

Micro Failure Region Models Inducing Massive Correlated Failures on Networks Topologies

Nicolás A. Boettcher, Yasmany Prieto, Jorge E. Pezoa

▶ To cite this version:

Nicolás A. Boettcher, Yasmany Prieto, Jorge E. Pezoa. Micro Failure Region Models Inducing Massive Correlated Failures on Networks Topologies. 3rd International Conference on Information Technology in Disaster Risk Reduction (ITDRR), Sep 2018, Poznan, Poland. pp.130-141, 10.1007/978-3-030-32169-7 10. hal-02799283

HAL Id: hal-02799283 https://inria.hal.science/hal-02799283

Submitted on 8 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Micro Failure Region Models Inducing Massive Correlated Failures on Networks Topologies

Nicolás A. Boettcher^{1,2}, Yasmany Prieto¹ and Jorge E. Pezoa¹

¹ Electrical Engeeneering Department, Universidad de Concepción, Concepción, Chile {nboettcher, yprietoh, jpezoa}@udec.cl

Abstract. Natural disasters, depending on both how many occur con- currently and their size, may produce large-scale correlated failures in data network infrastructure. These failures may cause service interruptions due to disconnections of nodes in the network. Proper fault modeling is crucial to calculate network damage, determine which data paths will remain active between a pair of nodes, and thus maintain a resilient network. While in the literature different sizes of circular shapes are used to model fault regions, in this work a new fault model is proposed. The model adjusts to the granularity level established by the network opera- tor to define the size and number of concurrent fault regions. Equipped with the failure model, it is possible to observe, through disjoint paths problem, the advantages of using micro failure region models to mitigate false positive failures associated when macro failure region is used.

Keywords: Telecommunications network, Correlated failures, Disjoint paths, Micro failure region, Threat model, Geolocated failures.

1 Introduction

Telecommunication networks are a fundamental backbone for computer systems. Its elements, such as fiber optics, routers, and switches, are responsible for providing connectivity between different points through established routes where data is transmitted. Telecommunication networks are supported by an infrastructure which is deployed over large geographical areas. This infrastructure may be vulnerable to natural disasters such as earthquakes, hurricanes, tsunamis, fires and so on, or man-made disasters such as terrorist attacks, mass destruction weapons, and even accidental ones.

Infrastructure damage caused by natural disasters can be mitigated or enhanced depending on the geographical location. Simulations have shown the existence of coral reefs where tsunami impacts, can reduce the damage run-up on land on the order of 50% [1]. Even areas 30m above sea level or far from the coastal zone [2] have been defined as safe tsunami zones and are entirely immune to the tsunamis effects.

We detected, when macro failure region is reported, also micro regions without failures are included within it. This process of reporting damage in macro failure region

² Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Santiago, Chile nicolas.boettcher@mail.udp.cl

induces false positives failures to the damage detection process, losing real information about which places within the region were not affected by the disaster.

While most of the work in the literature focuses on analyzing the effects produced by large-scale disasters on the network, in this work we consider the influence of multiple small-scale correlated faults, called micro failure regions. In previous works, damage radii of different magnitudes have been used, even for the same types of natural disasters, without reaching a consensus on which is more representative for faults model. Reported disaster areas can reach radii of up to 160 km [3] while simulations have used a radii of 50 km [4] up to 800 km to represent occurrence of events [5]. Researches have simulated radii of more than 1000 km [6] and 3000 km [7] to depict different geometric shapes representing correlated fault events, where it is obtained that, for networks of dense regions, shapes like squares and circles are more harmful than narrow shapes, such as rectangles, when one seeks to obtain disjoint paths to represent the damage caused by geometric shapes.

Different failure probability functions have been dealt with in the literature, where the deterministic model, the inverse function to distance and the Gaussian [8] functions stand out. The larger the shape radius that represents the failure region, the higher the difference in the results provided by different models, but the smaller the radius, the lower the difference of these, leading to the same result.

Disasters can then be classified according to their characteristics and impact on telecommunication networks into three categories: predictable, unpredictable, and intentional [9]. To properly prepare a network for disasters, especially predictable ones, a correct risk analysis must be carried out. Three elements must be taken into account: the geographical occurrence feasibility, occurrence probability, and damage probability.

