A survey of strategies for communication networks to protect against large-scale natural disasters

Teresa Gomes*, János Tapolcai†, Christian Esposito‡, David Hutchison§, Fernando Kuipers¶, Jacek Rak∥, Amaro de Sousa**, Athanasios Iosifidis††, Rui Travanca‡‡, João André*, Luísa Jorge††, Lúcia Martins‡‡, Patricia Ortiz Ugalde§§, Alija Pašić¶¶, Dimitrios Pezaros∥∥, Simon Jouet∥∥∥, Stefano Secci∥∥∥∥, and Massimo Tornatore∥∥∥∥∥

*∥∥∥∥∥Dept. of Electrical and Computer Engineering & INESC Coimbra, University of Coimbra, 3030-290 Coimbra, Portugal
†∥∥∥∥∥Budapest University of Technology and Economics (BME), Budapest 1117, Hungary
‡∥∥∥∥∥Dept. of Computer Science, University of Salerno, I-84084 Fisciano (SA), Italy
§∥∥∥∥∥Lancaster University, School of Computing and Communications, United Kingdom
¶∥∥∥∥∥Delft University of Technology, Fac. of Electrical Engineering, Mathematics, and Computer Science, The Netherlands
∥∥∥∥∥Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Poland
**∥∥∥∥∥Instituto de Telecomunicações - DETI, Universidade de Aveiro, 3810-193 Aveiro, Portugal
††∥∥∥∥∥Dept. of Electronics Engineering, Alexander T.E.I. of Thessaloniki, GR-57400 Sindos, Thessaloniki, Greece
‡‡∥∥∥∥∥Dept. of Civil Engineering, Universidade de Aveiro, 3810-193 Aveiro, Portugal
§∥∥∥∥∥National Laboratory for Civil Engineering, 1700-066 Lisbon, Portugal
∥∥∥∥∥Instituto Politécnico de Bragança, Bragança, Portugal
∥∥∥∥∥Asociación de Empresas Tecnológicas Innovalia, Bilbao, Spain
∥∥∥∥∥School of Computing Science, University of Glasgow, Glasgow, G12 8QQ, United Kingdom
∥∥∥∥∥Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005 Paris, France
∥∥∥∥∥University of California, Davis, CA, 95616, USA & Politecnico di Milano, Italy

Email: *teresa@deec.uc.pt, †tapolcai@tmit.bme.hu, ‡christian.esposito@dia.unisa.it, §d.hutchison@lancaster.ac.uk,
¶f.a.kuipers@tudelft.nl, ‡‡rui.pg.gda.pl, **asou@ua.pt, ††aiosifidis@el.teithe.gr, ‡‡rui.travanca@ua.pt, ‡‡jandre@lnc.ac.pt,
††ljorge@ipb.pt, †††lucia@deec.uc.pt, †¶¶portiz@innovalia.org, ‡∥∥∥parallel@tmit.bme.hu, ‡‡∥∥∥dimitrios.pezaros@glasgow.ac.uk,
∥∥∥∥∥simon.jouet@glasgow.ac.uk, ∥∥∥∥∥stefano.secci@upmc.fr, and ∥∥∥∥∥∥∥∥∥∥∥massimo.tornatore@polimi.it

Abstract—Recent natural disasters have revealed that emergency networks presently cannot disseminate the necessary disaster information, making it difficult to deploy and coordinate relief operations. These disasters have reinforced the knowledge that telecommunication networks constitute a critical infrastructure of our society, and the urgency in establishing protection mechanisms against disaster-based disruptions.

Hence, it is important to have emergency networks able to maintain sustainable communication in disaster areas. Moreover, the network architecture should be designed so that network connectivity is maintained among nodes outside of the impacted area, while ensuring that services for costumers not in the affected area suffer minimal impact.

As a first step towards achieving disaster resilience, the RECODIS project was formed, and its Working Group 1 members conducted a comprehensive literature survey on “strategies for communication networks to protect against large-scale natural disasters,” which is summarized in this article.

Index Terms—vulnerability, end-to-end resilience, natural disasters, disaster-based disruptions.

I. INTRODUCTION

One of the findings of a taxonomy of Internet failures [1] was that existing protection mechanisms may be adequate for single link or node failures, but they cannot deal with large-scale disasters. Recent natural disasters exposed the vulnerability of communication networks to those events.

Following meteorological observations, the risk of natural disasters is rising. Disaster-based failures are commonly implied by such natural factors as: hurricanes, tsunamis, floods, or earthquakes. Tens of hurricanes worldwide are observed every year leading to power outages affecting communication networks on a massive scale and for a relatively long time (10 days on average). Examples include e.g., hurricane Katrina which caused severe losses in Louisiana and Mississippi in Southeastern US in August 2005. The 7.1-magnitude earthquake in December 2006 in Southern Taiwan was responsible for disruption of international communications to China, Hong Kong, Korea, Japan, and Taiwan due to simultaneous failures of seven submarine links providing Internet connectivity between Asia and North America. The Greatest Japan Earthquake on March 11, 2011 (of 9.0-magnitude), in turn completely (or partially) destroyed the telecom switching offices and was responsible for a massive damage to undersea cables.

Emergency networks have failed to disseminate the neces-
sary disaster information, making it difficult to deploy and coordinate relief operations. This has reinforced our knowledge that telecommunication networks are critical to society, and that disaster-based disruptions must be addressed. Hence, there is a strong need to implement mechanisms ensuring end-to-end communications between operational network nodes, even if only at a degraded level (low rate and/or large delays) in disaster-stricken areas. The goal is to be able to secure the necessary communications to support first-responder activities in a damaged area, and to have a strategy to progressively restore network services in the aftermath of the initial shock.

Although the existence of an emergency network in a disaster area is important, it is also desirable that network connectivity is maintained among nodes outside the impacted area. Networks can be designed so that services for customers not in the affected areas suffer minimal impact. This can be achieved using proactive and reactive approaches [2], [3]. For example, geographically-diverse routing [4] will increase network resilience to geographically-correlated failures.

