ROBUSTNESS AGAINST LARGE-SCALE FAILURES IN COMMUNICATIONS NETWORKS

Juan SEGOVIA SILVERO

Dipòsit legal: GI-251-2012
http://hdl.handle.net/10803/70008

Department of Computer Architecture and Technology

PhD Thesis

Robustness against Large-Scale Failures in Communications Networks

Juan Segovia S.

Advisors: Dr. Eusebi Calle and Dr. Pere Vilà

Submitted in fulfillment of the requirements
of the degree of PhD in Computer Engineering

Department of Computer Architecture and Technology
University of Girona
Girona, Spain

October 2011

Universitat de Girona

El **Dr. Eusebi Calle**, Titular de universitat del Departament d'Arquitectura i Tecnologies de Computadors de la Universitat de Girona, i el **Dr. Pere Vilà**, Titular de universitat del Departament d'Arquitectura i Tecnologies de Computadors de la Universitat de Girona,

CERTIFIQUEM

Que aquest treball, titulat "Robustness against Large-Scale Failures in Communications Networks", que presenta *Juan Segovia Silvero* per a l'obtenció del títol de doctor, ha estat realitzat sota la nostra direcció i que compleix els requeriments necessaris.

Dr. Pere Vilà Dr. Eusebi Calle

Girona, 7 d'octubre de 2011.

Contents

Abstract

This thesis is devoted to the study of robustness against large-scale failures in communications networks. Society has come to rely on communications networks for business and leisure and there is high expectation on their availability and performance. Communications networks (or a part thereof) can experience failures, for example due to cables cuts or node breakdowns, but such isolated failures usually go unnoticed by users thanks to effective recovery mechanisms put in place to conveniently mask the failures by applying the required corrective measures. Nevertheless, such mechanisms are not effective when large-scale multiple failures arise, that is, when a significant portion of the network fails simultaneously.

Large-scale failures usually have serious consequences in terms of the economic loss they cause and the disruption they bring upon thousands or even millions of users. A key requirement towards devising mechanisms to lessen their impact is the ability to evaluate the robustness of the network, that is, be able to assess the performance degradation that should be expected as a consequence of the failure.

In this thesis, our focus is on multilayer networks featuring separated control and data planes, as in GMPLS. Thus, the unit of service is a connection (for example, a lightpath in an optical network). Unfortunately, the majority of the existing measures of robustness are unable to capture the true service degradation in such a setting, because they essentially rely on purely topological features.

One of the major contributions of this thesis is a new measure of robustness, whose distinguishing feature is that it performs functional assessment to overcome the aforementioned limitation of the existing measures. The failure dynamics is modeled from the perspective of epidemic spreading, for which a new epidemic model is proposed. This model also takes into account that each GMPLS node, due to the separation of planes, is subjected to a multi-state failure that affects its own functionality and that of the whole network.

Finally, another contribution is a taxonomy of multiple, large-scale failures, adapted to the needs and usage of the field of networking, which results from the comprehensive literature review carried out as part of this research.

Resumen

La presente tesis está dedicada al estudio de robustez contra fallos a gran escala (o masivos) en redes de comunicaciones. La sociedad actual tiene una gran dependencia de estas redes, tanto para llevar a cabo diversas actividades económicas como para el ocio, y los usuarios tienen altas expectativas respecto a su disponibilidad y desempeño. Las redes de comunicaciones experimentan fallos, por ejemplo debido a cortes de cables o a la avería de algún nodo. Sin embargo, tales fallos aislados a menudo pasan inadvertidos para los usuarios gracias a que incorporan efectivos mecanismos de recuperación, diseñados para ocultar convenientemente los fallos aplicando las medidas correctivas requeridas. En cualquier caso, tales mecanismos no son efectivos cuando surgen fallos masivos y múltiples, es decir, cuando una parte significativa de la red falla simultáneamente.

Los fallos masivos suelen tener serias consecuencias en términos de las pérdidas económicas que acarrean y los trastornos que causan a miles e incluso millones de usuarios. Un requisito esencial para avanzar hacia el diseño de mecanismos que permitan amortiguar los efectos negativos, es la capacidad de evaluar la robustez de la red, es decir, poder valorar la degradación de desempeño que cabe esperar a consecuencia del fallo.

En esta tesis nos enfocamos básicamente en redes con arquitectura multi-nivel que poseen un plano de control y un plano de datos separados, como ocurre en GMPLS. Por lo tanto, la unidad del servicio es una conexión (por ejemplo, un *lightpath* en una red óptica). Desafortunadamente, la mayoría de las medidas de robustez existentes no son capaces de capturar la verdadera dimensión de la degradación del servicio en un entorno como estos, ya que dependen esencialmente de propiedades puramente topológicas.

Una de las principales contribuciones de esta tesis es una nueva medida de robustez, cuya característica distintiva es que realiza una valoración funcional para superar las limitaciones observadas en las medidas de robustez existentes. La dinámica de los fallos es modelada desde la perspectiva de la propagación de epidemias, para lo cual un nuevo modelo epidémico es propuesto. Este modelo considera además que cada nodo GMPLS, a consecuencia de la separación de planos, está sujeto a un régimen de fallos multi-estado que afecta a su propia funcionalidad pero también a la de toda la red.

Finalmente, otra contribución es una taxonomía de fallos masivos, adaptada a las necesidades y usos del campo de las redes de comunicaciones, resultado de una revisión exhaustiva de la literatura respectiva, realizada como parte del presente trabajo de investigación.

Acknowledgments

First and foremost, I would like to thank my advisors, Dr. Eusebi Calle and Dr. Pere Vilà, for their guidance and constant support throughout my research. During these years, they have been a source of inspiration and encouragement for me, and their patience and experience greatly contributed to bring my work into fruition.

I also want to express my gratitude to Professor José L. Marzo and to Dr. Ramón Fabregat. They gave me the opportunity to join the Broadband Communications and Distributed Systems (BCDS) group and were instrumental in securing the financial support I needed while working towards my PhD. Many thanks also to Dr. Benjamín Barán, who encouraged me to start this adventure.

I am equally grateful to Dr. Janos Tapolcai from Budapest University of Technology and Economics, Dr. Sarah Ruepp from DTU Fotonik, with whom I had the opportunity to discuss ideas that have enriched my work. Many thanks also to all my colleagues at BCDS, for their support, friendship and the enjoyable time spent together.

Finally, I thank my family, whose love and support have been constantly with me, helping me to cope with the rigors of distance.

This work is partially supported by the Spanish Ministry of Science and Innovation projects TEC 2009-10724 and MTM 2008-06349-C03; by the Generalitat de Catalunya through the research support programs SGR-1202 and SGR-296, and through its predoctoral grant program.

List of Figures

List of Tables

List of Acronyms

AS	Autonomous System
ATM	Asynchronous Transfer Mode
GMPLS	Generalized Multiprotocol Label Switching
IP	Internet Protocol, or the TCP/IP protocol family
LER	Label Switch Edge Router
LSP	Label Switched Path
LSR	Label Switch Router
MPLS	Multiprotocol Label Switching
MTBF	Mean Time Between Failures
MTTR	Mean Time To Repair
OAM&P	Operation, Administration, Maintenance and Provisioning
OXC	Optical Cross-Connect
PDU	Protocol Data Unit
QoS	Quality of Service
RFC	Request for Comments
RSVP	Resource Reservation Protocol
SDH	Synchronous Digital Hierarchy
SONET	Synchronous Optical Network
SRLG	Shared-Risk Link Groups
TDM	Time Division Multiplexing
WDM	Wavelength Division Multiplexing

1 Introduction

The purpose of this chapter is to present the motivation for our research work, identify the main objectives and provide an overview of the structure the document.

1.1 Motivation

Communications networks have become an essential piece of infrastructure for society. They facilitate and promote the exchange of ideas, goods and services, make people's daily life easier and, in general, act as an enabling technology so that human societies can develop and prosper further. Around the world and without pause, networks carry all kinds of traffic, from cellular phone conversations and credit card transactions, to multimedia content shared either for business or leisure.

For networks to be effective and serve whenever they are called upon, they must be reliable. In fact, one obvious user expectation is that data arrives intact at its destination, no matter how far that is, or which combination of communications media and technologies are used along the path. As any engineering system, however, a network (or a part thereof) can fail for a number of reasons, for example because of faulty hardware, software bugs, breakage of physical medium (e.g., fiber cables), and even because of power outages. These are all examples of failures that affect specific and, generally, separated networks elements. It is customary to assume that such failures are independent events, and empirical evidence exists that very rarely do several such events overlap in time on one network, so that the resulting setting is commonly referred to as a *single failure* scenario.

A large number of techniques exist for dealing with failures, collectively known as network recovery techniques. The fundamental idea underlying recovery is that of redundancy, whereby network elements deemed to be unreliable are backed up with one or more spare resources that come into

play upon a failure. Almost all of the recovery techniques focus on single failures and seek to offer a specific trade-off between resilience guarantee and resource consumption. The recovery techniques are indisputably mature and their efficacy is proven. It is reported in the literature that while failures occur regularly, even daily, in the network of a typical telecommunications operator, they go largely unnoticed by users.

More recently, a different class of failures has been attracting attention, namely, the class of *large-scale* failures. The distinguishing trait here is that a significant portion of the network fails simultaneously, often due to a single cause, such as natural disaster or an intentional attack. From the point of view of the traditional recovery techniques, a large-scale failure is difficult to handle due to the fact that the redundancy-based approach that is effective for single-failures is no longer suitable: the cost of implementing massive redundancy for rarely occurring events is simply prohibitive.

Although large-scale failures may be relatively rare, they usually have serious consequences in terms of the economic loss they cause and the disruption they bring upon thousands or even millions of users. Therefore, it is vitally important to have methods and tools that can be used in the design and operation of the communications infrastructure so that essential services can be preserved as much as possible when large-scale failures occur. In this context, a key requirement is the ability to evaluate the *robustness* of the network, that is, be able to assess the performance degradation that should be expected as a consequence of the failure.

Multiple failures have long been studied from the topological perspective, for example by measuring the variation in connectivity when nodes are removed from the network. In recent years, research in the area known as "Network Science" has focused on studying the characteristics, evolution and behavior of complex systems, among them the Internet. Although such studies provide valuable knowledge that can be used to better understand large-scale failures, they tend to equate failure to node (or link) removal and rely on graph connectivity as the main measure of robustness.

This thesis is primarily concerned with path-oriented transport networks, that is, the core infrastructure of a telecommunications operator designed to carry large volumes of aggregated traffic organized in individualized channels called "connections". The concepts and methods developed here can be applied in any path-oriented network environment, but we choose to focus on GMPLS-controlled optical networks. Note that these types of networks present the following peculiarities with respect to failures:

- The architecture is usually multilayer, combining two or more transmis-

sion technologies. Thus, nodes are complex units consisting of several hardware and software components, providing well-defined independent functions.

Therefore, node state is not necessarily binary (is working/has failed), meaning that one node-level functionality might be disabled, due to a component failure, while others might not.

- Nodes and links are not equal in the role they play as communication intermediaries. For example, depending on topological structure as well as routing policy, link capacities, demand types and other factors, one node might concentrate a sizeable fraction of the total traffic, while an adjacent node is only lightly utilized.

  Thus, the failure of one element may lead to substantial changes in traffic flow and service quality, whereas the impact of losing some other node might be less dramatic.

Assessing the vulnerability to failures for these types networks solely through structural measures, e.g., size of the largest component, can be misleading. Consider, for example, the case in which a well connected node of a large network fails. The residual topology remains fully connected, therefore, from the structural point of view, the network suffered a small change. But from the functional point of view the situation might be quite different if several new connections have to be rejected because some required quality parameter or constraint cannot be satisfied any more. This can happen for example if there is a maximum admissible hop count per connection and new paths are significantly longer due to the absence of the failed node. Obviously, a structural robustness measure cannot take into account partial node failures either.

1.2 Objectives

The objective of this thesis is to study the vulnerability of communications networks to large-scale failures, and develop methods to measure and compare their functional robustness.

Our focus is on transport networks and we assume that they are operated through a control plane that is separate from the data plane, as in GMPLS. Therefore, partial as well as complete node failures should be taken into account.

1.3 Outline of the Thesis

This document is organized into 6 chapters, including this one, and additionally the bibliography and appendixes at the end.

In Chapter 2, we present a review of fundamental resilience-related concepts and background information on the technologies and protocols found in transport networks, among them an overview of GMPLS. The chapter also includes a taxonomy of network failures, both single and multiple failures, and an overview of the basic recovery techniques employed in GMPLS-based networks.

Chapter 3 is devoted to network models and measures of robustness. It begins with an introduction to basic definitions of graph theory commonly used in networking, then it describes the graph properties used for categorizing networks (network models), and finally gives a summary of the measures of robustness found in the literature.

In Chapter 4, we summarize the benefits and limitations of redundancy-based network recovery and then numerically evaluate topological damage through a number of well-known metrics. Then, we perform a numeric evaluation of functional damage in a multiple link failure scenario and propose heuristic-based strategies to increase functional robustness in that context.

Chapter 5 deals with propagating multiple node failures. It includes a description of our epidemic-based multiple failure propagation model called "SID" and presents an application of that model to the evaluation of availability on ring topologies. Finally, it defines a new robustness metric for path-oriented networks and demonstrates the use of the new metric to compare several topologies from the point of view of functional robustness.

In Chapter 6, we summarize the results of our work and then outline some ideas that may be worth exploring for future research.

There are two appendices: Appendix A contains the author's list of publications and Appendix B gives detailed information about the topologies used in this dissertation.

2 Background on Resilience and GMPLS Networks

This chapter starts with a review of fundamental resilience-related concepts from the perspective of their applicability to communication networks. Then, an overview of the technologies and protocols usually deployed in transport networks is presented, together with a brief introduction to GMPLS. Finally, we give a taxonomy of failures and summarize the approaches to recovery usually applied in GMPLS-based networks.

2.1 Fundamental Concepts of Resilience

The terminology concerning resilient networks is still evolving. It is common to find that two different terms are used to describe a single property or concept, or one term referring to partially overlapped or even different concepts. Furthermore, terms such as reliability, availability, dependability, and survivability are also used in other fields. To avoid confusion, we start this section by defining them from the perspective of communications networks.

2.1.1 Faults, Errors and Failures

A *failure* is an event that occurs when the delivered service deviates from correct service, that is, when the system does not perform its intended function, or does it without the required quality. It is a manifestation of an error observable from outside of a system [67]. A service fails either because it does not comply with the functional specification, or because the specification did not adequately describe the system function. A failure is a transition from correct service to incorrect service, and the period of delivery of incorrect service is a called *service outage*, while the transition from incorrect service to correct service is called *service restoration*.

Determining the exact cause of a failure in a system can sometimes be difficult, impractical or even impossible, due to the fact that it is usually a complex set of interacting components, where these in turn can be complete systems. The structure as a whole implements the required functions, maintains a set of internal and external states, and exposes a certain behavior, which, from the point of view of its user, constitute the *service* [7].

An *error* is a system state that is liable to lead to failure. It is the manifestation of a fault within a system. Since a service is a sequence of the system's external states, a service failure means that at least one external state of the system deviates from the correct service state [67],[7].

A *fault* is the cause of an error, determined or hypothesized. Faults can be internal or external to a system, and may have a multitude of physical and human causes, and be of various types. Many errors do not affect the system's external state, that is, do not develop into a failure, because the system can have mechanisms to detect and isolate the fault and thus neutralize its consequences, ability known as *fault tolerance*. Examples of fault tolerant techniques employed to overcome transient errors in communication are FEC (forward error correcting codes) and CRC (cyclic redundancy check), as well as retransmission. However, fault tolerance is of little use to cope with faults made at the design or specification stages, and the only practical alternative in such cases is *fault prevention*, that is, aim at avoiding faults. Fig. 2.1 illustrates these two complementary approaches. It also highlights the need for recovery planning when errors finally provoke a failure.

The failure of a component causes a permanent or transient fault in the system that contains it, which in turn provokes a fault in one or more system(s) for which service is provided, hence giving birth to a failure propagation. This propagation and its effect on service delivery of a simple two-component system is illustrated in Fig. 2.2.

A system operating in *degraded mode* offers a subset of its services after a failure. The specification may identify several modes, for example slow service, limited service, emergency service, and so on. As only a subset of its functionality, or only its performance is suffering, it is said that the system has suffered a *partial failure*.

2.1.2 The concept of Resilience

In the context of communication networks, *resilience* is defined as the capability of a network to operate and maintain an acceptable level of service in the presence of adverse conditions [129], whose cause can be varied, as discussed in Section 2.4. Underlying this concept is the requirement that,

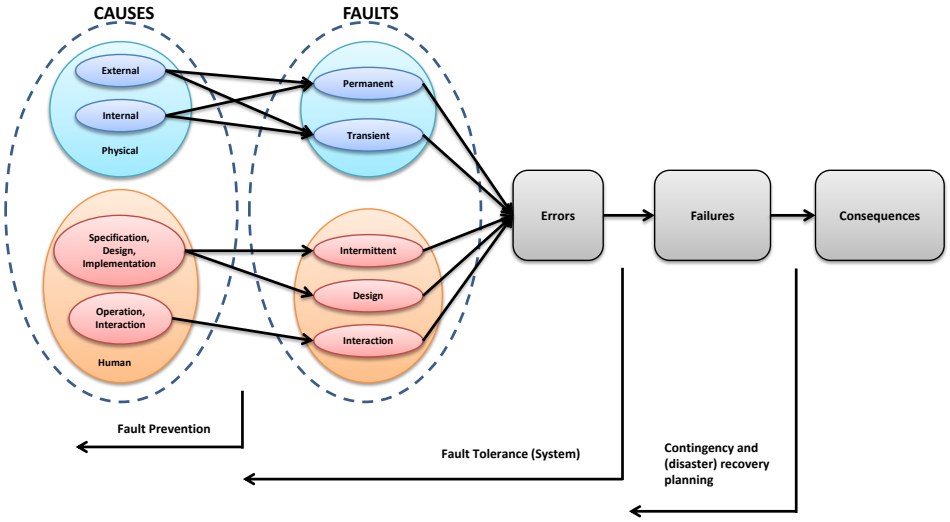


Figure 2.1: *Relationship between fault prevention, fault tolerance and recovery [67]*

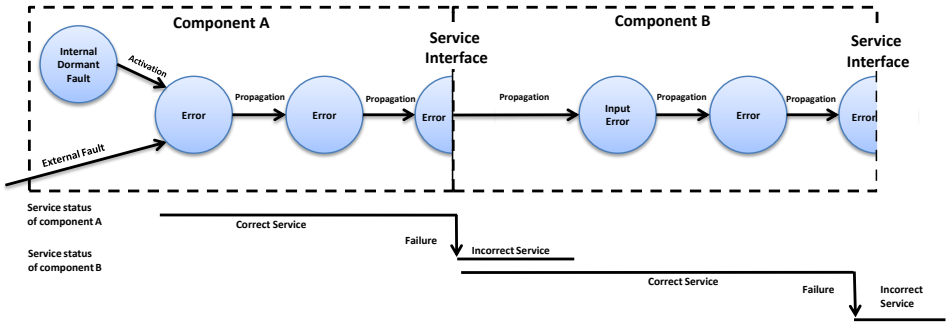


Figure 2.2: *An example of error propagation in a two-component system [7]*

when challenges arise, the service provided by the network must remain accessible, even if that means operating in a degraded mode, and that the network must start recovery actions to put it out of degraded mode rapidly and automatically [146].

Resilience is a desirable system capability not only in networking but in the majority of engineering fields. Given its broad definition and scope of application, different approaches have been developed over time, targeting specific challenges and problem domains, and giving rise to sometimes overlapping or diverging nomenclatures. In the networking literature it is common to find terms such as survivability, reliability, fault tolerance, robustness, and dependability. There have been efforts to produce a unified, integrating framework, but reaching a consensus is difficult given that these concepts have their own history and are well entrenched in their respective fields, where often they are viewed as end objectives themselves rather than attributes of some other concept [4].

The conceptual framework presented in [129] divides the all-encompassing abstract idea of resilience into two categories of disciplines (see Fig. 2.3). Included in the first category, called "challenge tolerance disciplines", are those that deal with the design and operation of systems capable of providing service continuity when faced with challenges. In the second category, "trustworthiness", are those that define measurable properties of resilient systems, such as reliability and availability. We can relate the first group of disciplines to design objectives, and the second to the assessment of their performance from different perspectives.

2.1.3 Fault Tolerance and Survivability

Fault tolerance and Survivability are two concepts widely used in the networking literature. According to the taxonomy proposed by [129], fault tolerance is part of the broader design objective of survivability.

Fault Tolerance

Fault tolerance is defined as the ability of a system to tolerate faults such that they do not provoke a service failure. Traditionally, the design technique used to implement fault tolerance is *redundancy*, whereby systems components deemed to be unreliable are backed up with one or more spare components that come into operation should the main component fails. The basic idea is to increase system reliability out of relatively less reliable parts. The component can be a physical element (a circuitry, a fiber link, a complete

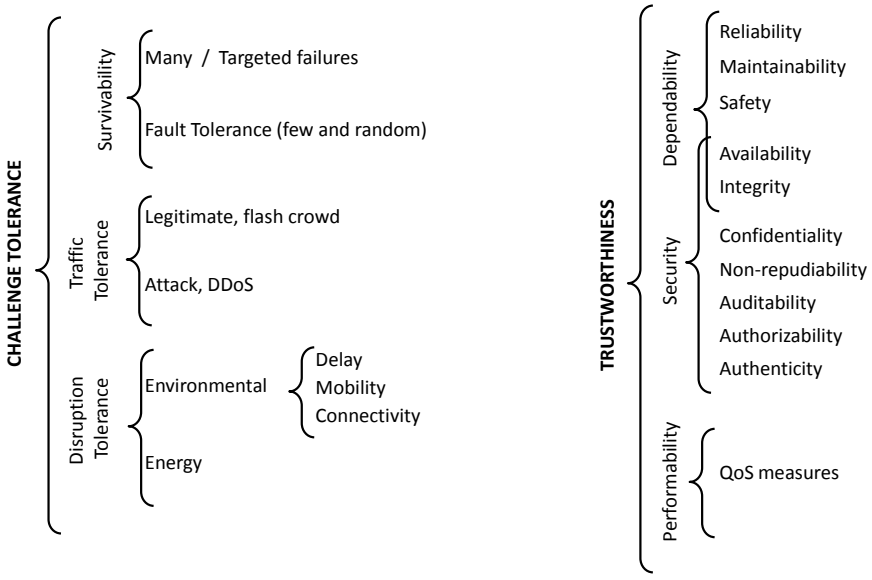


Figure 2.3: *Categories of resilience disciplines [129]*

node, etc.) or a logical one, such as a software module or a path.

Fault tolerance usually requires fault detection mechanisms, as well as the ability to reconfigure the system dynamically. To that end, the system must also incorporate the ability to perform fault localization, fault notification, fault containment and fault recovery [4].

Survivability

Survivability has been defined in several ways over time; a good summary can be found in [78]. In a recent proposal, it is defined as the capability of a system to fulfill its mission, in a timely manner, in the presence of threats such as attacks or large-scale natural disasters [129]. The primary design goal is the fulfillment of the mission, which implies performing only essential services (degraded mode operation) instead of attempting full service recovery after or during a failure. Thus, survivability applies to the entire system that offers well specified services, not to particular components of the system [4].

The given definition pays especial attention to multiple failures, and is what leads their proponents to consider survivability as a superset of fault tolerance.

On the other hand, the term *network survivability* is defined in [96] as "the set of capabilities that allows a network to restore affected traffic in the

event of a failure. The degree of survivability is determined by the network's capability to survive single and multiple failures." Note that this definition ignores the possibility that traffic can be restored selectively.

In essence, survivability, as a design objective, offers a trade-off between functionality and resource consumption. For example, if a large network is required to be highly available, and there exists the risk of it being the target of a coordinated attack, the resources required to implement redundancy to shield the services from the effects of such attack would be prohibitive due to the fact that large parts of the system might have to be duplicated in full. This trade-off means that, at design time, a decision must be made about which classes of faults are considered unrecoverable and which ones should be coped with. The consequence is that cost can be reduced but the potential for failure increases. If the rate of failure is below what is deemed acceptable, and if alternate service is provisioned for all the failure scenarios, then the network would be survivable under the given conditions [78].

2.1.4 Basic Assessment of Resilience

In this subsection, the most common measures of resilience employed in the networking literature are defined, namely reliability and availability. Additionally, the key concept of robustness, used throughout the thesis, is presented.

Reliability

The International Telecommunications Union (ITU-T) recommendation E.800 defines *reliability* as the "ability of an item to perform a required function under given conditions for a given time interval." [72]. The term "item" refers to any element of interest in the system, be it simple or compound. As a measurable attribute of resilience, reliability $R(t)$ is a function of time that calculates the probability $R(\tau)$ of uninterrupted service from $t = 0$ to $t = \tau$. In practice, the computation depends on several aspects which are specific to the item under consideration, such as the probability distribution to characterize system failure (e.g., exponential, normal, Weibull) and the reliability analysis technique to be employed (e.g., reliability block diagrams, fault trees, Markov models) [4].

In any case, in the networking literature, the term "reliability" is often used, despite its formal definition, with slightly different meanings by different authors, usually to convey the idea that some quantitative attribute is above or below a given threshold. Additional usage of the term can be found in

Section 3.4.3 concerning the measurement of network connectivity.

Availability

Another concept closely related to reliability is *availability*, defined as the "ability of an item to be in a state to perform a required function at a given instant of time or at any instant of time within a given time interval, assuming that the external resources, if required, are provided." [72].

Availability considers two equally valid ways for an item to be in the working state: either it has been working without any failure since it began operating, or it has failed once or several times but has been repaired each time. Thus, it reflects a statistical equilibrium between failure processes and repair processes in maintained repairable systems. The question that availability strives to answer is: given the frequency of failures and the rate at which repairs are conducted, what is the average fraction of time that one will find the system in the operating state? [40].

To numerically assess availability, the most common equation is

$$A = \frac{MTBF - MTTR}{MTBF}, \tag{2.1}$$

where $MTBF$ and $MTTR$ are the mean time between failures and the mean time to repair, respectively, which are usually provided by system vendors or obtained through observation over long periods of time [124].

Some authors propose an integrative concept called *dependability* that encompasses reliability and availability, together with other attributes not described here such as maintainability, safety, confidentiality and integrity [4],[129]. As a qualitative property, dependability defines the ability to deliver service that can justifiably be trusted, whereas from a quantitative point of view, it is the ability to avoid service failures that are more frequent and severe than is acceptable to its users [7]. Dependability is seldom found in the networking literature.

Robustness

According to Sterbenz et al. [129], *robustness* is neither a design objective nor a measure of trustworthiness, but an indicator of the behavior or performance of a system subjected to specific challenges. Therefore, it can be used to characterize the reaction to faults and failures of a given system (e.g., a whole network, a connection, a link). It must be noted, though, that in the literature it is also used as synonym for resilience or survivability, or to

indicate that a system is able to withstand a challenge without breaking down, as in a robust topology that can experience a certain number of random link removals and even then keep connectivity.

2.2 Overview of Transport Network Technologies

The demand for capacity, especially at the core of the network, has been growing tremendously. Optical network technologies has proven instrumental in coping with that demand, especially with the advent of Wavelength Division Multiplexing (WDM).

Fiber optic cable as a practical transmission medium was introduced in telecommunications in the early 1980s. In the beginning, the prevailing layer-1 technology developed to build fiber-optic based networks was SONET (Synchronous Optical Network). ITU sponsored a derivative called SDH Synchronous Digital Hierarchy. As they are closely related, they are sometimes referred to as SONET/SDH, but for simplicity we will use just "SDH" from now on.

SDH is a circuit-oriented, voice-optimized technology that was developed to supersede PDH (Plesiochronous Digital Hierarchy). It is a TDM (Time Division Multiplexing) system and its introduction represented a major revolution in the early 1990s because the previous generation was rather lacking in what is called *Operation, Administration, Maintenance and Provisioning* (OAM&P). SDH incorporated extensive OAM&P capabilities that allowed carriers to offset the cost of building a new backbone network with the substantial cost reduction offered by OAM&P, which represents their primary expense.

SDH networks can be configured in point-to-point, ring or mesh topologies, although most of them are configured as rings. SDH possesses a very rich set of OAM&P capabilities, and provide protection and restoration capabilities. The most common use of SDH nowadays is as a building-block (transport substrate) for other technologies. For example, it used to be common to transport ATM traffic over SDH, and the same happened when the TCP/IP protocol family and the Internet became widely deployed.

In the first stages of the development of optical transport technologies, one optical fiber cable provided exactly one data channel, but this changed with the advent of *Wavelength Division Multiplexing*.

2.2.1 Wavelength Division Multiplexing

WDM is a technology that enables the transmission of more than one optical signal by a single fiber at the same time. Its principle is essentially the same as frequency-division multiplexing (FDM), that is, several signals are transmitted using different carriers, each occupying non-overlapping parts of a frequency spectrum [30]. That possibility was foreseen very early in the development of optical communications, but it was only realizable with later advances in the technology of optical materials and components, such as erbium-doped fiber amplifiers (EDFAs), and photodetectors.

WDM was initially deployed as point-to-point systems to mitigate the problem of capacity exhaustion. Traditionally, increasing capacity needed the deployment of additional fibre and replacement of installed equipment with new higher-rate TDM systems. WDM opened the possibility that optical links could have their capacity multiplied manifold, without the need to lay new fibers. Of course, the equipment at the nodes connected to those fibers would need to be replaced, but that would cost less than installing new fibers.

With the current technology, several hundred optical channels — or *lambdas*, as they are also called — can be multiplexed into a single fiber, each one operating at 40 gigabits per second or more [128]. Notable features of WDM include the ability to amplify all the wavelengths at once without first converting them to electrical signals, and the ability to carry signals of different speeds and types simultaneously and transparently over the fiber. Naturally, WDM can be used to transport SDH traffic, which helps preserve investment in equipment.