Most works in literature consider only occurrence and damage probability in network components. Their main scope is detecting vulnerabilities in the network, to identify how to improve the network resilience, based on local graph metrics such as centrality, among others [6]. Besides that other works have focused on identifying additional graph characteristics that may affect its resilience. In [10] an algorithm is proposed to diversify network components, maintaining the same graph characteristics, and assign them a location in the topology to minimize the failure probability against 0-day attacks. Nevertheless, few papers consider geographically correlated faults and also focus on identifying the most critical network components [8, 11, 12]. While these works generate random failures on the network, losing information about which places are safe from disasters, according to the best of our knowledge there are no studies that only consider high resolution empirical events that can affect network components according to its geographical location.

As defined in [13], network resilience depends on its topology and how the disaster model is represented. Also, to the best of our knowledge, we are the first to consider the number and the size of failure regions, besides their geographical location. These considerations modulate the impact of a disaster in the network resilience.

On the other hand, network resilience requires fast response times influenced by the way post-disaster new routes are selected between a couple of nodes. Choosing the new path, in real time, requires generating control traffic, waiting times, and may not have

connectivity with the controller to set up a new route. A proactive strategy, to react faster to failures, is to predefine a pair of routes between two nodes. To do so there are several [14], but the predominant one is diverse paths selection.

The diverse paths selection problem consists in finding paths between a pair of nodes that do not have common properties. The more diversity there is among routes, the less likely it is that a feature that is susceptible to failure would affect all paths equally. Disjoint path problem, which never reutilizes links with the same properties between two nodes, is the most used to choose paths between a pair of nodes. For this reason, in our work, to compare micro with macro failure regions, we assess the failure regions size consequences through the sum of the number of links used by both primary and backup paths.

2 Problem Statement

This section presents the concepts that will be employed in this work to describe the link vulnerabilities within a geolocated telecommunications network.

The fault region corresponds to a geographical failure region, represented by a circular shape that represents the area affected by a disaster like an earthquake, flood, forest fire, attack by third parties, etc. The fault region is denoted by $r_j^{a_j}$, where the *j*th fault region has a diameter denoted by d_j . The set of fault regions is denoted as follows:

$$r = \left\{ r_1^{d_1}, r_2^{d_2}, \dots, r_j^{d_j}, \dots r_{\varsigma}^{d_{\varsigma}} \right\}$$
 (1)

where ς corresponds to the fault regions number.

The *adherence distance*, denoted by ad, represents the minimum Euclidean distance at which different fault regions can be separated. If the distance between their edges is smaller than ad, they must be joined together and form a single virtual fault region with a new diameter, denoted by d'_j . The new virtual fault region is created using the smallest enclosing-circle problem [15] which consists in creating a minimum bounding circle, as in [6]. Containing fault regions with a distance less than ad. If a fault region does not have neighboring fault regions at a distance less than ad, a virtual fault region identical to it is created. Each virtual fault region is denoted by $r'_j^{at_j,ad}$ and the set of virtual fault regions is denoted as follows:

$$r'^{ad} = \left\{ r_1'^{d'_1, ad}, r_2'^{d'_2, ad}, \dots, r_j'^{d'_j, ad}, \dots, r_\varsigma'^{d'_\varsigma, ad} \right\}, \tag{2}$$

where *ad* is the adherence distance used to create the set of virtual fault regions, ς' is the virtual failure regions number and $\varsigma' \leq \varsigma$.

Algorithm 1 illustrates how to calculate whether two fault regions should be adhered or not together.

```
Algorithm 1 Check adherence between two regions based on ad value.

1: function Adherence (r_i, r_j, ad)

2: if minimumDistance(r_i, r_j) \le ad then edges

3: return MinCircle(r_i, r_j) //return virtual failure region 4: else

5: -1 //do nothing
```

Each *j*th fault region has a default state, denoted by $r_{state}[j] = 1$, $\forall j$, that enable it to group with their neighbors. When all states have been disabled, it means all have already been considered in virtual fault regions for a specific *ad*. Algorithm 2 explains how *i*th virtual fault regions for an specific *ad* are obtained from *j*th fault regions.

As can be seen in Fig. 1, initially are six fault regions of equal diameter, with value four, and after applying ad = 0 we finally get three virtual fault regions with different widths.

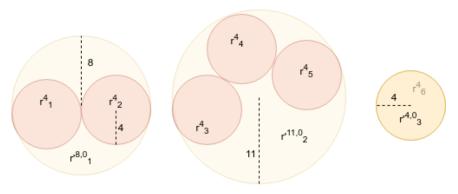


Fig. 1. Fault regions adherence for ad = 0, ζ = 6 and ζ ' = 3.