A systematic approach on how to build resilient network systems can be found in [5], an overview of algorithms for survivable planning and routing is given in [6], and a survey on disaster survivability in optical networks is presented in [2]. More recently, Miranda et al. [7] presented a brief overview of the requirements for rapidly re-establishing connectivity and for providing levels of service adequate for emergency services. The COST CA15127 (RECODIS) Action is aimed to develop appropriate solutions to provide cost-efficient resilient communications in the presence of disaster-based disruptions, considering both existing and emerging communication network architectures. As a first step towards achieving the goals of RECODIS, and within the context of the activities its Working Group 1 (Large-scale natural disasters), a survey of existing strategies for communication networks to protect against large-scale natural disasters is presented here.

The paper is structured as follows: In section II, an overview of the vulnerability of communication networks to disaster-based disruptions is presented. In section III, rules and techniques for making network architectures less vulnerable to disaster-based failures are analysed. In section IV, disaster-resilient routing algorithms are discussed. Section V concludes the paper.

II. VULNERABILITY OF COMMUNICATION NETWORKS TO DISASTER-BASED DISRUPTIONS

The efficient protection of communication networks against large-scale disasters requires, as a first step, the assessment of the vulnerability of communication networks and its supporting physical infrastructures to such events. In this section, we review the recent scientific literature on measures, methods, and systematic approaches dealing with network vulnerability assessment and with the identification of the most vulnerable regions of networks. We also review the current threats and trends concerning the vulnerability of the physical infrastructures supporting communication networks.

A. Measures of network vulnerability

Critical Node Detection (CND) is a valuable method to determine the vulnerability of networks to multiple failures. CND problems aim to optimally remove a subset of nodes (the critical nodes) of a given network in order to optimize or restrict a given metric of network degradation. The problem can be defined either by upper-bounding the number of critical nodes and maximizing the degradation metric, or by lower-bounding the degradation metric and minimizing the number of critical nodes. Veremyev et al. [8] address two CND variants defined on a simple undirected graph $G$. In the first variant, for a given integer $K$, the aim is to identify a set of $K$ critical nodes minimizing the pairwise connectivity (also referred to as the average 2-terminal reliability metric in other works). In the second variant, for a given integer $L$, the aim is to identify a minimum set of critical nodes, so that the largest connected component in the remaining graph contains no more than $L$ nodes. For both variants, the authors propose alternative more compact Integer Linear Programming (ILP) models, together with reformulations and valid inequalities that improve the performance of solvers, while computing the optimal solutions of given problem instances.

For a given graph with associated node costs and a given cost budget $C$, Veremyev et al. [9] consider a set of critical nodes as a subset of all nodes whose total cost is not higher than $C$ and whose removal maximally degrades the connectivity of the graph. In [9], the degradation aim is the maximization of a distance-based connectivity metric, which takes into account not only the pairwise connectivity but also the shortest path distance penalties between node pairs that remain connected. The paper proposes a general ILP model (that can be adapted to the different distance-based metrics by proper parameter definition) and an alternative exact algorithm that iteratively solves a series of simpler ILP models. The paper also compares the proposed approach with different node centrality-based greedy algorithms (the degree centrality, the closeness centrality, the betweenness centrality and the eigenvector centrality) showing that it provides much better solutions than the centrality-based ones.

For a given network, the objective of Dinh et al. [10] is to compute a minimum set of critical network elements whose removal results in a specific degradation target of the network pairwise connectivity. The minimized set of network elements is referred to as a $\beta$-disruptor, where $0 \leq \beta < 1$ denotes the fraction target of the pairwise connectivity degradation. Network elements can be either edges or nodes (vertices), resulting in two problem variants (the $\beta$-edge disruptor and the $\beta$-vertex disruptor). The paper proves that both problem variants are NP-hard and proposes an $O((\log n \log \log n)$ pseudo-approximation algorithm (for the $\beta$-vertex disruptor) and an $O((\log n)^{3/2})$ pseudo-approximation algorithm (for the $\beta$-edge disruptor). For the $\beta$-vertex disruptor case, the proposed method is compared with three node centrality-based greedy algorithms: (i) sequentially removing the node with maximum degree, (ii) sequentially removing the node with the
maximum betweenness centrality, and (iii) sequentially removing nodes in descending order of their eigenvector centrality values. The results show that the $\beta$-vertex disruptor sizes of the node centrality-based algorithms are much larger than the ones computed by the proposed pseudo-approximation algorithm. In [11], Dinh et al. assume a given set of link costs and another set of node costs and they extend the previous work [10] to the general case where the $\beta$-disruptor can be a mix of links and nodes. The paper proposes, first, a $O((\log n)^{1/2})$ bicriteria approximation algorithm and, then, a hybrid meta-heuristic that combines simulated annealing, variable neighbourhood search and spectral clustering to improve the efficiency of the bicriteria approximation algorithm.

Sterbenz et al. [5] propose ResiliNets, a framework intended to unify several disciplines, strategies and principles used for network survivability and resilience. The framework describes axioms for systematic resilience and includes a so-called $DR^2$+DR strategy with an inner control loop ($D^2R^2$ - defend, detect, remediate, recover) aiming for a system to rapidly adapt to challenges and attacks maintaining an acceptable service level, and an outer control loop (DR - diagnose, refine) enabling the longer-term evolution of the system. A set of design principles for resilient systems is proposed, including prerequisites, design trade-offs, enablers and behaviour required for resilience. For resilience analysis, [5] proposes a two-dimensional representation of the network state in the operational state and service parameters dimensions, describing the effects of challenge $\rightarrow$ fault $\rightarrow$ error $\rightarrow$ failure chains as state transitions in this space. In [12], the authors expand the previous analysis, presenting path diversity metrics and an (updated) analytical resilience framework, based on functional metrics to quantify network resilience in the presence of challenges, like (unspecified) disaster-based failures. A definition of the resilience space is provided, describing the states through which the system may evolve while recovering from a challenge. To evaluate vulnerability, the authors define a path diversity function, which measures (graph) similarities on both links and nodes, and a path diversity measure, defined as an aggregation of path diversities for a selected set of paths between a given node pair. The authors mention the need to include geographic diversity through a function of desired minimum distances between node pairs as a measure parameter to model area-based challenges. The system resilience is computed as the resilience space dimension (both in instantaneous/static and average/dynamic).

