Abstract—Guaranteeing a high availability of network services is a crucial part of network management. In this study, we show how to compute the availability of network services under earthquakes, by using empirical data. We take a multi-disciplinary approach and create an earthquake model based on seismological research and historical data. We then show how to integrate this empirical disaster model into existing network resiliency models to obtain the vulnerability and availability of a network under earthquakes.

While previous studies have applied their models to ground shaking hazard models or earthquake scenarios, we compute (1) earthquake activity rates and (2) a relation between magnitude and disaster area, and use both as input data for our modeling. This approach is more in line with existing network resiliency models: it provides better information on the correlation between link failures than ground shaking hazard models and a more comprehensive view than a fixed set of scenarios.

I. INTRODUCTION

Service-level agreements between backbone network operators and clients define a minimum required availability of network services. A violation of this agreed-upon service availability may lead to a financial penalty for the network operator; hence, network operators must carefully (under)estimate the availability of their services and, if necessary, reserve protection resources and implement recovery schemes to meet the availability demands. A typical availability value is “five nines” (99.999%), which translates to an average of at most 5.26 minutes downtime per year. However, a taxonomy of Internet failures [1] has revealed that many core network outages last much longer, and are often caused by disasters that are beyond the protection schemes deployed, or due to not correctly taking into account the co-dependency and hence correlation in tightly-coupled systems. As traditional availability estimation approaches (wrongly) assume link-failure events to be independent, they overestimate the availability of network services. This can potentially cause severe penalty costs for network operators and significant degradation of the quality of experience to their customers.

One of the primary sources of link-failure correlation is the physical structure of the network; network components located in the same geographical area may fail simultaneously. Geographic failures could, for instance, be caused by natural disasters, such as earthquakes, hurricanes, or tsunamis [2], [3]. Such geographically correlated failure events are called regional failures and, due to their significant impact, are receiving increased attention [3]–[13].

This paper deals with earthquakes—one of the most common natural disasters to cause significant network outages. We take a multi-disciplinary approach, combining insights from both seismology and network resiliency to provide accurate assessments of the resiliency of communication networks to earthquakes in the studied regions. This is done in two steps: First, we create an earthquake probability model based on insights from past earthquakes (i.e. historical data). Then, we integrate this empirical disaster model into existing network resiliency models to obtain the vulnerability and availability of a network under earthquakes. Integrating empirical disaster models is an important step in determining the resilience of networks to natural disasters, but is often ignored in the existing network resiliency literature. We have applied our model to historical data from Italy, to indicate the vulnerabilities of an Italian communication network to earthquakes.

While some previous studies have applied their models on empirical earthquake data, these studies either made use of the ground shaking hazard model (see Fig. 1a) or a set of potential earthquake scenarios.

Ground shaking hazard models give the probability that, for a given site and within a time window, a given intensity level is exceeded. These models are created by aggregating over every potential future earthquake. However, by doing so for the analysis of network vulnerability, we lose information on the correlation between ground motions. Namely, we cannot distinguish between a situation where a single potential earthquake affects two cells simultaneously and a situation when each cell is affected by a different potential earthquake, and hence, not simultaneously. This makes ground shaking hazard models unsuitable for the analysis of network vulnerability, where the correlation between component failures is a crucial issue.

Considering specific earthquake scenarios, on the other
hand, provides a detailed insight into the effects of these specific potential earthquakes (e.g., one can quickly determine which network components are likely to fail together). However, as the number of earthquake scenarios under consideration is limited, and no information is provided for other scenarios, this approach lacks the comprehensive view of probabilistic seismic hazard models that takes into account all potential earthquakes.

In this paper, we take a different approach by computing earthquake activity rates (Fig. 1b). Like in ground shaking hazard models, our novel data structure considers grid cells. However, for each grid cell $C$, and each moment magnitude range $(M_w - 0.1, M_w]$, the earthquake activity rate gives the rate, and so the occurrence probability, of earthquakes with magnitude within this range and epicenter in $C$. We consider all relevant magnitude ranges, where $M_w$ is 4.6, 4.7, 4.8, ..., 8.1, thus our collection of earthquake activity rates provides a comprehensive view of all earthquakes that are strong enough to damage communication equipment.