2.2.2 WDM Networks

WDM networks are constructed by linking together optical cross-connect (OXC) nodes following a certain topology of choice. The purpose of an OXC is to switch an optical data stream from an input port to an output port [119]. OXCs usually encompass wavelength multiplexers and demultiplexers, a switching engine, and wavelength converters. A schematic representation of an OXC is given in Fig. 2.4. The demultiplexer is responsible for decomposing an optical signal into its constituents wavelength channels. Once separated, they are sent to a bank of optical switches so that an appropriate output port for each signal is selected. Before being injected into the outgoing fibers for transmission, signals are multiplexed again. An OXC may utilize optical-electrical conversion at the input port, and electrical-optical conversion at

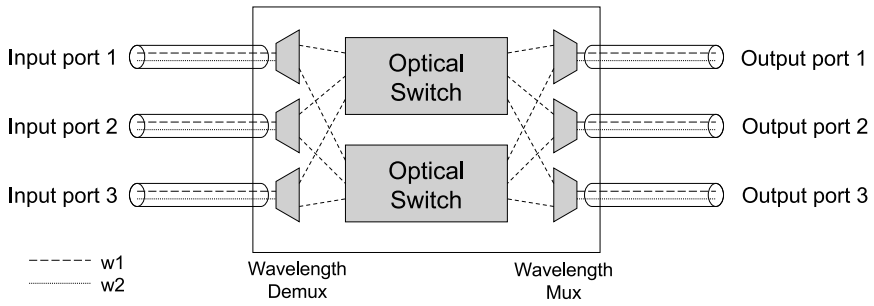


Figure 2.4: *A schematic representation of an 3x3 optical cross-connect (OXC)[121]*

the output port, a process called optical- electrical-optical conversion, or OEO, or it may be all-optical [30],[119].

A specific type of OXC usually found in optical networks is called optical Add-Drop Multiplexer (OADM). An OADM device is capable of filtering an incoming wavelength at an intermediate node along a path, removing it from the incoming signal and directing it to a drop port. Conversely, it can add one or more new wavelength channels to an existing WDM signal. An OADM device can either add/drop fixed wavelength(s) or dynamically select its target wavelength(s), in which case it is called reconfigurable OADM (ROADM) [52],[139].

All in all, what a WDM network offers to a client (any other network that uses it to transport traffic) is a logical connection between two designated source and destination points, realized by an end-to-end optical path, a *lightpath*. Lightpaths may traverse a number of fiber links in the optical network, ideally without being subjected to any OEO conversion. Given that a lightpath behaves as a clear channel between the source and destination, in theory there is nothing in the signal path to limit the throughput of the fibers [121]. If no wavelength converters are used, a lightpath is associated with the same wavelength on each hop. This is the so-called *wavelength continuity constraint*. By using wavelength converters, traffic that arrives on a certain wavelength might leave on a different one if the frequency used by the input wavelength is in use in the output link selected.

The procedure of setting up a lightpath between any source-destination pair involves choosing an appropriate route and assigning the required wavelength(s) on the route selected. This problem is referred to as the *Routing-and-Wavelength Assignment* (RWA), whose computational complexity is very high when the continuity constraint is present [31]. Wavelength conversion, in its most general form, removes the wavelength continuity constraint, making

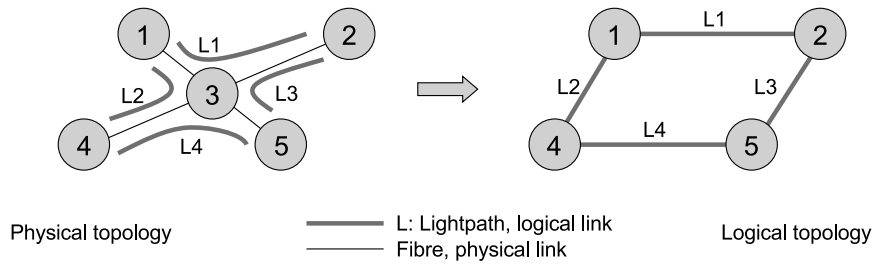


Figure 2.5: *Example of a logical topology defined over a physical network*

it possible to establish a lightpath as long as each link along the path from source to destination has a free wavelength and at the same time reducing the RWA problem to the classical routing problem [121].

Lightpaths can be established statically, before the network operation begins, or dynamically, on demand. In either case it gives greater flexibility to the network design, operation and maintenance. In fact, lightpaths make it possible to abstract away from the physical topology, as it allows for the definition of multiple and even concurrent logical topologies over a single physical network. Fig. 2.1 shows an example where a logical ring topology is defined over a physical mesh network.

Several factors contributed to the broad adoption that the WDM technologies have today. With them came greater flexibility to the optical core networks, greatly increased network capacity without requiring new fibers, and reduction of OAM&P costs. Moreover, its adoption can be considered as low-risk because it can be carried out in stages thanks to WDM's capability to interoperate with technologies of the previous generation such as SDH and ATM.

2.3 GMPLS-based Networks

Taking into account that transmission technologies evolve continually and that networks grow over time (in coverage, services, clients, etc.), it is not unexpected that their operation require combining several technologies at once. One popular organization is that of layering [67]. Layering is an approach that, in a sense, reduces complexity, but performing the OAM&P functions in such networks is complex, as each layer or technology comes with custom OAM&P mechanisms adapted to its purpose, and are rarely fully interoperable with the others.

Two efforts to provide a way to seamlessly operate layered optical

networks are Automatically Switched Optical Networks (ASON) [74] and Generalized Multiprotocol Label Switching (GMPLS). At their conception, ASON and GMPLS pursued different paths. Whereas ASON focused on the functional definition of transport networks, so that their protocol independence can be guaranteed, GMPLS followed a more practical approach, defining protocols and interoperation guidelines. However, nowadays there is a concerted effort to bring together both initiatives.

GMPLS provides a unified conceptual framework for signaling, routing and link management in different types of transports networks [95]. It is an extension and generalization of MPLS, hence its name. Multiprotocol Label Switching (MPLS) introduced to the IP protocol suite the ability to forward packets along semi-permanent, previously-provisioned paths, called Label Switched Paths (LSP). GMPLS extends the label switching architecture of MPLS to other types of non-packet based networks, so that it supports the following types of switching: packet switching (IP, ATM, and Frame Relay), wavelength switching in a wavelength-routed network, port or fiber switching in a wavelength-routed network, and time slot switching for a SONET/SDH cross-connect [121]. Its proposed features sparked great interest both in academia and in the industry, who quickly took advantage of the concepts to improve network management and advance the convergence of their transport networks.

Even tough technologies such as GMPLS can facilitate the operation of multilayer networks, there are several drawbacks, among them the capacity overhead introduced by the layering itself, that is, by the successive encapsulation of protocol data units, as in IP over ATM over SDH. Thus, a long-term objective discussed both in academia and the telecommunications industry is to progressively reduce the number of layers, with the desired final state being an IP-over-WDM network, as depicted in Fig. 2.6.

In this dissertation, GMPLS takes a paradigmatic role, so that networks whose operations follow the models advocated by GMPLS are considered fundamentally similar, in particular regarding the use of precomputed paths as a way to organize traffic flow. Below, we use this feature to characterize *connection-oriented networks* and then we present a summary of the main features of GMPLS.

2.3.1 Connection-oriented versus Connectionless Networks

Communications networks may be broadly divided in two groups based on their approach to routing, namely *connection-oriented* and *connectionless* networks [133]. In connection-oriented networks, before any user data can

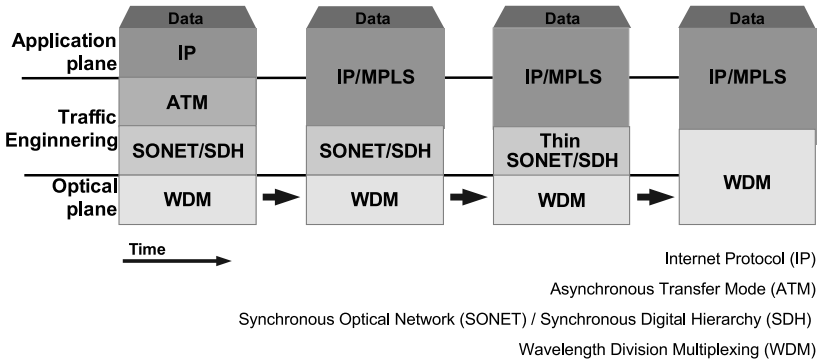


Figure 2.6: *The temporal evolution of multilayer networks. The expected final stage is a two-layer GMPLS-controlled IP-over-WDM network [8]*

be moved from source to destination, a virtual circuit is established between them. The virtual circuit is realized as one or more paths through the network and the data traffic between the two parties flows through them exclusively. This circuit is referred to as *connection* and explains the term *connection-oriented* (or its equivalent *path-oriented*). On the other hand, in connectionless networks, there are no virtual circuits, and the data stream is divided into smaller units called packets, which are routed through the network independently towards the destination, so that one packet could end up using a different path from its predecessor in the data stream.

Each approach has its advantages and disadvantages, and the election of one over the other has important consequences on network design and operation. For example, in connectionless networks it is difficult to provide guarantees about the performance of a given data flow when failures arise. In fact, in pure IP, which is the paradigmatic protocol suite of connectionless networks, routing is based on the best-effort principle, meaning that all the traffic receives services without any guarantee, and resources (for example, capacity) are not tied to specific data flows. In contrast, in connection-oriented networks, resources may be allocated, exclusively or otherwise, during path establishment so as to meet requirements on bit rate, resilience and other quality parameters.

In core networks, where traffic volume is huge and violations of service guarantees may bring on costly financial penalties, the preference has been to use the connection-oriented approach.

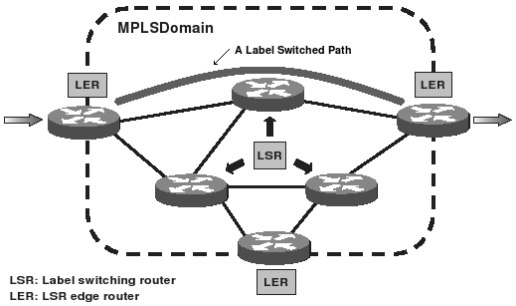


Figure 2.7: *Example of an MPLS domain*

2.3.2 Routing and Forwarding in GMPLS

A fundamental concept in G/MPLS (by G/MPLS we mean both bare MPLS and GMPLS) is that of *label*, an entity essential for forwarding. In MPLS, a label is a fixed-size numeric value, whereas in GMPLS, by necessity, it is a more abstract identifier, including lambdas and time slots. The "routers" in a GMPLS-based network forward data based solely on labels, so much so that any layer-3 identifiers (such as IP addresses) that may exist inside the protocol data units are completely disregarded. Note that MPLS forwarding nodes are usually called "switches" instead of "routers".

A network that consists entirely of MPLS-capable switches and administered by a single entity is called an MPLS domain. Inside such a domain, two types of forwarders are distinguished (see Fig. 2.7): Label Switch Router (LSR) and Label Switch Edge Router (LER). LERs are located at the edges, where traffic enters or exits the domain. The ingress LER has the responsibility of triggering the process to setting up and tearing down LSPs, thus they are responsible for coordinating the path selection process (i.e., routing). As they interface at the IP level with other equipment outside the MPLS domain, they must also be regular IP routers. On the contrary, LSRs need only support MPLS, although they can also include IP routing functionality.

It is the responsibility of the ingress LER to put a header containing the label before the received protocol data unit (PDU). The assembled PDU is then forwarded to the next LSR along the LSP inside the MPLS domain. When the PDU is about to leave the MPLS domain, the egress LER removes the MPLS header and forwards the data to the outside router.

The label together with the port on which the data was received determine the output port and the new outgoing label. Fig. 2.8 shows an example of the table that every LSR has to keep for each LSP that passes through it.

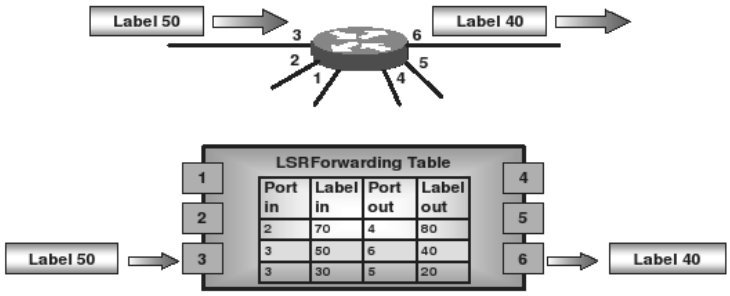


Figure 2.8: *Example of the forwarding table of an MPLS switch[139]*

Labels have only local meaning, that is, a label number 50 could appear many times in the MPLS domain, or even in a single LSR. However, the tuple (*port in, label in*) is unique in an LSR. The switching mechanism is called *label swapping* because the LSR reads the header of the incoming PDU, uses (*port-in, label-in*) as key in its lookup table and substitutes the current values of *port-in* and *label-in* with *port-out* and *label-out* from the table.

LSPs can be nested [18], which gives rise to a hierarchy of LSPs. This feature is useful to overcome the natural limitations of certain transmission mediums, such as the number of wavelengths in a WDM system. LSP hierarchy helps in dealing with the discrete nature of optical bandwidth [8], as available capacity can be locally allocated independently from the capacity handled natively by the transmission technology. Another aggregation technique available in GMPLS is *link bundling*, whereby the attributes and resources of several parallel links of similar characteristics are aggregated, creating a new (virtual) link which can be managed as any other physical link. In doing so, the size of the link state database can be reduced by a large factor [8].

LSPs are set-up through a combination of extensions to existing IP link-state routing protocols (e.g., OSPF and IS-IS) with signaling protocols such as Resource Reservation Protocol (RSVP) and Label Distribution Protocol (LDP). Entities in the network use the information obtained through the routing protocols to compute paths subject to specific resource or policy constraints, a procedure known as *constrained routing*. Examples of constraints are hop count, delay, path diversity and capacity.

2.3.3 GMPLS-specific Features

To meet the layer integration expectations, GMPLS addresses certain limitations found in MPLS. One is the already mentioned generalization of label, because GMPLS must be able to handle transmission technologies that do not use any explicit, value based, identification, as assumed by MPLS.

Thus, GMPLS adapts and extends MPLS in several ways [18]. One feature that has profound consequences on network design and operation is the formal *separation of control and data planes*, a feature particularly important to support technologies where control traffic cannot be sent in-band with the data traffic. This is in contrast to traditional IP and MPLS, where routing (and eventually signaling) messages are transported by IP datagrams together with data information.

Another salient feature is the support for *multiple types of switching*. The addition of new switching types has impact on basic LSP properties, how labels are requested and communicated, the directionality of LSPs (they are now bidirectional), how errors are propagated, and the information used for synchronizing the ingress and egress. Furthermore, links conforming an LSP can use *different label encodings* depending on the underlying switching type, such as a time slot, or a wavelength. They only need to match at both ends of the virtual path.

In GMPLS, LSPs can be *bidirectional*. This helps in reducing the possibility of resource contention when allocating reciprocal LSPs via separate signaling sessions, in simplifying certain failure restoration procedures, and in lowering the LSP set-up latency and the number of messages required.

Other new features are:

- Addition of the so-called *Forward Adjacencies*, a mechanism that may improve bandwidth utilization when bandwidth allocation can be performed only in discrete units, as well as a mechanism to *aggregate forwarding state*, thus allowing the number of required labels to be reduced;

- An upstream node can *suggest a label* when establishing LSPs. This may help in reducing the time to set-up an LSP through certain kinds of optical equipment where there may be a long delay in configuring the switching fabric.

- On path selection, a switch can *restrict the range of usable labels*. This feature is targeted at the optical domain, where in some cases wavelengths used by the path must be restricted either to a small subset of possible wavelengths, or to one specific wavelength.

- Support for requesting the use of a specific label on a specific interface.
- Allows for the inclusion of *technology-specific parameters* in signaling when using RSVP.

2.3.4 The GMPLS Functional Planes

From a functional point of view, GMPLS proposes organizing the network into three planes [75], as follows:

- **Data plane**, responsible for carrying user data between end nodes by employing the resources of the physical infrastructure (nodes, links, lambdas, regenerators, etc.)
- **Control plane**, tasked with facilitating the allocation and deallocation of the data plane resources assigned to connections along their paths.
- **Management plane**, responsible for the supervision of the whole system, including the other planes.

It is interesting to note that the emphasis of GMPLS is on the control plane. Nonetheless, the separation of planes is fundamentally functional and conceptual, not necessarily a physical separation. Its protocol's data units can be sent in-band or out-of-band, and it may be implemented in terms of channels that, albeit belonging to the data plane infrastructure, do not participate in data forwarding. Theoretically, it can also be implemented through a completely separate network. Nevertheless, the control plane continues being IP-based in the sense that the protocols used to fulfill its mission run atop the IP protocol, such as the Open Shortest Path First with Traffic Engineering Extensions (OSPF-TE) [81], or the Intermediate System to Intermediate System with Engineering Extensions (IS-IS-TE) [80].

OSPF-TE and IS-IS-TE are used to advertise the network status and allow the control plane to set up, delete, restore and maintain LSPs. The signaling protocol has also been further extended to support LSPs protection and restoration. For link management, GMPLS introduced the Link Management Protocol (LMP) [57]. LMP runs between adjacent nodes and is used for both link provisioning and fault isolation.

Given that the control plane is critical for the correct operation of the GMPLS network, much effort has been devoted to increasing its resilience. This can be tackled from the perspective of the procedures and protocols required for successful recovery in the control plane [75]. Alternatively, it can be viewed as a design problem, where the control plane is an separate network, as in [122].

2.4 Network Failure and Recovery

This section presents a taxonomy of failures in communications networks and summarizes the approaches to recovery usually applied in GMPLS. With regards to failures, the terminology is that of Section 2.1 Thus, we treat nodes and links as decomposable systems, that is, each instance comprises a set of interacting components. Nodes are composed of several hardware subsystems or modules which vary depending on their purpose, the technology they must support, their role (e.g., core switching versus traffic aggregator at the edge of the network), etc. Typical examples of hardware modules are line-cards, interface ports, and switching fabric. Additionally, they run software used for the operation, monitoring and control of its resources.

As for links, their physical nature and structure are determined by the communication medium employed (i.e., a tangible link does not exist in wireless communication). In optical networks, a link is basically composed of a fiber optic cable and zero or more ancillary devices such as repeaters, amplifiers and regenerators. Due to the multiplexing techniques used in optical networks, a single fiber optic cable is composed of several transmission subunits, each one called *channel*, and the transceivers and multiplexer/demultiplexers needed at both ends are also considered link components [120].

2.4.1 Types of Physical Failures

From the point of view of *what* physical element can fail, failures in an optical network can be classified as follows:

1. **Link failure**
   A link failure occurs when at least one of its constituents fail leading to a complete interruption in the data flow, or to an unacceptable degradation in the communication quality. *Cable cut* is its most common manifestation, caused by digging, construction work, and other human interventions. Additionally, fires, earthquakes, floods and even rodents also cause link failures.

   Cable cuts dominate all other sources of externally-imposed network failures because the deployed mileage is so great that cable-cutting events occur virtually every few days in large networks [97]. To mitigate the failure, traffic must be diverted to alternate links or nodes.

2. **Channel failure**
   A channel failure occurs when one or more wavelengths on a fiber in a

WDM network become unavailable due to the failure of specific laser(s). To resolve such a situation, from a hardware point of view, traffic can be switched to a backup laser emitting in the same wavelength, or employ a tunable laser as replacement [128].

3. **Node (hardware) failure**
   Node failures occur when it can no longer perform as expected due to malfunctions of its components, or it is completely taken down (due to power outage, or accidental human (operator) error for example). They are far less common than link failures [128], in part due to the fact that they are usually deployed in controlled environments (i.e., they are less exposed to external agents) as well as to the built-in redundancy of its components. Nonetheless, human (operator) error can trigger node failure (accidental power-down, for example) When a node fails, all of its links can also be considered to fail simultaneously.

4. **Node (software) failure**
   The software that allocate resources, implements routing algorithms or otherwise controls the hardware can have bugs or be attacked by virus, which can cause the node to stop fulfilling its functions. It can be argued that such failures are no more than node (hardware) failures due to the fact that the software runs on nodes. However, some peculiarities warrant its assignment it to a specific category.

   Firstly, a software failure does not provoke the automatic and simultaneous failure of the links of the affected node; some of them might fail, in the sense that they might be rendered useless somehow, while others do not. Secondly, a software error that manifests itself in one node can lead to faulty behavior in other nodes, thus extending the malfunction through the network. Lastly, a node with failing software can continue providing services which were correctly configured before the failure, leading to partial failure, i.e., a state in which the node becomes uncontrollable but partially functional.

2.4.2 A Failure Taxonomy

Faults, errors and failures can be classified in several ways, taking into account their time of occurrence, their duration, their extent and severity, etc. A broad categorization applicable to several engineering fields can be found in [7], where 16 elementary classes are given for faults alone. On the other hand, [4] focuses on computer networks and offers a framework that

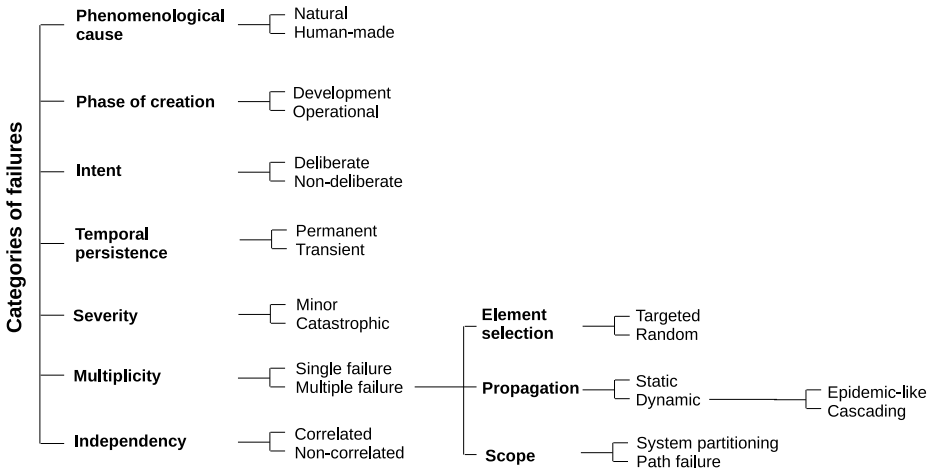


Figure 2.9: *A taxonomy of failures in data networks*

integrates the major reliability-related concepts and disciplines. It proposes taxonomies that complement and specialize the ones given in [7].

As shown in Fig. 2.9, we classify failures according to several criteria, which have been selected based on both their pertinence to networking and relevance to the objectives of this thesis. But for the sake of simplicity and coherence with the existing literature, we call them all *failures* and do not emphasize the difference between faults and failures. Thus, we refer to a "transient link failure" for example, although it is clear that a link failure is an event that can be handled by fault management procedures when the network is viewed as a single system.

The categorization is as follows:

1. **Phenomenological cause**

   (a) *Natural*, in other words, without human intervention, for example a wildfire or earthquake;

   (b) *Human-made*, for example erroneous configuration of routing tables, and unplugging of equipment at the wrong moment.

2. **Phase of creation**
   The stage of the system life at which the failure occurs.

   (a) *Development*: mistakes made during network planning/design, or during upgrades later on;

(b) *Operational*, encountered during network operation.

3. **Intent**

   (a) *Deliberate*: a malicious human act with the aim of harming the system. If this happens during operation, it is usually categorized as an *attack*. An attack can act upon physical elements (e.g., a deliberate fiber cut), or make use of software (virus, worms, trap doors, etc.).

   (b) *Non-deliberate*: introduced accidentally or without awareness. Includes those caused by incompetence as well as omissions (not acting when an action is expected).

4. **Temporal persistence**

   (a) *Permanent*, when no recovery is possible.

   (b) *Transient*, when the item is returned to an operational state after a certain time, following an automated or human intervention.

   Permanent failures can appear, for example, in satellite communication, but that is outside the scope of this thesis. In terrestrial communication, we assume that all failures are *transient* due to the fact that failed equipment can be replaced, broken links can be repaired, etc.

5. **Severity**
   The severity can be evaluated in terms of the relation between the benefit provided by the fully functional service and the consequences of the failure. In this context, benefit can be economic or otherwise.

   (a) *Minor*: few elements are affected or the harmful consequences are of similar cost as the benefit provided by correct service delivery;

   (b) *Catastrophic*: the cost of the failure is orders of magnitude higher than the benefit provided by correct service delivery.

6. **Multiplicity**

   (a) *Single failure*, which happens when exactly one network element fails at a given time or at most one element is failed.

   (b) *Multiple failure*: several concurrent, overlapping or sequential single failures are active (e.g., in non-repaired state) in the system. Although two or three concurrent failures constitute multiple failures, they are usually treated in the literature as special cases with their own name: *dual failures*, and *triple failures*.

7. **Independency**

   (a) *Correlated*, if several failures (i.e., a multiple failure event) can be related to a single cause. For example, an earthquake can take out of service several nodes and links of a large area (that is, a geographically correlated failure). Likewise, several fiber cables that potentially belong to separate networks can be cut by a single digging

   (b) *Non-correlated*, if the failures in a multiple failure event can be attributed to distinct causes.

2.4.3 Large-scale Failures

A large-scale failure is a special case of *multiple failure* in which a significant proportion of network elements are affected by failures which are all related to a single cause, for example an earthquake or an intentional attack [16]. Thus, referring to the taxonomy of Section 2.4.2, it is a multiple, correlated failure.

The impact of large-scale failures can be catastrophic. In 2006 there was a major earthquake in the Taiwan area. Several submarine cables were broken, and the communication infrastructure of countries in the region suffered either complete interruption or serious disruption for several days [77]. Although backup resources (multiple fiber cores installed together) were in place, and automatic restoration procedures were activated, the former proved useless as the earthquake affected them as well, and the latter caused even more trouble due to limitations of the management system to fully handle the multilayer network, which ultimately forced human intervention to complete the repairing. Additionally, Hurricane Katrina in 2005 caused damage to telecommunication networks worth several billion dollars and the ensuing outage lasted for days in some areas [54],[111]. Likewise, the cuts (apparently unintentional) experimented by submarine cables in the Mediterranean See in January 2008 resulted in more than 20 million Internet users of a dozen countries being affected [20].

Another example, this one not related to natural disasters, is the politically motivated attack on web sites of the government and of businesses of Estonia in 2007, which crippled the country's network for days [86]. In another recent incident, a highly sophisticated malicious software known as the *Stuxnet Worm* was reported to interfere with computers used to control industrial processes worldwide and especially in Iran [56]. In this case the target was not a communication infrastructure, but it is nonetheless an

example of attacks that are tuned to specific environments and that are able to effectively cause damage in a preselected region.

2.4.4 Categorizing Multiple Failures

Multiple failures present some unique characteristics with respect to propagation, element targeting and integrity scope, which are discussed below.

1. Element selection

1. *Targeted*: the attacker uses knowledge about the network to select those elements whose failure would cause the greatest damage, that is, the attacker has a strategy. Although natural disasters usually "select" elements of a given geographical area, it is not appropriate to label them as targeted since intention is absent.

2. *Random*: elements are hit with equal probability.

Note that under single failures it is also possible to distinguish between targeted and random attacks, but the severity would be serious only if the network exhibits a structure particularly susceptible to certain failures, for example a star or a bus topology.

2. Propagation

1. *Static*: the set of the elements whose failure are attributable to a specific cause does not vary with time, except that its cardinality may temporarily decrease due to repairs or replacements. In other words, the failure does not propagate.

2. *Dynamic*: the failure progresses over time, so that the set of affected elements is different from one instant to the next, varying in size or membership.

   Two broad subcategories can be defined for propagating failures in communication networks:

   (a) **Epidemic-like**: Caused mainly by computer viruses and worms, although erroneous implementation of protocols and other software bugs could potentially provoke them too. Its unique characteristic is that the agent provoking the failure in a given node replicates itself and spreads autonomously through the network.

Due to the strong analogy between computer viruses and infectious diseases, *epidemiological models* have been widely used in their study, considering a variety of network structures and propagation environments. For example, [152] and [36] study the spreading and prevalence of viruses on the Internet, while [32],[33] and [29] consider the problem of determining thresholds for the occurrence of endemic infections.

(b) **Cascading**: Cascading or induced failures are those in which two or more elements are failed, but all of them were driven to that state by the failure of a single element [147].

The subtle difference between epidemic-like failures and cascading failures is that in the former, the failure-inducing agent intervenes repeatedly on different elements, whereas in the latter the combination of the network's structural properties (e.g., nodal degree distribution) and dynamics (e.g., traffic flow) play a role in inducing the emergence of more failures, even to a complete collapse. This type of failure has happened several times on power grid networks; a summary of blackouts in the North American region and their catastrophic consequences can be found in [69]. It has also been noted on real and model communication networks [102]. Modeling cascading failures has been extensively studied; a comprehensive survey on the subject can be found in [21].

3. Scope

This category focuses on the integrity of two types of systems: *a)* the whole network; *b)* individual connections. It is loosely based on the categorization of dual-link failures given in [91].

1. *System Partitioning*: after the failures, the network is disconnected (or "partitioned"), i.e., there is at least one pair of nodes for which it is not possible to find a path that connects them.

   Disconnection can directly follow from the removal of failed elements, or be the consequence of potential capacity scarcity due to flow reassignments after the failures, or, in more general terms, to the unavailability of some required resource (minimum path length, wavelength continuity, etc.).

2. *Path failure*. When the system under consideration is not the whole network but individual services (i.e., *connections*), it is possible that a

multiple failure event affects simultaneously all the paths assigned to a given connection. Suppose a connection has two node-disjoint paths, where one is the primary path and other the backup. If the multiple failure event *hits* both paths, the service fails as no sane alternate path is left.

Given that this situation is specific for each connection (whether it is affected or not), it is quite possible that the network remains connected and only a subset of connections are affected.

2.4.5 Recovery in GMPLS-based Networks

Recovery is the term used to denote the sequence of actions after the detection of a fault or failure in a network so as to maintain the required performance level for existing services (e.g., according to service level agreements) and to facilitate the normalization of the network [83]. Many recovery techniques have been proposed and several of them have gained wide acceptance in the industry. A summary of those more relevant to next-generation optical networks can be found in [63] and [65].

Recovery typically involves the activation of a procedure to switch traffic off a working path (or working LSP) and to a recovery path (also called backup path) when a failure is detected on the former. Both the working path and the backup path share the same properties: an ingress interface, and egress interface, and a set of intermediate nodes and links. The difference is in how and when one or the other is used: The working path carries traffic under normal operation, while the backup path does so when the working path fails, although it can also carry *extra traffic* under normal operation. In any case, extra traffic can be preempted when failures occur, thus they are not protected.

2.4.6 Recovery Phases

In GMPLS, "recovery" is an umbrella term that includes several techniques and mechanisms which differ in several ways, but broadly speaking, they all follow a series of steps when an LSP, a link or a node fails. Each step is called a *phase* and their purposes are summarized below (more details can be found in [96],[115],[9]):

- **Phase 1. Failure detection**
  Failure detection is the action of detecting a defect or sensing an

unacceptable performance degradation, followed by the activation of signals that the control plane interprets as an intervention request.