The geolocated network topology is given by a graph G=(V,E) where V is a set of nodes and E is the set of links denoted by $E=\{e_1,e_2,\ldots,e_i,\ldots,e_{\epsilon}\}$, where e_i is defined as the ith link belonging to E and ϵ is the number of links in G.

Each node represents a network router in a city. Every link depicts a straight line that joins two nodes, that represents the optical fiber that communicates a couple of nodes. By the way, every link has a secure area assigned, called buffer, formed by a hippodrome around it [16], where the Euclidean distance from any point of the link to hippodrome edge is denoted as h. The buffer allows determining if a link has the risk of being compromised by failures that invade their security area. The h must be greater than zero, as every link is at least as thick. Figure 2 shows that there may be several fault regions intersecting the hippodrome. Each one has an Euclidean distance associated to the link, called radiation distance, denoted by rd, where the smaller the value, the greater the risk of link damage. Figure 2 shows the distance of $r_1^{\prime 4,0}$ and $r_2^{\prime 9,0}$ to the link are rd_1 and rd_2 respectively. The higher is h, the further away fault regions have to be found to keep the link free from fault regions.

Algorithm 2 Create virtual fault regions based on ad value.

```
1: function SearchEnableRegion(r_{state})
        for j \leftarrow 1, len(r_{state}) do
3:
         if r_{state}[j] == 1 then return r_i
return -1
                                                                  //All r_{state} values are 0
4: procedure Create r'(r, ad)
5:
     counter \leftarrow 0
      r_{state} \leftarrow ones (1, len(r))
6:
7: start
8:
      if \sum r_{state} > 0 then
9:
10:
         r'[counter] \leftarrow SearchEnableRegion(r)
11:
         for i \leftarrow 1, len(r) do
            if r_{state}[i] == 1 then
12:
                tmp \leftarrow adherence(r[i], r'[counter], ad)
13:
               if tmp \neq -1 then
14:
15:
                  r'[\text{counter}] \leftarrow tmp
                r_{state}[i] \leftarrow 0
16:
17:
          goto start
18:
      else
19:
         return r'
```

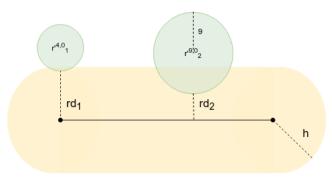


Fig. 2. Security area of the link formed by the hippodrome.

As well as considering a buffer for each link, we also considered a buffer associated to the topology equal to three times h. We defined the geographical area where the topology is located, with their respective topology buffer, as Z zone. All fault regions must be within the Z zone, otherwise they will not be considered as fault regions for the network. Figure 3 depicts the so-called Z zone, where dotted line represents the Z zone edge. The sum of the areas of r components, that are a subset of the Z zone, must be less than or equal to the Z area. This restriction allows to assure that at the moment of creating virtual fault regions, the number of these will diminish until converging in one that covers all fault regions.

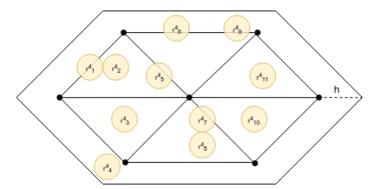


Fig. 3. Z zone elements.

3 Threat model

The simplest threat model indicates a fault must intersect the link in order for it to be out of service. The closer a failure region is to the link, due to radiation damage, the greater the failure risk. Because of this, we use a fault model where the link risk to go out of service depends on how close it is to the nearest fault region.

The use of probabilistic distribution-based models like Gaussian was not considered, because the geographic damage reported within a fault region depends exclusively on the geography in which it is located. If you have a large area whose geography is not homogeneous, regardless of the fault that occurs in the region, damage will not be symmetrical and this would require a specific distribution of damage for each reported fault region.

The failure risk of the ith link, affected by the jth virtual failure region for an specific ad is defined by $\Omega^{ad}_{i,j}$. The failure risk is zero if the virtual failure region is at a distance equal to or greater than h from the link, and increases proportionally as it approaches it, obtaining a value equal to one in case of intersecting the link.

$$\Omega_{i,j}^{ad} = \begin{cases} 1 - \frac{rd}{h} & si \ 0 \le rd \le h \\ 0 & otherwise \end{cases}$$
 (3)

Using the proposed model, the failure risk of the *i*th link product of all virtual fault regions corresponds to:

$$\Omega_i^{ad} = \max \Omega_{i,j}^{ad} \tag{4}$$

Based on failure risk value, two types of links were defined, according to their vulnerability:

- A *vulnerable link* is defined as any link with failure risk between 0 and 1 (starting at zero, and up to, but not including, one).
- An *out of service link* is defined as any link with failure risk equal to one. The out of service link concept allows remove it from *G*, so as not to consider their existence at the time of paths selection between two nodes.

