grid constant Δ (the squares' side length) which is set by *ComputeGrid(Rec, r, ϵ)* function at the beginning of the algorithm. The number of grid points in the rectangle is A/Δ^2 . Thus the running time is at most proportional to the number of pairs of grid points, which is $O(A^2/\Delta^4)$.

Now, from the given in subsection 3.3.1 and the proof of Theorem 3.3.3 we obtain that as $\Delta \rightarrow 0$

$$\epsilon = O\left(\mathcal{D}^2AT\sqrt{\frac{\Delta}{r}} + \Delta MA^2\right)$$

Thus, when $M = O\left(\frac{\mathcal{D}^4T^2}{\epsilon r}\right)$ (reasonable in practical usages), $\Delta = \Omega\left(\frac{\epsilon^2 r}{\mathcal{D}^4 A^2 T^2}\right)$, and the total running time of the algorithm is

$$O(A^2/\Delta^4) = O\left(\frac{A^{10}\mathcal{D}^{16}T^8}{\epsilon^8 r^4}\right).$$

Otherwise, if $M = \omega\left(\frac{\mathcal{D}^4T^2}{\epsilon r}\right)$, we obtain that $\Delta = \Omega\left(\frac{\epsilon}{MA^2}\right)$, and the running time of the algorithm is

$$O\left(\frac{A^{10}M^4}{\epsilon^4}\right).$$

□

3.3.3 Multiplicative Approximation

Since the constant in Theorem 3.3.3 depends on the maximum value of the functions f and g , which may be undesirable in case these maxima are high, we have the following theorem, giving a combined additive and multiplicative accuracy with the constants independent of the maxima of f and g . Using the following Theorem 3.3.5, the function $ComputeGrid(Rec, r, \epsilon)$ can be modified to a new function $ComputeGrid(Rec, r, \epsilon, \varepsilon)$ which can be implemented by selecting the value of Δ guaranteeing that the additive error will be at most ϵ and the multiplicative error will be at most ε , as described in the following theorem.

Theorem 3.3.5. *For a grid of constant Δ , a point $p \in Rec$, and the result \tilde{C} for $D = cut(p, r)$ obtained by Algorithm 1, it holds that $(1 - \varepsilon)C - \epsilon < \tilde{C} < (1 + \varepsilon)C + \epsilon$, where C is the actual TEC value for $cut(p, r)$, for $\epsilon = c_1 \cdot \sqrt{\Delta}$, $\varepsilon = c_2 \cdot \sqrt{\Delta}$, such that c_1 and c_2 depend only on Rec , r , and M the bound on the variation of $f(\cdot)$, $g(\cdot)$, and $f(\cdot)f(\cdot)g(\cdot, \cdot)$ over Rec , but are independent on their maximum values.*

Proof. From the proof of Theorem 3.3.3, the standard error in the numerical integration over the grid depends only on the grid constant Δ , the radius of the cut r , and the bounded variation rate M of the integrated function.

For a point $u \in Rec \setminus D$ and the closest grid point $w \in Rec \setminus D$ nearest to u (with a distance at most Δ from each other), denote by R' the segment of the tangent going out from u to one side of D within Rec , similarly, denote by R the segment of the tangent going out from w to the same side of D within Rec . Denote by $R(x)$ the point on R with coordinate x , and similarly for $R'(x)$, where the x -axis is taken to be the line that does not intersect with D and going through the angle bisector for the angle between R and R' ¹. By the proof of Lemma 3.3.2 we obtain that the euclidean distance

¹There are two pairs of vertical angles by formed by the intersection of R with R' and thus, two angle bisectors, the angle bisector of one pair is intersecting the cut and the other does not intersecting the cut, we refer to the pair of which their angle bisector does not intersect the cut.

$\|R'(x) - R(x)\|$ satisfies $\|R'(x) - R(x)\| < a\sqrt{\Delta}$ for any $x \in Rec$, where a is a constant depending on the sides of Rec (see subsection 3.3.1), this is obtained directly from the proof of Lemma 3.3.2 by using similarity of triangles properties. For convenience, for a point u , denote $\tilde{f}_u(v) = f(u)f(v)g(u, v)$, and write it in Cartesian coordinates $f_u(x, y)$ (with respect to the axis described above). The TEC from β - links and γ - links going out from w is bounded by

$$\iint_{K_u \cup D} \tilde{f}_u(x, y) dx dy + \int dx \int_{R(x)-a\sqrt{\Delta}}^{R(x)} \left(\tilde{f}_u(x, y) + M(a\sqrt{\Delta} + \Delta) \right) dy,$$

thus, the difference between the TEC values of β - links and γ - links emanating from u to β - links and γ - links emanating from w is bounded by

$$\int \int_{R(x)-a\sqrt{\Delta}}^{R(x)} \left(\tilde{f}_u(x, y) + M\sqrt{\Delta}(a + \sqrt{\Delta}) \right) dx dy \quad (3.5)$$

Taking strips of length $a\sqrt{\Delta}$ we obtain the following:

$$\begin{aligned} & \int \int_{R(x)-a\sqrt{\Delta}}^{R(x)} \left(\tilde{f}_u(x, y) + M\sqrt{\Delta}(a + \sqrt{\Delta}) \right) dx dy \leq \int \int_{R(x)-2a\sqrt{\Delta}}^{R(x)-a\sqrt{\Delta}} \left(\tilde{f}_u(x, y) + M\sqrt{\Delta}(2a + \sqrt{\Delta}) \right) dx dy \\ & \leq \int \int_{R(x)-3a\sqrt{\Delta}}^{R(x)-2a\sqrt{\Delta}} \left(\tilde{f}_u(x, y) + M\sqrt{\Delta}(3a + \sqrt{\Delta}) \right) dx dy \leq \dots \end{aligned} \quad (3.6)$$

Note that

$$\iint_{K_u \cup D} \tilde{f}_u(x, y) dx dy \geq \int \int_{R(x)-2r}^{R(x)} \tilde{f}_u(x, y) dx dy \quad (3.7)$$

Thus, when integrating over $K_u \cup D$ by summing integrations of strips with length $a\sqrt{\Delta}$, at least $\frac{2r}{a\sqrt{\Delta}}$ such strips are needed.

Now, taking the average of the sequence (3.6) of length $\frac{2r}{a\sqrt{\Delta}}$ (note that we allow a fraction of an element, e.g., the sequence can be a less than one long) and (3.7) we obtain that the difference between the TEC values given in (3.5) is bounded by

$$\begin{aligned}
& \frac{a\sqrt{\Delta}}{2r} \left[\iint_{K_u \cup D} \tilde{f}_u(x, y) dx dy + \int_{R(x)-a\sqrt{\Delta}}^{R(x)} M\sqrt{\Delta}(a + \sqrt{\Delta}) dx dy \right. \\
& \left. + \cdots + \int_{R(x)-2r}^{R(x)-2r+a\sqrt{\Delta}} M\sqrt{\Delta}\left(\frac{2r}{a\sqrt{\Delta}}a + \sqrt{\Delta}\right) dx dy \right] \\
& \leq \frac{a\sqrt{\Delta}}{2r} \iint_{K_u \cup D} \tilde{f}_u(x, y) dx dy + aM\mathcal{D}\sqrt{\Delta}(r + a\sqrt{\Delta} + \Delta), \tag{3.8}
\end{aligned}$$

where \mathcal{D} is the diagonal length of Rec .