Given the magnitude of an earthquake, we attach to its earthquake distribution model a damaged zone radius $R(M_w)$, which is based on the observed structural damage after an earthquake of magnitude in $(M_w - 0.1, M_w]$. The collection of earthquake activity rates (along with the corresponding damaged zone radii) are then plugged into a network resiliency model. The models we use assume that all links that intersect a damaged zone (namely, within radius $R(M_w)$ of the epicenter of an earthquake of magnitude $M_w$) fail, and calculate the sets of links that can fail simultaneously (and with these sets, the connectivity and availability of the network and its services).

Our main contributions are as follows:

- We argue that commonly used probabilistic ground-shaking hazard models are not suitable for assessing network resiliency to earthquakes and propose to use earthquake activity rates instead. Sections III and V describe how to calculate these earthquake activity rates, in particular given the data available for Italy (namely, a catalog of earthquakes in the last 1000 years and intensity prediction equations, correlating the damage to the distance from the epicenter, which were based on observations within Italy).

- We show, in Section IV, how to integrate earthquake activity rates into two existing network disaster resiliency models: (1) the stochastic geographically correlated link failures model in [5], and (2) the disaster set and failure states model from [4]. Although these models take a different approach towards computing the availability of network services under natural disasters, both approaches give exactly the same results when applied to our empirical earthquake model.

- We apply these methods to generate the earthquake activity rates of Italy and investigate the network resiliency of an Italian network under the different models (Section V).

- We have created an open repository containing our input and output datasets\(^1\).

To the best of our knowledge, we are the first to integrate seismological empirical data on earthquake activity rates into network resiliency models. We believe this multi-disciplinary approach is an important step towards assessing the real resiliency of communication networks to natural disasters, thus, based on our outputs, problems like resilient routing or resilient network design can be handled more efficiently.

II. RELATED WORK

Although considered as early as 1991 by D. Bienstock [15], research on geographically correlated failures, regional failures, or disasters has only really taken off in the last decade. A typical assumption is that the disaster area takes a fixed

\(^1\)https://github.com/jtapolcai/regional-srlg/tree/master/earthquake
shape, such as a disk with fixed radius [16]. By making this assumption, we can compute the amount of disasters required to disconnect two nodes [17], the most vulnerable spot(s) of the network against this type of disaster [8], or the impact after a randomly placed disaster [18]. More often than not, the disaster is also assumed to be able to occur anywhere around the network with uniform probability and link failures are happening independently of each other in the disaster zone. In reality, disasters such as earthquakes are not spread out uniformly over the whole network area, and link failures are correlated in the disaster zone.

Few papers on the resilience of networks against disasters apply their model to empirical earthquake data. L. Ma et al. considered the problem of optimally placing data center networks and content [12]. In their experiments, L. Ma et al. set the parameters based on the peak ground acceleration with an exceedance probability of 2% in 50 years of the USGS 2014 National Seismic Hazard Map\(^2\). The occurrence probabilities were set randomly.

P. N. Tran and H. Saito proposed algorithms for minimizing the sum of all end-to-end disconnection probabilities under a set of earthquake scenarios by adding new links [10], and by finding geographical routes for network links [11]. They applied their model to 5 potential future earthquake scenarios from J-SHIS\(^3\).

J. Oostenbrink and F. Kuipers proposed an efficient method to compute the distribution of a network performance measure, based on a finite set of disaster areas and occurrence probabilities [4]. To obtain these disaster areas, they made use of the same type of earthquake scenarios as in [10] and [11], but took a more deterministic approach by putting a threshold on the seismic intensity; links crossing any cell with an intensity above the threshold fail. By merging these cells, each scenario can be described by a single, potentially disjoint, disaster area.

J. Tapolcai et al. presented a tractable stochastic model for regional link failures taking into account the correlation between the failures [5]. In case of any link set \(S\), the output of the model tells the probability that when the next disaster happens, the set of failed links will be exactly \(S\).

### III. On Seismic Hazard

#### A. Earthquake Activity Rates

**Earthquake activity rates** are the occurrence rates of earthquake events as a function of space, time, and magnitude. These rates are a primary component of any seismic hazard analysis, including our study of network earthquake resiliency. To obtain the earthquake activity rates, we first need to define an earthquake source model, an area or an active fault that could host earthquakes as testified by the recent and instrumental seismic activity, historical seismicity, geomorphological evidence, and regional tectonics.