Based on [17], where four risk levels are established to represent the health of the elements belonging to a network of sensors, we map Ω_i^{ad} values to a group of n=4 vulnerability levels denoted by Λ_i^{ad} , according to (5). These *n*-risk levels are represented by an exponential cost $(2^0, 2^1, ..., 2^{n-1})$, to penalize more vulnerable links

$$\Lambda_i^{ad} = \begin{cases} 8 \left(0.75 < \Omega_i^{ad} < 1 \right) & \text{extremely vulnerable} \\ 4 \left(0.5 < \Omega_i^{ad} \le 0.75 \right) & \text{highly vulnerable} \\ 2 \left(0.25 < \Omega_i^{ad} \le 0.5 \right) & \text{moderately vulnerable} \\ 1 \left(0 \le \Omega_i^{ad} \le 0.25 \right) & \text{little vulnerable} \end{cases}$$
 (5)

The higher n, the higher risk levels number and the higher the cost of using more risky links.

4 Results

To test the proposed fault model, geolocalized records of empirical fire focus data reported by NASA satellites were used [18], which were associated to micro fault regions as they have a resolution of 1 km. The data set corresponds to a period of 48 h in which at the end of this period it is assumed that they were concurrent disasters events. The measurements were made with a real-world geolocalized topology, specifically AT&T telecommunications network, as it is one of the most referenced in the literature. The topology was obtained from InternetZoo [19], but we upload it to data world repository [20] where we leave it available in multiple formats including sql, which is compatible to work with Postgres database and its PostGIS module. Only fault regions that are a subset of Z zone, calculated from the above topology, were selected. They were sequentially loaded into the database until the indicated number of faults to be considered by the network operator was reached (i.e. if the network operator choose 50 failures regions, these will load the first 50 faults listed in database). The buffer size and n allows change the risk level of each link. In our work, we used a h equal to 50 km (half of what is considered in [13]) to represent the 4-risk levels, to assign costs to the link associated to the minimal radiation distance.

To select both primary and backup paths, we used the disjoint paths problem [14]. We select the pair of routes with the minimum risk cost and we not considered out of service links in the selection process. A pair of routes were selected for different adherence distance values, from 0 to 800 km, between San Francisco and Atlanta. These cities represent the most distant nodes in AT&T network and are the ones with highest connectivity degree, allowing both to communicate through a greater number of different routes and maximizing the number of different results.

Figure 4 depicts that the number of virtual fault regions decreases as *ad* increases. It is observed that the number of fault regions is always greater than or equal to the number of virtual fault regions. Some fault regions in the *Z zone* intersect each other and merge into a single virtual fault region. As the adherence distance increases, it is more likely different fault regions will merge, looking for all a convergence to one.

Figure 4a shows the number of virtual fault regions generated out of the fault regions as the *ad* increases from 0 to 300 km. We can observe that the number of virtual fault

regions in this interval is bigger when the number of fault region is also bigger. However, in Fig. 4b, in the interval of *ad* from 300 to 800 km, we observed that there is an apparent contradiction. In some cases the number of virtual fault regions is shorter when the number of fault regions is bigger. This phenomenon is attributed to the fact that the more fault regions exist, evenly distributed, the less distance will be required for all to merge.

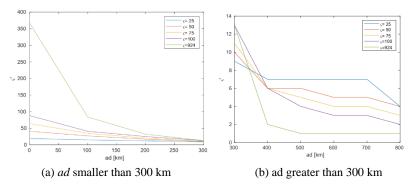


Fig. 4. Adherence distance versus virtual failure region number.

In Fig. 5 we observe the micro failure cluster effect. For $\varsigma = 924$ and values of *ad* from 200 km and greater, macro failure regions contain some nodes of the topology. When a fault region affects a node, both the node and the links attached to it fail. In this scenario, the communication with such nodes (ATLN, RLGH, CLEV) is immediately discarded.