Thus, in leading terms, as $\Delta \rightarrow 0$, we obtain an error with multiplicative factor $\frac{a}{2r}\sqrt{\Delta}$ and an additive term of $aMr\mathcal{D}\sqrt{\Delta}$ between the two TEC values. Using similar techniques, one can bound the error obtained by the areas of inaccuracy:

(i) Grid squares that are partially in cut D and partially outside. This error can be bounded using similar methods when representing the integrated function in polar coordinates with respect to the center of D . Since this area is bounded by an annulus of radii $[r - \Delta/\sqrt{2}, r + \Delta/\sqrt{2}]$ around the center of D , we can take strips of length Δ . Thus, this can be bounded by a multiplicative factor of $b\Delta$ and additive term of $c\Delta$ for b and c which are depended linearly on the radius r .

(ii) For a node $u \in Rec \setminus D$, grid squares that are partially in K_u and partially outside. This area is bounded by $2\sqrt{2}\mathcal{D}\Delta$. Thus, using similar methods, we can take strips of length Δ , obtaining an error bounded by a multiplicative factor of $b'\Delta$ and additive term $c'\Delta$, for b' and c' which are depended linearly on \mathcal{D} .

Overall, the total accumulated error obtained by the modified Algorithm 1, in leading terms as $\Delta \rightarrow 0$, is with multiplicative factor $\frac{a|Rec|}{2r}\sqrt{\Delta}$ and an additive term

of $aMr\mathcal{D}|Rec|\sqrt{\Delta}$, where $a \leq \text{const}\mathcal{D}^2\frac{1}{\sqrt{r}}$ (see subsection 3.3.1). □

The following theorem gives a bound on the running time of the modified *EDCC* algorithm, independent of the maxima of f and g .

Theorem 3.3.6. *For a Rec of area A with diagonal length \mathcal{D} , attack of radius r , a multiplicative accuracy parameter $\varepsilon > 0$ and an additive accuracy parameter $\epsilon > 0$, the total running time of the modified Algorithm 1 (*EDCC*) is $O\left(\frac{A^{10}\mathcal{D}^{16}}{\varepsilon^8 r^{12}} + \frac{A^{10}\mathcal{D}^{24}M^8}{\epsilon^8 r^4}\right)$ where M is a bound on the variation rate of $f(\cdot)$, $g(\cdot)$, and $f(\cdot)f(\cdot)g(\cdot, \cdot)$ over *Rec*.*

Proof. As described in the proof of Theorem 3.3.4, the number of grid points in the rectangle is A/Δ^2 . Thus the running time is at most proportional to the number of pairs of grid points, which is $O(A^2/\Delta^4)$.

From the proof of Theorem 3.3.5, we obtain that as $\Delta \rightarrow 0$

$$\varepsilon = O\left(\frac{\mathcal{D}^2 A^2}{r^{1.5}} \sqrt{\Delta}\right),$$

$$\epsilon = O\left(M\sqrt{r}\mathcal{D}^3 A\sqrt{\Delta}\right).$$

Thus, if $\Delta = \Omega\left(\frac{\varepsilon^2 r^3}{\mathcal{D}^4 A^2}\right)$, the total running time of the algorithm is

$$O(A^2/\Delta^4) = O\left(\frac{A^{10}\mathcal{D}^{16}}{\varepsilon^8 r^{12}}\right).$$

Otherwise, $\Delta = \Omega\left(\frac{\epsilon^2 r}{M^2 \mathcal{D}^6 A^2}\right)$ and the total running time of the algorithm is

$$O\left(\frac{A^{10}\mathcal{D}^{24}M^8}{\epsilon^8 r^4}\right).$$

□

Chapter 4

Find Sensitive Locations Scheme

4.1 Sensitivity Map for Circular Attacks and Maximum Impact

In the previous chapter we showed how to evaluate the damage of a cut D in a specific location. Using Algorithm 1 (*EDCC*) one can approximate the TEC for a circular attack at every point, and in particular, find an approximated worst case attack (one with the highest TEC value).

To achieve this goal, we divide Rec into squares, forming a grid. Then, we execute *EDCC* algorithm from the previous section for every grid point (squares center-points) such that it is a center-point of a circular cut of radius r . This leads to a “network sensitivity map”, i.e., for every point we have an approximation of the damage by a possible attack in that point.

The approximated worst cut is given by taking the point with the highest TEC value among all the centers of grid squares. The actual worst case cut can be potentially located at any point within the grid squares whose centers’ calculated TECs are approximated by the *EDCC* algorithm. To guarantee an attack location with TEC of

Algorithm 2: FindSensitiveLocations (FSL): Approximation algorithm for the network sensitivity map under a circular attack

- 1: For a network N , a cut of radius r , and accuracy parameter $\epsilon > 0$, apply the function $computeGrid(Rec, r, \epsilon/2)$ to find $\Delta > 0$ such that the accuracy of Algorithm 1, given by Theorem 3.3.3 is $\epsilon/2$.
 - 2: Form a grid of constant Δ (found in step 1) from Rec . For every grid point p , apply procedure $EDCC(N, cut(p, r), \epsilon/2)$. The grid point with the highest calculated TEC is the center of the approximated worst cut.
-

at least $C - \epsilon$ where C is the TEC of the actual worst cut and an accuracy parameter $\epsilon > 0$, we provide algorithm *FSL* (see pseudo-code in Algorithm 2).

We now prove the correctness of Algorithm 2 (*FSL*).

Theorem 4.1.1. *For an accuracy parameter $\epsilon > 0$, the attack with the highest TEC value \tilde{C} found by the above algorithm satisfies $\tilde{C} \geq C - \epsilon$, where C is the TEC value for the actual worst cut.*

Proof. By Theorem 3.3.3 for every $\epsilon' > 0$ one can find a grid constant $\Delta > 0$ such that for any point $p \in Rec$ the TEC value of $cut(p, r)$ obtained by algorithm *EDCC* is within ϵ' -accuracy (additive) from the actual TEC value of $cut(p, r)$.

For any cut located at a grid point, take a cut located at some other point within the grid square, so it is within a distance $d < \Delta$ from the center of the square. The difference between the TEC for these two cases is exactly the same as in the symmetric case, where the functions f , g and the grid, are shifted a distance d in the other direction. Thus, using similar arguments as in the proof of Theorem 3.3.3, we obtain that the difference is at most ϵ' . Taking $\epsilon' = \epsilon/2$ completes the proof. \square

The following theorem determines the running time of Algorithm 4.1.2 for an additive accuracy parameter $\epsilon > 0$.