In seismic hazard analysis, there are different approaches to defining earthquake source models, e.g.:

1. \(\lambda(M) = \frac{\exp (-\beta M_0) - \exp (-\beta M_c)}{\exp (-\beta M_0) - \exp (-\beta M_a)}\)
for all magnitudes $M$ between $M_0$ (minimum magnitude) and $M_u$ (upper or maximum magnitude); otherwise $\lambda(M)$ is 0. Additionally, $\lambda_0$ is the smoothed rate $S_r$, of earthquakes at $M_w = 4.5$ and $\beta = b\ln(10)$, where $b$ is the $b$-value of the magnitude-frequency distribution.

Once all possible earthquake source models that could affect the hazard of an investigated site are defined, it is possible to define the earthquake activity rates of each source by an earthquake probability distribution, known as magnitude-frequency distribution.

A magnitude-frequency distribution indicates the probability that an earthquake, of a size within the upper and lower bound of the distribution, may occur anywhere inside the source during a specified period of time. The upper and lower bound represent, respectively, the maximum magnitude, or the largest earthquake, considered for a particular source, and the minimum magnitude, or threshold value, below which there is no engineering interest or lack of data.

In this work, we focus on Italy (a European country with a high seismic hazard), and we use the historical catalog of earthquakes [14] depicted in Fig. 2 as input data to model the occurrence of moderate-to-large ($M_w > 4.5$) earthquakes for Italy via a standard smoothed approach [22] (for more details see Section V). While this does give us the rates for all combinations of epicenters and magnitudes, we still need the relation between magnitude and disaster area to be able to apply these rates to the network resiliency models.

B. The Radius of the Damaged Zone

The only earthquake effect that can be quantified at the scale of the whole country is ground shaking. Any other earthquake effect, such as induced landslides, the liquefaction potential, the coseismic fault displacement, or tsunamis, require a site investigation. Shaking intensity is localized and is generally diminishing with distance from the earthquake’s epicenter. Although the intensity can be amplified in sedimentary basins, with certain kinds of unconsolidated soils, and due to particular topographic features, at the scale of the whole country we can assume conditions are relatively homogeneous, and the seismic intensity only depends on the distance from the earthquake epicenter.

In Italy, the Mercalli-Cancani-Sieberg (MCS) scale [24] is used to measure the intensity of shaking at any given location due to an earthquake. Intensity scales empirically categorize the intensity of shaking based on the effects reported by observers and are adapted for the effects that might be observed in a particular region. The MCS scale ranges from I to XII; the lower degrees of the MCS scale generally deal with how the earthquake is felt by people, while the higher numbers of the scale are based on observed structural damage. Structural damage starts being observed at an MCS intensity of VI. Thus, we assume all links (and nodes) inside the area with an MCS intensity of at least VI are damaged, while all components outside of this area remain functioning. Note that as we assume the intensity only depends on the distance from the epicenter, this disaster area takes the form of a disk.

Thus, to obtain all disaster areas, we now only need the disaster area radius for each magnitude $M_w \in \{4.6, 4.7, \ldots, 8.1\}$. For this purpose, we used an intensity prediction equation for Italy [25], where the expected intensity $I$ at a site located at epicentral distance $R$ is:

$$I = I_E - 0.0086(D - h) - 1.037(\ln D - \ln h),$$

where $D = \sqrt{R^2 + h^2}$ is a sort of hypocentral distance, and $h = 3.91\text{km}$ represents the hypocentral depth, which may be viewed as the average depth of the apparent radiating source [25]. $I_E$ is the average expected intensity at the epicenter for a given earthquake and can be computed by using empirical relationships with the moment magnitude $M_w$. We computed the average expected intensity for a wide range of moment magnitudes (4.6 to 8.1), using the following equation taken from [25]:

$$I_E = 1.343 + 1.621 \times M_w$$

Then, for each $I_E$, we computed the expected intensity of $I$ with Eq. (3) for an epicentral distance $R$ that ranges from 2 to 200km. In this way, it is possible to compute for each $I_E$, and so $M_w$, the site-distance $R(M_w)$ from the epicenter of the desired intensity level (in our case equal to VI, Fig. 3).