This phase is handled at the layer closest to the failure; for optical networks this is the physical (optical) layer. Among the indicators of a fault condition in an optical network are the loss of light, optically measured bit error rate (BER), dispersion, crosstalk, and attenuation. This is the only recovery phase that the control plane cannot achieve by itself.

- **Phase 2. Failure localization**
  Failure localization provides information about the localization and the identity of the failed item, so that the control plane can evaluate the impact in terms of affected LSP(s) and trigger the necessary recovery actions. It can be performed either by the data plane or the control plane (e.g., by using LMP).

- **Phase 3. Failure notification**
  This phase is used for two purposes: *a)* to inform intermediate nodes that a failure has been detected; and *b)* to inform that the service is not available and that recovery procedures should be initiated. Such information is intended for nodes designated to trigger recovery, located at the LSP end-point or at some other intermediate node, depending on the planned recovery scheme.

- **Phase 4. Failure mitigation**
  Sometimes also called *recovery* or *redirection* [37], the purpose of this phase is to provide the alternate service that the network has committed to under the corresponding failure scenario. That could mean, for example, rerouting the the affected traffic flows so that in the future they will pass over parts of the network that remain unaffected.

  The specific steps performed as well as the guarantee of success are highly dependent on the approach taken to reserve the resources for recovery, namely protection or restoration. These are described in more detail in Section 2.4.7 below.

- **Phase 5. Normalization**
  Typically, backup paths are longer than working paths, thus they use more resources (capacity, wavelength, memory for state, etc). Normalization, also called *reversion*, is an optional phase which consists in switching traffic from the recovery path back to the working path once the failed element is repaired.

The first three phases (detection, localization, and notification) are referred to as *fault management*.

2.4.7 Protection and Restoration

There are basically two types of recovery, the difference being mainly the approach towards resource reservation, namely *Protection* and *Restoration*. If resources destined for recovery are allocated before the fault, for example at the time a particular connection is established, then the approach is called "protection". On the contrary, if resources are sought later on when the recovery procedure is in effect, the approach is called "restoration". Within these two groups, further variations exists based on scope, capacity usage, path setup methods, and so on.

Recovery can be applied to LSPs, segments (a subsequence of links on a path), or links. If LSP protection is used, one or more backup paths are fully established to protect one or more working paths, implying that route computation was completed, the paths were fully signaled all the way, and that resources were allocated and cross-connected between the ingress and egress nodes. In essence, protection means that no signaling takes place to establish the backup path(s) when a failure occurs because it had already been done at setup.

Restoration means that some paths may be pre-computed, signaled, and selected a priori as backup, but not cross-connected. The complete establishment of the backup path occurs only after the working path fails, and requires additional signaling.

The advantage of protection over restoration is its fast recovery time and, in some of its variants, the guarantee that backup resources will be available when needed. On the other hand, restoration techniques can be more flexible with respect to which failure scenarios can be handled and at the same time its capacity requirements can be lower compared to protection [142].

Given the diversity of transmission technologies, topologies, and protection strategies that GMPLS is called to support, a large number of recovery mechanisms have been proposed. Taxonomies and further references can be found for example in [24],[65],[37].

Types of Path Protection

Within the scope of protection, there are some types of recovery that are worth mentioning here, namely *dedicated* and *shared* path protection. Their

importance lies in the fact they are present in the all the major transport technologies.

As previously discussed, one drawback of protection compared to restoration is that it uses more capacity. One way to reduce this potentially wasteful consumption of resources is to break the one-to-one association between working path and backup path. This basic idea gives rise to the following usual combinations:

1. *1+1 Type*: *Dedicated Protection*, or DPP: One dedicated protection path protects exactly one working path, and the normal traffic is permanently duplicated at the ingress node on both the working and protection path. Both paths are link/node disjoint. The receiving end chooses which one is better. No extra traffic can be carried over the protection path.

2. *0:1 Type*: *Unprotected*: No specific recovery path protects the working path. However, the working path can potentially be restored through any alternate available route/link, with or without any pre-computed restoration route. No resources are pre-established for this recovery type.

3. *1:1 Type*: *Dedicated Recovery with Extra Traffic*: One specific recovery path protects exactly one specific working path, but the normal traffic is transmitted over only one LSP (working or backup) at a time. Both paths are link/node disjoint. Extra traffic can be transported using the recovery path resources.

4. *1:N Type*: *Shared Recovery with Extra Traffic*: A specific recovery path is dedicated to the protection of up to N working paths (with $N > 1$). The set of working paths is explicitly identified. Extra traffic can be transported over the recovery path. All these paths must start and end at the same nodes.

   Sometimes, the working paths are assumed to be resource disjoint in the network so that they do not share any failure probability, but this is not mandatory. If more than one working path in the set of N is affected by some failure(s) at the same time, the traffic on only one of these failed paths may be recovered over the recovery path. The choice of N is a policy decision. This type is applicable to both protection and restoration. This type of sharing is usually called Shared Path Protection (SPP).

5. *M:N Type*: A set of M specific recovery paths protects a set of up to N specific working paths (with $M, N > 1, N \geq M$). The two sets are explicitly identified. Extra traffic can be transported over the M recovery paths when available. All the paths must start and end at the same nodes.

   Similar to the *1:N Type*, sometimes the working paths are assumed to be resource disjoint in the network so that they do not share any failure probability, but this is not mandatory. If several working paths in the set of N are concurrently affected by some failure(s), the traffic on only M of these failed paths may be recovered. The choice of N and M is a policy decision. This type is applicable to both protection and restoration.

3

The Robustness of Complex Systems: A Review of Measurements

This dissertation is concerned with the robustness of large communications networks, of which the Internet is the paradigmatic example. In fact, the Internet has been identified as an instance of what has been termed a *complex network*, a subject that currently attracts tremendous interest and is being studied in connection with the so-called Network Science [87].

Although there is still no generally accepted definition of the essential characteristics of "complex networks", some authors [13] propose that their main traits are as follows: *First*, they are not globally-engineered systems but rather the spontaneous outcome of the interactions among the self-organized units that constitute them. *Second*, the system behavior is an emergent property that cannot be understood by studying each system subunit in isolation. *Third*, they exhibit self-similarity, so that the fluctuations and heterogeneities observable in their structure span all scales of the system. And *fourth*, they are usually dynamic systems, showing complex, evolving topological features, such as hierarchies and communities. As daunting as complex networks may seem, judging them by their characteristics, the concepts and tools used to study them are shared with other disciplines, such as graph theory, physics, chaos theory and even epidemiology [87].

This chapter aims to summarize the main existing measures of robustness applicable to data networks, from the purely structural (i.e. topological) point of view as well from the perspective of system function. To that end, we start the chapter with a review of graph concepts and properties. Then follows an overview of the main network models that researchers have used to study telecommunication networks. Afterwards, we discuss a number of measures of robustness found in the literature and, finally, give a summary of the tools available for network topology generation.

3.1 Fundamental Graph Concepts

This section introduces the basic concepts and terminology pertaining to graphs employed throughout the thesis. Graphs offer an abstract and convenient representation of networks in general, and of the topology of communications networks in particular. Despite this distinction, we use the terms "graph" and "network" interchangeably when referring to the topology of a given network.

3.1.1 Graphs and Paths

A *graph* $G = (V, E)$ is a tuple where V is a set of *vertices* or *nodes*, and E a set of *edges* or *links*. Each element (u, v) in E is a subset of V; u is called the edge source and v the target (or destination). Two vertices connected through an edge are said to be adjacent, or neighbors.

A customary representation of a graph is an *adjacency matrix* A, a two-dimensional matrix where A_{ij} is non-zero if vertices i and j are connected through an edge. Another popular representation, particularly suitable for sparse graphs in computer programs, is the *adjacency list*, an array of lists that contains, for each source vertex, the list of destination vertices.

When edges have orientation such that $(u, v) \neq (v, u)$, G is called a *directed* graph (or *digraph*), and *undirected* otherwise. A *weighted* graph associates one or more attributes to every edge, so that, for example, $w(e)$ is the value of that attribute w for edge e. Weights are usually numeric and are also referred to as "cost".

In a *simple graph*, at most one edge is allowed per pair of nodes. Furthermore, loops are forbidden, so that for all edges (u, v), $u \neq v$. A graph that is not simple is called a *multigraph*. A *dynamic graph* is one which experiences time-varying changes [87], so that a certain link or node may exist at a time t but not at some other.

A *path* is a sequence P of vertices $v_1, v_2, ..., v_n$, where every pair of contiguous vertices is an edge, i.e., $(v_i, v_{i+1}) \in E,\ i = 1, 2, \ldots, n-1$. In communications networks, paths are usually *simple*, that is, nodes may appear in the sequence only once. The number of edges in P is called *path length* or *hop count*.

A graph is *(strongly) connected* if there exists at least one path for every and all its node pairs. If this condition is not met, the graph has isolated subgraphs, each one called a *component*. G has components G_1 and G_2 if no path exists from any node belonging to G_1 to any node in G_2 [87]. A connected graph has exactly one connected component. As already noted,

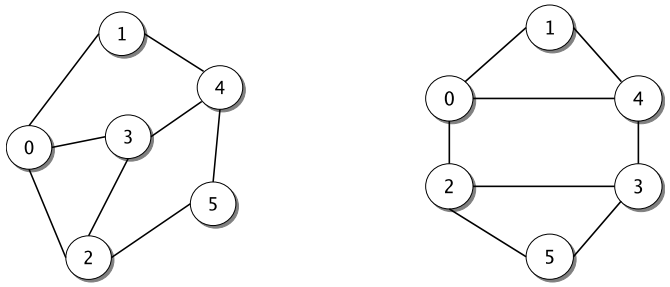


Figure 3.1: *Two isomorphic graphs.* $|V| = 6$ *and* $|E| = 8$

a disconnected graph is also said to be *partitioned*. Data communication normally deals with connected networks as disconnection is a very undesirable situation.

The *shortest path* ℓ_{uv} between a pair of nodes u and v, also known as the "geodesic path", is the path through the graph that connect u and v with minimum total "cost", where "cost" can be distance, delay, hop count, etc. Without further qualification, it usually refers to hop count. There may be several shortest paths for a given node pair, all having the same minimal cost. In an undirected graphs, this quantity is symmetric, that is, $\ell_{uv} = \ell_{vu}$.

Two graphs $G = (V, E)$ and $G' = (V', E')$ are *isomorphic* if they contain the same number of vertices connected in the same way. More formally, G and G' are said to be isomorphic, written $G \simeq G'$, if there exists a one-to-one function p, called an *isomorphism*, from V onto V' such that for all $u, v \in V$, $(u, v) \in E \Leftrightarrow (p(u), p(v)) \in E'$ [49]. Graph isomorphism defines equivalence classes of graphs, which allows us to distinguish properties that are inherent to structure rather than to graph representation. For example, if the diameter of graph G is, say, four, all the graphs isomorphic to it have the same diameter. Fig 3.1 illustrates two undirected isomorphic graphs.

Graphs can be used to represent both the physical arrangements between nodes in a network (the *physical topology* of a communications network, for example), as well as the logical relations, where the physical adjacency is not a requirement. The collection of Autonomous Systems (AS) and their peering agreements can be represented as a high-level (interdomain) *logical topology* of the Internet. The same applies to a group of people and their acquaintance in a social network.

In research, communications networks are usually modeled as simple, connected, weighted graphs, with link capacity and length (e.g., distance

in km) as the most common weights. Although the directionality of edges depend on the application, it is more common to treat links as undirected (implicitly bidirectional). When studying performance in failure-related scenarios, graphs are dynamic to account for the temporary remotion of nodes and links.

3.1.2 Basic Graph Features

The total number of vertices in a graph is denoted by N, or equivalently by $|V|$, the cardinality of the vertex set. Likewise, $|E|$ is the number edges. In Graph Theory, N is the graph *order*, and $|E|$ is its *size*. However, in many biological and physical contexts, N is usually referred to as the graph size as it identifies the number of distinct elements in the system [13]. This nomenclature is customary in the networking literature, and we adhere to it in this thesis as well.

Node degree, also called *nodal degree*, is the number of of edges attached to a node. In a digraph, *in-degree* and *out-degree* indicate the number of incoming and outgoing edges of a given node, respectively.

The *average node degree*, denoted by $\langle k \rangle$ or $\bar{k}$, is the average number of edges connecting the vertices in the graph. For an undirected graph, it is

$$\langle k \rangle = \frac{2|E|}{|V|}. \tag{3.1}$$

The *average shortest-path length* $\langle \ell \rangle$ is the arithmetic mean of the lengths of the shortest paths of all the node pairs in the topology,

$$\langle \ell \rangle = \frac{1}{N(N-1)} \sum_{ij} \ell_{ij}. \tag{3.2}$$

This property is also called the *characteristic path length* [144] or simply the *average distance*. In a disconnected network, $\langle \ell \rangle = \infty$.

The *diameter* of a topology is the greatest minimum distance between any pair of nodes, usually measured in hop count. Thus, the diameter d_G can be defined as

$$d_G = \max_{i,j} \ell_{ij}. \tag{3.3}$$

The *edge connectivity* of a graph is the smaller number of edges whose removal disconnects the graph. So, a *k-edge-connected* graph suffers disconnection only after k or more edges are removed. Similar definition exists for

vertex connectivity. A 2-connected graph is called *biconnected*. Note that, transport networks are usually designed at least as biconnected, so that any single node or link failure does not provoke disconnection

3.2 Metrics and Non-trivial Graph Features

Graphs can be compared and classified according to a series of distinguishing properties called metrics, which are in general function of the graph structure [99]. The number of metrics is in fact very large and a comprehensive list can be found in [43],[5]. We review here only the ones that are used in the characterization of network models and in the assessment of robustness.

Throughout this section, the graph $G = (V, E)$ is a simple connected undirected graph of N nodes.

3.2.1 Degree sequence and Degree distribution

The sequence $g = [d_1, d_2, \ldots, d_N]$, where each element is the degree of the nodes in G, is called the *degree sequence* of G (some authors require that g be non-decreasing [49] while others do not [87]). The degree sequence is a representation of the structure of the graph, although it is an incomplete one, for the information in g is not sufficient to uniquely reproduce G. In fact, all the graphs isomorphic to G will have the same degree sequence.

Due to its lossy and isomorphic nature, it can be useful when a family of graphs is needed. By randomly generating several graph instances from one degree sequence, all such graphs will share the same structure. Note, however, that certain sequences are not realizable. For example $g = [2, 2, 2, 1]$ cannot correspond to any graph.

An even more compressed representation of the structure of a graph is the *degree distribution*, which is the probability distribution of the degree of its nodes. Let $n(k)$ be the number of nodes of degree k in G. The probability that a randomly selected node will have exactly k edges is

$$P(k) = \frac{n(k)}{N} \quad k = 0, 1, \ldots, k_{max} \tag{3.4}$$

where k_{max} is the highest nodal degree in the graph. An equivalent definition states that P_k is the fraction of nodes in the graph that have degree k.

Nodal degree is one of the main properties that has been used to categorize network models, based on the degree distribution $P(k)$ or its decay. Network models are discussed in Section 3.3.

The *joint degree distribution* $P(k, k')$ is the probability than an arbitrary link connects nodes u and v whose degrees are k and k' respectively [43]. In terms of conditional probability, this is usually expressed as

$$P(k'|k) = \frac{\langle k \rangle P(k, k')}{kP(k)}, \tag{3.5}$$

that is, the probability that an arbitrary neighbor of a node of degree k has degree k'. The joint degree distribution gives information about the nodal degree correlation in the topology.

3.2.2 Path length distribution

As stated in 3.1.1, the hop count H_N of a path is the number of edges contained in it, alternatively called path length. The path length distribution $\Pr[H_N = k]$ is the histogram of the length of the shortest path between all possible node pairs in the graph. The average path length is a measure of central tendency that is used to characterize this distribution. It is sometimes referred to as the "degree of separation".

The study of the path length distribution has helped define the small-world network model. In large systems such as the Internet and social networks, the exact path length distribution can be difficult or impossible to calculate [2]. Thus, estimating methods are used instead.

3.2.3 Clustering coefficient

A graph is said to show clustering if the probability of two vertices being connected by an edge is higher when there is another vertex to which both are attached [144]. Thus, a *clustering coefficient* measures the degree to which vertices in a graph tend to cluster together.

The clustering coefficient of the whole graph is called the *global* clustering coefficient, and is defined in terms of the clustering coefficient of each node, the *local* clustering coefficient. The formulation given below accounts for the number of "triangles" present in the graph [144], but there are several alternative coefficients based on more complex approaches (k-neighbors, cycles of a certain order, etc.). See [21] and the references therein.

The *local* clustering coefficient characterizes the density of connections in the environment or neighborhood of a node [68]. Let $N_i = \{v_k : (i, k) \in E\}$ be the neighborhood of vertex i, that is, the set of nodes to which i is directly connected, and K_i the cardinality $|N_i|$ of this set. The local clustering coefficient C_i in the undirected graph G is

$$C_i = \frac{2|\{(u,v)\}|}{K_i(K_i-1)} \quad \forall u,v \in N_i, (u,v) \in E. \tag{3.6}$$

The denominator in (3.6) is the total number of edges that exists in a fully connected neighborhood, while the numerator is actual number of edges.

The global clustering coefficient $\bar{C}$ is the average of the local clustering coefficient over all the vertices:

$$\bar{C} = \frac{1}{N}\sum_{i \in V} C_i. \tag{3.7}$$

3.2.4 Measures of Centrality

Centrality measures the importance or prominence of a node (or link) in a given network. Common measures of centrality are degree centrality (which ranks nodes according to their nodal degrees), eigenvector centrality, closeness centrality and betweenness centrality [108]. The second one is defined in Section 3.2.6, while the last two are described next.

Closeness centrality

Closeness is the average shortest path between a vertex and all other vertices reachable from it [106]. Thus, for a vertex v of a connected graph G, the closeness $C_C(v)$ is

$$C_C(v) = \frac{1}{N-1}\sum_{t \in V,\ t \neq v} dist(v,t) \tag{3.8}$$

where $dist(v,t)$ is the length of the shortest path between vertices v and t. Some authors define closeness as the reciprocal of the summation in (3.8), so that a value closer to 1.0 indicated greater centrality.

Betweenness centrality

Betweenness centrality determines how often a vertex on a given graph lies along the shortest path between all possible pair of nodes [60]. More formally, the betweenness centrality $C_B(v)$ of vertex $v \in V$ is defined as follows:

$$C_B(v) = \sum_{s,t \in V, s \neq t} \frac{\sigma(s,t|v)}{\sigma(s,t)} \tag{3.9}$$

where $\sigma(s,t)$ is the number of shortest paths that exist between vertices s and t, and $\sigma(s,t|v)$ is the number of shortest paths between vertices s and t that pass through vertex v. A similar definition can be given for edges, in which case the measure is called *edge betweenness centrality*, $C_B(e)$ for a given edge e. If there is more than one shortest path between a given pair of vertices, each one is given equal weight so that the weights sum to unity [106],[22].

$C_B(v)$ can be normalized so that a value close to 1.0 would mean that v is present in almost all the shortest paths in the network. Likewise, a value close to zero would mean that the role of vertex v as intermediary in the communication is marginal.

Vertex betweenness centrality has been found to be strongly correlated with node degree in most types of networks (see for example [106],[15] and the references therein). Possible correlations between edge weight and betweenness has also been explored but in this case the relationship depends on more than one factor; see [143].

Fig. 3.2 is a visual representation of the edge betweenness centrality of the well-known Cost266 reference topology, which has 37 nodes and 57 undirected links. More details on this and other European reference transport networks can be found in [92],[112]. In this example, the graph is unweighted. The higher the value of $C_B(e)$, the thicker the link line is. One can see that thick lines appear scattered on the topology, so that neither node position (periphery versus "core"; see the *closeness* and *eccentricity* measures) nor node high degree warrant a high $C_B(e)$.

3.2.5 Assortativity coefficient

In assortative networks, nodes tend to connect to other nodes of similar degree, while in disassortative networks nodes with low degree are more likely connected with highly connected ones [21]. One way to measure the degree of assortativity is through the *average degree of the nearest neighbors*, which is defined as follows for a given degree k,

$$k_{nn}(k) = \sum_{k'} k' P(k'|k), \tag{3.10}$$

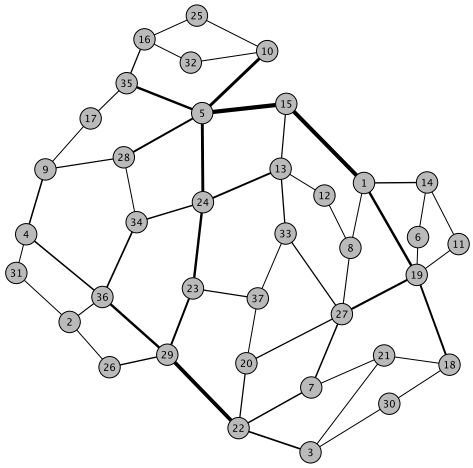


Figure 3.2: *The Cost266 topology. Link thickness indicates link importance in terms of edge betweenness centrality.*

where $P(k'|k)$ is the joint degree distribution. If high degree nodes tend to connect with equally high degree nodes, $k_{nn}(k)$ is an increasing function of k. Conversely, a decreasing function of k indicates that high degree nodes tend to connect with low degree nodes. If there are no correlations, $k_{nn}(k) = \langle k^2 \rangle / \langle k \rangle$ [21], i.e., it is a constant and thus independent of k.

The global assortativity coefficient r of an undirected graph G can be determined by taking into account the Pearson correlation coefficient of the degrees at both ends of the edges [105]:

$$r = \frac{\left[\frac{1}{M}\sum_e (j_e k_e)\right] - y^2}{\left[\frac{1}{2M}\sum_e (j_e^2 + k_e^2)\right] - y^2} \tag{3.11}$$

where j_e and k_e are the degrees of the vertices at the end of edge $e \in E$, $M = |E|$ and

$$y = \frac{1}{2M}\sum_e (j_e + k_e).$$

Many social networks have significant assortative structure, while technological and biological networks seem to be disassortative [105]. The Internet in particular seems to be disassortative at the AS-level [94].

3.2.6 Algebraic Connectivity and other spectral measurements

The properties of a graph in relation to the characteristics of its associated matrices, such as its adjacency matrix or Laplacian matrix, are the subject of study in the spectral graph theory.

Spectral analysis has applications in determining community structures, in characterizing models of real networks, in extracting information on connectivity, in epidemic spreading, among many others (further applications and references are given in [21]). In the following paragraphs we briefly review the spectral measurements that appear in relation to the assessment of network robustness.

The (normal or ordinary) *spectrum* of a graph G corresponds to the set of eigenvalues λ_i $(i = 1, 2, \ldots, N)$ of its adjacency matrix A [43], which is sometimes written as λ_A. When G is simple and undirected, A is real and symmetric, and thus the graph has real eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$.

The *spectral radius* of G is the largest eigenvalue $\lambda_{1,A}$ of A, often referred to simply as λ_1, which plays an important role in determining epidemic thresholds, but has also proven useful in traffic engineering, see [94] and the references therein.

Spectral measurements has also been used to assess centrality. Whereas degree centrality ranks nodes based on their nodal degrees, *eigenvector centrality* acknowledges that the weight of connections among nodes are not equal, since connections to nodes that are themselves influential bring more influence than connections to less influential nodes. In such a setting, the centrality of a node can be defined as a value proportional to the average of the centralities of its neighbors. Let N_i be set of nodes adjacent to i. The eigenvector centrality x_i of node i is

$$x_i = \frac{1}{\lambda} \sum_{j \in N_i} A_{ij} x_j, \tag{3.12}$$

where x is an eigenvector of the adjacency matrix with eigenvalue λ. One property of the eigenvector centrality is that it takes into account both the number and the quality (or degree of influence) of connections, so that a high nodal degree does count towards a higher centrality, but the contribution of a smaller number of influential contacts may be more "valuable" overall [108].

Another useful mathematical tool in this category is the *Laplacian spectrum*, defined as $Q = \Delta - A$, where $\Delta = diag(d_1, d_2, \ldots, d_N)$, with $d_i =$ the degree of node i. Thus, the entires of Q are as follows:

$$Q_{i,j} = \begin{cases} d_i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise.} \end{cases} \tag{3.13}$$

Alternatively, Q can be normalized [39], so that the entries are defined instead as

$$Q_{i,j} = \begin{cases} 1 & \text{if } i = j \text{ and } d_i \neq 0 \\ -\frac{1}{\sqrt{d_i d_j}} & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

One peculiarity of the normalized Laplacian is that its eigenvalues are between 0 and 2 (inclusive), whereas in the unnormalized one there is no such constant limit.

The Laplacian spectrum offers information about network connectivity. The number of times zero appears as an eigenvalue in Q equals the number of (connected) components in G. Thus, this count is exactly one in a connected graph.

The first non-zero eigenvalue of Q (i.e., the second smallest eigenvalue $\lambda_{2,Q}$ in a connected graph) is called the *algebraic connectivity* [58]. The magnitude of the algebraic connectivity reflects how well connected the overall graph is. Therefore, the farther λ_2 is from zero, the more difficult it is to separate a graph into independent components. However, the algebraic connectivity is equal to zero for all disconnected networks. Therefore, as soon as the connectedness is lost, due to failures for example, this measure becomes less useful by being too coarse.

3.3 Network Models

The study of topological properties of a variety of large, complex systems found in the real world, such as biological systems and social and communications networks, has made it possible to observe that, despite their inherent differences, they can be grouped and characterized in terms of certain key properties, such as clustering coefficient, average path length and degree distribution [21]. Based on this observation, several network models have been proposed that reproduce, at varying degrees of fidelity, the topological properties observed empirically in real world systems. The models make

it possible to randomly generate network instances with which to study, through simulation or analysis, the behavior or performance of complex systems.

In the following subsections, an overview of the models more relevant to communications systems is presented. Comprehensive review of other models as well as variations of the ones given here can be found in [21],[43],[62], [87],[13].

3.3.1 Erdős-Rényi Networks

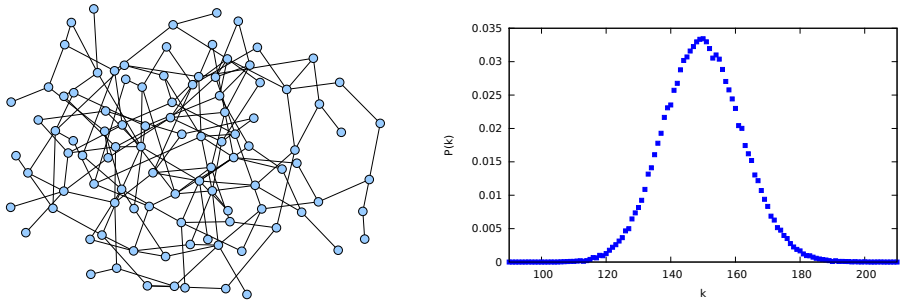
One of the first type of networks for which a model was formulated is called the *random network of Erdős-Rényi* (ER), *ER network*, or simply *random network*. The model was proposed by Erdős and Rényi [53] and, independently, by Gilbert [61], both at the end of the 1950s. The term "random" refers here to the fact that, during the generation of a graph instance, the arrangement of links is disordered.

In the Erdős-Rényi model, the graph generation process starts with N disconnected nodes, and links are added between two randomly selected nodes (avoiding self-loops and duplicate edges) until the number of links equals some desired value K. On the other hand, in Gilbert's model, the generation procedure consists of selecting links with probability $0 < p < 1$ from a complete graph of N nodes, such that the resulting graph has, on average, $m = p\frac{N(N-1)}{2}$ undirected links, with average nodal degree $\langle k \rangle = p(N-1)$ [43],[87]. This variant is usually referred to as the $G_{N,p}$ model.

The procedures outlined above do not guarantee that the outcome will be a connected graph. However, it has been proved that the structure of ER networks vary as a function of p. In fact, there exists a critical probability $p_c = \frac{1}{N}$ which substantially alters the resulting structure, that is, the model exhibits a phase transition. Assuming a graph in $G_{N,p}$, with $N \to \infty$,

- if $p < p_c$, almost surely, the graph has no component of size greater than $O(\ln N)$.
- if $p = p_c$, the largest component of the graph has, almost surely, size $O(N^{2/3})$.
- if $p > p_c$, the graph has, almost surely, a unique giant component, and no other component contains more than $O(\ln N)$ nodes.

The distance between nodes in ER networks is small. The average path length is $\langle \ell \rangle \sim \frac{\ln N}{\ln \langle k \rangle}$, while the clustering coefficient is $\bar{C} = p = \frac{\langle k \rangle}{N}$ [21],[87], which, for large N, is also a small quantity. Note that in ER networks,



(a) **(b)**

Figure 3.3: *Random networks: (a) An example with* $N = 100$ *and* $p = 0.03$ *(b) Average degree distribution over 30 ER instances of* $N = 5000$ *and* $p = 0.03$

the probabilities of node pairs being connected by links are, by definition, independent. So, the probability that two nodes are connected is not higher when they share a mutual neighbor than when they do not [107].

ER networks are sometimes called Poisson random graphs due to the fact that, for large N and fixed $\langle k \rangle$, the degree distribution is approximated by a Poisson distribution. Fig. 3.3a shows an example ER network of 100 nodes, while 3.3b illustrates the degree distribution for this type of network.

It is generally accepted that ER networks play a role more as a benchmarking tool than as a model of real-world networks [87], which have been found to exhibit structures different to the ones obtainable through the ER model. Nevertheless, in the case of the global Internet, some authors [141] suggest that the subgraph formed by the union of all shortest paths in the network (at the AS level), which they call the "observable part of the network" and that concentrates 80% of the total traffic, is (ER) random, despite the fact that the underlying network in full is not.

3.3.2 Generalized Random Networks

Generalized random networks are extended ER networks that aim at obtaining a better fit between model and reality, for example with respect to clustering or node degree distribution. *Configuration model*, introduced initially in [17], allows the generation of networks with arbitrary degree distributions. To that end, the generation process requires a degree sequence $g = [d_1, d_2, \ldots, d_N]$ from which $K = \frac{1}{2}\sum_i d_i$ links are created. The values in g must be chosen in such a way that the fraction of nodes with degree k will tend to the desired degree distribution $P(k)$ as N becomes large.