Theorem 4.1.2. *For a Rec of area A with diagonal length \mathcal{D} , attack of radius r , and an additive accuracy parameter $\epsilon > 0$, the total running time of Algorithm 2 (*FSL*)*

is $O\left(\frac{A^{15}\mathcal{D}^{24}T^{12}}{\epsilon^{12}r^6} + \frac{A^{15}M^6}{\epsilon^6}\right)$, where $T = \max_{u,v \in Rec} \{f(u)f(v)g(u,v)\}$, and M is the supremum on the variation rate of $f(\cdot)f(\cdot)g(\cdot,\cdot)$ over Rec .

Proof. The algorithm first determines a grid constant Δ such that the accuracy of Algorithm 1, given by Theorem 3.3.3 is $\epsilon/2$. Then samples a circular cut of radius r at the center point p of each grid square. For each such cut, the algorithm executes $EDCC(N, cut(p, r), \epsilon/2)$ in $O(A^2/\Delta^4)$ time. The grid has $O(A/\Delta^2)$ points. Thus the total running time is at most $O(A^3/\Delta^6)$.

Now, from the proof of Theorem 3.3.4 we obtain that when $M = O\left(\frac{\mathcal{D}^4T^2}{\epsilon r}\right)$,

$$\Delta = \Omega\left(\frac{\epsilon^2 r}{\mathcal{D}^4 A^2 T^2}\right),$$

and thus, the total running time of the algorithm is

$$O(A^3/\Delta^6) = O\left(\frac{A^{15}\mathcal{D}^{24}T^{12}}{\epsilon^{12}r^6}\right).$$

Otherwise, if $M = \omega\left(\frac{\mathcal{D}^4T^2}{\epsilon r}\right)$, we obtain that $\Delta = \Omega\left(\frac{\epsilon}{MA^2}\right)$, and the running time of the algorithm is $O\left(\frac{A^{15}M^6}{\epsilon^6}\right)$. \square

Algorithm 2 can be modified to give a multiplicative approximation which is independent of the maxima of f and g (over Rec), by using the modified function $computeGrid(Rec, r, \epsilon/2, \epsilon/2)$ and the modified $EDCC$ algorithm (as described in subsection 3.3.3). The correctness of this algorithm is obtained similarly as in Theorem 4.1.1. For the running time we provide the following theorem.

Theorem 4.1.3. *For a Rec of area A with diagonal length \mathcal{D} , attack of radius r , a multiplicative accuracy parameter $\varepsilon > 0$ and an additive accuracy parameter $\epsilon > 0$, the total running time of the modified FSL algorithm is $O\left(\frac{A^{15}\mathcal{D}^{36}M^{12}}{\epsilon^{12}r^6} + \frac{A^{15}\mathcal{D}^{24}}{\varepsilon^{12}r^{18}}\right)$,*

where M is a bound on the variation rate of $f(\cdot)$, $g(\cdot)$, and the product $f(\cdot)f(\cdot)g(\cdot, \cdot)$ over Rec .

Proof. As described in Theorem 4.1.2, the total running time is at most $O(A^3/\Delta^6)$. Now, for the modified *EDCC* algorithm described in subsection 3.3.3, we obtain by the proof of Theorem 3.3.6 that the grid constant Δ satisfies $\Delta = \Omega\left(\frac{\varepsilon^2 r^3}{\mathcal{D}^4 A^2}\right)$ or (the asymptotic minimum) $\Delta = \Omega\left(\frac{\varepsilon^2 r}{M^2 \mathcal{D}^6 A^2}\right)$

Thus, the total running time of the modified *FSL* algorithm is

$$O\left(\frac{A^{15} \mathcal{D}^{36} M^{12}}{\varepsilon^{12} r^6} + \frac{A^{15} \mathcal{D}^{24}}{\varepsilon^{12} r^{18}}\right).$$

□

4.2 Random Attacks

The impact of a *random circular cut* to *deterministic* network was recently studied (see section 1.2). The random location of a disaster can model failure resulting from a natural disaster such as a hurricane or collateral (non-targeted) damage in an EMP attack. An interesting question is to evaluate the expected impact of a random circular cut to our *stochastic* network model.

Algorithm 1 (*EDCC*) gives an approximation for the expected damage caused by a circular cut which is located at a specific point. We can use it to develop an algorithm for evaluating the damage caused by a random circular cut, using a similar concept as in Algorithm 2 (*FSL*).

For a random circular cut distributed *uniformly* over Rec , an additive approximation of the expected damage caused by such a random cut to our stochastic network model is given by

Algorithm 3: RandomCircularCutEvaluation (RCCE): Approximation algorithm for evaluating the expected damage caused by a random circular attack.

- 1: For a random network N , a random cut of radius r , distributed uniformly over Rec , and an additive accuracy parameter $\epsilon > 0$, apply the function $computeGrid(Rec, r, \epsilon/2)$ to find $\Delta > 0$ such that the accuracy of Algorithm 1, given by Theorem 3.3.3 is $\epsilon/2$.
 - 2: Form a grid of constant Δ (found in step 1) from Rec . Denote by $Grid$ the set of grid points (square centers).
 - 3: **return** $\frac{1}{|Rec|} \sum_{p \in Grid} EDCC(N, cut(p, r), \epsilon/2)$
-

$$\frac{1}{|Rec|} \iint_{Rec} EDCC(N, cut(p, r), \epsilon') dp. \quad (4.1)$$

For a given additive accuracy parameter $\epsilon > 0$, the idea is to evaluate the above equation numerically over a grid, with a grid constant Δ small enough, and an appropriate ϵ' , such that the total additive error will be at most ϵ . This is provided in Algorithm *RCCE* (see pseudo-code in Algorithm 3).

Similarly as in Algorithm 2 (*FSL*), by the proof of Theorem 4.1.1, we obtain that the grid constant Δ , chosen in Algorithm 3 guarantees that for two cuts of radius r , located at points u and v within euclidean distance at most Δ from each other, the difference in the TEC value for these two cuts is at most ϵ . Thus, when evaluating (4.1) numerically in Algorithm 3, the total error is at most $\frac{1}{|Rec|} \cdot |Rec| \cdot \epsilon = \epsilon$. Thus, the correctness of Algorithm 3 is obtained similarly as for the *FSL* algorithm, given by Theorem 4.1.1. The running time of Algorithm 3 is also the same as for the *FSL* algorithm, given by Theorem 4.1.2.

Using the multiplicative approximation with the modified *EDCC* algorithm and the modified function $computeGrid(Rec, r, \epsilon/2, \epsilon/2)$ (as described in subsection 3.3.3), a similar result with running time as in Theorem 4.1.3 can be obtained for a combined multiplicative and additive approximation of the worst-case cut, independent of the

maxima of f and g over Rec .

A random circular cut can be modeled more generally when its location is distributed with some distribution function with density ψ over Rec . Similarly to the uniform case, an approximation of the expected damage caused by such a random cut to our stochastic network model is given by

$$\iint_{Rec} \psi(p) EDCC(N, cut(p, r), \epsilon') dp. \quad (4.2)$$

For a random circular cut distributed with density function $\psi(p)$ over Rec , such that ψ is a function of *bounded variation* over Rec , A *multiplicative* approximation algorithm can be obtained by applying similar techniques as in Algorithm 3 and Algorithm 2, choosing the grid constant Δ small enough, such that multiplicative error factor will be at most ϵ . The approximation and running time in this case depends also on the *supremum* on the variation rate of ψ over Rec (which was zero in the uniform case), due its contribution to the (multiplicative) accumulated error when evaluating numerically the integral (4.2) over the grid.