Palmieri et al. [13] aim to study the stability and survivability of the Internet on the occurrence of a catastrophic event. Stability is defined at the routing level as a measure of the number and frequency of topological information exchanges within the Border Gateway Protocol (BGP), while survivability refers to the ability to quickly recover the service levels offered before the catastrophic event. Three large-scale events were analysed: the Taiwan Earthquake on December 27, 2006, the Japan Earthquake on March 11, 2011, and the USA Blackout on September 8, 2011. For each event, the authors used BGP data collected by the Routing Information Service (RIS) project from RIPE NCC to study the number of routes and geo-localization of IP addresses that became unreachable. The conclusion is that disasters cause catastrophic consequences to the current Internet settings and architecture, in terms of reachability of network prefixes, Internet global connectivity, and recovery time. Moreover, the consequences are not limited to the disaster area, but they also affect farther areas due to the interdependencies between Autonomous Systems and the damages of major transmission links between countries.

B. Identification of vulnerable regions

A large-scale disaster typically affects an area which can be represented by a certain shape. Therefore, a problem of interest is the identification of those areas in which the network is embedded that, upon failure, would cause gravest disruption to network performance.

Neumayer et al. [14] focus on assessing the level of vulnerability of geographical networks to natural disasters or human attacks. The physical topology is modelled as a bipartite graph in which nodes and links are geographically located on a plane. The model has distinct height and width parameters, representing the north-south and east-west geographic capacity of the network. Then, a disaster results in a vertical line segment cut in the bipartite graph which removes all links that intersect it. A worst-case cut is defined as a cut with the maximum total capacity of intersected links, and a Mixed Integer Linear Programming (MILP) model is formulated for identifying the worst-case cut that minimises the maximum flow between its two sides. The authors also formulate a lower bound on the worst-case cut and develop a polynomial time algorithm for finding the worst-case cut in a bipartite graph.

In [15], Neumayer et al. study how to find the most vulnerable parts of a network, subject to geographically correlated failures. The network model considers nodes located in a plane and connected via straight lines representing optical fibres. Two geographical failure models are considered: (1) line segments of a given length which may cut links, and (2) circles of a given radius, which destroy all nodes and links contained within, including links whose endpoints lie outside the circle. Despite the infinite number of positions where the line segments and circles can be placed, Neumayer et al. [15] demonstrate that only a polynomial number of positions needs to be considered. However, the proposed algorithms are of high complexity, namely $O(N^6)$ or even $O(N^8)$, where $N$ represents the number of nodes. By using a slightly relaxed failure model in which only links that have at least one endpoint in the failure region are destroyed, Trajanovski et al. [16] are able to determine the most vulnerable region for a failure of circular shape at a reduced complexity than that of [15]. Moreover, they can exactly and in polynomial time determine such vulnerable regions also for failures of elliptical shape and for shapes represented by a polygon.

Motivated by scenarios where nodes closer to the attack (or disaster) central point have higher failure probabilities than nodes farther away, Agarwal et al. [17] consider probabilistic geographically-correlated failure models. The components
(nodes, links, or lightpaths) have a failure probability function and the focus is on finding the network parts that are expected to be most vulnerable. In particular, three metrics are used: (1) the expected component damage, (2) the average two-terminal reliability, and (3) the expected maximum flow after the attack. In addition, single as well as multiple simultaneous attacks are considered. Agarwal et al. propose approximation algorithms that are based on finding those intersections (called faces) of the component failure probabilities that have the highest value.

Iqbal et al. [18] study the resilience of a network, where geographical information is available for both nodes and links. They focus on, and provide polynomial-time algorithms for, finding the links that are spatially close to each other. The definition of closeness is an input parameter and hence, if very large, can represent a disaster scenario or, if small, can represent a construction-related failure scenario. The proposed algorithms make use of the R-Tree data structure used in Geographical Information Science. Iqbal and Kuipers [19], in addition to taking geographical information of nodes and links into account, also include the notion of time and a risk profile of the area in which the network is embedded. The rational is that some disasters, like hurricanes, may traverse an area and hence may lead to different component failure probabilities at different points in time. They provide polynomial-time algorithms to assess the most vulnerable connections.

Gardner and Beard [20] define a geographic vulnerability as a geographic region, such that if the nodes (and links incident on those nodes) in that region fail, the network becomes disconnected. Using the Two-Terminal and All-Terminal methods, and depending on the radius of a threat, geographic vulnerabilities are identified. These methods identify node cut sets that fall within the threat radius. The consideration of the threat radius allows the reduction of the search space, but the computation time still grows exponentially.

Long et al. [21] propose the use of weighted spectrum (WS) to evaluate network survivability regarding geographic correlated failures. Furthermore, a comparative analysis is conducted by solving an optimization problem to determine the cut with the largest impact for a number of measures in the literature (namely, Algebraic Connectivity, Network Criticality, Average Shortest Path, Network Diameter) as well as WS. The experiments show that WS is the most versatile measure to evaluate geographically correlated vulnerable links and nodes of backbone networks.

Critical services tend to be supported by virtual networks, which increases the importance of considering the robustness of multi-layered networks. Gardner et al. [22] define a geospatial event as an occurrence that can cause a geographic vulnerability, which in turn is defined as a geographic area of a network that, if damaged, can cause significant impact to the function of the entire network. The threat radius is used, as in [20], to define a geographic vulnerability and a state-space analysis method is proposed suitable for multi-layer networks [22]. A network state is defined by the set of failed and operational nodes. If a network state causes the network to function below a given performance measure threshold, then that state is said to cause the network to be non-functional. In order to avoid the space state explosion, the multi-layered Self-Pruning Network State Generation (SP-NSG) model [22], for a given threat radius, selects for analysis only admissible network states. Furthermore, and to make large networks tractable, a K-means clustering algorithm is employed. This work was extended in [23] where a new metric designated Network Impact Resiliency (NIR) (inspired in performability) is proposed. Network impact is an indication of the inability of the network to perform its function in the network state resulting from a failure and NIR combines network impact with state probability.