IV. MODEL OF CORRELATED LINK FAILURES CAUSED BY EARTHQUAKES

The network is modeled as an undirected connected geometric graph $G = (V, E)$, with $n = |V|$ nodes and $m = |E|$ links embedded on the Earth’s surface. The links can be either geodesics or a series of adjacent geodesics. Note that our algorithms are mostly linear in the network size, thus a link represented by adjacent line segments can be modeled as a series of 2-dimensional points (see Fig. 4).

A. Stochastic Modeling of Earthquakes

To model regional failures caused by an earthquake, we apply the findings of [4] and [5]. As in the present setting, both approaches translate to the same model, we will only present the common way of modeling earthquakes. In this setting, we are investigating the failures caused by the next earthquake, thus we assume there is exactly one earthquake in the investigated time period. The earthquake is identified with its epicenter and moment magnitude:
We index the cell grid such that \( E \) point in \( c \) which are independent Poisson arrival processes with parameters \( \lambda_i,j,M \). We claim that under the assumption of exponential distribution with parameter \( \lambda_i,j,M \): the radius of the area where network elements fail at magnitude \( M \) (see Fig. 3). In this example, we set the intensity threshold to a relatively high \( \lambda \) to grant space for the outputs (it is \( \lambda \) in the rest of the paper).

**Regional Failure model:**

After each earthquake \( i,j,M \), the physical infrastructure in an area of a circular disk is destroyed. Its center point is the center of \( c \), while \( R(M) \) is its radius. Each link fails having a point in the disaster area, the rest remain intact.

**Output:**

**Complete structure**

\[
JLFP(S) : \text{the probability that at least } S \text{ will fail}
\]

\[
JLFP(a) = 4.07 \times 10^{-2} \quad JLFP(b) = 3.53 \times 10^{-2} \\
JLFP(c) = 1.13 \times 10^{-2} \quad JLFP(d) = 2.91 \times 10^{-3} \\
JLFP(e) = 1.46 \times 10^{-2} \quad JLFP(f) = 2.60 \times 10^{-2} \\
\]

**Simple structure**

\[
LSFP(S) : \text{the probability that exactly } S \text{ will fail}
\]

\[
LSFP(a) = 3.45 \times 10^{-2} \quad LSFP(b) = 2.96 \times 10^{-2} \\
LSFP(c) = 1.00 \times 10^{-2} \quad LSFP(d) = 2.30 \times 10^{-3} \\
LSFP(e) = 1.33 \times 10^{-2} \quad LSFP(f) = 2.52 \times 10^{-2} \\
\]

**Input:**

\[
\text{Input:} \quad \text{Failure model:}
\]

**Model parameters:**

\[
p_{i,j,M} \text{, the probability of } E_{i,j,M} \text{, the earthquake having a magnitude } M \in \{4.6, 4.7, \ldots, 8.1\} \text{ and centre point in } c_{i,j}, \text{ where } c_{i,j} \text{ represents a latitude-longitude cell on the Earth's surface, taken from a grid over the network area } R(M) : \text{the radius of the area where network elements fail at magnitude } M \text{, see Fig. 3.}
\]

In the example, we set the intensity threshold to a relatively high \( \lambda \) to grant space for the outputs (it is \( \lambda \) in the rest of the paper).

**Fig. 4.** An illustration of the problem inputs and outputs.
Both representations have their advantages and drawbacks: while JLFP(S) can be queried directly to obtain the joint failure probability of a link-set, the number of LFSPs needed to describe the stochastic effect of the next earthquake can be significantly smaller.

V. Simulation Results

We have used the backbone topology of Interoute’s network, which is the largest privately owned Europe-wide IP cloud. They also act as an infrastructure provider, and the locations of their fibers are shown on their web page\(^3\). We created a copy of their network in the Italian region using the yEd graph drawing tool. We created at most 10 corner points of each fiber link such that the error in mapping is less than a few kilometers. Fig. 5 shows the network topology we created.

To compute the earthquake activity rates \((r_{i,j,M_u})\), we followed the smoothed approach described in Section III, and we used the Italian historical catalog, which covers approximately the last 1000 years (from 1 January 1005 to 28 December 2014) and consists of 4427 events. Before using the catalog, we removed all events not considered main shocks via a declustering filter \([27]\). This process resulted in a catalog composed of 1839 independent events. We adopted the completeness magnitude thresholds \(M_c\) over different periods given by \([28]\), and, following \([29]\), used a correlation distance \(d_c\) of 30km.