4.3 Simulations and Numerical Results

In order to test our algorithms, we estimate the expected impact of circular cuts on communication networks in the USA based on a population density map. We implemented the algorithm as a C program. Data for the population density of the USA taken from [20] was taken as the intensity function $f(u)$. The supplied data is the geographic density of the USA population at a resolution of $30''$, which is approximately equivalent to 0.9km. The matrix given was of dimensions 3120×7080 . See Fig. 4.1.

In order to achieve faster running times, the data was averaged over 30×30 blocks, to give a resolution of approximately 27km. This gave a matrix of dimensions 104×236 .

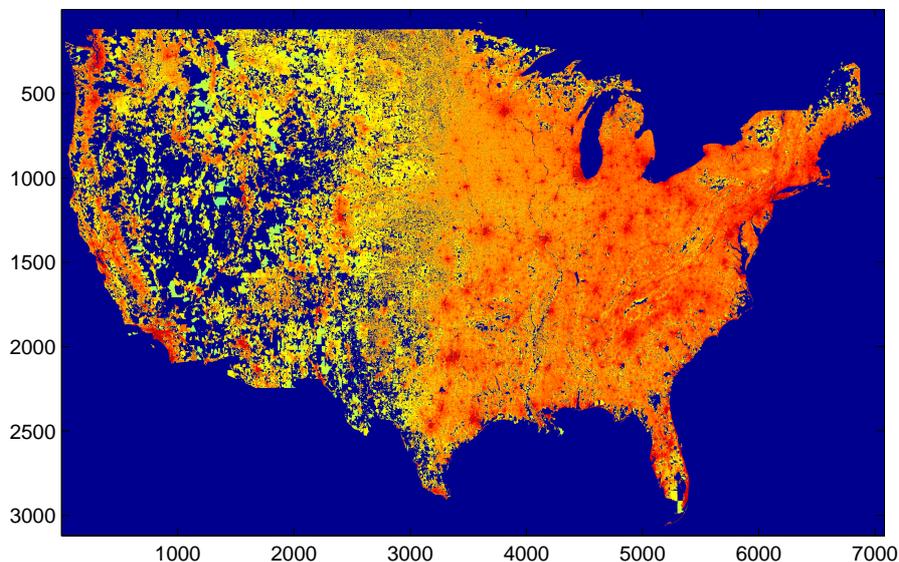


Figure 4.1: Color map of the the USA population density in logarithmic scale. Data is taken from [20].

The algorithm was then run over this intensity function. The function $y(u, v)$ was taken to be $1/\text{dist}(u, v)$, where dist is the Euclidean distance between the points, based on observations that the lengths of physical Internet connections follow this distribution [5]. The capacity probability function h was taken to be constant, independent of the distance, reflecting an assumption of standard equipment. Each run took around 24 hours on a standard Intel CPU computer.

Results for different cut radii are given in Fig. 4.2, Fig. 4.3 and Fig. 4.4. As can be seen, the most harmful cut would be around NYC, as expected, where for the larger cut radius, a cut between the east and west coast, effectively disconnecting California from the north-east, would also be a worst case scenario. Lower, but still apparent maxima are observed in the Los-Angeles, Seattle and Chicago areas.

Comparing the results to the results obtained for the fiber backbone in [32] (see Fig. 1.2) it can be seen that some similarities and some dissimilarities exist.

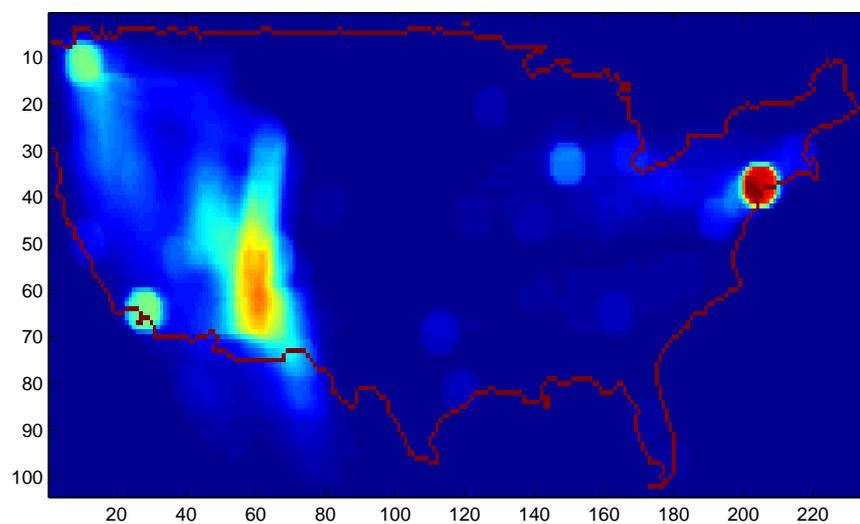


Figure 4.2: Color map of the centers of circular cuts with radius $r = 5$ (approximately 130km). Red is most harmful.

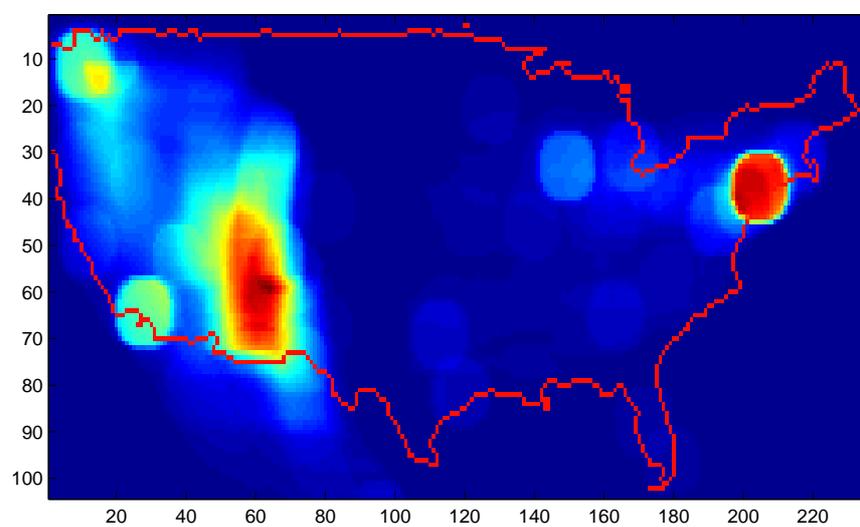


Figure 4.3: Color map of the centers of circular cuts with radius $r = 8$ (approximately 208km). Red is most harmful.