C. Vulnerability of physical infrastructures: threats and trends

Over the last few decades, the demand for communication structures has increased due to the requirements for their use in the telecommunications sector and, with the advent of mobile communications, this demand became even greater. The telecommunications sector is becoming increasingly important in our modern society. Communication within social, economic and industrial systems is becoming increasingly digital, wireless and interdependent (e.g. with the power grid), the consumer market is globally expanding, and there is an escalating offer/demand input. The increased significance of these systems comes at the same time when the life-span of existing key physical infrastructures is reaching its maturity, the market has changed from being state managed to being fragmented, privatised and/or publicly regulated, and there is not enough information available and adequate communication within and between infrastructures sectors concerning vulnerabilities. Unfortunately, the number of failures observed in communication structures is high compared to other structures of equal economic and societal importance [24], [25]. A great number of failures observed are due to poor design, which results in unsafe structures that can suffer from full collapse [24], [25]. Mainly for economic and functional reasons, communication structures, e.g. masts and towers, are lightweight structures with structural characteristics such as high slenderness and high flexibility. With the desire to install wireless hubs specifically in locations of high population density, new structural forms have appeared, i.e. monopoles that challenge the limits of conventional pole design. Though monopoles are widely used, the current methods for their analysis and design are outdated and/or inappropriate. Therefore, a detailed review of both the methods of analysis and loads definition becomes imperative [24], [26]. In the last decade, there has been a growing interest in the field of structural health monitoring, resulting in the development of new techniques and equipment, such as fibre-optic sensors based on Fibre Bragg Gratings (FBG). As underlined in recent studies, FBG-based accelerometers are presented as an excellent tool to better understand the response of this type of structures [24], [27]. As an example of the failures observed in communication structures, the collapse of a 40-metre-high monopole with a tubular cross-section is shown in Fig. 1 and Fig. 2.
On August 29, 2005, Hurricane Katrina struck the United States Gulf Coast and caused catastrophic damage to the combined telecommunications and power infrastructure. Devastating effects on both infrastructures hampered rescue efforts, blocked attempts to coordinate early responses, and made calls for aid impossible from the hardest-hit areas [28],[29]. White House Katrina Report described the results: “The complete devastation of the communications infrastructure left responders without a reliable network to use for coordinating emergency response operations” [28]. While it is tempting to view a catastrophe such as Katrina as an once-in-a-lifetime event, doing so would be an exercise in wishful thinking. Eight months before Hurricane Katrina, on December 26, 2004, the Indian Ocean tsunami highlighted the heavy human cost of communications breakdown during extreme events. While seismic monitoring stations throughout the world detected the massive sub-sea earthquake that triggered the tsunami, a lack of procedures for communicating these warnings to governments and inadequate infrastructure in the regions at risk delayed the transmission of warnings. However, based on the successful evacuation of the handful of communities that did receive adequate warning through unofficial channels, it is clear that better communications could have saved thousands of lives.

Climate change, natural catastrophes and failure of critical infrastructures are ranked at the top of the 2015 Global Risks database prepared by the World Economic Forum [30] and for which less progress has been made. The latest UK National Risk Register of Civil Emergencies also considers failure of critical infrastructures to have a high risk level and thus a priority risk [31]. Irrespective of the success of our mitigation efforts, the impact of climate change will increase in the coming decades. While efforts must continue towards mitigating its effects, there is no other choice but to take adaptation measures to deal with the unavoidable climate impacts that are anticipated and their economic, environmental, and social costs. Extreme weather and climate changes leave worldwide infrastructure systems exposed to different and more extreme conditions. Since the available amount of resources is finite, it is highly likely that degradation and interruption of vital services will occur at certain times. It is essential that stakeholders can turn the page on inefficient past practices and commit themselves to comprehensive and continuous planning and management policies of critical infrastructure assets, with the goal of reducing uncertainties, risks, and magnitude of adverse consequences, increasing sector and society safety, resilience and sustainability. Doing so requires a mix of technical and policy changes that, together, will serve to mitigate damage and accelerate restoration.

III. Enhancing the Disaster-Resistance of Existing Networks and the Deployment of Emergency Networks

Currently, several networking infrastructures are already deployed and functioning around the world as part of the global Internet. In order to have survivable networks, it is infeasible to dismantle these existing networks and realize novel ones that are properly equipped with means to support suitable service and connectivity levels after the occurrence of a disaster. The more realistic approach is to augment these existing networks with proper means to increase their degree of survivability. To this aim, there are two complementary ways to proceed: network robustness improvements introduced before the disaster occurrence, and recovery actions after the disaster takes place. The following sub-sections describe the available literature for these two directions.

A. Pre-disaster network robustness improvement

Current networks are characterised by intrinsic vulnerabilities within their design and deployment that affect their survivability in case of a disaster. A first way to make existing networks robust is to have network operators proactively adopt proper means to minimize network disruptions and data loss in the case of a disaster and to introduce a appropriate
redundancy within the network. Figure 3 classifies proactive techniques for network robustness and survivability.

Two of the main issues affecting the robustness of a network in case of a disaster are connectivity availability and traffic overload. Specifically, a disaster may be able to compromise some network elements, such as routers, gateways or links, reducing the connectivity within the network. This may cause network partitions and a reduction of the possible paths among the network nodes. The effect of this issue is that certain nodes are no longer able to connect with the other ones and/or to access the Internet, and/or traffic has to be forwarded only along certain few links, causing congestion. Such phenomena are further exacerbated by humans intensifying their activity on the Internet after a disaster, by trying to obtain information about the effects of the disaster or to contact their dearest ones. For example, cellular networks are typically overloaded and unable to provide any service after a natural and/or human-made disaster within the affected area.