Several graph generating procedures have been proposed; the most important ones are outlined in [21],[43]. The basic procedure assigns a number of "half-edges" to each node i equal to its expected degree d_i, and pairs of nodes with available "half-edges" are selected randomly, creating a new edge that links them, thus converting two "half-edges" in one complete edge. New links are repeatedly added in this way until all "half-edges" are used up, yielding a random graph that is a member of the ensemble of graphs with the given degree sequence.

A different generation procedure is given in [107] which, contrary to the one just outlined, takes into account not only a degree sequence but also a target clustering coefficient, thus generalizing even more the graph generation capability.

It has been established [101] that a graph generated through the configuration model for an arbitrary degree distribution $P(k)$ almost surely has a giant component when $Q > 0$,

$$Q = \sum_{k \geq 1} k(k-2)P(k) \tag{3.15}$$

provided that the maximum degree k_{max} is small. Approximating expressions for other properties (diameter, average path length, etc.) can be found summarized in [21].

3.3.3 The Watts-Strogatz Small-World Networks

In a variety of real networks, the path between any two nodes is relatively short, despite the large size of such networks. The existence of links that function as bridges or shortcuts between different areas of the network creates this so-called *small-world property*, which mathematically is characterized by an average path length $\langle \ell \rangle$ that grows proportionally to the logarithm of the number of nodes N in the network,

$$\langle \ell \rangle \propto \ln N. \tag{3.16}$$

This property was first investigated in the context of the social sciences in the 1960s but later on its presence was confirmed in diverse networks, including biological and technological ones. In fact, it also appears in ER random networks, as was already mentioned briefly in Section 3.3.1. However, and in contrast to ER networks, real world networks have not only small average path lengths but also high clustering coefficients.

The most popular model of random graphs having both small $\langle \ell \rangle$ and large $\bar{C}$ was proposed in [144]. This model, known as the *Watts-Strogatz* (WS) small-world model, offers a procedure to construct such networks as follows: It starts with a pre-existing regular ring network of N nodes, in which each node is connected to its m nearest neighbors in each ring direction, for a total of $2m$ neighbors per node. Then, all the $K = mN$ links are subjected to rewiring by randomly reassigning its target node with a probability p. This results in the introduction of the "bridges" that connect initially distant nodes and leads to the shortening of paths in the network.

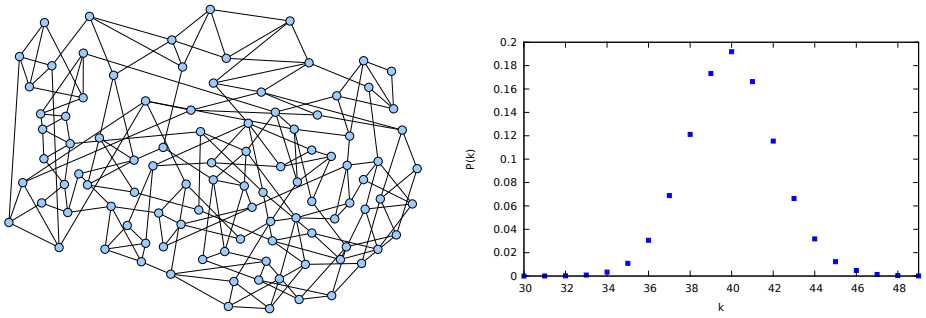
Interestingly, the rewiring procedure has a different effect on average path length than on clustering coefficient: small changes in p can result in drastic non-linear reductions of $\langle \ell \rangle$, whereas the effect on $\bar{C}$ is linear, so that clustering decreases slowly compared to $\langle \ell \rangle$ [62]. Thus, by assigning a non-zero value to p, it is possible to obtain networks that simultaneously have short path lengths and high clustering, as shown by Watts and Strogatz [144]. Note than when $p = 0$, the outcome is the same regular network, whereas when $p \to 1$, it is a connected random graph with short average distances but almost no clustering. It has been observed, however, that the abrupt decrease in $\langle \ell \rangle$ depends not on p alone but also on network size (number of nodes) and other properties of the original regular network. Further details on the subject can be found in [14],[5],[21] and the references therein.

An estimation of the clustering coefficient of a WS network as a function of p is provided in [12] as follows:

$$C_{WS}(p) = \frac{3(m-1)}{2(2m-1)}(1-p)^3 \tag{3.17}$$

where the first factor (the fraction involving m) is basically $C_{WS}(0)$, that is, the clustering of the network before the rewiring. As for the degree distribution, its shape resembles that of a random graph with a peak at $\langle k \rangle = 2m$ although with a smaller variance, which increases progressively as $p \to 1$. With regards to degree-degree correlations, several real networks whose structure are thought to be small-world like (mainly social networks) exhibit assortative character [105].

Fig. 3.4a shows an instance of a small-world network generated with the WS procedure just outlined, with $p = 0.12$ and $m = 2$. The degree distribution shown in Fig. 3.4b is the average of 30 realizations.



(a) **(b)**

Figure 3.4: *Small-world networks: (a) An example (WS) with* $N = 100$*,* $m = 2$*,* $p = 0.12$ *(b) Average degree distribution over 30 WS instances of* $N = 5000$*,* $m = 20$ *and* $p = 0.12$*.*

3.3.4 Scale-free Networks

One feature of the ER and WS networks is that their degree distributions are essentially Poisson, as discussed in the preceding subsections. However, large networks, both naturally occurring and man-made ones, seem to deviate significantly from that. In fact, they tend to have a few highly-connected nodes to which a large number of low degree nodes are attached. Furthermore, they lack a characteristic degree, thus their designation as *scale-free networks*. More specifically, their degree distribution has been found to follow a *power law*,

$$P(k) \sim k^{-\gamma} \tag{3.18}$$

for large k, with $2 < \gamma < 3$ [21].

Barabási and Albert introduced in 1999 a model of scale-free networks, known as the *BA model* [10]. It relies on a graph-augmentative process to generate network instances featuring the essential properties that the authors attributed to the networks they analyzed in the study. Whereas with the ER and WS models the emphasis was on correctly reproducing a certain graph structure, with the BA model the main interest shifted towards capturing the underlying dynamics of complex networks, on the premise that those dynamics were ultimately responsible for the emergence of the scale-free property [5].

Two fundamental ideas in the BA model are *incremental growth* and *preferential attachment*. The first addresses the need to model evolving

systems, as already mentioned, while the latter recognizes the fact that the likelihood that any two nodes are connected is not uniform across a real network, where new links tend to connect to nodes which already have a large number of connections instead. One example is the "citation network" of scientific publications. Well-known, highly cited papers are more likely to be cited by new papers than lesser known ones. Thus, preferential attachment can lead to the emergence of so-called *hubs*, a phenomenon also known as the rich-gets-richer.

The network generation procedure of the BA model starts with a small set of m_0 connected nodes, to which others are added in successive steps to mimic network growth. For each new node u, m new links ($m \leq m_0$) are created so that u can be connected to old nodes by following the (linear) preferential attachment rule. This rule states that the probability of adding a link $u \to v$ depends on the degree k_v of node v as follows,

$$P(u \to v) = \frac{k_v}{\sum_{j \in V} k_j} \tag{3.19}$$

where the denominator is the sum of the degrees of all nodes in the network, which serves as a normalization factor. The higher the degree of a node, the greater the likelihood of it being selected as the target of a new link. After t timesteps, the resulting graph will have $N = t + m_0$ nodes and mt links, with $\langle k \rangle = 2m$. Furthermore, for large t, this procedure generates graph instances whose exponent in the power-law degree distribution is $\gamma = 3$, independent of m [5],[21].

Interestingly, concepts similar to preferential attachment had already appeared in the literature several times and well before the introduction of the BA model, the first one [149] dating back to 1925 (a review with historical perspectives can be found in [89]). But it was after Barabási and Albert's work that the interest on the subject grew tremendously, leading to the development of several other models of scale-free networks, either as alternatives to the BA model or as enhancements (see surveys in [62],[21]).

A large number of systems have been studied and labeled as scale-free, pertaining to as diverse fields as ecology, biology, social sciences, information systems and telecommunication. The Internet in particular has received much attention from researchers, who have studied its structure and postulated that scale-free features can be observed at different layers, leading to initial speculations about its alleged lack of tolerance to attacks [11]. However, the Internet's true structure is a topic on which consensus has not yet been reached.

The large size of the Internet and its peculiar organization (or lack thereof) have motivated the use of sampling techniques to study it, with the drawback that bias may be introduced by the discovery methods employed or the data sets used, so that measurements might find a power-law degree distribution where the true underlying topology is, for example, random or even regular [35],[84],[1]. New, more accurate mapping techniques are under development, see for example the recent work [44] on neighborhood discovery, however, another problem remains, namely, agreeing on which are the defining features of scale-free networks. Currently, several overlapping definitions are in use, so that systems identified as scale-free are attributed differing properties and behaviors. An in-depth discussion of the problems caused by this lack of agreement is presented in [89], which also offers formalisms that seek to clarify the meaning of "scale-free" in the context of complex systems and help settle the issue.

Regarding the properties of BA networks, the average path length is shorter than in ER and WS networks of the same number of nodes and edges, scaling as [34]

$$\langle \ell \rangle \sim \frac{\log N}{\log(\log N)} \tag{3.20}$$

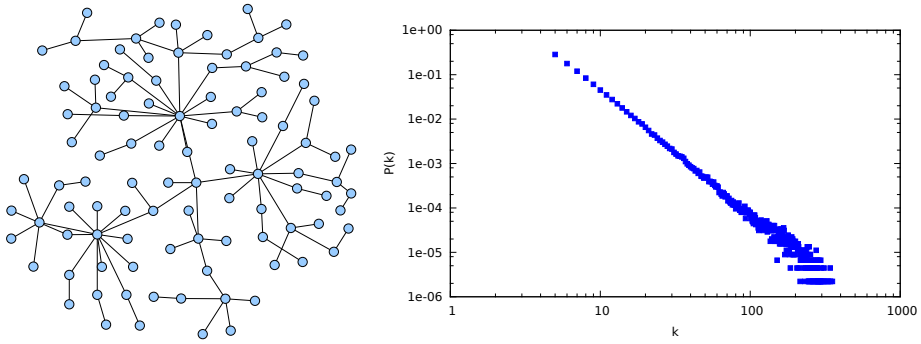
while the clustering coefficient vanishes as N grows large [13],

$$C_{BA} = \frac{m}{8N}(\ln N)^2, \tag{3.21}$$

therefore it is longer than in ER networks but still shorter than in WS networks. Another peculiarity of the BA networks is the absence of degree-degree correlations, that is, they are disassortative [21]. Fig. 3.4a is an example of a small scale-free network, while Fig. 3.4b is the degree distribution of BA networks of size $N = 5000$, with $m = 5$ (degrees averaged over 90 realizations). It shows the typical quasi-straight line in log-log scale.

3.3.5 Tools and Models for Internet-like Topologies

One aspect often overlooked when studying network models and their properties is that real-world systems constitute physical spatial entities, that is, their nodes and links occupy precise positions in the Euclidean space. The location of each element is usually the result of a conscious decision process in which several factors are pondered, including economic and technical aspects. Commonly, this planning is carried out with high autonomy at a local scale, for example within the boundaries of the administrative domain of a service provider.



(a) (b)

Figure 3.5: *Scale-free networks: (a) An example (BA) with* $N = 100$ *and* $m = 1$ *(b) Degree distribution averaged over 90 BA instances with* $N = 5000$ *and* $m = 5$

A necessary goal of this planning stage is to produce an economically feasible design, i.e., that the required capital expenditure and estimated operating costs remain within budget. Along with economic considerations, the designers must take into account constraints stemming from the use of specific technologies, protocols and products, find ways to overcome barriers created by geographical features (mountains, rivers, etc.), make room for foreseeable expansions, ensure compatibility with any existing installation, and comply with government-mandated regulations. One has to wonder then what effect do these restrictions and practical considerations have on the structure of global networks emerging from the interconnection of smaller entities, of which the Internet is the prime example. But even more important is the question of how to incorporate into the network models those planning issues that influence localization, so that synthetic topologies can more faithfully represent real world networks.

Historically, one of the first topology generators to be used for network simulation was based on the Waxman model [145], which is itself an extension of the ER model. Under the assumption that long-range links are expensive, this model places routers at random in a two-dimensional space and adds links between them so that the probability of connecting two routers decays exponentially with their distance. The "flat" Waxman model and its variations were soon deemed inappropriate, however, as further research on real network topologies suggested that hierarchy played an important role in determining the relation between nodes as well as between groups of nodes (i.e., domains). The models developed afterwards to address the issues of hierarchy and locality are known as *structural*, of which the Transit-stub

model [150] and the Tier model [27] are salient examples. The transit-stub model and other simpler ones are implemented in the software package called GT-ITM (Georgia Tech Internetwork Topology Models). A step-by-step description of algorithm to implement Tiers is given in [50].

A generator based on a structural model is more complicated than one that implements the simple Waxman model because reproducing a complete realistic hierarchy implies dealing with the different structural and functional roles in the network (that is, distinguishing the parts that correspond to access, core, LAN, MAN, etc.), as well as adhering to the requirements of the technologies typically used to implement them (SDH rings, WDM mesh networks, etc.). Moreover, the proponents of structural models argued that generators should take into account the design principles applied in real environments, for example with regard to connectivity and redundancy [6]. All these require the user to provide several parameters to control the generation process.

Structural topology generators were in turn superseded when complex network research provided further insight into the large-scale statistical properties of the Internet topologies, in particular concerning the power-law nature of the degree distribution [6]. This motivated the introduction of *degree-based* generators, with the consequence that focus shifted towards large-scale topology features and away from the local properties that the structural approach tried to mimic [134]. Two well-known degree-based generators are PLRG (Power-Law Random Graph) [3] and Inet (Internet Topology Generator) [76]. For a given target number of nodes N and an exponent β, PLRG assigns a degree to each node, drawn randomly from a power-law distribution, and then proceeds to also randomly create links to complete the required node degrees. The main purpose of Inet is to generate AS-level topologies. It also preassigns degrees and then use a preferential attachment rule to create the links. Another generator model in this category is GLP (Generalized Linear Preference), introduced in [23]. GLP creates the topology through incremental growth and preferential attachment. At each step, one of two operations is probabilistically chosen: adding a new node and m links, or adding just m new links.

The degree-based generators are certainly popular but have also received criticism, see for example the conclusions of the qualitative comparison of power law topology generators performed in [68]. One of the criticisms refers to the fact that two topologies may have the same nodal degree distribution and still be almost opposites from the network engineering point of view [6], which suggests that there can be several explanations for the

emergence of a particular structure. Along this line of thought, Fabrikant et al. [55] speculated that the power laws observed in the Internet are perhaps the manifestation of trade-offs, that is, complicated optimization problems with multiple and conflicting objectives. The proposed model is called the *heuristically optimized trade-off* (HOT) model, which is itself a generalization of a previous class of models called *highly optimized tolerance*. HOT's guiding design principle is that the core and the edge of a good ISP network have different characteristics: the former is constructed as a sparsely connected mesh of high-speed, low-connectivity routers capable of carrying aggregated traffic over high-capacity links. At the edges, on the other hand, there are tree-like structures of highly connected nodes that act as traffic concentrators [6].

According to the HOT model, the network grows over time, gaining a new node at each time step. This new node is added at a random position in an Euclidean space, and a link is created to some existing node so that the distance of this new link is as short as possible and, at the same time, the location of the new node is as central as possible in the sense of closeness centrality. The fact that the application of the HOT model also results in power-law distributed degrees is not surprising, as its rule can be viewed as a form of preferential attachment [13]. The appeal of the HOT model is that it is closer to reality, since on one hand it takes into account, at least partially, the economic and technological aspects of network engineering whereby topologies are not purely random, and on the other hand, it produces the expected power-law structure.

Several publicly available software tools implement the network generation models just described as well as others. Table 3.1 is a partial list of the most popular ones. As our list favors those that offer a battery of generators combined in a single package or library, we omit further reference to generators such as Inet (whose last version appeared in 2002) and KU-LocGen [130]. The software can be obtained from the URL given in the table. Note that BRITE (Boston University Representative Internet Topology Generator) [98] has been in unsupported mode for several years now (i.e., it is neither being further developed nor maintained) but its comprehensive set of features has made it a classic generator that continues to be relevant. Note that BRITE, aSHIIP and GT-ITM are specifically tailored to communication networks, while the rest are general, that is, the generated graphs contain basically just nodes and links and are not concerned with link capacity, delay, geographical location, etc.

Table 3.1: *Topology generation software*

	Name	Description
1	BRITE [98]	Supports AS level or router level topologies. Models: Waxman, BA (two variants), GLP. http://www.cs.bu.edu/brite/
2	aSHIIP [138]	Generates hierarchical, multidomain (AS) topologies. Models: ER, BA (two variants), Waxman, GLP. http://wwwdi.supelec.fr/software/ashiip/
3	GT-ITM [150]	Generates flat and multilevel topologies. Models: Waxman, ER, hierarchical (n-level), transit-stub. http://www.cc.gatech.edu/projects/gtitm/
4	IGen	Generates topologies based on network design heuristics. Belongs to the family of structural topology generators; does not rely on probabilistic methods. http://informatique.umons.ac.be/networks/igen/
5	iGraph	Library for complex network research. It is a C language library, with bindings for Python and R. Models: ER, BA (several variants), WS (several variants), based on given degree sequence, among others. http://igraph.sourceforge.net/
6	NetworkX	Python package for the study of complex networks. Models: ER (several variants), WS (several variants), based on degree sequence, among many others. http://networkx.lanl.gov/

3.4 Measures of Network Robustness

This section presents a number of measures which have been explicitly proposed for the assessment of network robustness. It is of interest to note that few of them (in fact, the most recently proposed ones) offer an assessment of system function (such as throughput or number of accepted requests), the majority focusing on other purely graph theoretical evaluations.

3.4.1 Network criticality

Network criticality aims to compare different networks based on their robustness to changes in traffic demand and topology [136]. The authors exploit the fact that on an undirected graph the ratio of the random-walk betweenness [106] of a node to the node weight (e.g., the sum of the capacity of a node's incident links) is the same for all nodes in the graph, and so can be used as a single, characterizing measure. This ratio is named the *network criticality*.

Network criticality can also be viewed from the perspective of resistance distance in electrical circuits. If τ_{sd} denotes the resistance distance between nodes s and d, then the network criticality is $\sum_{s,d} \tau_{sd}$, i.e., the total resistance distance. As the objective is to compare different topologies, the normalized network criticality $\hat{\tau}$ is defined finally as

$$\hat{\tau} = \frac{1}{N(N-1)} \sum_{s,d} \tau_{sd}. \tag{3.22}$$

The lower the value of $\hat{\tau}$ the better. As paths (e.g., LSPs in a MPLS network) underlie the concept, $\hat{\tau}$ is meant to measure the robustness of core/transport networks only.

A simulation based comparison of several topologies (regular and small-world) from the perspective of $\hat{\tau}$ and algebraic connectivity can be found in [19],[137]. The conclusion is that the former better reflects the changes in performance that results from the addition/removal of nodes or links, as it carries more information about the structure of the network.

3.4.2 Symmetry ratio

Symmetry ratio is a measure introduced in [47] and is based on previous studies by the same authors on the robustness of networks against targeted attack on nodes The term symmetry refers to the invariance observed in vertex adjacency (i.e., the edge set) when the vertex set is permuted. The

studies suggested that, in general, "regular" networks fared much better than disordered networks when subject to targeted attacks. The term "regular" in this context means that all nodes (and links) offer more or less the same value to an attacker.

Incidentally, a similar observation is made in [100], which studied the effects of attacks on path-oriented networks, and concluded that in order to reduce the extent of the losses, one has to either design the topology to be more regular or deliberately discourage the use of "central" nodes during routing, if the former is not possible.

A specific formulation to compute the degree of symmetry of a network is given in [47], which relates the number of distinct eigenvalues ϵ of the adjacency matrix of the network to its diameter $diam(G)$, as follows:

$$r = \frac{\epsilon}{diam(G) + 1}, \tag{3.23}$$

where the closer to 1.0 the value of r, the more robust the network.

For random (ER) networks, they conjecture that $\epsilon = N$ as $N \to \infty$. Given that ER networks have relatively small diameter, they should have high values of r.

3.4.3 Connectivity and Average two-terminal reliability

If the criterion to declare a system functional is that the underlying network remains connected as node or links are removed, then connectivity (see Section 3.1.2) can be an appropriate measure of robustness. However, if the transient disconnection of small sections of the network is acceptable, as happens in large data networks such as the Internet, it does not offer a meaningful measure.

One option is to watch the size of the giant component, but a more flexible alternative is the so called *k-terminal reliability*. This metric is the probability that a randomly chosen subset of k nodes are connected, which requires computationally expensive calculations in the general case. However, if a global measure of the pair-wise connectivity is sufficient, the much simpler *average two-terminal reliability* A_{2TR} can be used instead, which is defined as follows [104]:

$$A_{2TR} = \frac{\sum_{i=1}^{c} K_i(K_i - 1)}{N(N-1)} \tag{3.24}$$

where c is the number of components and K_i is the number of nodes in component i. This ratio gives the fraction of node pairs which are accessible through a path at a given instant.

When the network is fully connected, exactly one component exists and A_{2TR} is 1. Successive removal of nodes or links will bring it closer to zero. If failures affect two topologies in the same proportion (same percentage of nodes, for example), the one that takes longer to reach a given critical A_{2TR} can be considered the more robust.

Note that A_{2TR} can be reformulated so that node pairs have a weight, for example to take into account that certain pairs are more valuable because their share in the total traffic is comparatively larger, either as end-nodes or as intermediaries.

3.4.4 Elasticity

Elasticity relates total throughput to node removal to evaluate the robustness of complex networks (see [131],[132]) The fundamental idea is to successively remove a certain fixed number of nodes r (in the original definition, $r = 1\%$) and measure the ensuing throughput degradation. The more pronounced and abrupt is the throughput drop experienced by a given topology, the lower its robustness.

More formally, the elasticity E at the ζ-th iteration, when $r\zeta$ nodes have been removed from the topology, is the area under the curve of throughput T_G versus the percentage of nodes removed,

$$E(\zeta) = \frac{1}{2N} \sum_{k=0}^{\zeta} \left(T_G(rk) + T_G(r(k+1)) \right) \tag{3.25}$$

$T_G(x)$ is the maximum throughput in the "residual" graph consisting of $(N - x)$ nodes. By definition, $T_G(0) = 1$. To evaluate T_G, [132] proposes a linear programming optimization formulation as part of the definition of E itself. Note that in this case, routing is optimal from the point of view of total throughput, but not necessarily so with respect to path length. Additionally, the authors present numerical examples of the evaluation of robustness against several forms of attacks, where the routing involved in the computation of T_G is heuristic-based (Dijkstra).

3.4.5 Viral conductance

The aim of *viral conductance* is to measure the ability of a topology to contain the spread of epidemics [148]. The use of the spectral radius as an epidemic threshold is well established in the epidemics literature. If the effective spreading rate is below the threshold, the virus contamination dies out, but becomes prevalent above it. Thus, among two topologies, the one with the higher threshold is considered the more robust.

However, it is easy to find instances in which two topologies (a ring and another less regular one, for example) have the same spectral radius yet they perform differently on a case of epidemic outbreak (see [148] for illustrative examples). Moreover, it may happen that in a low intensity epidemic, one topology shows higher resistance than the other, but as the intensity varies, their comparative performance is reversed. To more faithfully account for such behaviors, viral conductance proposes using the average fraction of infected nodes for all effective epidemic intensities. More formally, viral conductance VC is defined as the area under the curve of the fraction of infected nodes in stead-state $y_\infty(s)$,

$$VC(G) = \int_0^\lambda y_\infty(s)ds = \lambda\overline{y_\infty} \tag{3.26}$$

where s is the effective curing rate (i.e., the ratio of the virus' cure rate δ to the birth rate β), λ is the largest eigenvalue of the adjacency matrix of the graph G, and $\overline{y_\infty}$ is the average value of the fraction of infected nodes at stead-state for all $0 \leq s \leq \lambda$. Robustness is inverse to VC, i.e., the lower the value of VC the better.

Closed-form heuristic-based approximating expressions for VC are given in [148] as well as numerical examples that compare the robustness of different types of network structures (small-world, power-law and others). The epidemic is assumed to conform to the SIS model (see Chapter 5 for further details on epidemic models).

It must be noted that the application domain of VC is epidemics, therefore it can disregard aspects of practical importance for communications networks, such as traffic, demand and routing. Furthermore, improving robustness may lead to contradictory requirements between both domains: from the point of view of epidemic spreading, limited connectivity (low nodal degrees) may be desirable, whereas a poorly connected topology can lead to unacceptable service in communications networks when failures occur.

4

Multiple Uncorrelated Link Failures

Communications networks, especially at their core, are designed to withstand several types of failures, such as accidental or intentional fiber cuts, loss of switching capabilities caused by power outages, malfunctioning due to equipment aging, and even operator mistakes. This ability to maintain service continuity in the presence of failures is due to efficient recovery techniques incorporated in their design, as well as to diverse technologies that have been developed to that end. Several methods and techniques are reported in the literature for dealing with failures, see for example surveys [65] and [38].

One feature that most of the existing approaches to network recovery have in common is the assumption that at any given time, only one failure is outstanding, which is known as the *single-failure assumption* [142]. However, networks are equally prone to multiple-failure events, that is, the concurrent failure of several communication elements. For instance, earthquakes, flooding and natural disasters have the potential to disrupt a large number of network elements simultaneously.

Although large-scale failure events may be relatively rare [151], that fact does not lessen the economic loss they cause, or the disruption they can bring upon thousands or even millions of users. Unfortunately, in such failure scenarios the redundancy-based recovery techniques that are effective under the single-failure assumption are not suitable anymore, simply because the cost of implementing massive redundancy for rarely occurring events is prohibitive [70].

In light of the fact that redundancy becomes impractical with multiple failures, understanding the robustness of a given topology to such failures becomes crucial. In fact, this topic has long been studied from the topological perspective, for example by measuring connectivity and fragmentation when a certain proportion of nodes are removed. But as we shall see in this chapter, such measures give limited information, often even misleading, about the network's functional robustness.

In this chapter we consider a failure scenario in which a large number of links in a GMPLS-based network fail simultaneously and thus disrupt the LSPs which constitute the network service. We study the extent to which the network function is affected by such failures and explore and compare two simple protective actions that network operators may consider when deciding how best to prepare their network for large-scale failures. Using our taxonomy of Chapter 2, the failures considered here can be classified as non-deliberate, transient, random, static, multiple link failures.

4.1 Resilience through Redundancy: Benefits and Limitations

As we have seen, the central idea underlying the recovery mechanisms is that of redundancy. The backup path is an abstract entity comprising of all the additional components that are put in place in case the working path fails. This is particularly so in protection-based recovery, but it is less clear with restoration-based recovery. However, the use of restoration can be less attractive in core networks due to the stringent requirements on recovery time and recovery guarantees faced by network operators. In that sense, protection is a "safer bet" and thus our remaining discussion on failures assume that type of recovery.

Let us examine now the general operation of two of the most basic path protection schemes, namely Dedicated Path Protection (DPP) and Shared Path Protection (SPP), which will serve us to illustrate the effectiveness of these protection techniques in single failure scenarios and their inapplicability once multiple failures are taken into account. Both DPP and SPP can be implemented through a variety of transmission technologies. DPP has long been supported by SONET/SDH's 1+1 (or 1:1) Automatic Protection Switching and is offered as an option in several WDM products. Additionally, MPLS Fast Reroute [114] provides support for both.

As discussed in Chapter 2, in DPP, two (link- or node-) disjoint paths are exclusively assigned to an arriving connection, one acting as the working path and the other as the backup path. Thus, it can survive any single (link or intermediate node) failure. Fig. 4.1 illustrates the basic operation of DPP. In this example, one (bidirectional) connection exists between nodes 1 and 9. Now let's assume that links 2–3, 3–5 and 5–7 are geographically close to each other and that an event such as an earthquake affects them all so that their failure is concurrent in time. Although the network would continue being fully connected, DPP loses its efficacy because both paths have failed and,

by design, dynamic recovery is not attempted. Note that the failed elements need not be geographically related; any failure that touches both paths will equally suffice. A possible solution is to assign more than two paths, say k disjoint paths. However, finding k such paths for arbitrary source-destination pairs is not always possible, as it depends on the topology and even on the routing strategy employed. Furthermore, even if there were k paths, they might not comply with QoS constraints, for example on maximum hop count [140]. Note that in our example, we could not have three node-disjoint paths between nodes 1 and 9.

Nevertheless, DPP offers the fastest recovery and the best protection in single-failure scenarios. In fact, it works as expected even when more than one concurrent failure network wide exists, for it operates at the connection level. The drawback is the total capacity required to support it, which is at least more than twice that of what is necessary for unprotected connections. This disadvantage is alleviated by using SPP instead, which, in contrast to DPP, shares a single backup path among $n > 1$ separate connections, thus leading to savings on capacity at the cost of lowering the expectation of a successful recovery. This trade-off has been studied in [124] for a number of reference transport topologies, some of whose results we discuss briefly here.

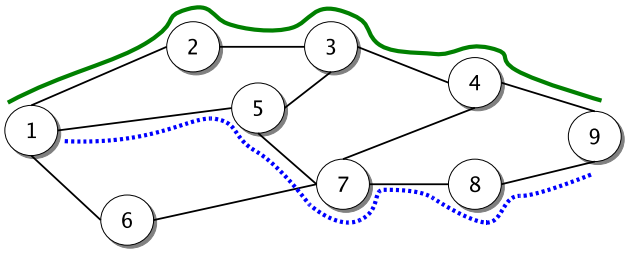
The aforementioned study focused on connection availability, that is, the probability that individual services would be in the operating state at any given instant, under the assumption that the network could experience failures originated in hardware malfunctioning or fiber cable cuts. The computation of connection availability is based on the intrinsic reliability data (i.e., standardized or vendor-provided MTTRs and MTBFs) of the components sustaining the LSPs, such as optical cross-connects, transponders, amplifiers, and optical cables. Essentially, an LSP can be viewed as a series system so that it is available if all its components are available, is

$$A_p = \prod_{1}^{k} A_{c_i} \tag{4.1}$$

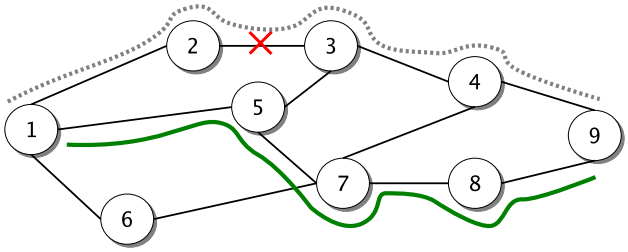
where A_{c_k} is the i-th component of the LSP p. Therefore, the greater the number of components involved in an LSP, the higher the risk of failure. Likewise, the longer the physical distance of links (in kilometers), the greater the probability of fiber cuts.