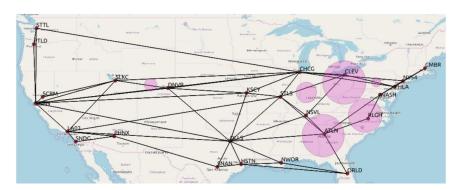


Fig. 5. Paths selection influenced by false positives failures at nodes.

Figure 6 shows the pair path length (PPL) for each ad when $\varsigma = 924$. An increase in the PPL is observed again as ad increases, where for an ad equal to 75 km the longest PPL of those obtained was reported. For an ad equal to 80 km solutions are no longer found represented by a PPL equal to zero.

Figure 7 shows the changes in pair routes selection between two points when *ad* increases in 5 km. While in Fig. 7a the route to communicate Dallas (DLLS) and Atlanta

(ATLN) is through the link DLLS-ATLN, in Fig. 7b it does so through Nashville (NSVL), increasing by one link the total number of links used by both primary and backup routes. This change happens because in the link DLLS-ATLN, by increasing the *ad* value, the risk level increases.

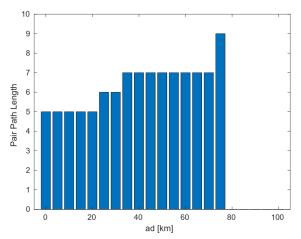


Fig. 6. Adherence distance versus Path Pair Length for each ad.

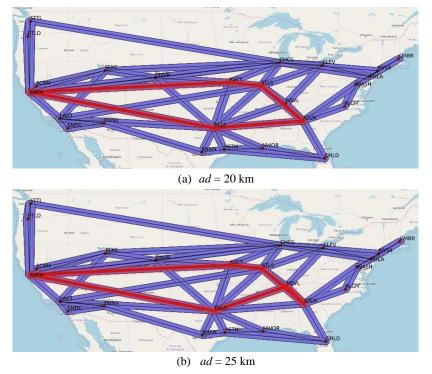


Fig. 7. Primary and backup paths selection influenced by ad.

5 Related work

In the literature there are works that use similar techniques to some of those used here [4, 6, 13], but mainly focused on determining critical region of the network to improve their infrastructure unlike us, we keep the network intact and improve the technique to choose the best routes based on the failure model precision. A technique about how to cluster failure regions is address in [6]. They propose a critical region identification model using the smaller circle problem to, once identified, improve the performance of GeoDivRP, a resilient routing protocol that considers geodiversity.

In our work we used the smallest enclosing-circle problem to cluster fault regions that were at a shorter adherence distance to the predefined by the network operator, keeping the same size faults for geographically fault regions isolated from the neighbors. We also demonstrated, through the metric PPL, that the choice of the best pair of routes is affected by the adherence distance. In [13], to detect the most vulnerable points in the network, they use hippodromes with a radius of 60 miles to represent the area of damage of each link and circles of different sizes, from 60 to 300 miles, which represent the areas of failure. They conclude, the smaller the attack radius, the smaller the number of pairs of hippodromes intersected by the fault areas affecting a smaller number of network links. Even so, the problem about how to choose the correct fault regions size is not solved.

In our work, we obtained that, for very close value ad ranges the same results can be obtained (e.g. among 0 and 20 km or among 35 and 70 km). Results depend on where the fault regions and topology are geographically located. Notwithstanding in [4] they propose a new model of regional probabilistic faults, which besides considering the distance to the epicenter also considers the value of the network components, taking into account the availability of the links, according to their repair time. Unlike them we consider all faults occur simultaneously, but on the same way out of service links are not considered for disjoint paths selection between a pair of nodes.

6 Conclusion

The main problem identified in the literature, when modeling faults, is that there is no consensus to define the fault region radius and which it repercussions brings to the routing problem. In this paper we provide a novel failure model that takes into account the decision of the network operator to determine the number and region failures size in Z zone. In this work it was evidenced that to increase in only 5 km the adherence distance, this can cause changes in the communication routes between two points. Also, for both AT&T topology and NASA dataset, we obtained that it is possible to cluster failure regions up to ad = 20 km obtaining the same results as for 1 km. Increasing ad over 20 km may generate false negatives at to the moment to select the best pair route between nodes. False negatives in macro failure regions affect to the best pair path selection.

Acknowledgments

This work was supported by CONICYT: FONDECYT Regular 2016 Folio 1160559, PCHA/Doctorado Nacional Folios 2015-21150775 and 2015-21150313.