The natural solution to deal with the connectivity issue is to provide redundancy within the network, i.e., networks should be designed in such a way that any node of the network can reach all other nodes of the network after a failure event. This is typically obtained by improving fault tolerance via proactive failure recognition and having several backup links when the primary one is unavailable due to the disaster, as described in the standardised guide for the design of survivable networks in [32], or in [33] for the case of optical networks. A redundant network design implies the use of backup resources and high costs. A different solution is the one presented in [34], which investigates new methods for lessening the impact of large-scale failures (uncorrelated multiple failures) in terms of the number of affected connections, by deploying immunization strategies (i.e., by fortifying a certain set of links). This is achieved by identifying the links to which extra network protection can be applied so that the impact of such failure events, in terms of the number of connections affected, is minimized. This is done by introducing two heuristic-based link prioritization strategies for improving network resilience: one is built upon the concept of betweenness centrality, while the other one adopts the measure that the authors named as the observed link criticality. Zhang et al. [35] considered shielding critical links (e.g. strengthening cables), under general and geographical failure models. A MILP formulation to minimize shielding cost to ensure the connectivity of a given source-destination pair, was developed, then extended to guarantee network connectivity. Simulated annealing was also used to solve larger problems. Another different approach for network connectivity is to equip the network and its nodes with some methods to rearrange the network resources and services on a partially damaged network (self-organizing network) [36], so as to mitigate the effects of disasters. A similar solution is presented in [37], which focuses on the ability of devices – called Stem Nodes (SN) – to reconfigure or install new components to be able to cover multiple network roles (e.g., gateway or relay). This work proposes distributed algorithms, based on swarm intelligence principles, through which each SN can autonomously select its role, so that end-to-end performance is maximized while the lifetime of the spontaneous emergency network (so called STEM-NET) is prolonged. Last, Software-Defined Networking and Network Functions Virtualisation are useful enabling technologies to reconfigure logical networks or how they are mapped to the physical network [38], [39].

When it comes to protect a network infrastructure against extensive disruptions as those arising during a disaster, ensuring network connectivity might result in a very costly design of the network. Therefore, an alternative proactive survivability approach that is currently attracting a lot of attention and still represents an open field of research is the one known as content connectivity, i.e., ensuring the reachability of content from any node in the network without necessarily guaranteeing network connectivity. The main idea is that even if a network gets disconnected into multiple separate components due to failures, the replication of content can be planned in such a way that at least one copy of each relevant piece of content is still reachable in each disconnected component. The concept of content connectivity was first introduced in [40], where the authors developed an Integer Linear Program (ILP) model that guarantees content connectivity against single-link failures. Following the same methodology, the authors in [41] formulated a more practical version of the Survivable Virtual Network Mapping (SVNM) problem, where network connectivity is guaranteed against any single failure, while content connectivity is guaranteed against any double-link failures. Finally, a generalization of the content connectivity concept, referred to as a k-content connectivity, has been proposed to ensure that any node in the network can still reach the content after k − 1 link failures in [42], where a flexible and cost-saving algorithm of the problem is also presented.

The second issue we have mentioned above is that of traffic congestion happening after a disaster. The first solution to this problem is the one described in [43], where the authors propose a mechanism for off-loading traffic to lightly loaded neighbours, thereby increasing both handover success rate and leftover power. A User Equipment (UE) controlled and Base Station assisted process is described to allow handover of equipment calls to lighter loaded base stations in a disaster scenario. The main contribution is claimed to be the ability of UE to self-detect the onset of a natural disaster and to act accordingly (by selecting a less loaded base station from the several available). The natural disaster detection is contingent on base stations being able to provide the UE information associated with load and power or battery capacity, which should show distinctive patterns at the onset of a natural disaster. Each base station would in turn use its own knowledge on distance to base station (estimated from signal attenuation) and direction of movement to select the suitable base station to use at each moment. Another solution is the one in [36], which starts from the consideration that telecom networks usually have some unused capacity to accommodate traffic fluctuations and avoid capacity exhaustion. Such capacity can be exploited to provide better protection against disasters by alleviating the traffic deluge and to relieve the rescue operations after
A vulnerability of a network may not be intrinsic to the network itself; in fact, the correct behaviour of a network and its survivability depend also on the correct functioning of the power grid. If there is an energy outage due to a failure within the power distribution infrastructure caused by the disaster, the network may become unavailable or seriously compromised. Such a problem can be treated by introducing redundant power sources that the network may use in case of a disaster so as to avoid any compromises to its correct behaviour. In the design of a Never Die Network (NDN) presented in [44], self-powered fixed wireless network stations, cognitive mobile stations and wireless balloon stations represent a set of solutions to tolerate energy outage without affecting the network. Also in [45], there is a discussion of a planning framework to reduce telecommunication network power supply vulnerability during natural (earthquakes, hurricanes, storms and blizzards for example) and man-made disasters. Such a framework suggests several different solutions to improve energy effectiveness in case of disasters, such as coordinating portable generator set deployment among different network operators, and by installing permanent photovoltaic systems at sites where long electric outages are likely.

A last proactive robustness technique is related to the internal organization and deployment of a network, which falls within the so-called disaster avoidance control, presented in [46]. It mainly consists in relocating software objects from a high-risk region to a low-risk region, as described in [46].

B. Post-disaster network recovery

Infrastructure failures as well as traffic overload arising in post-disaster areas have to be dealt with in future communication systems as effectively as possible, in order to provide connectivity among governmental and non-governmental emergency management teams, first responders and victims, victims and families, etc. Noting that the subject of communication varies as time elapses after a disaster [47], i.e., disaster alarm, evacuation programs and orders, safety confirmation, first aid support, lifeline information, shelters and traffic information, etc., there are two pillars in post-disaster service recovery: (i) rapid emergency communication network deployment, most usually on top of surviving network infrastructure to allow critical service provisioning in the first period after the disaster, and (ii) effective maintenance of disconnected network infrastructure that will allow full service recovery – see Fig. 4. This subsection summarizes the most recent and representative efforts in this context. Further techniques and approaches may be found in the references of the reviewed papers.