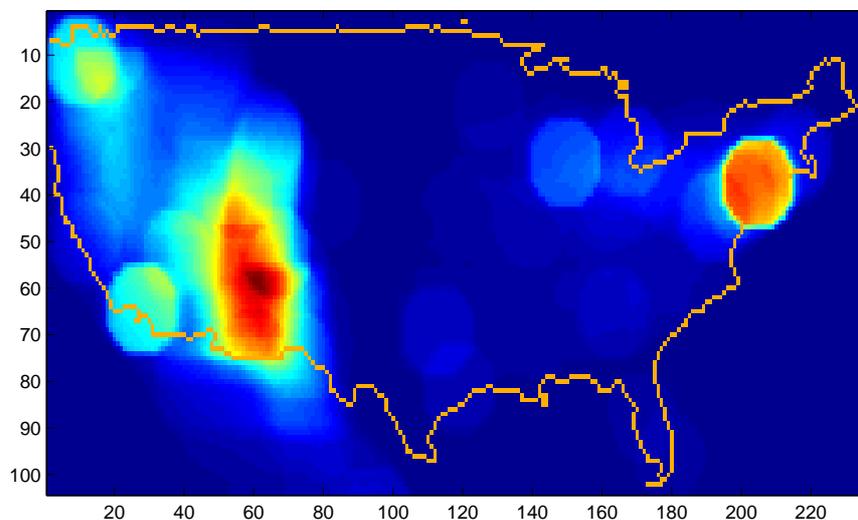


Figure 4.4: Color map of the centers of circular cuts with radius $r = 10$ (approximately 260km). Red is most harmful.

While the NYC maxima is apparent in all measures, California seems to be missing from the maxima in [32], probably reflecting the low density of fibers in that area in the map studied in [32]. As many fibers may exist that are not represented on the map in [32], it may be reasonable to assume that a cut around California would have a more significant effect than reflected there. On the other hand, several worst case cuts in Texas, and especially in Florida are apparent in [32] and are missing in the current simulation results. It seems that the effect of the narrow land bridge in Florida makes cuts there especially harmful, whereas our simulation assumes links are straight lines, which will make links to both east and west coast pass through the ocean, thus making cuts less harmful.

As a full map of communication lines is not available, it is still unclear how good of an approximation the results here supply. However, the tool can be used in conjunction with more complicated modeling assumptions, including topographic features and economic considerations to give more exact results.

Chapter 5

Probabilistic Stations' Locations Model

5.1 Background

This section relates to a different model than the model we studied in previous chapters. To avoid confusions, this chapter should be seen as another research result, not directly connected to previous chapters of this work. This section is based on a model that involves randomness only on the locations of the nodes with very specific assumptions, such as that the number of nodes is known, and it is given whether there is a link between two nodes and its capacity value. Furthermore, each node is probabilistically distributed *independently* in a specific given geographical area. The model was proposed by *Eytan Modiano* and his student *Jianan Zhang*.¹ When the model and problems were proposed, it was very preliminary and under very specific assumption, yet algorithms for the problems were not known.

¹The model was presented to me when I was visiting Prof. Eytan Modiano at the *Massachusetts Institute of Technology (MIT)* in Feb. 2013, for presenting my results about the random model we proposed (see chapter 2) as given in the previous chapters of this thesis.

We formalized a more generalized version of their model, loosing some of the assumptions, and showed that with few modifications, the algorithm proposed in [2] for the *Probabilistic Failures* model, can be used for finding worse-case cuts with respect to nodes under our model. Until the finalization of this thesis, and to the best of our knowledge, the problem of finding worst-case cuts with respect to links under this model remains open.

We note that our work on random networks as given in this thesis in the previous chapters is based on a different model (as stated in chapter 2) which present different type of randomness than the model we describe in this chapter, as in previous chapters we do not assume the number of nodes is given, and the existence of links and capacity rates are also probabilistically distributed. The model in previous chapters is motivated by the problem of assessing the vulnerability of the network when the attacker does not possess complete information about the network's geographical layout, but can model the network from data such as demographic maps, topographic maps, economy maps, etc., while the model described in this chapter is more limited by the assumptions made, assuming randomness in a more restricted manner by assuming that the attacker possesses *specific* significant information on the geographical layout of the network, but does not possess other *specific* significant information (nodes' locations).

Although our model from previous chapter may be adapted to solve the goals of the model described in this chapter, we find it interesting to tackle their model as is, which is not assuming distribution of nodes through a spatial Poisson point process, as it may rise an interesting algorithmic challenge.

In the next section we present our generalized version of their model, then in section 5.3, we demonstrate an approximation algorithm for finding worse-case cuts with respect to nodes for this model, based on a variation on the algorithm presented in [2].

5.2 Probabilistic Stations' Locations Model and Problem Statement

We consider a Communication Network in the plane \mathbb{R}^2 with probabilistic stations' locations. Consider a finite set of nodes in the plane as stations and links as cables that connect stations. Links are represented by a line segment. It is given whether there is a link between two stations i and j , and each such link is associated with its capacity c_{ij} .

Let w_u be the weight associated with a node u , which allows us to define various measures, such as the total capacity going through u , the load u can carry or other measures correlated with the geographical importance of u . In this probabilistic model, the number of nodes n is known, and each node u is probabilistically distributed *independently* within a circular region $C(u) \subseteq \mathbb{R}^2$ centered in c_u . Each circle $C(u)$ induces a spatial probability distribution with density $g_u()$, specifying the probability for the existence of u at each location.

A circular attack p_r is defined by its center-point p and its radius r . For a node u with $C(u)$ centered in c_u and an attack p_r , let $f_u(u, p_r)$ be the probability that the node u is affected by the attack p_r . One likely example is where $f_u(u, p_r)$ depends on the distance from p to c_u .

We consider only non-increasing functions $f_u(\cdot, \cdot)$ of the distance between the attack's epicenter and a network's component, and with constant description complexity. Note that $f_u(\cdot, \cdot)$ depends on the spatial probability distribution of u , thus $g_u()$ satisfies this assumption as well. Intuitively, these are functions that can be expressed as a constant number of polynomials of constant maximum degree, or simple distributions like the Gaussian distribution.

For example, for an attack p_r and node u denote the intersection area $R = p_r \cap C(u)$, so $f_u(u, p_r) = \iint_R g_u(u) du$ is the probability that u is affected by p_r . It is reasonable to

assume that $f_u(u, p_r)$ can be computed easily as a function of the distance from c_u to p_r .

Our goal is to find an attack location which has the highest expected impact on the network, where the impact is measured either by the number of failed components (e.g. nodes or links), or by the total *weight* of failed components.

For a set of components Q and an attack p_r , let $\Phi(Q, p_r)$ denote the expected number (or alternatively, weighted sum) of components of Q affected by the attack p_r . By linearity of expectation, we get

$$\Phi(Q, p_r) = \sum_{u \in Q} w_u f_u(u, p_r).$$

Our goal is to find a point $p \in \mathbb{R}^2$ that maximizes $\Phi(Q, p_r)$.

5.3 FPTAS for the Expected Vertices-Worst-Cut

We describe the idea for developing a FPTAS for finding a location with the highest impact with respect to the networks' nodes. The algorithm is based on an algorithm proposed in [2] for the *Probabilistic Failures* model. We show that with few modifications it can adapt to our model. Thus, we first give an overview about *Probabilistic Failures* model and an algorithm proposed in [2].