References

- 1. Kunkel, C. M., Hallberg, R. W., Oppenheimer, M.: Coral reefs reduce tsunami impact in model simulations. Geophysical research letters 33(23) (2006).
- 2. Arias, J. P., Bronfman, N. C., Cisternas, P. C., Repetto, P. B.: Hazard proximity and risk perception of tsunamis in coastal cities: Are people able to identify their risk? PLoS one 12(10), e0186455 (2017).
- 3. Weems, T. L.: How far is far enough. Disaster Recovery Journal 16(2) (2003).
- Pa'si'c, A., Girao-Silva, R., Vass, B., Gomes, T., Babarczi, P.: FRADIR: A Novel Framework for Disaster Resilience. In: 10th International Workshop on Resilient Networks Design and Modeling (RNDM), pp. 1–7. IEEE, Norway (2018). https://doi.org/10.1109/RNDM.2018.8489828
- Rahnamay-Naeini, M., Pezoa, J. E., Azar, G., Ghani, N., Hayat, M. M.: Modeling stochastic correlated failures and their effects on network reliability. In Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN), pp. 1–6. IEEE, Hawaii (2011). https://doi.org/10.1109/ICCCN.2011.6005789
- Cheng, Y., Sterbenz, J. P.: Critical region identification and geodiverse routing protocol under massive challenges. In The 7th edition of RNDM (International Workshop on Reliable Networks Design and Modeling), pp. 14–20. IEEE, Germany (2015). https://doi.org/10.1109/RNDM.2015.7324303
- Trajanovski, S., Kuipers, F. A., Ili'c, A., Crowcroft, J., Van Mieghem, P.: Finding critical regions and region-disjoint paths in a network. IEEE/ACM Transactions on Networking (TON) 23(3), 908–921 (2015).
- 8. Agarwal, P. K., Efrat, A., Ganjugunte, S. K., Hay, D., Sankararaman, S., Zuss man, G.: The resilience of WDM networks to probabilistic geographical failures. IEEE/ACM Transactions on Networking 21(5), 1525–1538 (2013).
- 9. Mukherjee, B., Habib, M. F., Dikbiyik, F.: Network adaptability from disaster disruptions and cascading failures. IEEE Communications Magazine 52(5), 230–238 (2014).
- Boettcher, N. A., Prieto, Y., Restrepo, S. E., Pezoa, J. E.: Resilient Multiculture Network Design in the Presence of Exploit-Triggered Correlated Failures. IEEE Latin America Transactions 17(9), 2336–2344 (2018).
- Neumayer, S., Zussman, G., Cohen, R., Modiano, E.: Assessing the vulnerability of the fiber infrastructure to disasters. IEEE/ACM Transactions on Networking, 19(6), 1610–1623 (2011).
- 12. Neumayer, S., Modiano, E. Network reliability under geographically correlated line and disk failure models. Computer Networks 94, 14–28 (2016).
- Agarwal, P. K., Efrat, A., Ganjugunte, S. K., Hay, D., Sankararaman, S., Zussman, G.: Network vulnerability to single, multiple, and probabilistic physical attacks. In: Military Communications Conference (MILCOM), pp. 1824–1829. IEEE, United States (2010). https://doi.org/10.1109/MILCOM.2010.5679556
- 14. Iqbal, F., Kuipers, F. A.: Disjoint paths in networks. In: Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1–11. Wiley, (1999). https://doi.org/10.1002/net.3230040204

- 15. Skyum, S.: A simple algorithm for computing the smallest enclosing circle. Information Processing Letters 37(3), 121-125 (1991).
- Agarwal, P. K., Efrat, A., Ganjugunte, S. K., Hay, D., Sankararaman, S., Zussman, G.: The resilience of WDM networks to probabilistic geographical failures. IEEE/ACM Transactions on Networking 21(5), 1525–1538 (2013).
- 17. Zeng, Y., Zheng, G.: Delay-bounded and robust routing protocol for emergency applications using wireless sensor networks. In 2nd International Confer- ence on Advanced Computer Control (ICACC 2010), vol. 4, pp. 37–41. IEEE, China (2010). https://doi.org/10.1109/ICACC.2010.5487268
- 18. NASA's Moderate Resolution Imaging Spectroradiometer (MODIS), Active Fire Data https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data, last accessed 2018/07/10.
- 19. The Internet Topology Zoo Homepage, http://www.topology-zoo.org, last accessed 2018/12/07.
- 20. DataWorld, Author's Homepage, https://data.world/nicoboettcher/internet-zoo-topologies, last accessed 2018/12/07.