1) Rapid emergency communication network deployment: Emergency communication networks must ideally fulfil a set of requirements to provide and maintain sustainable communications [7]: resilience, basic service set provision, self-capabilities (such as self-organization, -optimization, -healing), node mobility, inter-operability, and compatibility with other heterogeneous undamaged network systems as well as low SWaP (size, weight and power). Two major categories of emergency networks can be identified that, in any case, may need to set up and operate simultaneously over a disaster area. The first one includes networks based on vehicular or transportable network nodes, e.g., base stations (BS) and access points (AP), while the second one is based on user mobile devices with enough energy to set-up ad-hoc/mesh networks, acting as relay or gateways towards surviving network nodes.

a) Emergency networks based on transportable nodes: The relevant literature presents solutions with a wide range in set-up complexity and capabilities offered. A complete wireless mesh network based on MDRU (movable and deployable resource unit) nodes was developed and tested by Sakano et al. [48]. MDRUs are network nodes (carried in a van) that offer connectivity to the Internet using satellite, pre-installed optical fibre cables or surviving wireless access gateways. The van-type MDRU can get the power supply from three types of sources, namely gasoline electric generator, lithium-ion battery unit, and electrical power input from outside. In general, it can operate for five days without an external power supply. In addition, portable WiFi modules (battery powered – replenished by a solar panel) spread around the van area offer AP gateway services via fixed wireless access connection with the MDRU. The MDRU can provide voice service using an IP-PBX server to one hundred users (that use their real phone number) simultaneously. Field experiments showed that a distance of 700 m from the MDRU can be covered with three APs. This can be remarkably extended with a relay-by-smartphone system. The overall system seems strong, however in addition to the necessity of moving the van and the WiFi equipment to the disaster area, energy efficiency and MDRU inter-operability need to be further investigated.

Among the easiest to deploy transportable emergency networks is the “EmergeNet” [49]. This is a rapidly deployable, small-scale cellular network based on the well-tested OpenBTS open-source platform that uses a software-defined radio transceiver to enable GSM transmission/reception. With an additional set of software tools, inbound and outbound VoIP calling and messaging is enabled through Skype for end users with or without Skype clients. Features for automatic reconfiguration of the BSs to maximize the functionality in the face of power, network and/or hardware failures are also
provided. Due to the fact that each BS can serve up to 7 concurrent calls, an automatic SMS-based call queueing system is proposed to provide fair access, while avoiding user recalls. EmergeNet can be AC or DC power supplied (e.g. by power grids, portable generators, vehicle batteries). A three-cell battery bank can provide 4 days of autonomous operation and solar panels can be used as well to offer 24/7/365 uptime in most locations. EmergeNet, based on the experimental tests provided, seems to be a promising solution. However load and security issues have to be further investigated.

Cheng et al. [50] describe a similar (based on OpenBTS) small-cell system offering basic GSM services. The cells, with a predefined cell ID, support registration of rescue team members and victims through SMS. Furthermore, a pre-installed to the mobile phone Android APP automatically detects the emergency network after an early warning message, provides the position of the victim and supports the rescue process with a predefined set of features and information transfer. This is more complex than EmergeNet, however it may be promising if extended testing is performed.

b) Emergency networks based on end-user devices: Network node transportation may not be feasible in certain disaster cases. Therefore, in order to provide fast and infrastructure-free service recovery, a great number of techniques engaging mobile user equipment to set-up ad-hoc/mesh networks up to a surviving node have been presented. Among them, the above mentioned STEM-NET [37], providing self-configurable SN implementation, allows the set-up of an ad-hoc network on top of existing end-user devices.

However, while STEM-NET requires a new software architecture to be pre-installed in the mobile nodes, Minh et al. [51], [52] propose a tree-based multi-hop WiFi network that involves a downloadable (upon set-up of the network) software-based implementation of network functions. Each mobile acts as a virtual AP (VAP) to provide service to the rest of the mobiles and as a STA (IEEE 802.11 wireless station) communicating with the VAP to which it belongs. A mechanism for auto-reconfiguration on link failures is also proposed. The authors demonstrated experimentally that the multi-hop network can be established in a few minutes over an area of 600 m to 1 km. A drawback is that only the end user can initiate a connection, due to private addressing of mobile nodes, while load-balancing and security issues have to be further considered. A similar (yet preliminary) approach involving smartphone hotspots by WiFi tethering is proposed by Ray et al. [53]. The authors provide algorithms for hotspot selection based on the end-user’s direction of mobility and the number of terminals that the hotspot allows, based on calculation of its leftover energy. However no implementation tests and results are given.

The extended use of mobile devices as relays in the above-mentioned configurations has driven research efforts in optimal relay placement in disaster scenarios. Herlich and Yamada [54] simulated how disaster survivors can place mobile devices as stationary relay chains to interconnect evacuation centres and Internet gateways. Results show that among the strategies considered, the most promising are: (i) link every evacuation centre to the closest gateway with a relay chain (Direct strategy), and (ii) link each evacuation centre to the 3 closest evacuation centres or gateways with relay chains (Neighbour 3 strategy). In the same context, Król et al. [55] formulated a modified k-connectivity algorithm (k paths connecting each disconnected evacuation centre with a connected one) and simulated the proposed algorithm over a realistic disaster scenario. Both approaches are interesting, but further analyses are required.

Finally, the potential of IoT-enabled devices to enhance network resilience in face of disaster has been introduced by Petersen et al. [56]. The authors claim that IoT devices may be used as relays based on their inherent ability to leverage the spontaneous wireless networking paradigm, their battery-powered operation capability, and the ability of sensors to monitor various environmental parameters that can be exploited to provide real-time data around disaster areas. Nevertheless, major challenges have to be met towards this end, i.e., limitations and inter-operability issues due to the heterogeneous PHY and logical network connectivity of IoT devices, traffic prioritization, social acceptance, as well as security issues.

2) Effective network maintenance: Post-disaster situations present high variability both in the infrastructure failures and the communication needs as time evolves from minutes to days to weeks after the occurrence of the first event. Therefore, efficient planning of communication network maintenance is vital in order to support the required services. Heegaard et al. [57] propose a survivability quantification framework that is used to model a multi-phase recovery procedure. In this model, which is constructed by combining continuous time Markov chain performance models, the system performance is gradually changed throughout the multi-phase recovery, where each phase models a specific recovery action or a set of parallel actions. Each action may depend on the outcome of previous actions. A numerical example, considering escalated levels of recovery and deferred repair, is also presented.

A model to provide an optimal repair schedule over an optical network after failure of multiple network elements is discussed in [58]. The model is based on the travelling repairman problem (TRP), and proper algorithms are examined for scheduling repair tasks in order to achieve minimum damage and low repair time. A Mixed Integer Linear Program (MILP) solution is given, and heuristic algorithms, that is, a greedy algorithm (GR), a dynamic programming (DP), and a simulated annealing (SA) one, are proposed for solving the problem. Simulation results over some illustrative examples show that the heuristic solutions only deviate slightly from the optimal solution and, due to their low complexity, can be used for larger-scale problems.