Probabilistic Failures Model and Algorithm Overview With difference from our proposed *Probabilistic Stations' Locations* model where each node is probabilistically distributed over a specific region, in [2] they formalized the *Probabilistic Failures* model where the network is deterministic, but the attack induces a spatial probability distribution on the plane, specifying the damage probability at each location. For a network component (i.e. node or link) q , given at attack location $p \in \mathbb{R}^2$, let $f(q, p)$

be the probability that q is affected by p . Usually $f(q, p)$ is given, or can be computed as a function of the distance from p to q . Considering probability functions $f(\cdot, \cdot)$ with similar assumptions as we stated on the probability function in the *Probabilistic Stations' Locations* model. (e.g., constant description complexity). Similarly as stated in the *Probabilistic Stations' Locations* model, each network component q is associated with weight w_q , which allows to define various measures.

The goal is the similar to other problems, to find an attack location (or a set of locations) which has the highest expected impact on the network.

For a set of components Q and an attack p_r , let $\Phi(Q, p_r)$ denote the expected number (or alternatively, weighted sum) of components of Q affected by the attack p_r . By linearity of expectation, we get

$$\Phi(Q, p_r) = \sum_{u \in Q} w_u f(u, p_r).$$

Now we want to find a point $p \in \mathbb{R}^2$ that maximizes $\Phi(Q, p_r)$. The *super-level-set* of a function $f(q, \cdot)$ with respect to a real value y is the set of points $p \in \mathbb{R}^2$ such that $f(q, p) \geq y$. Given an accuracy parameter $\epsilon \in (0, 1)$, and a set of components Q , the algorithm first determines a monotonically decreasing vector Y of level values. For each level value in Y , and each network element q , the super-level-sets of $f(q, \cdot)$ are constructed. Note that a point can be in multiple super-level-sets. Next, each super-level-set is a geometric region surrounding q . Note that the region corresponding to some level value y contains all super-level-sets of values $y' < y$. Intuitively, the number of level values determines the accuracy and the running time of the algorithm: the more level values we have, the more accurate the results are at the price of additional time to complete.

The next step is to find a point which maximizes Φ (up to some error). In the case when all weights correspond to 1, this is the point that belongs to the largest number of

super-level-sets, each might correspond to a different network element (depending on the vector Y , we might count the number of super-level-sets as a weighted sum). In the case of different weights, each region of component q can conceptually be viewed as w_q coinciding regions. Thus, having essentially the same problem. Finding a point which belongs to the largest number of regions (usually, referred to as a point of maximum *depth*) is a well-studied problem in *computational geometry*.

The algorithm presented in [2] returns a point p such that $\Phi(Q, p_r) \geq (1 - \epsilon)\Phi(Q, p^*)$ where $p^* = \operatorname{argmax}_{p \in \mathbb{R}^2} \Phi(Q, p_r)$ (namely, the optimal attack location), and $0 < \epsilon < 1$.

The basic idea of the algorithm presented in [2] is to define the *arrangement* of super-level-sets which is the subdivision of the plane into vertices, arcs and faces. Then, for every vertex p of the arrangement computing $\Phi(Q, p_r)$ and taking the maximum.

The technique appears in details in [2], and takes total time of $O(m \log m + \chi)$ where m is the number of components and χ is the total number of vertices, arcs, and faces in the arrangement. For an accuracy parameter $\epsilon > 0$, $\chi = O(\frac{m^2}{\epsilon^2} \log^2 m)$.

We note that recently, *Agarwal, Kaplan, and Sharir* presented in [3] an improved Monte Carlo algorithm for this problem that for $0 < \epsilon < 1$ returns a location p such that $\Phi(Q, p_r) \geq (1 - \epsilon)\Phi(Q, p^*)$ where $p^* = \operatorname{argmax}_{p \in \mathbb{R}^2} \Phi(Q, p_r)$ with runtime $O(\epsilon^{-4} m \log^3 m)$ where m is the number of links. As it turns out, the problem can be reduced to the problem of estimating the maximal depth in an arrangement of ‘random Minkowski sums’ of the form considered in [3].

In [2] they also gave a more generalized algorithm, allowing *compound components*. A compound component is a network component compounded from simple components such as nodes or links. Mainly motivated for assessing *lightpaths* vulnerability. A lightpath is a *compound component* consists of an ordered sequence of links, and considered to be unusable if at least one of its component links is affected by the attack. This generalization is mentioned in section 5.4, where we discuss about finding a “worst-cut” with respect to links.

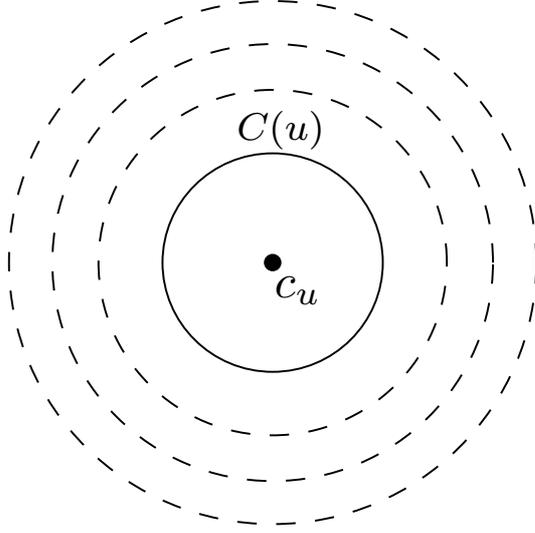


Figure 5.1: Example of the region $C(u)$ of node u with 3 super-level-sets.

Expected Vertices-Worst-Cut Algorithm under the Probabilistic Stations' Locations Model For finding a worst-case circular cut with respect to the network's nodes, i.e. an attack location which affect the largest number of nodes, or for the weighted case, the highest weighted sum of the affected nodes. The modifications needed is in the definition of the super-level-sets, as now we construct them around each node's distribution region (whereas in the original algorithm they were built around the links). Also, each node u has its own probability function $f_u(\cdot, \cdot)$ which may be different for each node. Thus, we have n probability functions $f_{u_i}(\cdot, \cdot)$, $i \in [n]$ and not necessarily identical. Therefore the definition of Φ is a bit different as stated previously.

For a node u , the *super-level-set* of a function $f_u(u, \cdot)$ with respect to a real value y is the set of points $p \in \mathbb{R}^2$ such that $f_u(u, p) \geq y$. The vector Y is being set similarly as for the *Probabilistic Failures* model. For a node u , each super-level-set induces a geometric region around $C(u)$. Figure 5.1 depicts an example of 3 super-level-sets induced by a node's u region $C(u)$. Figure 5.1 can be interpreted where $C(u)$ induces a uniform distribution for the location of u or Gaussian distribution; each contour defines

the edge of a super-level-set.

Next step is to find a point which maximizes Φ (up to some error) That is done similarly as in the original algorithm described previously.