Under DPP, a connection is available if at least one of its constituent paths (working (w) and backup (b) paths) is available,

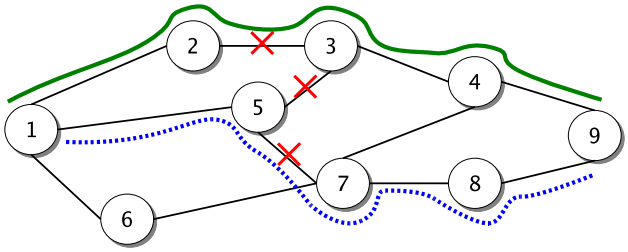
$$A_{DPP} = A_w + (1 - A_w)A_b. \tag{4.2}$$



(a) *Failure-free state.*



(b) *Link 2–3 fails. Traffic is diverted to the backup path.*



(c) *Links 2–3, 3–5 and 5–7 fail. Both paths are affected and the service fails for the affected connection.*

Figure 4.1: *Basic operation of DPP. A working path (solid line) and a backup path (dotted line) are provisioned for a connection between nodes 1 and 9.*

Note that the equivalent formulation for SPP is different because the risk induced by sharing must additionally be taken into account. These two protection schemes were compared in the study from several points of view, of which we mention two, namely *restoration overbuild* and expected *downtime*. The comparison was performed through simulation, assuming a dynamic traffic scenario on four topologies of varying size and geographical coverage, with shortest-path routing. The values in Fig. 4.2 correspond to the average of 80000 connection requests.

Restoration overbuild measures how much extra capacity is devoted to protection, compared to the capacity used by working paths alone, while *downtime* refers to the total time per year in which the service (e.g., connection) is not operative. As can be seen, the extra capacity required with SPP is consistently lower than with DPP (between approximately 18 and 30% lower). In contrast, the expected downtime is much higher with SPP, though the difference depends on the topology (for instance, the relatively poor performance of the Janos-US-CA topology is related to it having both a larger diameter and longer physical links than the others).

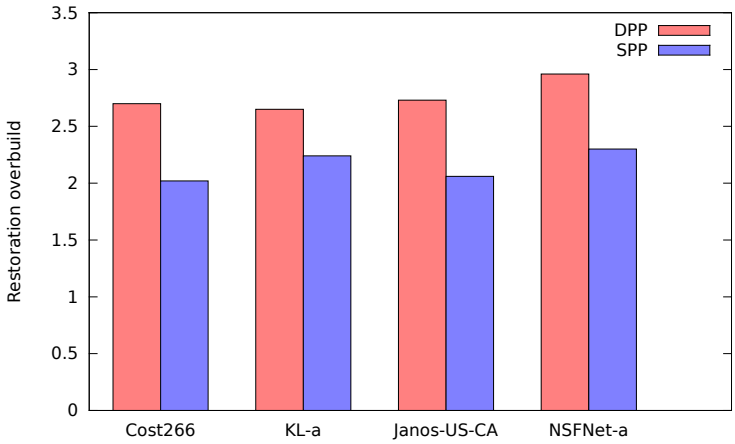
4.2 Evaluation of topological damage

In Section 3.4 we have summarized a number of measures found in the literature for the assessment of network robustness. In this section, we evaluate numerically the robustness of five networks using three topological metrics, as follows:

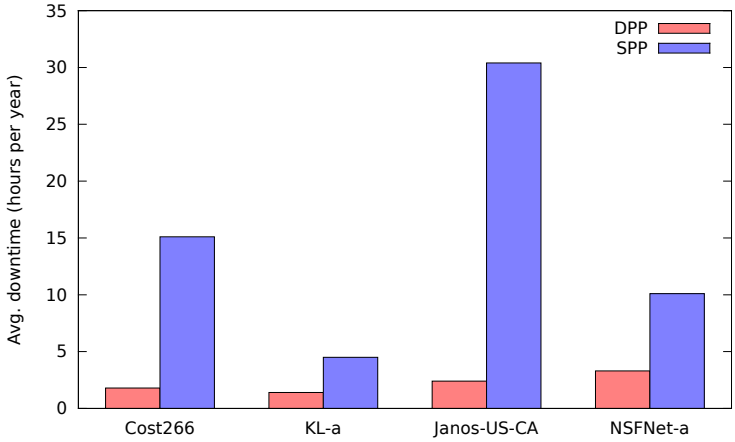
1. S_{LC}, the size of the largest component,
2. A_{2TR}, the average two-terminal reliability, and
3. A_Q, the algebraic connectivity.

All the test topologies are synthetic: two random (ER) (er400d3 and er400d6), one power-law (eba400n), one exhibiting community structure (bt400) and one semi-regular (cost266x6). The first four were randomly generated and the latter was derived from the reference topology Cost266. Their main properties are summarized in Table 4.1; see Appendix B for further details and visual representations.

For each topology, we evaluate the respective metric as function of the fraction of failed elements r. A failure is materialized as the removal of either links or nodes (with all their incident links), selected randomly with uniform



(a) *Restoration overbuild in DPP versus SPP*



(b) *Expected yearly downtime*

Figure 4.2: *Performance comparison of DPP and SPP on four reference transport topologies with respect to restoration overbuild and downtime*

Table 4.1: *Main properties of the topologies used in this section. $|E|$ denotes the number of undirected edges*

Topology	N	$\|E\|$	$\langle k \rangle$	$diam$	$\langle \ell \rangle$	$\bar{C}$
cost266x6	222	371	3.34	20	8.42	0.00
bt400	400	749	3.75	19	9.06	0.17
er400d3	400	618	3.09	12	5.48	0.20
er400d6	400	1205	6.03	7	3.56	0.05
eba400h	400	609	3.05	11	4.61	0.44

probability. These removals represent arbitrary and uncorrelated multiple failures.

Unless stated otherwise, each point in the figures that follow is the average of 1000 repetitions at selected values of $0 \leq r \leq 1$. Thus, the lines between consecutive points are there only to facilitate visualizing the trends. The evaluation is started anew at each r, that is, the elements removed at r_2 are not in addition to the ones removed at r_1, but are selected again from the original topology. We use the subscripts v and e to distinguish between node and link removal. Thus, $r_v = 0.1$ indicates the removal of 10% of nodes, whereas r_e would be used when dealing with link removal.

4.2.1 Size of the largest component

The size of the largest component (alternatively, the size of the giant component) is one of the most frequently used metrics, especially in complex networks, see for example [42],[51],[66],[64] and the recent survey [93].

Fig. 4.3 shows the variation observed in the relative size of the largest connected component as more and more links are removed. These results correspond to $r = 1, 2, \ldots, 9, 10, 15, 20, 30, \ldots, 100$ (in percentage). Initially, when the network is still intact, the size of the largest component equals the network size, i.e., $S_{LC}(0) = N$. Overall, the topology least affected by link removal is erd400d6. It performs well even when r is high, which is seemingly a consequence of its good connectivity. The cost266x6 topology also behaves remarkably well up to $r = 0.2$, but from there on it quickly goes down. Interestingly, for $r < 0.3$, eba400h is the worst performing in this group. As a power-law topology, eba400h has many poorly connected nodes, of which more than 60% have at most degree two (see Appendix B). Therefore, when a number of links are removed, very often many nodes are left isolated in small components (usually of size one or two), thus decreasing the size of the giant component. In any case, it is remarkable that S_{LC} decreases smoothly

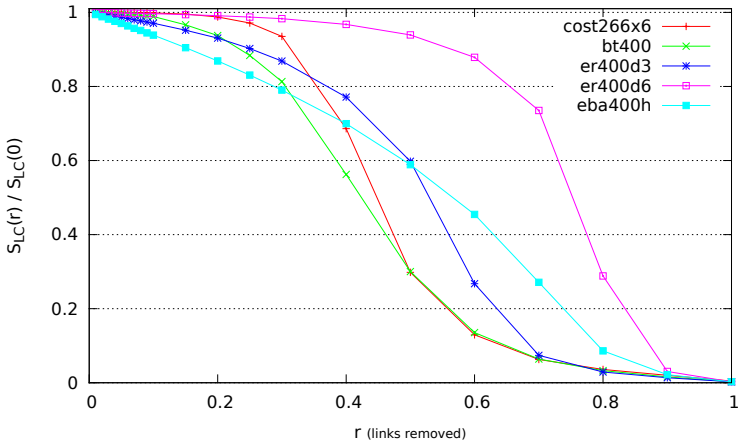


Figure 4.3: *Effect of link failure on the size of the largest component*

for eba400h, and to a certain degree for er400d3, whereas it drops quite sharply for the other topologies.

The effect of node removal on S_{LC} is shown in Fig. 4.4. As expected, the impact is stronger because node removal implies link removal. Thus, neighboring nodes that are not selected for failure are left isolated, leading to further topology fragmentation. Overall, the difference between the topologies for $r < 0.15$ is small: $S_{LC}(r_v)$ is almost linear in all the cases, although with differing slopes.

4.2.2 Average two-terminal reliability

The relative size of the largest component gives a rough idea of the integrity of the network, but it does not really measure as to what extent the network is still able to sustain communication between node pairs. A better measure is A_{2TR}, the average two-terminal reliability which has been defined in Section 3.4.3 as the fraction of node pairs which are accessible through at least one path.

Fig. 4.5 shows the evolution of A_{2TR} when links are removed. The shapes and relative positions of the curves are similar to what is shown in Fig. 4.3 regarding the S_{LC}. There is an important difference, however: $S_{LC}(r_e) > A_{2TR}(r_e)$ for all $r \leq 0.8$ and, except for cost266x6 and er400d6, the curves differ significantly even for small r. This means that S_{LC} is too optimistic; the impairment to the overall ability to communicate is much worse.

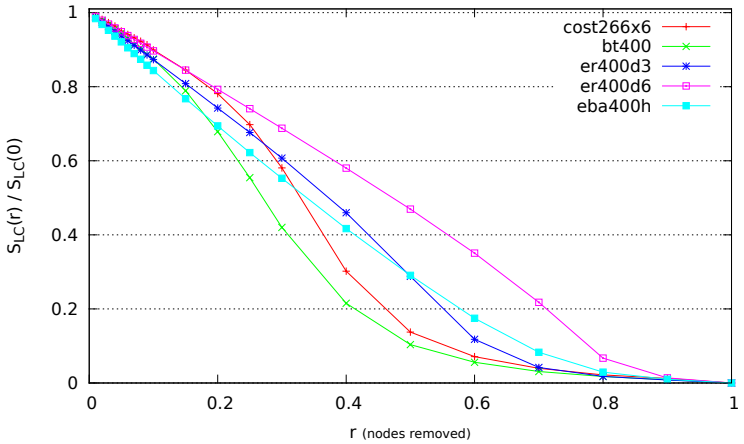


Figure 4.4: *Effect of node failure on the size of the largest component*

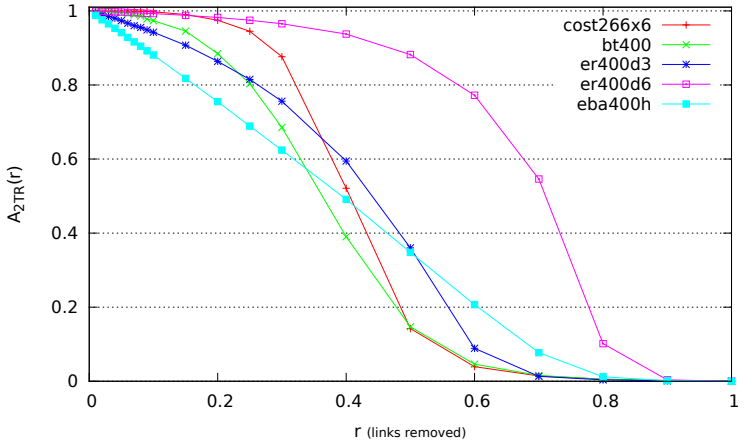


Figure 4.5: *The effect of link failure on* A_{2TR}

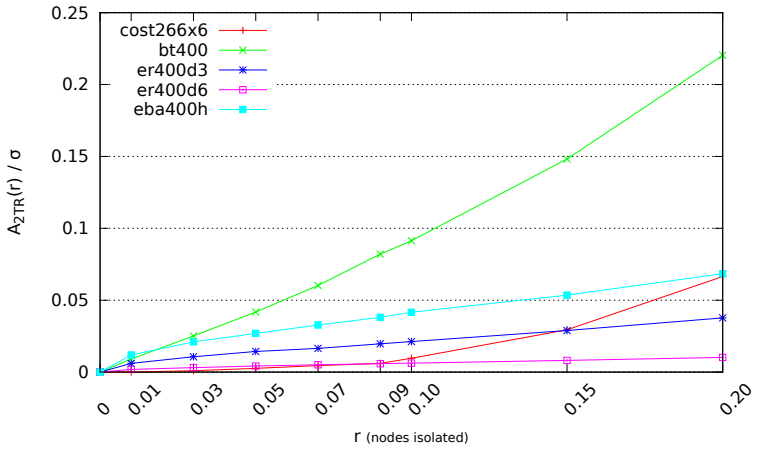


Figure 4.6: *Coefficient of variation of A_{2TR} as nodes are isolated*

It is also possible to measure A_{2TR} under node removal, but that alters N and thus the results cannot be related to the original full topology. An alternative is to substitute "remove node x" with "isolate node x" (by removing all its links). We use that approach to explore another aspect not mentioned up to now: the variability of the measurements. The less variation the better because that means that no matter which specific elements fail, the reduction in performance is expected to be approximately the same. Fig. 4.6 shows the coefficient of variation $A_{2TR}(r_v)/\sigma$ observed throughout the repetitions, at some selected $r \leq 0.20$. Three topologies show good stability: the two random ER (er400d3, er400d6) and the power-law eba400h. The bt400 topology is exceptionally bad, which in this context means that there are certain crucial elements whose failure substantially alter the A_{2TR}. On the other hand, cost266x6 is again peculiar in that it is quite stable for $r < 0.1$, but after that point the variation increases rapidly.

4.2.3 Algebraic Connectivity

The second smallest eigenvalue of the Laplacian is one of the most widely used measures of topological strength. We use here the normalized algebraic connectivity A_Q, as defined in Section 3.2.6, to assess the state of the topology as nodes are removed. We focus only on node removal because A_Q is designed to be insensitive to variations in the number of links. Furthermore, as the second smallest eigenvalue of Q is also zero when the number of components is more than one, we compute A_Q on the largest component only, which is

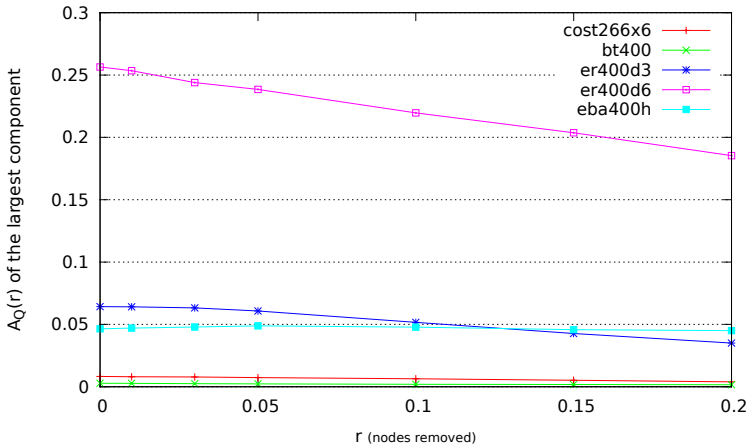


Figure 4.7: *Average algebraic connectivity of the largest component*

quite reasonable for small r, considering the results of Section 4.2.1 above. Due to the high cost of computing the eigenvalues, each point is the average over 200 repetitions instead of 1000.

As we can see in Fig. 4.7, the decline of the average A_Q is slow for all the topologies, but er400d6 clearly stands for its high algebraic connectivity. In the other extreme there are cost266x6 and bt400 which are, according to A_Q, extremely fragile. Note also that eba400h is fairly fragile but stable. However, these differences are not reflected in A_{2TR}. Compare for example figures 4.5 and 4.7 at $r = 0.1$ to see that either the differences are not that big or that the relative performance among the topologies is not coincident.

4.3 Evaluation of functional damage

The evaluations of damage performed in the previous section, while clearly useful in the general case of topology analysis, show some limitations when applied to data transport networks. This stems from the fact that they largely depend on topological features, thus disregarding traffic dynamics and operational constrains such as network load, heterogeneity in link capacity, and routing strategy. Moreover, as we have already pointed out, it is quite possible for the topology to remain connected after a multiple failure event, but even so the number of lost connections can reach unacceptable levels. Thus, instead of observing the state of the abstract topology, it might be better to wonder about the fate of the units of service of the transport

network, i.e., connections. The appeal of this approach for path-oriented networks is that each connection embodies in its path the influence of the structural properties of the topology on the traffic flow, as well as the network operator's policy on resource allocation, as implemented through routing.

We carry out this evaluation on the cost266x6 topology, which performed well in the previous section for small values of r and is close in structure to reference transport networks. We focus on link failures, which are assumed to be independent and equiprobable random events.

For simplicity, we assume that links are capable of carrying an arbitrary number of LSPs as long as free capacity is available, and that all nodes support full wavelength conversion. To introduce a minimum degree of heterogeneity in capacity, links have either C or $2C$ units of total capacity. Half of them belong to the first group and the other half to the second, where the membership to either set was decided on a random basis.

In order to simulate the provision of service in the network and the occurrence of large-scale failures, an event-driven simulator that reproduces the process of route selection in a path-oriented transport network was developed. The simulator handles the reception of connection requests between node pairs, triggers and coordinates the proper routing and capacity allocation based on the demand, keeps track of the usage of resources (the residual capacity on each link), releases connections when their holding time has expired, and collects the required statistical data.

With respect to the traffic demand, the simulator accepts a series of connection requests between randomly-selected source and destination nodes, which arrive according to a Poisson process. As we are interested in a dynamic traffic scenario, the capacity allocated to an accepted connection, whose holding time is an exponentially distributed random variable, is released as soon as the connection terminates. The aggregated traffic between any pair of nodes is either zero or a fixed value greater than zero, meaning that either they do not communicate, or contribute the same amount of traffic as the other pairs. This traffic matrix is randomly generated. The capacity requested by connections is a uniformly distributed random variable in the range 1–10 units. Given the large diameter of the topology, in this experiment we chose to discard traffic that can be considered "local" in the sense of node neighborhood. Thus, connection requests between nodes that are four hops or less apart are completely ignored.

For routing, a capacity-constrained minimum-hop shortest path is used. Thus, links that do not have enough residual capacity to satisfy the arriving demand are filtered out before the exploration begins. To avoid creating

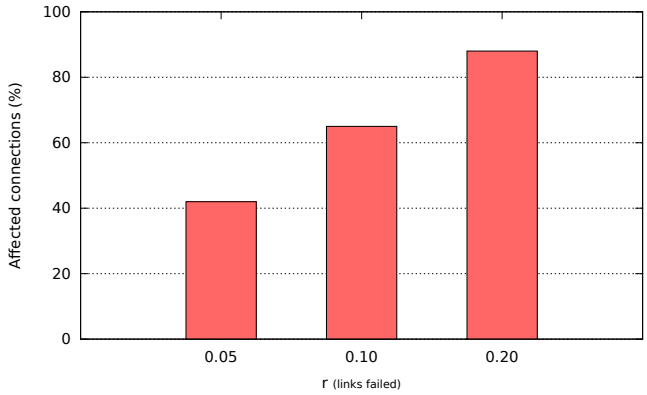


Figure 4.8: *Percentage of connections affected at given fraction of failed links*

unrealistically long paths, any request whose feasible path exceeds 24 hops is also rejected. Note that the average minimum path length in this topology is about 9 hops, while the diameter is 20. The blocking ratio is approximately 0.01 in all the experiments. For each accepted connection, one path from source to destination is created. As no protection is provided at the connection level, no additional path is created in addition to this working path.

Exactly one failure event is triggered during the simulation, whose time is chosen randomly once a stable state had been reached. At that time, r links are randomly chosen and deemed as failed. From now on, we refer to a specific value of r as the fraction of failed links. The simulations are performed with $r = 5\%, 10\%, 20\%$. The number of connections affected by the failure are averaged over 30 runs, each one processing a new set of 95000 randomly-generated demand set. A connection is considered affected if: a) its duration has not yet reached its declared lifetime, and b) at least one link included in its path is hit by the failure.

As can be seen in Fig. 4.8, even when r is relatively low (5%), almost half of the LSPs (42%) active at the time of failure are affected, figure that jumps to about 90% when $r = 20\%$. It is clear that these numbers will be different if connections are provisioned with protection, but in any case they highlight that functional damage, in this case from the perspective of active connections, is far more serious that can expected by just looking at purely topological measures, for example the ones evaluated in the previous section. Nevertheless, A_{2TR} and S_{LC} can be used to estimate the number of future connections that would be blocked by lack of connectivity, so both classes of measurements can be complementary.

4.4 Limiting functional damage through Link Prioritization

Let us suppose that all the links in a network have equal probability of being hit by a certain failure, and that it is possible to make them invulnerable at a fixed cost per link. Let us suppose also that several links can be affected at once — that is, there exists a fraction r of failed links, as before — and that a budget is available for shielding a limited number of them so as to reduce the total number of affected connections when such a large-scale failure event occurs. Which links should be part of this set of invulnerable links? Which criteria can be used for selecting them in the best possible way?

The combinatorial and non-deterministic nature of this problem make it difficult to offer a computationally simple and exact solution, and thus call for approximate solutions instead. In this section we discuss two heuristic-based approaches to the problem, which are discussed in more detail in [126] and [123]. The first one takes advantage of the concept of betweenness centrality, which from now on we will identify as EdgeBC. The second is based on link usage statistics collected as part of the connection set-up phase, identified from now on as OLC, for Observed Link Criticality. The idea is that they produce a prioritized list of links that we can choose from to satisfy the maximum number of links that are to become invulnerable. Thus, they offer a criterion for link prioritization.

We proceed now to explain both approaches, highlight their strengths and limitations, and compare, through simulations, their performance at different fractions of failed links.

4.4.1 EdgeBC: The betweenness centrality approach

The concept of betweenness has been used in a variety of settings, for example to find communities in networks [109], to test tolerance to targeted attacks [46], and to reduce connection blocking in path oriented networks [125]. As previously stated in Section 3.2.4 that the edge betweenness centrality $C_B(e)$ determines how often a link e lines along the shortest paths in the whole topology. Thus, it gives an estimation of the importance of a link as a mediator in the communication.

Given that $C_B(e)$ depends only on the topology, it can be computed just once as long as the topology remains unchanged. However, this very fact is also the source of its weakness, for it cannot fully take into account some fundamental aspects of an operational network. For instance, from the point of view of routing, the network topology suffers recurrent virtual and

transient changes. There is a virtual link removal when the corresponding residual capacity reaches zero. Conversely, the link is re-inserted later on when the connections that use it are torn down. Therefore, the shortest path at any instant depends on the network state: one particular connection request might be assigned the ideal shortest path, but the next one might not. Furthermore, links need not have all the same capacity and there may be imbalances in the traffic matrix, as the contribution to the total traffic of certain node pairs can be substantially different from other pairs. Thus, C_B should be viewed as a rough estimation of the effective centrality.

4.4.2 OLC: The Observed Link Criticality approach

This measure is based on the concept of criticality in minimum-interference routing [25]. The difference is that, instead of relying on an approximation based on static data, we can directly take advantage of dynamic information about resource usage that can be collected in the GMPLS control plane. Specifically, each link e can have associated a counter c_e for the number of LSPs going through it, and that counter can be updated as connections are accepted and released. That way, the relative importance of e is $M_e = \frac{c_e}{N}$, where N is the number of active connections at a certain instant. From this, an estimation of the link importance can be obtained as a simple moving average of M_e:

$$I_e = \frac{1}{k} \sum_{i=0}^{k-1} M_{e_i} \tag{4.3}$$

where k is a constant for the number of consecutive samples to use, and M_{e_0} is current value of M_e, $M_{e_{-1}}$ is the immediately preceding value and so on.

The disadvantage of this approach compared to EdgeBC is that the network must already be in operation, and preferably in a steady state, before it can be applied.

4.4.3 Performance Comparison

We compare the performance of EdgeBC and OLC through simulation, in much the same way as in Section 4.3. Besides the fraction of failed links r, there is an additional parameter, the size of the set of invulnerable links z, also a fraction.

The simulations are performed with $r = 5, 10, 20$, and $z = 10,\ 20, 30$ (both as percentages). Thus, one simulation run corresponds to a specific

Table 4.2: *Frequency distribution of connection path length of a representative simulation run*

Path length	Frequency (%)	Cumulative frequency
5	8.5	8.5
6	9.4	17.9
7	10.2	28.1
8	10.3	38.4
9	10.2	48.6
10	9.4	58.0
11	8.4	66.4
12	7.6	74.0
13	6.4	80.4
14	5.2	85.6
15	4.3	89.9
16	3.5	93.4
17–24	6.6	100.0

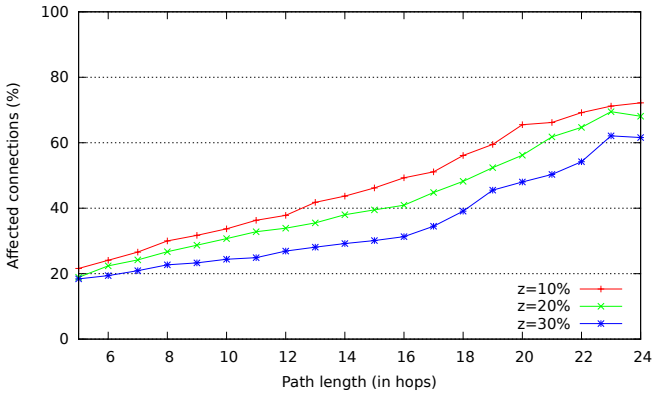
combination of r, z and the procedure for the selection of invulnerable links, which for simplicity is called *selection strategy* from now on. For comparison purposes, a third selection procedure is included, which chooses links randomly (uniform distribution). This gives 27 cases in total. As before, the results are the average of 30 runs per case. Two figures of merit are considered:

- the percentage of active connections affected by the failure.
- the frequency distribution of path lengths of the affected connections.

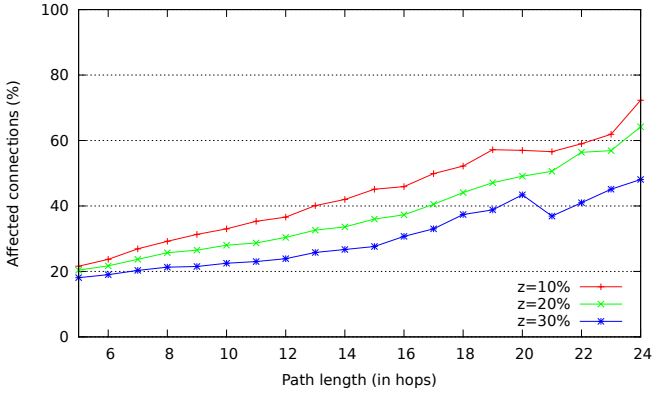
Table 4.2 shows the distribution of the frequency of connection path length of a representative simulation run. It can be observed that the frequency distribution is rather wide, although the very long paths are infrequent: only about 10% of them are longer than 15 hops.

As a single average value may not be very sufficiently representative, Fig. 4.9 presents the percentage of affected connections discriminated by path length in the case of have $r = 5\%$. The subfigures correspond to the three strategies when $z = 10, 20$ and 30.

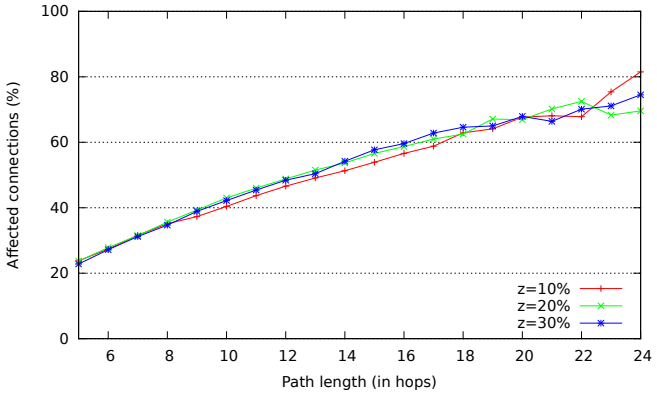
As can be expected, RANDOM is essentially insensitive to the fraction of invulnerable links. It is interesting to note in Fig. 4.9c that there exists a case in which z is six times the value of r, but even then the positive effect is negligible. This behavior is similar for all path lengths. Only for the longest paths in Fig. 4.9c does performance vary with respect to z, which is due to



(a) EdgeBC



(b) OLC



(c) Random

Figure 4.9: *Fraction of connections affected by the failure when* $r = 5\%$

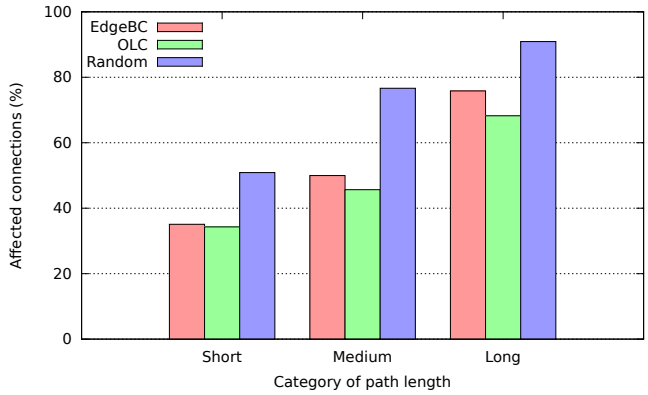


Figure 4.10: *Affected connections when $r = 10\%$ and $z = 30\%$, grouped by category of path length*

the fact that the number of connections of such long lengths is very small compared to the rest (see Table 4.2).

In the remainder of the cases, the behavior clearly depends on r and z. For instance, when $r = 5\%$ and $z = 10\%$ the difference with respect to RANDOM at path length = 10 is about 8%, but it jumps to almost 20% when $z = 30\%$. In general, both EDGEBC and OLC offer similar results, but when z is raised to 30%, OLC is the one whose reaction is more visible, producing the lowest values for the number of affected connections.