Random failures and recoveries of network elements as time evolves during post-disaster situations mandate scheduling and rescheduling of user needs and surviving infrastructure that serves them. Algorithms like the robust fault-tolerant version of the uncapacitated facility location problem (RFTFL)
can be used in assigning user demands to surviving datacentres, so that latency is minimized in case of multiple failures.

IV. DISASTER-RESILIENT ROUTING ALGORITHMS

In this section, we review several techniques that have been proposed since 2012 to route and protect traffic against network failures. For a survey of work before 2012 we refer the reader to [6]. We first consider tunable survivability against single failures. Then we consider protection schemes for two instantaneous failures or several failures all belonging to the same risk group. We proceed by considering disaster-aware routing, which demands considering the geo-information on the network as well as temporal characteristics. We conclude this section by discussing how to possibly route traffic when a disaster has manifested.

A. Tunable survivability against single link failure

Yallouz and Orda [60] argue that providing full protection against single link failures by establishing link-disjoint paths, uses excessive resources in practice. Thus, tunable survivability is proposed, which provides a quantitative measure for survivability and offers flexibility for the service provider to select paths for the connections by allowing some common links along the two paths. While a given level of survivability has to be satisfied by the path-pair, some bottleneck (e.g., bandwidth) or additive (e.g., delay) metric is minimized in the optimization problem. Previous works already addressed bottleneck QoS metrics (defined by the weakest component in the path), thus, this paper considers the important and much more complex class of additive metrics, where the QoS is the result of the sum of link metrics along the paths. The most important observation in the paper is that given a connection that, consists of two paths per source-destination pair under the single-link-failure model, only a failure on a link that is common to both paths can disrupt the connection. There are two ways of considering the weight of such common links in the optimization problem. Namely, when in the weight of the survivable path-pair, the weight of the common link is counted once (CO) or twice (CT). One of the main findings of the paper is that if the CT problem is considered, the links that may affect the survivability of the optimal solution are restricted to a very small subset of the network links. An algorithm is proposed to identify those links efficiently. Note that, as only this small set of links has to be considered as common links, the computational complexity of the proposed algorithms can be significantly reduced. However, both CO and CT versions of the problem are shown to be NP-hard.

Fortunately, through a graph transformation that reduces the problems under investigation to the restricted shortest path (RSP) problem, the existing efficient fully polynomial approximation schemes proposed for RSP can be applied.

Yallouz, Rottenstreich, and Orda [61] extend this line of work by computing tunable survivable trees. Again, the single-link failure model is considered and the research question is how to build a set of \( k \) spanning trees, such that the probability that at least one of the spanning trees in the connection is operational and/or the routing bandwidth is maximized. All links are considered to have the same failure probability, else the problem becomes NP-hard. The authors have established a novel polynomial-time algorithm for providing an optimal set of (any) \( k \) spanning trees that maximizes its survivability level while ensuring a guaranteed bandwidth. Additionally, they have provided tight bounds on the number of spanning trees that may be needed in order to achieve a maximum level of survivability. Finally, through simulations, they have showed that the maximum level of survivability can be well-approximated by establishing just two spanning trees.

B. Resilience against multiple failures

Rohrer et al. [62] seek to close the gap between the fragility of the current Internet and the notion of maximum flow reliability. They focus on the ability of a topology to remain connected if multiple simultaneous node and link failures occur and devise a mechanism – path diversification – to instantiate a unified interface that is as reliable as the underlying physical graph. Aiming to achieve a resilient multipath mechanism, the authors define several path diversity metrics that can be applied to both node pairs and complete networks.

They develop an algorithm for selecting the best subset of available paths, in the sense that these paths are maximally diverse and have minimal stretch. Through simulations, the authors analyse the extent to which the proposed metrics are correlated to both graph theoretic properties and network survivability.

For the case where backup paths can be shared, Liu and Tipper [63] address dual-link-failure protection within the context of IP/MPLS or WDM networks. By using a spare provision matrix method, the authors collect information for each individual flow and thus are able to compute the shared spare capacity for dual-link failures. In order to minimize network redundancy, the authors present a non-linear integer programming model. By partitioning this model into two linear sub-models which are solved sequentially, the authors obtain upper bounds on the required spare capacity. For large networks, an iterative heuristic is proposed to solve those sub-models. The authors consider several variations of shared backup path protection, combining the extensions of the 1+1 and 1:1 protection mechanisms for dual-link failures with active or passive sharing, allowing for capacity sharing schemes of different complexity.

Liu [64] subsequently addresses protection coordination between existing single-link-failure protection schemes (independently) implemented at each layer of two-layer networks (e.g., IP/MPLS over OTN/ASON). Making use of existing single-link-failure protection mechanisms at both layers, Liu is able to compute working and backup paths on both layers in a coordinated way by (1) capturing all dual-link failures on the bottom layer; and (2) passing that information to the top layer, where the aim is to minimize spare capacity. For those networks that are not three-connected, partial disjoint paths
can be used in an attempt to still guarantee the required level of protection.

Bermond et al. [65] address the problem of finding diverse paths between a pair of nodes when multiple correlated link failures are modelled as Shared Risk Link Groups (SRLG). This problem is NP-complete for general SRLGs. In this paper, the special case is considered where all the links in an arbitrary SRLG share a common endpoint: the “star property.” The authors investigate the problem of finding $k$-SRLG-disjoint source-destination paths, and also the problem of finding the maximum number of SRLG-disjoint paths between two nodes. Also polynomial-time algorithms are proposed for graphs satisfying some special properties. A multi-coloured graph is introduced, where each SRLG is represented as a different colour. Thus, a colour belongs to multiple edges (having a common node) and an edge can have multiple colours. The authors prove that the $k$-diverse coloured source-destination path problem ($k$-DCP) is NP-complete. Polynomial cases are presented for the $k$-DCP problem: (1) When the number of colours (i.e., SRLGs) is bounded by a constant, a polynomial algorithm exists for every topology. (2) If the nodal degree is bounded, a polynomial-time algorithm exists for $k$-DCP when the maximum degree is at most three, and for 2-DCP even when it is four. (3) In Directed Acyclic Graphs (DAG), where a polynomial algorithm is only possible for constant $k$ values. Finally, the problem of finding the maximum number of colour disjoint paths (MDCP) is considered. It is shown that the MDCP problem is NP-complete in the strong sense, and it is hard to approximate within a constant factor.