5.4 Links and Compound Components

In previous section we talked about finding a “worst-cut” with respect to vertices. However, note that using the same technique for links (i.e. fibers) is harder, as the link is determined by its two endpoints which their locations is probabilistically distributed, thus, the description complexity of f may not be a constant, thus computing the super-level-sets for the links might be difficult. Moreover, super-level-sets of links with a common endpoint are dependent, which might make the problem even harder.

In some cases fiber links are not damaged directly (e.g. EMP attacks). Thus, an idea that can be proposed is to consider a model where fibers are not affected directly by an attack, but each fiber consists of several amplifiers which are probabilistically distributed similarly as we described for vertices. Destroying an amplifier along the fiber makes it unusable. In such cases, a link can be represented as a *compound component* of finite number of nodes. Thus a link t can be represented by a set of nodes (i.e. amplifiers) $V_t = \{t_1, t_2, \dots, t_k\}$ where each node $t_i \in V_t$ is distributed within the region $C(t_i)$ through $g_{t_i}()$ independently. Note that the geometry of the actual fibers is not considered (e.g. the actual fiber-link can resemble a curve). As for the single component, we can associate weight w_t to the compound component t .

We say that a link t is destroyed if at least one node in V_t is affected by the attack p_r . Thus, the probability that t is affected by an attack p_r is

$$f(t, p_r) = 1 - \prod_{t_i \in V_t} (1 - f_{t_i}(t_i, p_r)).$$

A problem in applying directly the algorithm from previous section for this case is that the description complexity of f may not be a constant, thus computing the super-level-sets might be difficult. However, in [2] they provide a generalized algorithm which allows compound components, their algorithm is based on applying the algorithm for the single component case we described in the previous section, and thus, it can be varied to use the modified algorithm we described in the previous section to solve this case as well.

We note that in [2] the network is deterministic and thus computing the super-level-sets for links is easier, giving the case where links are affected directly. In their paper, the generalization of the algorithm to allow compound components was mainly motivated for lightpaths investigation.

An interesting open problem rises in the model described in this chapter is if the problem of finding a “worst-cut” with respect to links (where links are affected directly) can be approximated by a FPTAS. If it is possible to develop a FPTAS, it seems that additional techniques are required, as computing or approximating the super-level-sets for the links seems to be difficult (a link is determined by its two endpoints which their location is probabilistically distributed). Furthermore, super-level-sets of links with a common endpoint are dependent, which might make the problem even harder.

Chapter 6

Conclusions and Future Work

Conclusions Motivated by applications in the area of network robustness and survivability, we focused on the problem of geographically correlated network failures. Namely, on assessing the impact of such failures to the network. While previous works in this emerging field focused mainly on deterministic networks (as described in section 1.2), we studied the problem under non-deterministic network models.

We proposed a method to stochastically model the geographical (physical) layout of communication networks, i.e. the geographical locations of nodes and links, as well as the capacity of the links. This applies to both the case where the network is derived from a random model, or to the case where the physical topology of the network is unknown (or partially known) to the attacker, who possesses some statistical information (e.g., population density, topography, economy) about the geographical area of the network, or the probability of having link between various locations.

Using tools from geometric probability and numerical analysis, we developed approximation algorithms for finding the damage caused by physical disasters (modeled by circular cuts) at different points and to approximate the location and damage of the worst-case cuts for this model. We also provided an algorithm to assess the impact of a random circular disaster (i.e. non-targeted) to our random network model, motivated

by modeling a failure resulting from a natural disaster such as a hurricane or collateral (non-targeted) damage in an EMP attack. We proved the correctness of our schemes and mentioned the trade between running time and accuracy, for both additive and multiplicative error terms.

In order to test the applicability of our model and algorithms to real-world scenarios, we applied our algorithms to approximate the damage caused by cuts in different locations to communications networks in the USA, where the network's geographical layout was modeled probabilistically, relying on demographic information (i.e. population density) only. We found a strong correlation between locations of cuts that cause high relative damage to the population density distribution over the network's region. Some of the results agree with the exact results obtained before about the fiber backbone of a major network provider in the USA and some do not (as described in section 4.3). It is yet to be determined if taking into account more links would lead to results closer to our scheme's prediction or whether the results are fundamentally different due to an inexact link model.

Our results imply that some information on the network sensitivity and vulnerabilities can be deduced from the population alone, with no information on any physical links and nodes. Thus, our schemes allows to examine how valuable is public information (such as demographic, topographic and economic information) to an attacker's destruction assessment capabilities, and examine the affect of hiding the actual physical location of the fibers on the attack strategy. Thereby, the schemes can be used as a tool for policy makers and engineers to design more robust networks by placing links along paths that avoid areas of high damage cuts, or identifying locations which require additional protection efforts (e.g., equipment shielding).

In chapter 5 we discussed about a different model than the one we studied in chapters 2, 3, 4, presenting a different type of randomness, where nodes are not distributed through a spatial Poisson point process, but the number of nodes is given

and each node is distributed in a specific region. Additionally, the model in chapter 5 assumes randomness only on the locations of the nodes but the existence of a link between two nodes and its capacity is given, and thereby assuming the attacker possess partial but significant information on the geographical layout of the network. Although our model from chapter 2 may be adapted to serve the goals of the model in chapter 5 by selecting the probability functions to resemble the model in chapter 5, we tackled the problem as is. We give an approximation algorithm to find a worst-case cut with respect to vertices for this model, based on few modifications on a recent algorithm by *Agarwal et al.* which was originally proposed for the probabilistic failures model [2].

Future Work The discussion about finding vulnerable geographic locations to physical attacks naturally leads to the question of robust network design in the face of geographical failures. Several questions are proposed, one is to investigate the effect of adding minimal infrastructure (e.g., lighting-up dark fibers) on network resilience, and determine how to use low-cost shielding for existing components to mitigate large-scale physical attacks. Another question is on designing the network's physical topology under some demand constraints (e.g., nodes that should be located within a specific region, capacity and flow demands) such that the damage by a large-scale physical attack is minimized.

Another related research direction is to develop a framework for attack and defense strategies for opponents having no knowledge of each other's strategy. Using a game-theoretic approach, study a two player zero sum game where one player (the defender) attempts to design a network as resilient to physical attacks as possible under some demand constraints, while the other player (the attacker) picks a location for the cut, without having complete knowledge about the actual network's physical structure.

Bibliography

- [1] P. K. Agarwal, A. Efrat, S. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, “Network vulnerability to single, multiple, and probabilistic physical attacks,” in *MILCOM*, 2010, pp. 1824–1829.
- [2] P. K. Agarwal, A. Efrat, S. K. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, “The resilience of wdm networks to probabilistic geographical failures,” in *INFOCOM*, 2011, pp. 1521–1529.
- [3] P. K. Agarwal, H. Kaplan, and M. Sharir, “Union of random minkowski sums and network vulnerability analysis,” in *Proceedings of the twenty-ninth annual symposium on Computational geometry*, ser. SoCG ’13. New York, NY, USA: ACM, 2013, pp. 177–186.
- [4] P. Agarwal, A. Efrat, S. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, “The resilience of wdm networks to probabilistic geographical failures,” *Networking, IEEE/ACM Transactions on*, vol. 21, no. 5, pp. 1525–1538, 2013.
- [5] A. L. Barabasi and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, October 1999.
- [6] R. Bhandari, *Survivable networks: algorithms for diverse routing*. Kluwer, 1999.