We took the maximum magnitude \(M_u\) assigned to each node of the grid from the European Seismic Hazard Model \([30]\). The ESHM13 project evaluated the maximum magnitudes of large areas of Europe based on a joint procedure involving historical observations and tectonic regionalization. We adopted the lowest value of the maximum magnitude distributions proposed by ESHM13. We calculated the \(b\)-value of the Gutenberg-Richter (GR) distribution on a regional basis using the maximum-likelihood method from \([31]\), which allows for multiple periods with varying completeness levels to be combined. Following the approach proposed in \([32]\) and based on \([29]\), we used a penalized likelihood-based method for the spatial estimation of the GR \(b\)-values based on the Voronoi tessellation of space without tectonic dependency. Finally, we used Eq. 2 to compute the earthquake activity rates for each cell and each magnitude \(M_u \in \{4.6, 4.7, \ldots, 8.1\}\).

We have repeated the simulation of \([5]\) using the earthquake data we have discussed in this paper. Instead of evaluating the protection schemes we focus on probabilistic Shared Risk Link Groups (SRLGs) which are the key input to survivable routing, network planning algorithms and network service availability computations. We evaluated two data structures: Complete \([5]\) and Simple \([4]\). In the Complete data structure, each set of links subject to joint failure because of a single earthquake is stored, while in Simple only the disaster set is stored \([4]\). Note that the Simple data structure is smaller, while it may take longer to compute the availability of an actual network service.

Fig. 6a corresponds to Fig. 6a of \([5]\) and shows a distribution of the probabilities of stored items in the associative arrays. The number of link sets with probability more than \(10^{-4}\) is around 1000, which was about 5000 in \([5]\) when a uniform distribution was assumed for the epicenter of the natural disasters. We repeated the simulation with two grid parameters \(0.1 \times 0.1\) decimal degree in latitude and longitude,

Fig. 5. Interoute network topology with 25 nodes and 35 links.

![Fig. 5. Interoute network topology with 25 nodes and 35 links.](image-url)

Fig. 6. The size of the data structures representing regional failures caused by an earthquake.

![Fig. 6. The size of the data structures representing regional failures caused by an earthquake.](image-url)

\(^3\)https://network-map.interoute.com
and 0.05 \times 0.05. There was no significant difference between the two resolutions. To analyze the assumptions we used in earthquake modeling, we have computed the probabilistic SRLGs for three scenarios: Uniform location distribution: The probabilities of earthquakes are uniformly distributed over the grid cells. However, we still have multiple possible magnitudes at each location. Roughly speaking, we rescale their probability based on their rate compared to the total rate at the cell. Same shape: The epicenter distribution is not uniform. However, we only consider one event per cell: a circle with the (weighted) average of the radii of all possible/considered earthquakes. Uniform location distribution & same shape: Earthquakes are uniformly distributed over the grid cells, and each grid cell only has one event: a circle with the (weighted) average of the radii of all possible/considered earthquakes.

Comparing the graphs on Fig. 6b and 6a, we can deduce that the assumption of the uniform location distribution has little effect, since we keep the same link sets of LFSPs, and only their probabilities change. In contrast to this, fixing the disaster shape drastically reduces the number of LFSPs. Fig. 6c shows the data set if we change the MCS intensity threshold representing the resistance of the network infrastructure against earthquakes. In other words, MCS VIII means the network is built with more earthquake-proof buildings and infrastructure. Fig. 6d investigates the dependency between the failure probability of a set of links and the set cardinality. We grouped the elements S of Complete by their size |S|.

VI. Conclusion

In this study, we focused on computing the vulnerability of networks to earthquakes, using empirical data. To capture the inter-correlation of link failures, traditional ground-shaking hazard maps are insufficient. We have thus created an earthquake model – applied to Italy in this paper – based on an historical earthquake catalogue over the last 1000 years. Based on this catalogue, we generated earthquake activity rates and investigated how the ground-shaking intensity diminishes with the distance from the epicenter and identified the network links that were subject to structural damage in case of possible earthquakes. Combining these, we analyzed the vulnerabilities of the Italian portion of Interoute’s fiber network.

This paper is a step towards a deeper understanding of the risk of failures that backbone networks are subject to by bringing two fields together: Seismology and Telecommunications.

REFERENCES