C. Spatio-temporal disaster-aware routing

A large-scale disaster may affect an area of a certain shape, e.g., a circle representing the radius of an EMP attack. A problem of interest is therefore finding disjoint paths that cannot be separated by the failure of an area of a predefined size and shape.

Dikbiyik et al. [66] deal with the static network planning problem of survivable optical backbone networks. Input to the network planning problem are the topology graph, the number of wavelengths, the set of connections, and the set of possible disaster failures. First, a probabilistic risk model is developed to analyse the loss/penalty, given the set of possible disasters. It considers the physical locations of network equipment (e.g., physical routes of fibre links), their distances from the disaster’s epicentre, and the type of disaster. Second, a proactive traffic engineering solution for disaster protection is given, where valuable connections are routed on no-(or low-) risk regions. Third, the authors investigate a reactive traffic engineering solution, where disrupted connections are re-provisioned. The problem is formulated as an Integer Linear Program (ILP). Also, heuristics are developed to deal with large networks.

Trajanovski et al. [16] address the problem of finding two region-disjoint paths that (with the exception of the regions around the source and destination) cannot both be cut by a failure of given diameter. In this case, the shape of the failing region is not important, only its diameter is (as it could be rotated in any direction). The problem of finding region-disjoint paths is shown to be NP-hard, after which a heuristic algorithm is proposed and compared via simulations on realistic topologies to an exact ILP formulation and a naive approach.

Izadoost et al. [67] focus on the problem of large-scale failures in backbone networks with a dynamic probabilistic model that not only considers the time-varying dynamics of regional disasters, but also takes into account the probabilistic nature of failures resulting from such events. The authors propose a novel approach in probabilistic large-scale failure scenarios, which aims to increase the network survivability level and mitigate the effects of a disaster (connections disruption). The proposed survivability scheme is a preventive protection method that allows the network control plane to receive notifications about the current impact range of a disaster, to estimate the probability of failure for each path in the study, and to reroute the traffic from the endangered routes to the more reliable paths prior to the failure.

Iqbal and Kuipers [19] consider a similar spatio-temporal rerouting problem, for which they provide polynomial-time algorithms.

D. Post-disaster routing

As already mentioned in paragraph III-B1a, Ngo et al. [68] consider a post-disaster communication network that is based on MDRUs. To deal with the critical demand of ICT services, spectrum-efficient methods should be considered in MDRU-based networks. Furthermore, to solve the power supply problem, renewable energy functions should be used together with energy-efficient methods. In this context, the paper addresses the issue of combined optimization of both spectrum and energy efficiency in order to provide better system performance in a post-disaster situation. The authors introduce a new metric, namely, the spectrum-energy efficiency, to measure how many transmissions can be carried out with a limited frequency band and limited energy resources. They propose a scheme that is composed of two phases, namely, topology formation and transmission division. The topology formation phase creates a topology by using the top $k$ spectrum-efficient disjoint paths from each sender. The gateways that are not in the resulting topology are not used. In the transmission division phase, the traffic is split from each gateway to the neighbours in the topology by using a max-flow-with-vertex-capacities algorithm. The authors prove that a value of $k$ exists that leads to the maximum spectrum-energy efficiency of the MDRU-based network and that the proposed algorithm has polynomial complexity with respect to $k$.

Also the use of decentralized mobile wireless networks, such as delay and disruption tolerant networks (DTNs) that do not require end-to-end connectivity between source and destination, is a possible way to deal with disasters. However, DTN routing protocols were not designed for that purpose, and hence may suffer from performance degradation. DTN protocols, like the Spray and wait flooding based routing
protocol that attempts to gain the delivery ratio benefits of replication-based routing as well as the low resource utilization benefits of forwarding-based routing, may have to be adapted. Huda et al. [69] start by pointing out the limitations of DTN routing algorithms in areas affected by large-scale disasters: either they result in excessive energy consumption or they perform poorly because their underlying assumptions are no longer valid. To significantly reduce the call surge in a disaster-stricken area, the authors propose a Location-aware Message Delivery (LMD) approach to provide short message communications among family members, friends, and co-workers. The objectives of this system are to save power at the nodes (which are battery powered) and to ensure a high message delivery ratio. To achieve these objectives, LMD requires that communication devices must have location awareness and that the exchange of statistical location information and respective time of day must take place automatically among authorized parties (family members, friends, etc.). LMD uses a single copy of the message, makes locally optimal decisions and ensures an inherently loop-free forwarding rule.

V. CONCLUSION
Recent natural disasters have highlighted the relevance of communication systems for effective disaster mitigation. Emergency networks must be able to operate in challenging scenarios and allow to transmit the information necessary to deploy and coordinate relief operations. Moreover, the network architecture should be designed so that services for customers not in the affected areas suffer minimal impact.

In this work, we gave an overview on the state of the art in large scale regional failures. Approaches for network vulnerability assessment, strategies for enhancing the robustness of an existing network, and solutions for achieving resilient routing, including disaster-aware routing were presented.

ACKNOWLEDGMENT
We would like to thank the participants of WG1 (Large-scale natural disasters) of COST Action CA15127 who indirectly collaborated in this task: Michał Aibin, Péter Babarczi, Vitoria Bueno Delgado, Marco Casazza, Anna Fogtuen, David Hay, Bjørne E. Helvik, Rita Girão-Silva, Róza Gościen, Yuming Jiang, Peter Kieseberg, Ioannis Krikidis, Konstantinos Manousakis, Maria do Carmo Medeiros, Cemalettin Ozturk, João Patricio, Maria Potop-Butucaru, Luis Quesada, Sarah Ruepp, Dorabella Santos, Noor Shirazi, Krzysztof Walkowiak, and Zhongliang Zhao.

This article is based upon work from COST Action CA15127 (“Resilient communication services protecting end-user applications from disaster-based failures – RECODIS”) supported by COST (European Cooperation in Science and Technology).

REFERENCES