- [7] D. Bienstock, “Some generalized max-flow min-cut problems in the plane,” *Math. Oper. Res.*, vol. 16, no. 2, pp. 310–333, 1991.
- [8] J. Borland, “Analyzing the Internet collapse,” *MIT Technology Review*, Feb. 2008. [Online]. Available: <http://www.technologyreview.com/Infotech/20152/?a=f>
- [9] CenturyLink, Network Map. [Online]. Available: <http://www.centurylink.com/business/resource-center/network-maps/>
- [10] R. L. Church, M. P. Scaparra, and R. S. Middleton, “Identifying critical infrastructure: the median and covering facility interdiction problems,” *Ann. Assoc. Amer. Geographers*, vol. 94, no. 3, pp. 491–502, 2004.
- [11] G. Clapp, R. Doverspike, R. Skoog, J. Strand, and A. V. Lehmen, “Lessons learned from coronet,” in *Optical Fiber Communication Conference*. Optical Society of America, 2010, p. OWH3. [Online]. Available: <http://www.opticsinfobase.org/abstract.cfm?URI=OFC-2010-OWH3>
- [12] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, “Resilience of the Internet to random breakdowns,” *Phys. Rev. Lett.*, vol. 85, pp. 4626–4628, Nov. 2000.
- [13] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, “Breakdown of the Internet under intentional attack,” *Phys. Rev. Lett.*, vol. 86, no. 16, pp. 3682–3685, Apr 2001.
- [14] O. Crochat, J.-Y. Le Boudec, and O. Gerstel, “Protection interoperability for wdm optical networks,” *IEEE/ACM Trans. Netw.*, vol. 8, no. 3, pp. 384–395, Jun. 2000.
- [15] A. Fabrikant, A. Luthra, E. N. Maneva, C. H. Papadimitriou, and S. Shenker, “On a network creation game,” in *PODC*, 2003, pp. 347–351.
- [16] W. R. Forstchen, *One Second After*. Tom Doherty Associates, LLC, 2009.

- [17] J. S. Foster, E. Gjelde, W. R. Graham, R. J. Hermann, H. M. Kluepfel, R. L. Lawson, G. K. Soper, L. L. Wood, and J. B. Woodard, “Report of the commission to assess the threat to the United States from electromagnetic pulse (EMP) attack, critical national infrastructures,” Apr. 2008.
- [18] L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin, “Stability and topology of scale-free networks under attack and defense strategies,” *Phys. Rev. Lett.*, vol. 94, no. 18, 2005.
- [19] O. Gerstel and R. Ramaswami, “Optical layer survivability: a services perspective,” *IEEE Commun.*, vol. 38, no. 3, pp. 104–113, Mar. 2000.
- [20] Global Rural-Urban Mapping Project, Version 1, Network Data, 2000. [Online]. Available: <http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-density>
- [21] IETF Internet Working Group, “Inference of shared risk link groups,” Nov. 2001, Internet Draft. [Online]. Available: <http://tools.ietf.org/html/draft-many-inference-srlg-02>
- [22] Level 3 Communications, Network Map. [Online]. Available: <http://www.level3.com/interacts/map.html>
- [23] G. Liu and C. Ji, “Scalability of network-failure resilience: Analysis using multi-layer probabilistic graphical models,” *Networking, IEEE/ACM Transactions on*, vol. 17, no. 1, pp. 319–331, 2009.
- [24] D. Magoni, “Tearing down the internet,” *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 6, pp. 949–960, 2003.
- [25] J. Manchester, D. Saha, and S. K. Tripathi (eds.), “Protection, restoration, and disaster recovery,” *IEEE Network, Special issue*, vol. 18, no. 2, Mar.–Apr. 2004.

- [26] E. Modiano and A. Narula-Tam, “Survivable lightpath routing: a new approach to the design of WDM-based networks,” *IEEE J. Sel. Areas Commun.*, vol. 20, no. 4, pp. 800–809, May 2002.
- [27] A. Narula-Tam, E. Modiano, and A. Brzezinski, “Physical topology design for survivable routing of logical rings in WDM-based networks,” *IEEE J. Sel. Areas Commun.*, vol. 22, no. 8, pp. 1525–1538, Oct. 2004.
- [28] S. Neumayer, A. Efrat, and E. Modiano, “Geographic max-flow and min-cut under a circular disk failure model,” in *INFOCOM*, 2012, pp. 2736–2740.
- [29] S. Neumayer and E. Modiano, “Network reliability with geographically correlated failures,” in *INFOCOM*, 2010, pp. 1658–1666.
- [30] S. Neumayer and E. Modiano, “Network reliability under random circular cuts,” in *GLOBECOM*, 2011, pp. 1–6.
- [31] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, “Assessing the impact of geographically correlated network failures,” in *Proc. IEEE MILCOM’08*, Nov. 2008.
- [32] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, “Assessing the vulnerability of the fiber infrastructure to disasters,” in *Proc. IEEE INFOCOM’09*, Apr. 2009.
- [33] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, “Assessing the vulnerability of the fiber infrastructure to disasters,” *IEEE/ACM Trans. Netw.*, vol. 19, no. 6, pp. 1610–1623, 2011.
- [34] C. Ou and B. Mukherjee, *Survivable Optical WDM Networks*. Springer-Verlag, 2005.

- [35] C. A. Phillips, “The network inhibition problem,” in *Proc. ACM STOC’93*, 1993.
- [36] A. Pinar, Y. Fogel, and B. Lesieutre, “The inhibiting bisection problem,” in *Proc. ACM SPAA ’07*, 2007.
- [37] J. Spragins, “Dependent failures in data communication systems,” *Communications, IEEE Transactions on*, vol. 25, no. 12, pp. 1494–1499, 1977.
- [38] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications, 2nd Edition*, 2nd ed. Wiley, Jul. 1996.
- [39] K. Trivedi, D. S. Kim, and R. Ghosh, “Resilience in computer systems and networks,” in *Computer-Aided Design - Digest of Technical Papers, 2009. ICCAD 2009. IEEE/ACM International Conference on*, 2009, pp. 74–77.
- [40] C. Wilson, “High altitude electromagnetic pulse (HEMP) and high power microwave (HPM) devices: Threat assessments,” CRS Report for Congress, Aug. 2004. [Online]. Available: <http://www.fas.org/man/crs/RL32544.pdf>
- [41] W. Wu, B. Moran, J. Manton, and M. Zukerman, “Topology design of undersea cables considering survivability under major disasters,” in *Proc. WAINA ’09*, May 2009.
- [42] D. Zhou and S. Subramaniam, “Survivability in optical networks,” *IEEE Network*, vol. 14, no. 6, pp. 16–23, Nov.-Dec. 2000.