The performance of the three strategies is summarized in Table 4.3. The column "% Affected connections" gives the average percentage of connections affected by the failure. The remaining columns put connections into three categories based on their path length, and show what fraction of each group was adversely affected. The categories are as follows: a) *Short* (5–8 hops), b) *Medium* (9–18 hops), and c) *Long* (19–24 hops). Each value is an average of the individual results in the range. Every combination of r, strategy and z considered in this section has an entry in the table. The results grouped in this way are shown graphically in Fig. 4.10 for the case of $r = 10\%$ and $z = 30$. As can be seen, the difference between RANDOM and the other two is almost 20% for short paths. That difference jumps to around 30% for paths of medium length, and shrinks back to around 20% for long paths.

These results show the high sensitivity of a connection-oriented network to large-scale failures, because the failure of even a relatively small fraction of links (5%) causes disruption to almost half the connections in the studied scenario. Moreover, the RANDOM strategy, which represents a prioritization

Table 4.3: *Performance of* EdgeBC, OLC *and* Random, *discriminated by category of path length*

r	Strategy	z	% Affected connections	Path length		
				Short	Medium	Long
5	EdgeBC	10	34.6	25.6	42.8	67.3
		20	30.8	23.1	37.3	62.1
		30	25.1	20.4	29.2	53.6
	OLC	10	33.7	25.4	41.1	60.7
		20	28.6	22.9	33.8	54.1
		30	23.4	19.7	27.2	42.2
	Random	10	40.5	29.5	50.1	70.8
		20	42.0	29.7	52.1	69.1
		30	41.6	29.0	52.4	69.2
10	EdgeBC	10	54.5	43.1	64.5	86.2
		20	49.2	40.0	57.3	82.5
		30	42.7	35.1	50.0	75.8
	OLC	10	53.9	45.0	65.4	85.1
		20	47.5	39.0	55.7	77.1
		30	40.2	34.3	45.6	68.2
	Random	10	64.1	49.1	76.2	90.4
		20	64.5	50.2	75.8	90.1
		30	65.2	50.9	76.6	90.9
20	EdgeBC	10	79.9	69.9	88.2	98.5
		20	71.6	62.6	79.6	96.0
		30	65.2	56.6	73.2	93.9
	OLC	10	78.8	70.4	87.9	96.4
		20	70.7	61.8	78.6	94.7
		30	62.2	55.4	68.9	90.0
	Random	10	87.1	76.7	94.6	99.4
		20	87.4	77.0	94.9	99.6
		30	86.4	75.9	94.1	99.0

without a specific criterion, shows that little or no benefit is obtained in terms of robustness by choosing links disregarding their role in the overall traffic flow.

With respect to the strategies EdgeBC and OLC, results show that both are capable of minimizing the impact of these failures with the appropriate election of z. Of the two, we can see that OLC is the best overall performer: the number of connections affected is lower than with the EdgeBC and, at the same time, the number of surviving connections whose path lengths are medium and long is higher.

5

Large-scale propagating failures in GMPLS networks

In this chapter we address the problem of assessing the robustness against multiple node failures in GMPLS networks. Our focus is on failures that can propagate simultaneously on two axis: horizontally in the control plane (that is, from node to node), and vertically, from the control plane towards the data plane.

5.1 Multiple failures in GMPLS-based networks

Given that current networks integrate multiple transport technologies, systems as a whole usually follow a stacked multilayer architecture, whereby the upper layers operate on virtual topologies built successively upon structures realized in the lower layers [118]. This multi-layered architecture can improve network resilience due to the fact that it brings flexibility to fault management and recovery [37]. Unfortunately, it also introduces an undesirable effect known as *failure propagation*, whereby failures at the bottom layer may disrupt services in higher-level layers. Furthermore, by the very nature of the architecture, one failure at the bottom layer can manifest itself as several concurrent failures in higher layers.

The negative effects of failure propagation can be avoided or limited by having the network's lower layers automatically find or use new paths or subpaths after a failure, provided it has its own protection mechanism. Thus, recovery procedures can be automatically activated upon failure, making it invisible to the upper layer [45]. A different approach is to design the higher-level network topology taking into account the capabilities and constraints of the lower-layer network. Suppose for example a two layer system, as in IP over WDM. In such a system, the objective of this second approach is to place the demands (LSPs) of the IP layer on the WDM infrastructure in such a way

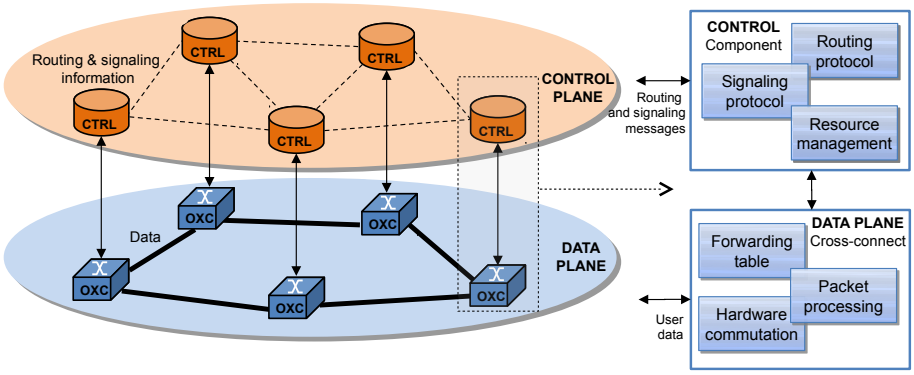


Figure 5.1: *The Control and Data planes in the GMPLS architecture*

that failures will not leave the IP topology disconnected and that capacity will be available to successfully complete the recovery at that same layer. This is a design problem known as *network mapping* (or survivable mapping), studied in the context of optical network design for survivability, on which there is ample literature, see for example [103],[118],[73],[113],[85] and [71]. As the network mapping problem is known to be NP-complete [45], several heuristic algorithms has been proposed either to find mappings or to augment a given topology until a desired mapping can be found, see for instance [82],[90] and [135]. In any case, we must remember that these approaches are effective when failures are localized, not for arbitrary large-scale multiple failures.

Furthermore, even if the architecture of the network under consideration is not multi-layer, very similar issues regarding failure propagation arise if it is a GMPLS-controlled network. As previously stated, GMPLS clearly distinguishes two different parts in every node (see Fig. 5.1), which is a consequence of the separation of planes. First of all there is a forwarding component, where specially designed hardware is dedicated to process as fast as possible incoming data streams towards the corresponding output ports, according to a forwarding table. Above this component, there is a generic control hardware executing a specific network operating system that runs the routing and signaling protocols and configures the forwarding table (when connections are established or released). Although both components are usually located in the same device, they have some degree of isolation from one another.

In such scenarios, it can happen that an attack or failure only affects the control component or only the forwarding component for a short period of time. It is even possible that, due to a virus, targeted attack or software

configuration error, the failure affects only to a single control plane mechanism (i.e. signaling protocol or routing protocol). In the case that the signaling module fails and the routing module is still working, connections cannot be established or removed through that node. In this case, it is possible to use the routing module to advertise the neighbors that there is no free capacity available so they do not attempt to establish new connections through the partially failed node. On the other hand, if the signaling module is still operational but the forwarding module fails, changes in the local state (e.g. capacity being allocated/released) will not be advertised to the neighbors and they will be working with out-of-date information. However, the failed node could still be able to process new connection requests and tear down existing connections.

It is of major importance to establish some mechanisms in order to recover the functionality of the failed control component as soon as possible and re-synchronize the control and forwarding components. One way to achieved this is by having the nodes implement re-synchronization mechanisms like "Non-Stop Forwarding" and "Graceful Restart" [110],[28]. Nevertheless, this process can be complex and may take some time to complete due to a first stage of reinstalling or rebooting the control component, and the message exchange procedure that must be performed to achieve re-synchronization [88]. In any case, the broader issue of resilience of the control plane has not been neglected by the research community, see for instance [75], [117], [122] and [79], but it is outside the scope of this thesis.

Now that we have established that failures can propagate "vertically" (from control to data plane), we can consider failures in the other axis, that is, "horizontally", from node to node. Certain types of failures, for example those originated in software bugs, intentional attacks and even configuration errors, make this scenario at least conceivable.

In this thesis, it is assumed that this type of multiple failures propagate through the control plane exclusively. This restriction, however, does not diminish the danger, for a failure that reaches another node through the control plane can trigger the inter-plane (vertical) failure propagation.

We have studied this type of failure scenarios from the perspective of epidemic networks [24]. The following sections explain the main ideas and the results obtained.

5.2 Basic terminology of epidemic networks

Epidemic networks is a general term that describes how an epidemic spreads when new cases of a certain disease, in a given population and during a given period, substantially exceed what is expected, based on recent experience. The rise and decline in epidemic prevalence of an infectious disease is a probability phenomenon dependent upon the transfer of an effective dose of the infectious agent from an infected individual to a susceptible one. Research in this area involves different aspects, such as modeling how an epidemic evolves or how to immunize part of the population to minimize or control the effect of the epidemic. Power supply networks, social networks, neural networks or computer networks are some cases where this subject is of special relevance. Furthermore, it is possible to generalize from virus (or diseases) to failures, for there are certain types of failures whose propagation dynamics resemble that of epidemics.

An epidemic network is usually modeled as a graph in which vertices (or nodes) represent the individuals and edges their relationship (for example a disease). Several types of nodes and failures can be represented. For instance, in a medical context when a failure affects a node, it refers to a biological virus infecting a cell. Just as when, in power supply networks, a failure refers to a power station stopping providing service.

The problem of virus propagation has attracted huge interest among the scientific community. Several models have been proposed for epidemic dynamics, the most common being *Susceptible-Infected* (SI) , *Susceptible-Infected-Susceptible* (SIS) *Susceptible-Infected-Removed* (SIR) models. Embodied in their names usually the stages of the disease for each individual in the network. In the SIS model, for example, an individual is initially health but susceptible to contagion, then it becomes infected and remain infective for some period of time, and returns to the initial susceptible state afterwards. An individual is "infective" when it can pass the disease onto its neighbors, which happens at some rate. In fact, all transitions between states have associated a rate. Fig. 5.2 is the typical representation of the stages of an epidemic model. In this case, it is the state-transition diagram of the SIS model, where β and δ are birth and death rates respectively.

Unlike the SIS model, an infected individual in the SI model will remain infected (and infective) forever, as in fact happens with some diseases in the biological world. On the other hand, the SIR model introduces the "removed" state, in which the individual dies after the infection. Thus, the epidemic will also die out over time (when all individuals had died). Note that "SIR" also identifies another variant, the *Susceptible-Infected-Recovered* model, in

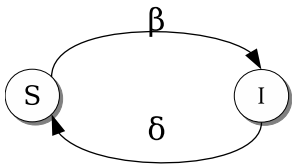


Figure 5.2: *The state-transition diagram of the SIS model*

which an individual can be infected just once because when it recovers, it becomes immune and will no longer pass the infection onto others.

These are the most common models, but the interested reader can find several others in the literature on epidemics and complex networks, see for instance the books [41] and [13]. Additional concepts and terminology pertaining to epidemics will be introduced as needed when we discuss our SID model in Section 5.3

5.3 A new model of failure propagation: The SID model

This section introduces the *Susceptible-Infected-Disabled* (SID) failure propagation model. SID can be considered an extension of the SIS model, where the addition of a new state is a consequence of our need to take into account failure propagation in the two directions previously discussed: node-to-node in an epidemic-like fashion, and from control plane to data plane.

Fig. 5.3 shows the transition-state diagram of the SIS model, embedded in a larger diagram corresponding to the SID model. Each node, at each time-step t, is either susceptible [S] or infected [I]. A susceptible node can be infected with probability β by receiving the infection from a neighbor. An infected node can be repaired with probability δ_1. Remember the assumption made regarding the failure of a node's control plane: it means that no new connections requests can be accepted by that node. It is interesting to note that the larger the value of $1 - \delta_1$, the greater the number of path requests blocked by the node. δ_1 can be evaluated by taking into account the time to detect the failure and the time to repair (and sometimes update) the modules affected by the failure/virus attack. This process can be performed without disrupting the ongoing connections in the data plane.

Strictly speaking, and following the example of other models, SID should be called SIDS, because nodes always return to the initial state. But as in [26] it was called "SID", we prefer to keep using that name for consistency.

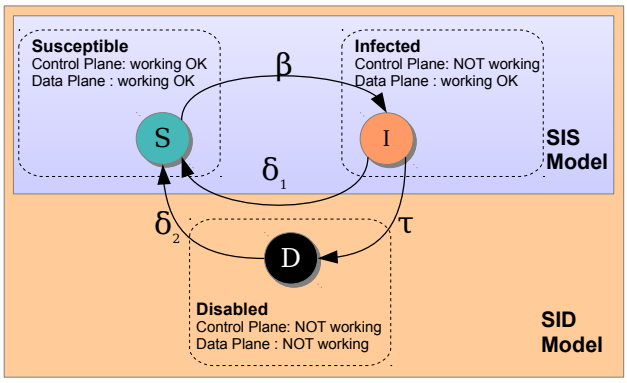


Figure 5.3: *State-transition diagram of the SIS and SID models and the relationship to the operational states of the GMPLS planes*

The added sate Disabled (D) is meant to takes into account the fact that an infected node may degrade to a complete nodal failure (i.e. control and data plane failure). When a node becomes disabled, all connections crossing that node are removed (i.e., lost). In that case, the node needs a process in order to be repaired, and the time needed is directly proportional to the $MTTR$ (Mean Time To Repair), that is, δ_2 can be computed as $1/MTTR$. It is worth mentioning that in the SID model an infected has to possible transitions, which makes it different from existing models.

5.3.1 SID epidemic thresholds

This model can be described by a *Markov chain* in either continuous time or discrete time with a small enough time step. The*basic reproduction number*, usually denoted by R_0 and defined as the average number of infections produced by an infective individual (infected node) in a wholly susceptible population is

$$R_0 = \frac{\beta}{\delta_1 + \tau} \lambda_1 \tag{5.1}$$

where $\lambda_1 > 0$ is the largest eigenvalue of the non-negative irreducible symmetric adjacency matrix of the network (see [48] for expressions of R_0 in a wide range of models), For the particular case of a homogeneous network, the largest eigenvalue is equal to the average nodal degree, see [33] and [116].

The formula for R_0 above can be interpreted as follows: $\beta\,\lambda_1$ is the transmission rate across an infective contact times the expected number of contacts (connexions), whereas $\frac{1}{\delta_1+\tau}$ gives the expected lifetime of an infected

node (i.e. the mean infectious period). Therefore, the following epidemic threshold can be stated:

- if $R_0 < 1$, equivalently, $\frac{\beta}{\delta_1+\tau} < \frac{1}{\lambda_1}$, then the infection dies out over time, that is, the number of infected and disabled nodes goes to zero.
- if $R_0 > 1$, equivalently, $\frac{\beta}{\delta_1+\tau} > \frac{1}{\lambda_1}$, then there is an epidemic outbreak affecting ultimately a fraction of the network nodes.

The spread of the infection in the network depends on the topology of the network through the single parameter $\lambda_1 > 0$. This phenomenon follows on from the systematic approach of considering the linearization of the model around the disease-free steady state, where the adjacency matrix of the network appears and its largest eigenvalue λ_1 determines the (un)stability. Analogous results have been reported in [33] and [116] for the case of the SIS model.

Finally, for a homogeneous network we have explicit expressions for the endemic steady state: the fraction of susceptible nodes is $\frac{1}{R_0}$, the fraction of infected nodes is

$$(1 - \frac{1}{R_0})\frac{1}{1+R_1} \quad \text{with} \quad R_1 = \frac{\tau}{\delta_2}, \tag{5.2}$$

and the fraction of disabled nodes is

$$(1 - \frac{1}{R_0})\frac{R_1}{1+R_1}. \tag{5.3}$$

Fig. 5.4 shows these proportions varying the parameter $\beta/(\delta_1 + \tau)$. Moreover, this endemic equilibrium is asymptotically stable whenever it exists ($R_0 > 1$). It also shows analytically the values for the number of nodes in the susceptible, infected and disabled states. Two important points are highlighted in the figure: the intersection of the infected and susceptible curves, and the intersection of the disabled and susceptible curves. The mathematical expressions for both intersection points of our model are also given.

5.3.2 Empirical validation of the model

To validate empirically the analytical model, we proceeded to simulate the spreading of an epidemic according to the SID model on different topologies and rates. A simple approach was used: at each time step and for each and every node in the topology, the next state is determined based on the

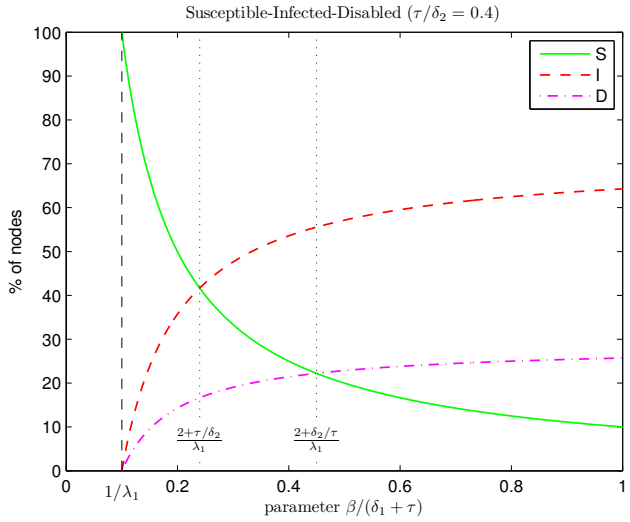


Figure 5.4: *SID model: Analytical values for the number of nodes per state*

current state and provided the rates. Thus, as time advances we can have the number of nodes per state and compare them with the analytical estimation.

Fig. 5.5 shows the temporal evolution of the epidemic on the t65 topology on a typical run. The rates chosen are shown in the figure. The expected fraction of infected nodes is so the analytical value is 37.5% (see Eq. 5.2). As can be seen in the figure, once the epidemic reaches a steady state, the fraction of infected nodes is close to the analytical value. Values for the number susceptible and disabled nodes can be obtained in a similar way. Several simulations have been performed to confirm that the analytical and the simulation values are always equal.

5.4 Failure propagation on Rings

This section studies the propagation of failures on ring topologies to understand what effect it has on network availability. It is assumed that failures propagate basically according to the SID model, with some adjustments to account for the peculiarities of ring topologies.

Rings are interesting for two reasons. Firstly, studying their reliability is important as they are widely deployed as part of several transport technologies, for example SONET/SDH, and are commonly found in metropolitan area networks [103],[59]. Secondly, their simple structure makes it possible

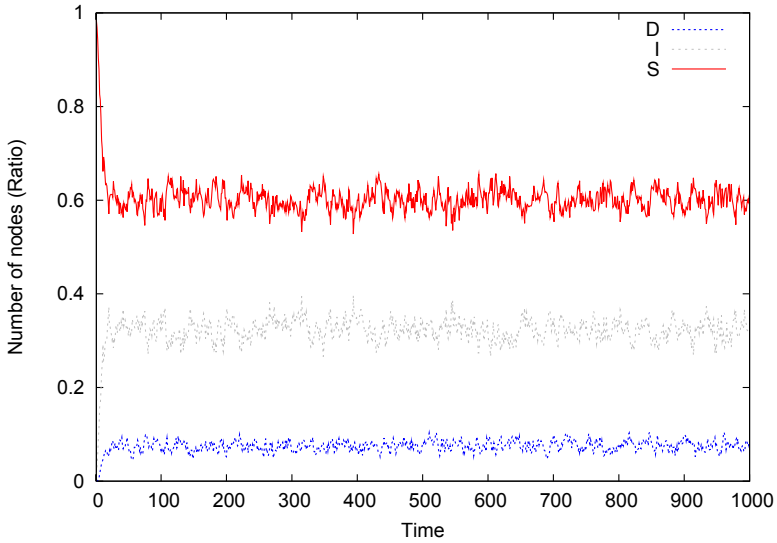


Figure 5.5: *Epidemic spreading on the* t65 *topology when* $\delta_1 = 0.3$, $\delta_2 = 0.3$, $\tau = 0.1$ *and* $\beta = 0.167$.

to use enumerative approaches that are impractical with arbitrary topologies due to the state space becoming exceedingly large. More specifically, if we assume that failure events in an individual node occur independently from one another and that they exhibit the memoryless property, that is, the inter-failure times are exponentially distributed, then a Continuous-Time Markov Chain (CTMC) can be used to model the propagation of failures on them and assess their reliability numerically.

This section is based on [127], which introduced CTMC models for two rings, one consisting of eight nodes and the other of thirty two nodes[1]. The focus in this section is on the results for the eight-node ring; the interested reader can find the aforementioned article the formulation for rings of arbitrary size as well as the precise mathematical formulations employed.

5.4.1 Assumptions

A CTMC is characterized by the so-called *state-transition-rate diagram*. In our case, this can be constructed based on the SID model and the associated rates, see Fig. 5.3. The values on the arrows refer to the transition rates between states, that is, the failure or repair events per unit of time. Thus, a

[1]The CTMC-based performance analysis presented in [127] was a joint work undertaken by researchers from *Universidad Carlos III de Madrid* and *Universitat de Girona*.

fully operational node (state S), which by definition means being susceptible, becomes infected at rate β. An infected node may become again operational (state S) or disabled (state D). The first case occurs at rate δ_1, which is the rate at which the network administrator fixes the problem, whereas the second case occurs at rate τ. The network operator may also repair disabled nodes at rate δ_2, returning it to state S.

One rate that is not included in Fig. 5.3 but necessary for developing the CTMC model is the spontaneous infection rate β_F. This value refers to the rate at which a given node none of whose neighbors are infected may spontaneously become infected. Its purpose is to account for the appearance of new infections occurring without external intervention. The value of β_F is assumed to be much smaller than β. Thus, for simplicity, it does not appear in the calculation of the corresponding infection rate. Furthermore, spontaneous infections should be rare under the following behavioral hypothesis: when a node has just had a control plane failure, no more isolated nodes are allowed to have spontaneous control plane failures, but only by infection propagation.

It is important to remember that in this scenario, the epidemic-like spreading of failures happens only among entities of the control plane, that is, the inter-plane failure propagation (from the control plane to the data plane) is not epidemic; instead, it is the consequence of assuming that a certain ratio of nodes in state I cannot be repaired, at which point the data plane (in fact the whole node) also fails, until it is returned to the susceptible state by manual intervention.

One peculiarity of ring-based networks as deployed in data communications (e.g., double ring configurations with traffic flowing in opposite directions on the ring) is that single node failures do not break connectivity. This can be thought of as if node removal generates a new smaller connected ring. In fact, several nodes can be removed and full connectivity preserved among the remaining nodes as long as the failed nodes are adjacent in the original topology. In this section, it is assumed that the system (the whole ring) is disconnected (or unusable, unavailable) when there are at least two nodes in state D, as well as when the number of nodes in state S is zero. Such a "disconnection" state represents a major failure requiring the urgent intervention on the part of the network operator.

5.4.2 A CTMC model for a small ring

Let us consider the eight-node topology of Fig. 5.6. It represents the case in which all nodes are fully operational, that is, in state S. The dynamics of the epidemic dictates that, over time, one or more nodes will become

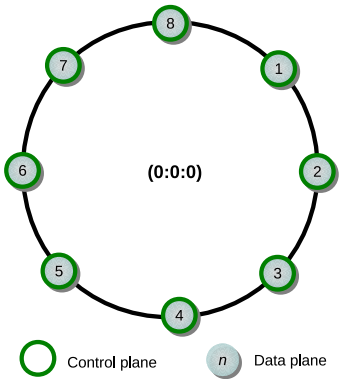


Figure 5.6: *The eight-node GMPLS-based ring example*

infected, some will be disabled for a certain time period and then return to the healthy state, etc. Each specific combination of node states constitute a ring *configuration*. To fully characterize the possible configurations, let us enumerate explicitly the rules governing the transitions:

1. Already infected nodes may infect only neighboring nodes. A node may be infected only if it has at least one neighboring node already infected. The first infection occurs spontaneously at rate β_F.

2. Already infected nodes may become disabled. Disabled nodes cannot infect other nodes, nor can they propagate their disabling state to other nodes.

3. Both infected and disabled nodes may be repaired by the administrator, but only if they are adjacent to a susceptible node. In other words, node repair strategies occur at the edges of the infected/disabled area.

With these rules in place, a triplet notation (N_{Il}:N_D:N_{Ir}) can be adopted for the ring configuration. The number of nodes in the $\boxed{\text{D}}$ state is indicated by N_D, while N_{Il} and N_{Ir} are the number infected nodes to the left and to the right of the disabled area, respectively. The "left" and "right" are identified by look towards the center of the ring. In the figures, the ring configuration triplet, also called the "ring state", is shown at the center.

Note that when $N_D = 0$ (that is, no nodes are in state $\boxed{\text{D}}$), the notation may be reduced to **(0:0:**N_{Ir}**)**. Furthermore, given the symmetry of rings, configurations **(0:1:1)** and **(1:1:0)** are equivalent and must be treated as one. By convention, $N_{Il} \leq N_{Ir}$ in the triples.

An example of a series of ring state transitions is given in 5.7. Initially, the ring is in state **(0:0:0)** as all nodes are susceptible. At some point in time, one node becomes spontaneously infected (this is node number 1 in Fig. 5.7b, but it may be any of them). This occurs with rate $8\beta_F$ and brings the ring to the state **(0:0:1)**. From there on, that infected node may cause a transition to one of the following states:

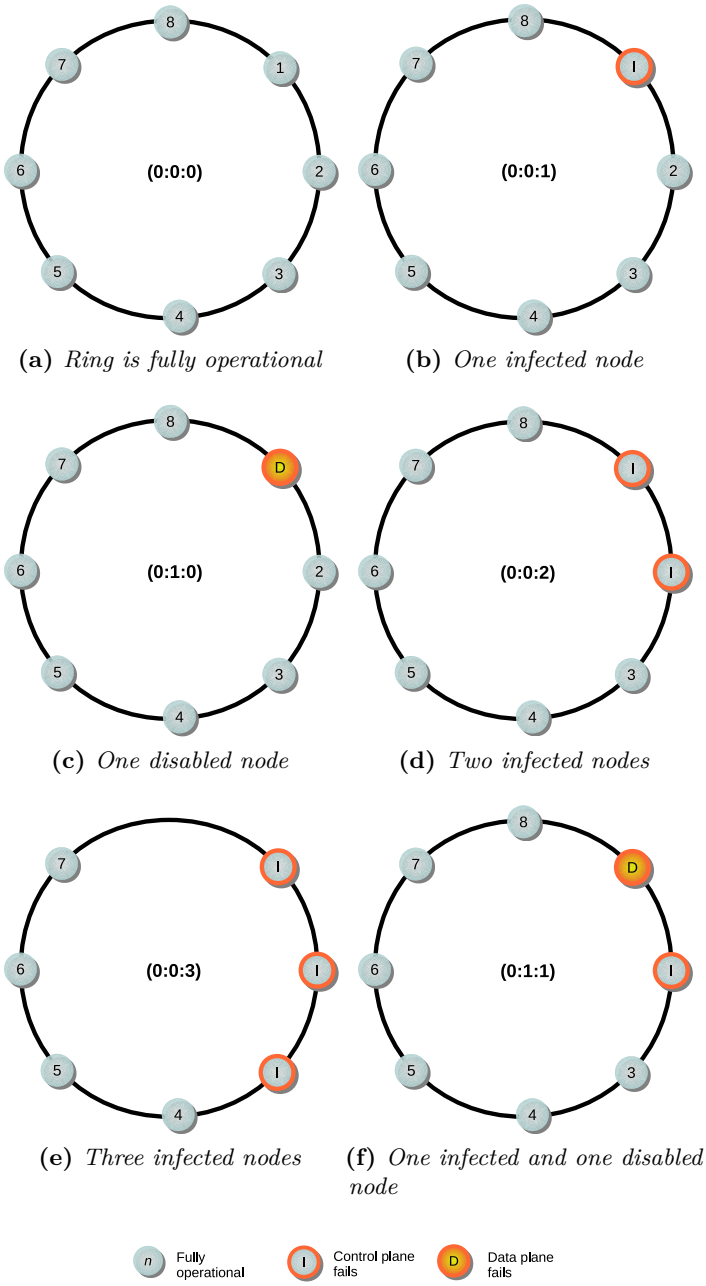
- the state **(0:0:2)** if the infection is passed on to a neighbor, which occurs with rate 2β since the infected node may infect any of its two neighboring nodes.
- the state **(0:1:0)** if it becomes disabled, which occurs with rate τ (see Fig. 5.7c).
- the state **(0:0:0)**, if the network operator repairs the node and returns it to the susceptible state. The rate of this transition is δ_1.

By proceeding in this way, all the possible transitions and rates can be enumerated, obtaining a full state-transition-diagram for the topology as a whole (the corresponding diagram is given in [127]). The infinitesimal generation matrix Q can be computed from the state-transition-rate diagram. This matrix characterizes the transient behavior of the CTMC. so that the steady-state probabilities (that is, the percentage of time that the ring is in a given configuration) can be obtained, as well as the first-passage times of a given state (that is, the amount of time on average to reach a given state).

5.4.3 Guidelines for the assignment of repair rates

By solving the steady-state probabilities of the CTMC-based model, it is easy to find the percentage of time that the ring stays in every state as a function of the two repairing rates δ_1 and δ_2. Remember that δ_1 is the rate at which the control plane of a node is repaired, and δ_2 the rate at which nodes are returned to the fully operational state after a complete failure. The units of all the rates are normalized as the amount of transitions events (failures or repairs) that occur in an infinitesimal period of time in the CTMC model.

Operators usually have no control over the failure rates (i.e., β and τ) but they can choose the repair rates. Therefore, the goal here is to offer guidelines for selecting the appropriate repair rates to attain a given network availability, say 99.999% availability. To that end, a sensitivity analysis of the two repair rates is performed, seeking to identify trends from which to derive the desired guidelines.



(a) *Ring is fully operational* **(b)** *One infected node*

(c) *One disabled node* **(d)** *Two infected nodes*

(e) *Three infected nodes* **(f)** *One infected and one disabled node*

Figure 5.7: *Examples of system states on the eight-node ring topology*

The repair rates influence the time spent in each network state. As there are many states even for small topologies, it can be more convenient to group the results into a number of categories:

- **Fully operational state**, that is, the percentage of time at which the ring has all its nodes fully functional.
- **Moderate Infection**: includes all states in which at most one node is disabled and at most $N/2$ nodes are infected.
- **Severe Infection**: includes all the states in which at most one node is disabled and more than half of the nodes are infected.
- **Disconnection**, which happens whenever the ring has more than one disabled node, or all of its nodes are infected.

These categories are symbolized in the figures that follow by $P_{(0:0:0)}$, P_{lowI}, P_{highI} and P_{DISC} respectively. Note that the boundary between moderate and severe infection can be defined arbitrarily. Here, $N/2$ was chosen for simplicity.

5.4.4 Numerical results

Fig. 5.8 shows the stationary probability of each category of states as a function of δ_1. Fig. 5.8a corresponds to the case when $\beta = 1$, $\tau = 1$ and $\delta_2 = 0.5$. As can be seen, severe infection and disconnection are clearly related; they are essentially the same curve. In them, the variation with respect to δ_1 exhibits basically two stages: it is relatively stable up to certain point, but then tends to zero very quickly. By comparing the two subfigures we can also see that δ_2 practically has no influence on the stationary probability.

In Fig. 5.9 there are the same two cases as before, expect that now β is 20 times larger. The trends are, in general, the same as in the previous figure. However, the specific values are quite different. For example, for the disconnection probability to approach 10^{-4}, δ_1 needs to be around 10^2 in Fig. 5.8a but around 10^3 in Fig. 5.9a.

In summary, we can say that in order to achieve a disconnection steady-state probability below 10^{-5}, it is required that $\delta_1 > 4 \times 10^2 \beta$ when $\beta = 1$, and $\delta_1 > 4 \times 10^3 \beta$ when $\beta = 20$.

Regarding δ_2 (the rate at which completely failed nodes are taken back to the susceptible state), the corresponding sensibility analysis shows that its influence over the stationary probabilities is also residual. The interested

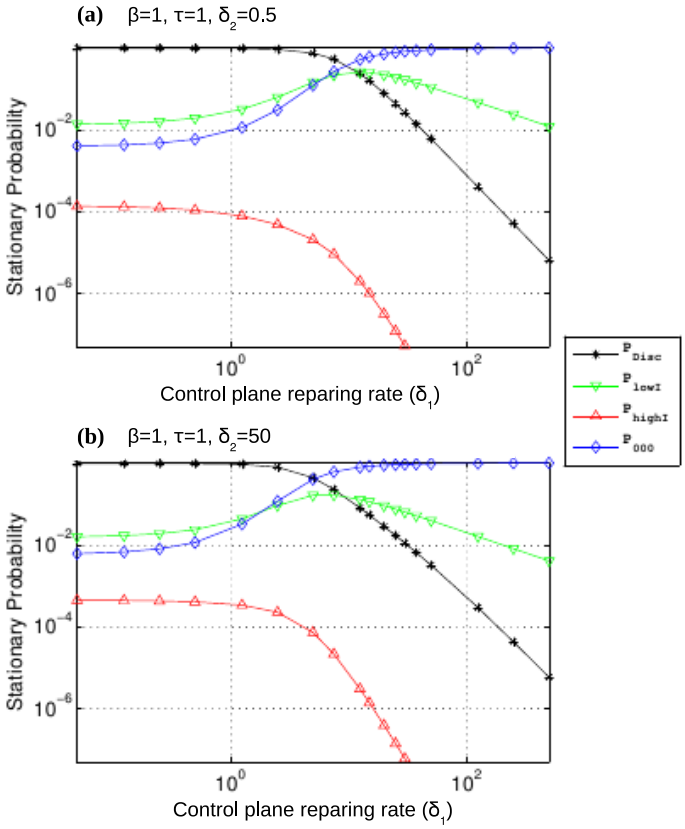


Figure 5.8: *Impact of δ_1 on the steady-state probabilities of the CTMC for the eight-node ring when $\beta = 1$*

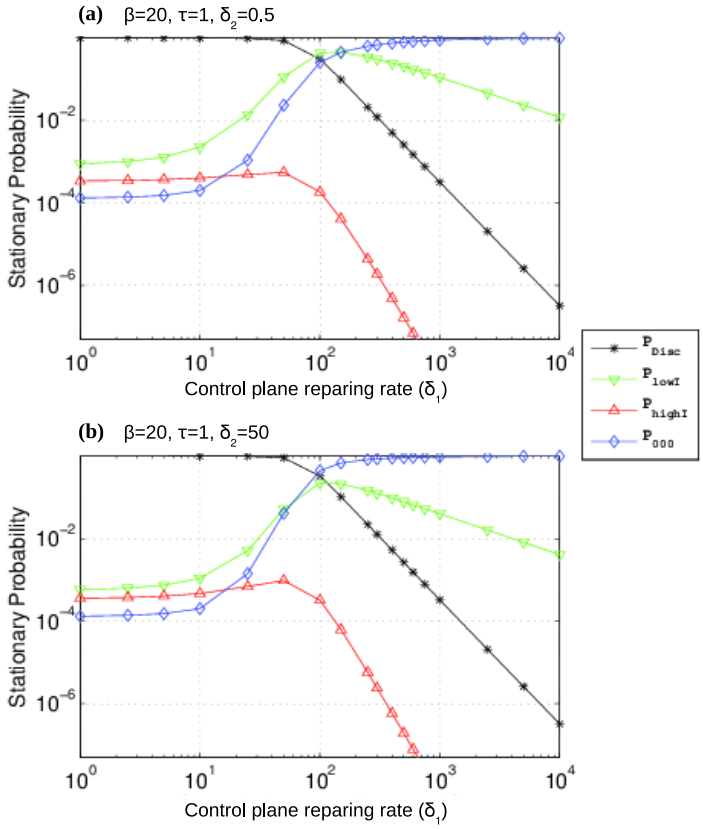


Figure 5.9: *Impact of δ_1 on the steady-state probabilities of the CTMC for the eight-node ring when $\beta = 20$*

reader can find the details in [127], as well as estimation of the mean time to failure as a function of the repair rates.

In essence, the key to having a highly available network is the ratio β/δ_1. The results show that it is safe to have a repairing rate $\delta_1 > 10^3\beta$ to guarantee 99.999% network availability.

5.5 Comparing robustness against propagating failures

This section shows an application of the SID model to the assessment of robustness when failures propagate from node to node. The starting point is the following: given two arbitrary topologies "A" and "B", if both are subjected to a large-scale failure of similar intensity (e.g., a SID-based epidemic spreading), topology "A" can be considered more robust than "B" against that failure if the effects of epidemic spreads more slowly on it than on "B".

It must be noted that this is a novel approach to measuring robustness, which we introduced in [26]. In the context of epidemic-based failures, the metric most commonly used in the literature is the *largest eigenvalue* of the adjacency matrix . The assumption is that larger this value, the more robust the topology is [33]. However, there are drawbacks in using structural measures for the type of networks and failure scenarios addressed in this thesis, as already discussed in previous chapters This new metric has been called "Topology Robustness against epidemics in GMPLS networks", or TRG.

We propose using blocking ratio as the main indicator of system (that is, network) performance. Basically, the goal is to identify the topology on which performance degrades more gracefully (more slowly). To use this indicator, it is necessary to load the networks with connections requests, let the failure spread and observe and record the effects. That is, a combined simulation of service provisioning and epidemic spreading must be performed.

5.5.1 Simulation environment

A dynamic traffic scenario is chosen for the simulations, that is, connections have random durations. The rest of the simulation parameters (connection inter-arrival times, routing policy, number of repetitions, traffic matrix, etc.) are similar what is used in Section 4.3. The differences are as follow. Firstly, three topologies are used here instead of one, namely `t204`, `t65` and `bt400`. Secondly, it is assumed that resources (e.g., capacity), are always available and no other path quality constraint is imposed (e.g., maximum hop count,

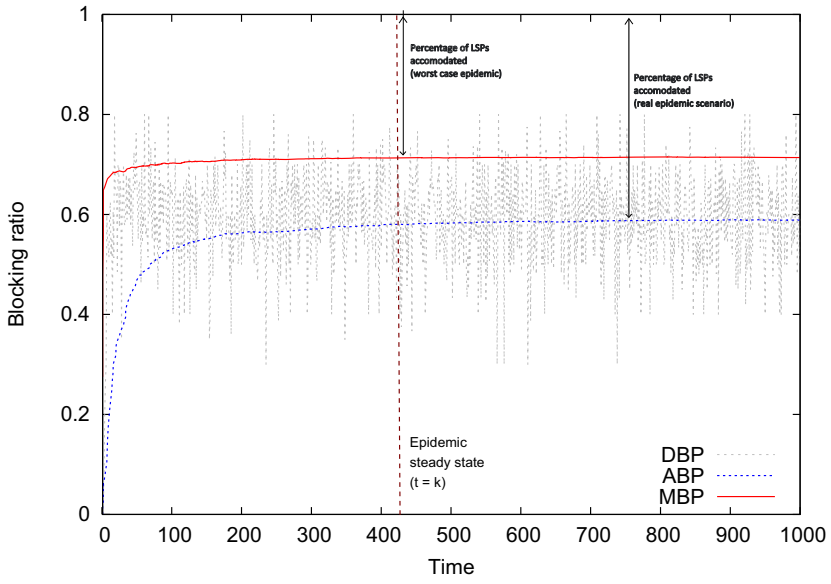


Figure 5.10: *Blocking ratio on the T65 topology when* $\delta_1 = 0.3$, $\delta_2 = 0.3$, $\tau = 0.1$ *and* $\beta = 0.167$.

delay). Thus, the blocking ratio depends only on the effects of the epidemic, that is, connections will be blocked only when no feasible paths exists because the necessary intermediary nodes are not available (are disabled).

With respect to the rates of the SID model, they are chosen so that the topologies to be compared are subjected to infections of similar intensity, which can be calculated easily through the formulations given as part of the definitions of the SID model.

5.5.2 Measuring the performance degradation

Fig. 5.10 contains three curves, all related to the temporal evolution blocking ratio (and the spread of the epidemic) on the `t65` topology (for brevity, the figures for other topologies are omitted, which in any case are essentially similar to this one). At that any discrete simulated time t, a certain number (say Q_t) of new connection requests arrive. Some of them (say B_t) are rejected due to the effects of the epidemic (one or more required nodes are disabled). The ratio B_t/Q_t is the instantaneous blocking ratio identified as

"DBP" in the figure. As can be seen, it oscillate but a trend is clearly visible.

A more stable value is the global or accumulated blocking ratio, identified as "ABP" in Fig. 5.10,

$$ABP(t) = \sum_{i=0}^{t} (B_i/Q_i), \tag{5.4}$$

that is, $ABP(t)$ is the accumulated blocking registered from time 0 up to time t. As expected, ABP grows rapidly and then stabilizes (in fact, it becomes the average DBP in the steady state). Thus, ABP is an average of the performance degradation.

However, there remains the question of measuring speed of the degradation. Let us first introduce the idea of maximum (or worst possible) blocking ratio for a specific epidemic scenario. This happens when the network is in a state in which the number of disabled nodes is the maximum possible. The model definition gives us the expression to estimate that number, which means that a complementary simulation can be run with that many nodes already disabled. The reported results are average values given that the simulations are repeated several times and the nodes which will be permanently disabled are chosen randomly. The curve "MBP" is this maximum blocking ratio, which is practically a straight line, as expected.

Now TRG can be defined in terms of ABP and MBP: TRG is the area between ABP and MBP, once the epidemic reaches its steady-state at some $t = k$, that is,

$$TRG = \int_{t=0}^{k} (MBP - ABP)\, dt. \tag{5.5}$$

However, as MBP and ABP are accumulated values, a simple approximation is the difference between them. The larger this area, the more robust the topology.

5.5.3 Topology comparison through TRG

Fig. 5.11 illustrates the use of the TRGs to compare the three topologies used in this section. Note that instead of giving a single TRG value for each topology, we opted for defining four infection levels, which are defined by assigning appropriate values to the basic reproduction number R_0.

Let us focus on the extreme infection scenario. It can be seen that the `t65` topology is the most robust of the three according to the TRG (its TRG bar is the longest in the illustration). However, if the topologies were ranked according to their largest eigenvalue, the most robust one would be `bt400`.

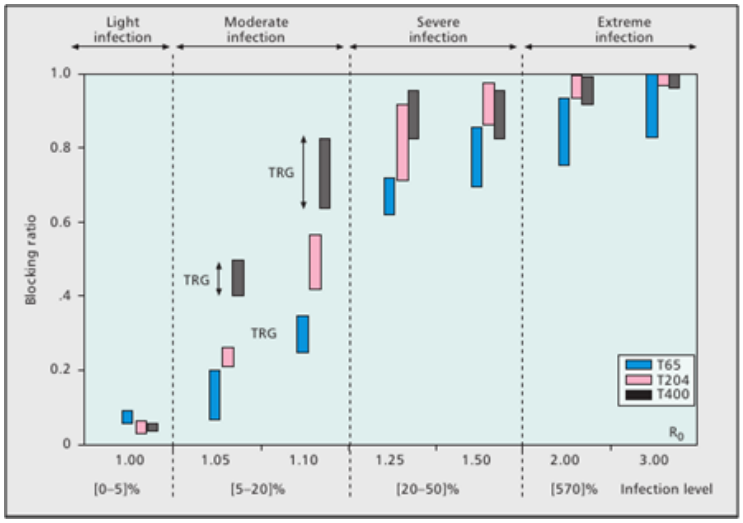


Figure 5.11: *Robustness comparison of the three studied topologies under different epidemic scenarios*

Contrary to the measure based on largest eigenvalue, with TRG it is also possible to observe that the topologies perform differently when the epidemic scenarios change. For instance, `t65` outperforms the other under light and extreme infections, but it is not the best in the rest of the cases. This behavior can be explained by the fact TRG considers not only the speed of the epidemic but also connections path lengths. Note that the probability of a connection being rejected is higher as its path length increases. As can be seen in Appendix B, `t65` has much shorter average path length than the others, and its largest eigenvalue is in between the other two.

5.6 Summary

In this chapter we have focused on propagating failures that affect basically nodes of GMPLS-controlled networks. Given the specific failure scenario defined, an epidemic-based approach was chosen for modeling the propagation. In that context, a new epidemic model was introduced, called SID.

This model served as the build block of a Continuous-time Markov chain defined to study the robustness of a GMPLS-controlled ring topology. Besides, a new measure of functional robustness tailored to path-oriented networks was introduced. It is called TRG, and we have shown through examples how this measure can be used to compare different topologies.

6 Conclusion and Future work

In this chapter, a summary of the main contributions of this work is given, together with possible directions for future research.

6.1 Conclusion

The aim of this thesis was to study the vulnerability of communications networks to large-scale failures, and to develop methods to measure and compare functional robustness. To that end, a new robustness metric was introduced that captures the peculiar features of transport networks regarding partial failures and uses service units (i.e., connections) as the reference point in the measurements. The main contributions of this work are summarized in the following paragraphs.

The SID model. A new epidemic-based failure propagation model was devised, and whose states characterize the different failure situations that a GMPLS-controlled node can experience when its functionality is divided between control plane and data plane. The dynamics of failure propagation in the control plane is approached from the perspective of epidemics, although the failures need not be related to computer viruses; instead, more plausible scenarios are those that can be created by software bugs or sophisticated attacks.

The TRG metric. The Topology Robustness in GMPS networks (TRG) measures how quickly a multiple failure event degrades the performance of the system in terms of its ability to accept connection requests. The underlying intuition is that if two topologies are subjected to failure events of equivalent intensity, the topology that degrades more gracefully (more slowly) is the more robust.

Multiple failures and availability analysis on ring topologies. We have used the SID model as a building block for deriving a Continuous-Time Markov Chain (CTMC) that characterizes the propagation of failures on ring

topologies. This CTMC model was subsequently used to produce guidelines for selecting repair rates so that a target network availability can be attained.

Link prioritization for limiting network functional damage. We have performed a simulation-based numeric evaluation of the number of affected LSPs in a multiple link failure scenario, and compared it with the average two-terminal reliability of the residual network, in order to illustrate that purely topological metrics are unable to capture the extent of the damage suffered at the service level. Then, two simple heuristic-based rules of link prioritization that can be used to improve connection survival to multiple failures were proposed.

Conceptual framework on resilience Additionally, in this thesis an extensive review of the terminology on resilience was presented, adapted to the needs and usage of the field of networking. The aim was to offer a coherent conceptual framework in order to avoid the confusing terminology that quite often appears in the networking literature concerning resilience.

6.2 Future work

There are several issues that have been left as future work throughout this thesis. We want to highlight the following:

- With respect to failure propagation, our work has focused basically on failures that can be modeled by epidemic dynamics. However, there are potentially other types of failures, for example targeted attacks, for which a different approach would be more appropriate.

- Although the topologies were carefully selected so that they correspond to different network models, little attention was paid to the effects that community structure might have on failure dynamics. For example, does a topology with strong community structure fare better than another that lacks that feature with regards to failure propagation?

- A formal validation of the SID model would be welcome, as well as empirical validation on larger topologies.

- In our simulations, we have always used a routing policy that dictates the use of capacity-constrained shortest paths. It would be interesting to explore what effect other routing policies have on functional robustness, assuming, for example a proactive routing approach towards the evolution of failures.

- For simplicity, it was assumed that connections are unprotected. However, scenarios in which the operator may offer different recovery guarantees depending on users' needs are conceivable. In such situations, interesting new problems arise, involving routing, recovery techniques and service differentiation.

Bibliography

[1] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling: Or, power-law degree distributions in regular graphs. *J. ACM*, 56:21:1–21:28, July 2009. ISSN 0004-5411.

[2] Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on World Wide Web*, pages 835–844. ACM, 2007.

[3] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 171–180, New York, NY, USA, 2000. ACM. ISBN 1-58113-184-4.

[4] M. Al-Kuwaiti, N. Kyriakopoulos, and S. Hussein. A comparative analysis of network dependability, fault-tolerance, reliability, security, and survivability. *Communications Surveys & Tutorials, IEEE*, 11(2): 106–124, 2009. ISSN 1553-877X.

[5] R. Albert and A. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[6] D. Alderson, L. Li, W. Willinger, and J. C. Doyle. Understanding internet topology: principles, models, and validation. *IEEE/ACM Trans. Netw.*, 13:1205–1218, December 2005. ISSN 1063-6692.

[7] A. Avižienis, J. Laprie, and B. Randell. Dependability and its threats: A taxonomy. In R. Jacquart, editor, *Building the Information Society*, volume 156 of *IFIP International Federation for Information Processing*, pages 91–120. Springer Boston, 2004. ISBN 978-1-4020-8156-9.

[8] A. Banerjee, J. Drake, J. Lang, B. Turner, K. Kompella, and Y. Rekhter. Generalized multiprotocol label switching: an overview of routing and management enhancements. *Communications Magazine, IEEE*, 39(1): 144–150, Jan. 2001. ISSN 0163-6804.

[9] A. Banerjee, L. Drake, L. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, and Y. Rekhter. Generalized multiprotocol label switching: an overview of signaling enhancements and recovery techniques. *Communications Magazine, IEEE*, 39(7):144–151, July 2001. ISSN 0163-6804.

[10] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[11] A. Barabási and E. Bonabeau. Scale-free networks. *Scientific American*, 288(5):50–9, 2003.

[12] A. Barrat and M. Weigt. On the properties of small-world network models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(3):547–560, 2000. ISSN 1434-6028.

[13] A. Barrat, M. Barthlemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008. ISBN 9780521879507.

[14] M. Barthélémy and L. A. N. Amaral. Small-world networks: Evidence for a crossover picture. *Phys. Rev. Lett.*, 82:3180–3183, Apr 1999.

[15] M. Barthélemy. Betweenness centrality in large complex networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):163–168, 2004. ISSN 1434-6028.

[16] B. Bassiri and S. Heydari. Network survivability in large-scale regional failure scenarios. In *Proceedings of the 2nd Canadian Conference on Computer Science and Software Engineering*, C3S2E '09, pages 83–87. ACM, 2009. ISBN 978-1-60558-401-0.

[17] E. Bender and E. Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, 1978. ISSN 0097-3165.

[18] L. Berger. Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description. RFC 3471 (Proposed Standard), Jan. 2003.

[19] A. Bigdeli, A. Tizghadam, and A. Leon-Garcia. Comparison of network criticality, algebraic connectivity, and other graph metrics. In *Proceedings of the 1st Annual Workshop on Simplifying Complex Network for Practitioners*, SIMPLEX '09, pages 4:1–4:6. ACM, 2009. ISBN 978-1-60558-704-2.

[20] T. Bilski. Disaster's Impact on Internet Performance - Case Study. In A. Kwiecień, P. Gaj, and P. Stera, editors, *Computer Networks*, volume 39 of *Communications in Computer and Information Science*,

pages 210–217. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-02671-3.

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5): 175–308, 2006. ISSN 0370-1573.

[22] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008. ISSN 0378-8733.

[23] T. Bu and D. Towsley. On distinguishing between Internet power law topology generators. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 638–647, 2002.

[24] E. Calle. *Enhanced fault recovery methods for protected traffic services in GMPLS networks*. PhD thesis, University of Girona, 2004.

[25] E. Calle, A. Urra, J. Marzo, G.-S. Kuo, and H.-B. Guo. Minimum interference routing with fast protection. *Communications Magazine, IEEE*, 44(10):104–111, oct. 2006. ISSN 0163-6804.

[26] E. Calle, J. Ripoll, J. Segovia, P. Vilà and, and M. Manzano. A multiple failure propagation model in GMPLS-based networks. *Network, IEEE*, 24(6):17–22, November-December 2010. ISSN 0890-8044.

[27] K. Calvert, M. Doar, and E. Zegura. Modeling internet topology. *Communications Magazine, IEEE*, 35(6):160–163, June 1997. ISSN 0163-6804.

[28] A. Capello, S. Milani, C. Moriondo, G. Rossi, P. Salamandra, M. Perrone, and M. Barone. Non-stop forwarding behaviour and performance in high-end IP routers for ISP's backbone networks. In *Design of Reliable Communication Networks, 2005. (DRCN 2005). Proceedings.5th International Workshop on*, pages 279–285, oct 2005.

[29] C. Castellano and R. Pastor-Satorras. Thresholds for epidemic spreading in networks. *Physical review letters*, 105(21):218701, 2010. ISSN 1079-7114.

[30] D. Cavendish. Evolution of optical transport technologies: from SONET/SDH to WDM. *Communications Magazine, IEEE*, 38(6): 164–172, 2000. ISSN 0163-6804.

[31] D. Cavendish and B. Sengupta. Routing and wavelength assignment in WDM rings with heterogeneous wavelength conversion capabilities. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1415–1424. IEEE, 2002.

[32] D. Chakrabarti, J. Leskovec, C. Faloutsos, S. Madden, C. Guestrin, and M. Faloutsos. Information survival threshold in sensor and P2P networks. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications*, pages 1316–1324. IEEE, 2007.

[33] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.*, 10(4):1–26, 2008. ISSN 1094-9224.

[34] F. Chen, Z. Chen, X. Wang, and Z. Yuan. The average path length of scale free networks. *Communications in Nonlinear Science and Numerical Simulation*, 13(7):1405–1410, 2008. ISSN 1007-5704.

[35] Q. Chen, H. Chang, R. Govindan, and S. Jamin. The origin of power laws in internet topologies revisited. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings*, volume 2, pages 608—617, 2002.

[36] Z. Chen, L. Gao, and K. Kwiat. Modeling the spread of active worms. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 3, pages 1890–1900. IEEE, 2003.

[37] P. Cholda and A. Jajszczyk. Recovery and its Quality in Multilayer Networks. *Lightwave Technology, Journal of*, 28(4):372–389, 2010. ISSN 0733-8724.

[38] P. Cholda, A. Mykkeltveit, B. Helvik, O. Wittner, and A. Jajszczyk. A survey of resilience differentiation frameworks in communication networks. *Communications Surveys Tutorials, IEEE*, 9(4):32–55, 2007. ISSN 1553-877X.

[39] F. R. K. Chung. Spectral Graph Theory. *CBMS Regional Conference Series in Mathematics*, 92:212, 1997. ISSN 0160-7642.

[40] M. Clouqueur and W. Grover. Availability analysis of span-restorable mesh networks. *Selected Areas in Communications, IEEE Journal on*, 20(4):810–821, 2002. ISSN 0733-8716.

[41] R. Cohen and S. Havlin. *Complex Networks: Structure, Robustness and Function*. Cambridge Univ Pr, 2010. ISBN 978-0-521-84156-6.

[42] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Physical Review Letters*, 85(21): 4626–4628, 2000.

[43] L. Costa, F. Rodrigues, G. Travieso, and P. Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007. ISSN 0001-8732.

[44] C. Crespelle and F. Tarissan. Evaluation of a new method for measuring the internet degree distribution: Simulation results. *Computer Communications*, 34(5):635–648, 2011. ISSN 0140-3664. Special Issue: Complex Networks.

[45] O. Crochat, J.-Y. Le Boudec, and O. Gerstel. Protection Interoperability for WDM Optical Networks. *IEEE/ACM Trans. Netw.*, 8(3): 384–395, 2000. ISSN 1063-6692.

[46] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda. Error and attack tolerance of complex networks. *Physica A: Statistical Mechanics and its Applications*, 340(1-3):388–394, 2004. ISSN 0378-4371.

[47] A. H. Dekker and B. Colbert. The symmetry ratio of a network. In *Proceedings of the 2005 Australasian symposium on Theory of computing*, volume 41 of *CATS '05*, pages 13–20. Australian Computer Society, Inc., 2005.

[48] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley Series in Mathematical & Computational Biology. Wiley, 1 edition, May 2000. ISBN 0471492418.

[49] R. Diestel. *Graph Theory (Graduate Texts in Mathematics)*. Springer, 4th edition, Feb. 2010. ISBN 3642142788.

[50] M. Doar. A better model for generating test networks. In *Global Telecommunications Conference, 1996. GLOBECOM '96. 'Communications: The Key to Global Prosperity*, pages 86–93, Nov 1996.

[51] J. Duch and A. Arenas. Effect of random failures on traffic in complex networks. In *Proc. SPIE*, volume 6601, page 66010O, 2007.

[52] M. Ellanti. *Next generation transport networks: data, management, and control planes*. Springer Verlag, 2005. ISBN 0387240675.

[53] P. Erdős and A. Rényi. On random graphs, I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[54] S. Erjongmanee, C. Ji, J. Stokely, and N. Hightower. Large-Scale Inference of Network-Service Disruption upon Natural Disasters. In M. Gaber, R. Vatsavai, O. Omitaomu, J. Gama, N. Chawla, and A. Ganguly, editors, *Knowledge Discovery from Sensor Data*, volume 5840 of *Lecture Notes in Computer Science*, pages 134–153. Springer Berlin / Heidelberg, 2010.

[55] A. Fabrikant, E. Koutsoupias, and C. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In P. Widmayer, S. Eidenbenz, F. Triguero, R. Morales, R. Conejo, and M. Hennessy, editors, *Automata, Languages and Programming*, volume 2380 of *Lecture Notes in Computer Science*, pages 781–781. Springer Berlin / Heidelberg, 2002. ISBN 978-3-540-43864-9.

[56] N. Falliere, L. Murchu, and E. Chien. W32. Stuxnet Dossier. *Symantec Security Response*, Nov. 2010.

[57] D. Fedyk, O. Aboul-Magd, D. Brungard, J. Lang, and D. Papadimitriou. A Transport Network View of the Link Management Protocol (LMP). RFC 4394 (Informational), Feb. 2006.

[58] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

[59] J. M. Finochietto, J. Aracil, Ángel Ferreiro, J. P. F.-P. Giménez, and Óscar González de Dios. Migration Strategies Toward All Optical Metropolitan Access Rings. *J. Lightwave Technol.*, 25(8):1918–1930, 2007. ISSN 0733-8724.

[60] L. C. Freeman. A set of measures of centrality based upon betweenness. *Sociometry*, 40(1):35–41, 1977. ISSN 0038-0431.

[61] E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959. ISSN 0003-4851.

[62] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. A Survey of Statistical Network Models. *Found. Trends Mach. Learn.*, 2: 129–233, February 2010. ISSN 1935-8237.

[63] W. Grover, J. Doucette, M. Clouqueur, D. Leung, and D. Stamatelakis. New options and insights for survivable transport networks. *Communications Magazine, IEEE*, 40(1):34–41, Jan. 2002. ISSN 0163-6804.

[64] J.-L. Guillaume, M. Latapy, and C. Magnien. Comparison of Failures and Attacks on Random and Scale-Free Networks. In T. Higashino, editor, *Principles of Distributed Systems*, volume 3544 of *Lecture Notes in Computer Science*, pages 186–196. Springer Berlin / Heidelberg, 2005. ISBN 978-3-540-27324-0.

[65] A. Haider and R. Harris. Recovery techniques in next generation networks. *Communications Surveys and Tutorials, IEEE*, 9(3):2–17, 2007. ISSN 1553-877X.

[66] S. He, S. Li, and H. Ma. Effect of edge removal on topological and functional robustness of complex networks. *Physica A: Statistical Mechanics and its Applications*, 388(11):2243–2253, 2009.

[67] B. Helvik. Perspectives on the dependability of networks and services. *Telektronikk*, 100(3):27–44, 2004. ISSN 0085-7130.

[68] J. Hernandez, T. Kleiberg, H. Wang, and P. Van Mieghem. A Comparison of Topology Generators with Power Law Behavior. In *Proceedings of SPECTS'2007*, pages 484–493, 2007.

[69] P. Hines, J. Apt, and S. Talukdar. Large blackouts in North America: Historical trends and policy implications. *Energy Policy*, 37(12):5249–5259, 2009. ISSN 0301-4215.

[70] T. Horie, G. Hasegawa, S. Kamei, and M. Murata. A new method of proactive recovery mechanism for large-scale network failures. In *Advanced Information Networking and Applications, International Conference on*, pages 951–958, Los Alamitos, CA, USA, 2009. IEEE Computer Society.

[71] S. Huang, M. Xia, C. Martel, and B. Mukherjee. A Multistate Multipath Provisioning Scheme for Differentiated Failures in Telecom Mesh Networks. *Lightwave Technology, Journal of*, 28(11):1585–1596, jun. 2010. ISSN 0733-8724.

[72] ITU-T Rec. E.800. Terms and definitions related to quality of service and network performance including dependability. ITU-T Rec. E.800, Aug. 1994.

[73] A. Jaekel, S. Bandyopadhyay, and Y. Aneja. Logical Topology Design for WDM Networks Using Survivable Routing. In *Communications, 2006. ICC '06. IEEE International Conference on*, volume 6, pages 2471–2476, jun. 2006.

[74] A. Jaiszczyk. Automatically switched optical networks: benefits and requirements. *Communications Magazine, IEEE*, 43(2):S10–S15, feb. 2005. ISSN 0163-6804.

[75] A. Jajszczyk and P. Rozycki. Recovery of the control plane after failures in ASON/GMPLS networks. *Network, IEEE*, 20(1):4–10, jan.-feb. 2006. ISSN 0890-8044.

[76] C. Jin, Q. Chen, and S. Jamin. Inet: Internet topology generator. Technical Report CSE-TR-433-00, EECS Department, University of Michigan, 2002.

[77] Y. Kitamura, Y. Lee, R. Sakiyama, and K. Okamura. Experience with restoration of Asia Pacific network failures from Taiwan earthquake. *IEICE Transactions*, 90-B(11):3095–3103, 2007. ISSN 0916-8516.

[78] J. Knight, E. Strunk, and K. Sullivan. Towards a rigorous definition of information system survivability. In *DARPA Information Survivability Conference and Exposition, 2003. Proceedings*, volume 1, pages 78–89. IEEE, 2003. ISBN 0769518974.

[79] O. Komolafe and J. Sventek. Impact of GMPLS Control Message Loss. *Lightwave Technology, Journal of*, 26(14):2029–2036, jul. 2008. ISSN 0733-8724.

[80] K. Kompella and Y. Rekhter. Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS). RFC 4202 (Proposed Standard), Oct. 2005.

[81] K. Kompella and Y. Rekhter. OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS). RFC 4203 (Proposed Standard), Oct. 2005.

[82] M. Kurant and P. Thiran. Survivable routing of mesh topologies in IP-over-WDM networks by recursive graph contraction. *Selected Areas in Communications, IEEE Journal on*, 25(5):922–933, 2007. ISSN 0733-8716.

[83] W. Lai and D. McDysan. Network Hierarchy and Multilayer Survivability. RFC 3386 (Informational), Nov. 2002.

[84] A. Lakhina, J. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 1, pages 332–341. IEEE, 2003.

[85] K. Lee and E. Modiano. Cross-Layer Survivability in WDM-Based Networks. In *INFOCOM 2009, IEEE*, pages 1017–1025, apr. 2009.

[86] M. Lesk. The new front line: Estonia under cyberassault. *Security & Privacy, IEEE*, 5(4):76–79, 2007. ISSN 1540-7993.

[87] T. Lewis. *Network Science: Theory and Applications*. Wiley Publishing, 2009. ISBN 0470331887.

[88] G. Li, J. Yates, D. Wang, and C. Kalmanek. Control plane design for reliable optical networks. *Communications Magazine, IEEE*, 40(2): 90–96, Feb. 2002. ISSN 0163-6804.

[89] L. Li, D. Alderson, J. Doyle, and W. Willinger. Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*, 2(4):431–523, 2005. ISSN 1542-7951.

[90] C. Liu and L. Ruan. A new survivable mapping problem in IP-over-WDM networks. *Selected Areas in Communications, IEEE Journal on*, 25(3):25–34, 2007. ISSN 0733-8716.

[91] S. Lumetta and M. Medard. Towards a deeper understanding of link restoration algorithms for mesh networks. In *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 1, pages 367–375. IEEE, 2002.

[92] S. Maesschalck, D. Colle, I. Lievens, M. Pickavet, P. Demeester, C. Mauz, M. Jaeger, R. Inkret, B. Mikac, and J. Derkacz. Pan-European optical transport networks: an availability-based comparison. *Photonic Network Communications*, 5(3):203–225, 2003. ISSN 1387-974X.

[93] C. Magnien, M. Latapy, and G. Jean-Loup. Impact of random failures and attacks on Poisson and power-law random networks. *ACM Comput. Surv.*, 43:13:1–13:31, April 2011. ISSN 0360-0300.

[94] P. Mahadevan, D. Krioukov, M. Fomenkov, X. Dimitropoulos, K. C. Claffy, and A. Vahdat. The internet AS-level topology: three data sources and one definitive metric. *SIGCOMM Comput. Commun. Rev.*, 36:17–26, January 2006. ISSN 0146-4833.

[95] E. Mannie. Generalized Multi-Protocol Label Switching (GMPLS) Architecture. RFC 3945 (Proposed Standard), Oct. 2004.

[96] E. Mannie and D. Papadimitriou. Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS). RFC 4427 (Informational), Mar. 2006.

[97] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C.-N. Chuah, Y. Ganjali, and C. Diot. Characterization of failures in an operational IP backbone network. *IEEE/ACM Trans. Netw.*, 16:749–762, August 2008. ISSN 1063-6692.

[98] A. Medina, A. Lakhina, I. Matta, and J. Byers. Brite: an approach to universal topology generation. In *Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 2001. Proceedings. Ninth International Symposium on*, pages 346–353, 2001.

[99] P. V. Mieghem. *Performance Analysis of Communications Networks and Systems*. Cambridge University Press, 2009. ISBN 9780521108737.

[100] W. Molisz and J. Rak. Impact of WDM network topology characteristics on the extent of failure losses. In *Transparent Optical Networks (ICTON), 2010 12th International Conference on*, pages 1–4. IEEE, 2010.

[101] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995. ISSN 1098-2418.

[102] A. Motter and Y. Lai. Cascade-based attacks on complex networks. *Physical Review E*, 66(6):65102, 2002. ISSN 1550-2376.

[103] A. Narula-Tam, E. Modiano, and A. Brzezinski. Physical topology design for survivable routing of logical rings in WDM-based networks. *Selected Areas in Communications, IEEE Journal on*, 22(8):1525–1538, 2004. ISSN 0733-8716.

[104] S. Neumayer and E. Modiano. Network Reliability with Geographically Correlated Failures. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, mar. 2010.

[105] M. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002. ISSN 1079-7114.

[106] M. Newman. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005. ISSN 0378-8733.

[107] M. Newman. *Handbook of Graphs and Networks*, chapter Random graphs as models of networks, pages 35–68. Wiley-VCH Verlag GmbH & Co. KGaA, 2005. ISBN 9783527602759.

[108] M. Newman. *Networks – An Introduction*. Oxford University Press, 2010. ISBN 9780199206650.

[109] M. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2):026113, Feb 2004.

[110] K. Nguyen, B. Jaumard, and A. Agarwal. A distributed and scalable routing table manager for the next generation of IP routers. *Network, IEEE*, 22(2):6–14, 2008. ISSN 0890-8044.

[111] G. O'Reilly, A. Jrad, R. Nagarajan, T. Brown, and S. Conrad. Critical Infrastructure Analysis of Telecom for Natural Disasters. In *Telecommunications, Network Strategy and Planning Symposium, 2006. NETWORKS 2006. 12th International*, pages 1–6, 2006.

[112] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski. SNDlib 1.0—survivable network design library. *Networks*, 55(3):276–286, 2010. ISSN 1097-0037.

[113] P. Pacharintanakul and D. Tipper. The effects of multi-layer traffic on the survivability of IP-over-WDM networks. In *ICC'09: Proceedings of the 2009 IEEE international conference on Communications*, pages 2354–2359, Piscataway, NJ, USA, 2009. IEEE Press. ISBN 978-1-4244-3434-3.

[114] P. Pan, G. Swallow, and A. Atlas. Fast Reroute Extensions to RSVP-TE for LSP Tunnels. RFC 4090 (Proposed Standard), May 2005.

[115] D. Papadimitriou and E. Mannie. Analysis of Generalized Multi-Protocol Label Switching (GMPLS)-based Recovery Mechanisms (including Protection and Restoration). RFC 4428 (Informational), Mar. 2006.

[116] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E*, 65(3):035108, Mar 2002.

[117] J. Perello, S. Spadaro, J. Comellas, and G. Junyent. An Analytical Study of Control Plane Failures Impact on GMPLS Ring Optical Networks. *Communications Letters, IEEE*, 11(8):695–697, Aug. 2007. ISSN 1089-7798.

[118] M. Pickavet, P. Demeester, D. Colle, D. Staessens, B. Puype, L. Depre, and I. Lievens. Recovery in multilayer optical networks. *Lightwave Technology, Journal of*, 24(1):122–134, jan. 2006. ISSN 0733-8724.

[119] B. Rajagopalan, J. Luciani, and D. Awduche. IP over Optical Networks: A Framework. RFC 3717 (Informational), Mar. 2004.

[120] R. Ramaswami, K. Sivarajan, and G. Sasaki. *Optical networks: a practical perspective*. Morgan Kaufmann Pub, 2009. ISBN 0123740924.

[121] G. Rouskas and H. Perros. A tutorial on optical networks. In E. Gregori, G. Anastasi, and S. Basagni, editors, *Advanced Lectures on Networking*, volume 2497 of *Lecture Notes in Computer Science*, pages 496–500. Springer Berlin / Heidelberg, 2002. ISBN 978-3-540-00165-2.

[122] M. Ruiz, J. Perello, L. Velasco, S. Spadaro, J. Comellas, and G. Junyent. Gmpls control plane network design with resilience guarantees. *Lightwave Technology, Journal of*, 29(1):37–47, jan.1, 2011. ISSN 0733-8724.

[123] J. Segovia, P. Vilà, E. Calle, and J. Marzo. Improving the resilience of Transport Networks to large-scale failures. *Journal of Networks*, Acepted for publication. ISSN 1796-2056.

[124] J. Segovia, E. Calle, P. Vilà, J. Marzo, and J. Tapolcai. Topology-focused availability analysis of basic protection schemes in optical transport networks. *Journal of Optical Networking*, 7(4):351–364, 2008. ISSN 1536-5379.

[125] J. Segovia, E. Calle, and P. Vilà. An improved method for discovering link criticality in transport networks. In *Broadband Communications, Networks, and Systems, 2009. BROADNETS 2009. Sixth International Conference on*, pages 1–8, sept. 2009.

[126] J. Segovia, E. Calle, P. Vilà, and J. Marzo. A heuristic analysis of resilience to multiple failures in GMPLS networks. In *Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2010 International Symposium on*, pages 258–264. IEEE, 2010.

[127] I. Seoane, E. Calle, J. Hernández, J. Segovia, R. Romeral, P. Vilà, M. Urueña, and M. Manzano. Failure propagation in GMPLS optical rings: CTMC model and performance analysis. *Optical Switching and Networking*, In Press:–, 2011. ISSN 1573-4277.

[128] M. Sivakumar, R. K. Shenai, and K. M. Sivalingam. A survey of survivability techniques for optical wdm networks. In K. M. Sivalingam and S. Subramaniam, editors, *Emerging Optical Network Technologies*, pages 297–331. Springer US, 2005. ISBN 978-0-387-22584-5.

[129] J. P. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith. Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines. *Computer Networks*, 54(8):1245–1265, 2010. ISSN 1389-1286.

[130] J. P. Sterbenz, E. K. Çetinkaya, M. A. Hameed, A. Jabbar, Q. Shi, and J. P. Rohrer. Evaluation of network resilience, survivability, and disruption tolerance: Analysis, topology generation, simulation, and experimentation (invited paper). *Springer Telecommunication Systems*, 2011. accepted March 2011.

[131] A. Sydney, C. Scoglio, P. Schumm, and R. E. Kooij. Elasticity: topological characterization of robustness in complex networks. In *Proceedings of the 3rd International Conference on Bio-Inspired Models of Network, Information and Computing Sytems*, BIONETICS '08, pages 19:1–19:8, 2008.

[132] A. Sydney, C. Scoglio, M. Youssef, and P. Schumm. Characterizing the Robustness of Complex Networks. *Int. J. Internet Technology and Secured Transactions*, 2(3/4):291–320, 2010.

[133] A. Tanenbaum. *Computer Networks*. Prentice Hall Professional Technical Reference, 4th edition, 2002. ISBN 0130661023.

[134] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: degree-based vs. structural. *SIGCOMM Comput. Commun. Rev.*, 32:147–159, August 2002. ISSN 0146-4833.

[135] K. Thulasiraman, T. Lin, M. Javed, and G. Xue. Logical topology augmentation for guaranteed survivability under multiple failures in IP-over-WDM optical networks. *Optical Switching and Networking*, 7 (4):206–214, 2010. ISSN 1573-4277.

[136] A. Tizghadam and A. Leon-Garcia. Autonomic traffic engineering for network robustness. *Selected Areas in Communications, IEEE Journal on*, 28(1):39–50, Jan. 2010. ISSN 0733-8716.

[137] A. Tizghadam and A. Leon-Garcia. Betweenness centrality and resistance distance in communication networks. *Network, IEEE*, 24(6): 10–16, November-December 2010. ISSN 0890-8044.

[138] J. Tomasik and M.-A. Weisser. Internet topology on AS-level: model, generation methods and tool. In *29th IEEE International Performance Computing and Communications Conference (IPCCC'10)*, pages 1–8, Albuquerque, NM, USA, December 2010.

[139] A. Urra. *Multi-layer survivability: routing schemes for GMPLS-based networks*. PhD thesis, Universitat de Girona, 2006.

[140] P. Van Mieghem and F. Kuipers. Concepts of exact QoS routing algorithms. *Networking, IEEE/ACM Transactions on*, 12(5):851–864, oct 2004. ISSN 1063-6692.

[141] P. Van Mieghem and H. Wang. The observable part of a network. *Networking, IEEE/ACM Transactions on*, 17(1):93–105, 2009. ISSN 1063-6692.

[142] J. Vasseur, M. Pickavet, and P. Demeester. *Network recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*. Morgan Kaufmann Publishers, 2004. ISBN 012715051X.

[143] H. Wang, J. Hernandez, and P. Van Mieghem. Betweenness centrality in a weighted network. *Physical Review E*, 77(4):46105, 2008. ISSN 1550-2376.

[144] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998. ISSN 0028-0836.

[145] B. M. Waxman. Routing of multipoint connections. *Selected Areas in Communications, IEEE Journal on*, 6(9):1617–1622, 1988. ISSN 0733-8716.

[146] L. Xie, P. Smith, M. Banfield, H. Leopold, J. Sterbenz, and D. Hutchison. Towards resilient networks using programmable networking technologies. In D. Hutchison, S. Denazis, L. Lefevre, and G. Minden, editors, *Active and Programmable Networks*, volume 4388 of *Lecture Notes in Computer Science*, pages 83–95. Springer Berlin / Heidelberg, 2009. ISBN 978-3-642-00971-6.

[147] T. Yellman. Redundancy in designs. *Risk Analysis*, 26(1):277–286, 2006. ISSN 1539-6924.

[148] M. Youssef, R. Kooij, and C. Scoglio. Viral conductance: Quantifying the robustness of networks with respect to spread of epidemics. *Journal of Computational Science*, 2011. ISSN 1877-7503. (in press).

[149] G. U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, F.R.S. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 213:21–87, 1925.

[150] E. W. Zegura, K. L. Calvert, and M. J. Donahoo. A quantitative comparison of graph-based models for internet topology. *IEEE/ACM Trans. Netw.*, 5:770–783, December 1997. ISSN 1063-6692.

[151] J. Zhang and B. Mukherjee. A review of fault management in WDM mesh networks: basic concepts and research challenges. *Network, IEEE*, 18(2):41–48, 2004. ISSN 0890-8044.

[152] C. Zou, W. Gong, and D. Towsley. Code red worm propagation modeling and analysis. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 138–147. ACM, 2002. ISBN 1-58113-612-9.

A

Publications and Projects

Publications

Journals and books

- I. Seoane, E. Calle, J. Hernández, **J. Segovia**, R. Romeral, P. Vilà, M. Urueña, and M. Manzano. Failure propagation in GMPLS optical rings: CTMC model and performance analysis. *Optical Switching and Networking*, In Press, 2011. ISSN 1573-4277.

- **J. Segovia**, P. Vilà, E. Calle and J. Marzo. Improving the resilience of Transport Networks to large-scale failures, Journal of Networks, Accepted for publication, ISSN 1796-2056.

- E. Calle, J. Ripoll, **J. Segovia**, P. Vilà and M. Manzano. A multiple failure propagation model in GMPLS-based networks. *Network, IEEE* 24(6):17-22, November-December 2010. ISSN 0890-8044

- **J. Segovia**, E. Calle, P. Vilà, J. Marzo, and J. Tapolcai. Topology-focused availability analysis of basic protection schemes in Optical Transport Networks. *Journal of Optical Networking*, 7(4):351-364, 2008. ISSN 1536-5379.

- J. Marzo, T. Stidsen, S. Ruepp, E. Calle, J. Tapolcai, **J. Segovia**. Network survivability: End-to-end recovery using local failure information, chapter in *Graphs and Algorithms in Communication Networks*, Koster, Arie M.C.A., Muñoz, Xavier (Eds.), pp. 137-160, Springer, December 2009.

Conferences

- **J. Segovia**, E. Calle, P. Vilà, and J. Marzo. A heuristic analysis of resilience to multiple failures in GMPLS networks. In *Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2010 International Symposium on*, pages 258-264. IEEE, 2010.

- **J. Segovia**, E. Calle, P. Vilà and Y. Donoso. Protection to Multiple Failures in GMPLS Networks: optimization and heuristics, in *Proc. II Workshop de Redes Multinivel*, June 2010.

- **J. Segovia**, J. Marzo, E. Calle, and P. Vilà. Robustness Analysis to Multiple Failures in GMPLS Networks, in *Proc. International Conference on transparent optical networks (ICTON)*, June 2010 (invited).

- **J. Segovia**, E. Calle, and P. Vilà. An improved method for discovering link criticality in transport networks. In *Broadband Communications, Networks, and Systems, 2009. BROADNETS 2009. Sixth International Conference on*, pages 1–8, sept. 2009.

- **J. Segovia**, E. Calle and P. Vilà. New applications of the betweenness centrality concept to reliability-driven routing, in *Proc. VIII Workshop in G/MPLS networks*, June 2009.

- **J. Segovia**, E. Calle and P. Vilà. Availability Analysis of GMPLS Connections based on Physical Network Topology, in *Proc. VII Workshop in G/MPLS Networks/IEEE International Conference on Optical Network Design and Modeling*, March 2008.

Projects

During my time as PhD student, I have collaborated in the following projects:

- **Project name**: Multilevel mechanisms for the allocation of resources and routing with recovery in backbone networks based on GMPLS (M2R3).
  Funding entity: Ministerio de Educación y Ciencia y (MEC)
  Reference: TEC2006-03883/TCM
  Funding: 90750 €
  Dates: 1/12/2006 to 30/11/2009
  Main researcher: Dr. José Luis Marzo

- **Project name**: Thematic network on virtual-circuit-oriented IP networks management (MPLS)
  Funding entity: Ministerio de Educación y Ciencia y (MEC)
  Reference: TEC2006-27633-E
  Funding: 10000 €
  Dates: 01/2007 to 12/2007
  Main researcher: Dr. José Luis Marzo

- **Project name**: Graphs and algorithms in communication networks (GRAAL)
  Funding entity: European Commission (COST)
  Reference: COST 293
  Funding: 320000 €(whole consortium)
  Dates: 11/2004 to 11/2008
  Main researcher: Dr. A. Koster (U. Warwick), secretary & grant holder Dr. J.L. Marzo (UdG)

- **Project name**: Transparent and reliable interdomain overlay network (TRION)
  Funding entity: Ministerio de Ciencia e Innovación (MICINN)
  Reference: TEC2009-10724
  Funding: 44649 €
  Dates: 01/01/2010 to 31/12/2012
  Main researcher: Dr. José Luis Marzo

- **Project name**: Experimental UpdateLess Evolutive Routing (EULER), http://www.euler-fire-project.eu/
  Funding Entity: European Comission FP7 STREP (Challenge 1 "Technologies and systems architectures for the Future Internet").
  Reference: No.258307
  Funding: 3.15 M€
  Dates: 10/2010 to 09/2013
  Main researcher: Dimitri Papadimitriou (Alcatel-Lucent Bell, Belgium)

Appendix B
Topologies

The properties of the topologies used in this dissertation are shown in Table B.1. The topology named t65 is from http://sndlib.zib.de/ [112], available under the name "ta2". Topology cost266x6 is the result of the juxtaposition of several near identical copies of the reference topology "Cost266", also available from the SNDlib web site.

The remaining topologies are synthetic, as follows:

- er400d3 and er400d6 are random (ER), generated through the Python interface to the iGraph library.
- eba400h is power-law (Barabási-Albert), generated with NetworkX.
- bt400 and t204 exhibit community structure and were generated with the help of BRITE using a two-layer configuration.

See Chapter 4 for more details on the aforementioned topology generators.

A visual representation of each topology follows, together with the corresponding nodal degree distribution.

Table B.1: *Properties of the topologies used in this dissertation.*

Property	cost266x6	bt400	t204	er400d3	er400d6	eba400h	t65
Number of nodes N	222	400	204	400	400	400	65
Number of undirected links	371	749	327	618	1205	609	108
Average node degree $\langle k \rangle$	3.34	3.75	3.21	3.09	6.03	3.05	3.32
Diameter	20	19	18	12	7	11	11
Average minimum path length $\langle \ell \rangle$	8.42	9.06	7.82	5.48	3.56	4.61	3.91
Clustering coefficient $\bar{C}$	0.00	0.17	0.17	0.20	0.05	0.44	0.25
Largest eigenvalue of A λ	3.5657	5.1951	3.8028	4.1158	7.1677	6.9956	4.7891
Assortativity coefficient r	-0.043	-0.296	-0.234	-0.144	0.000	-0.069	0.040

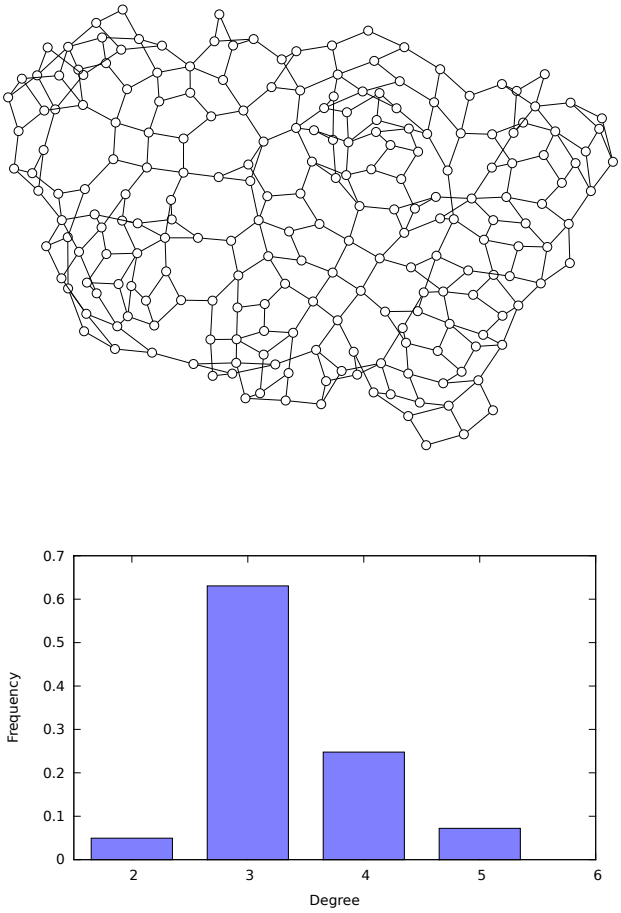


Figure B.1: *The cost266x6 topology*

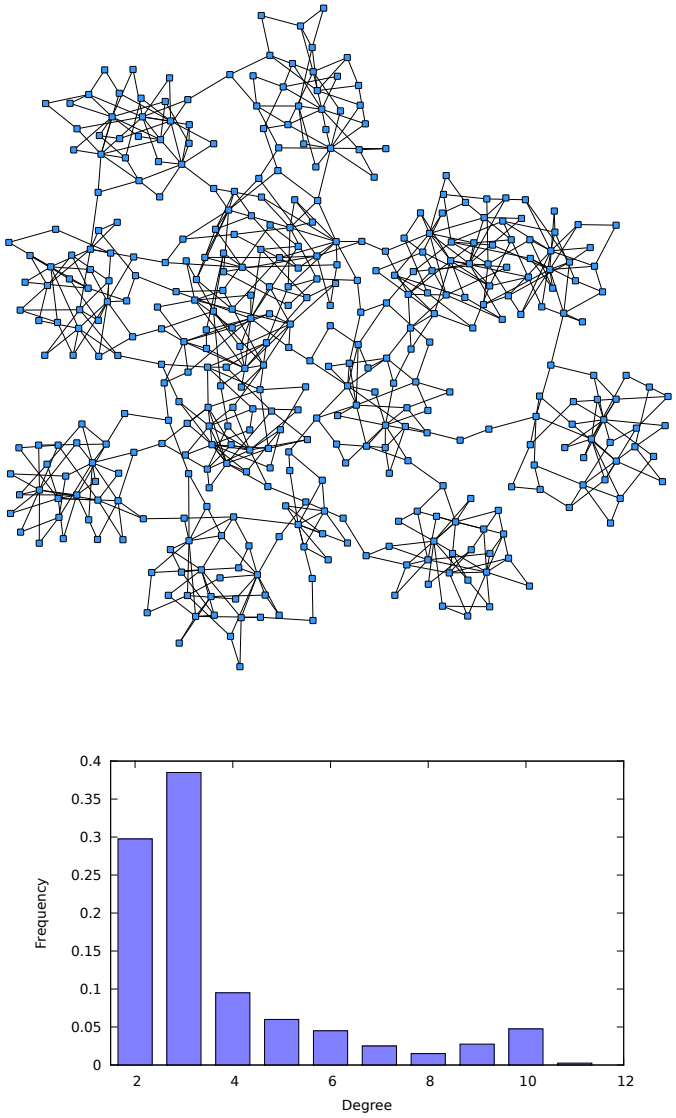


Figure B.2: *The bt400 topology*

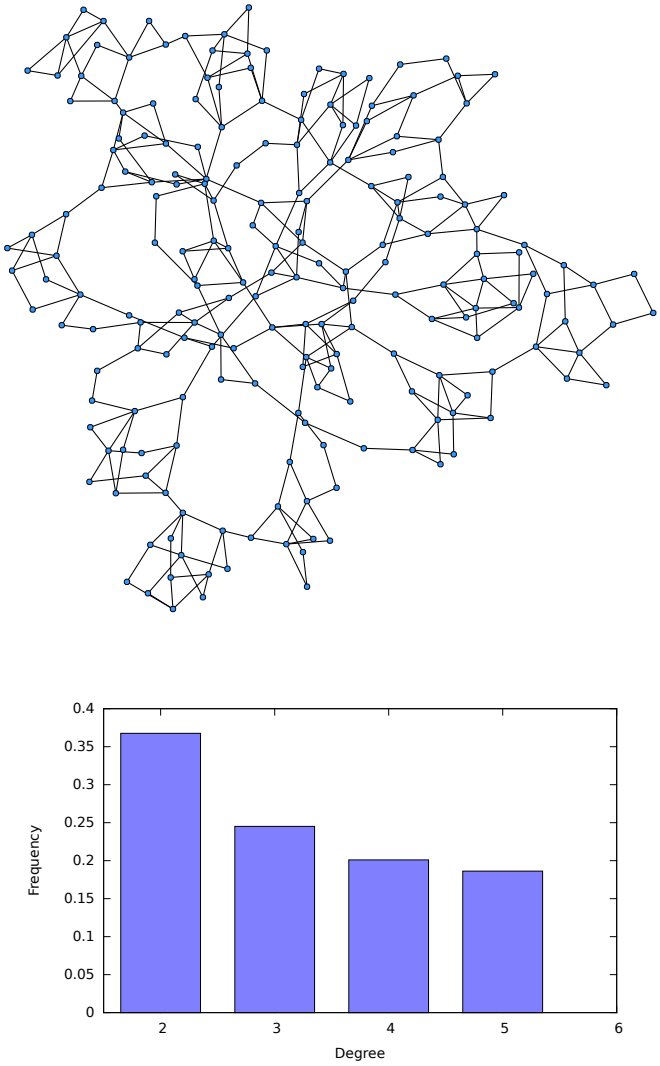


Figure B.3: *The t204 topology*

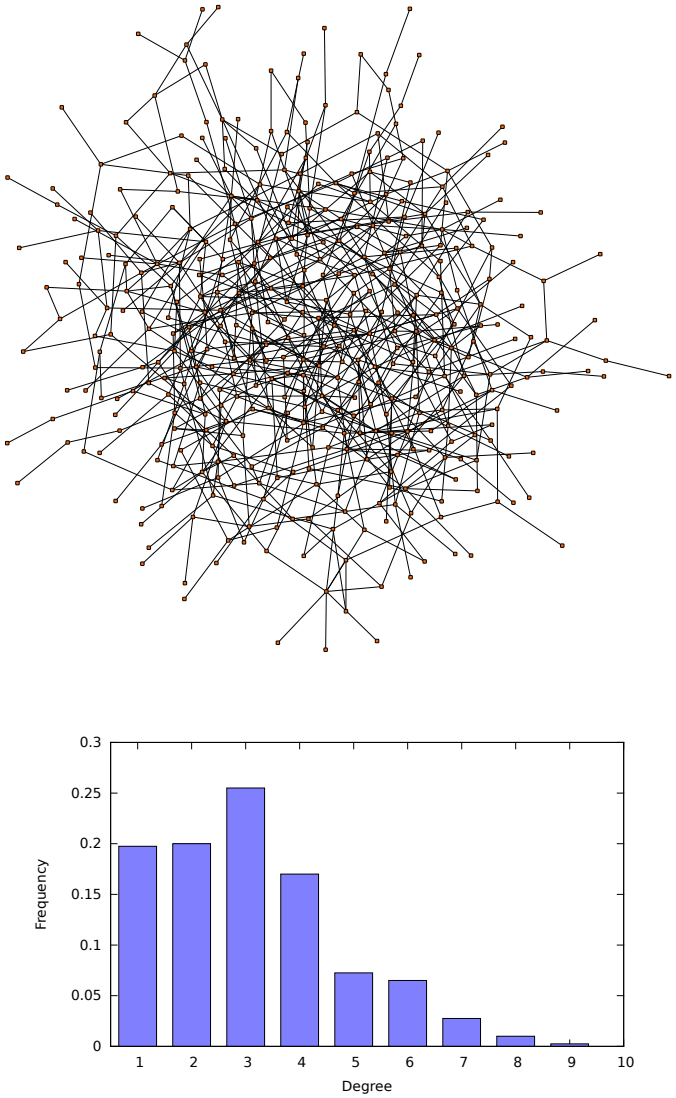


Figure B.4: *The er400d3 topology*

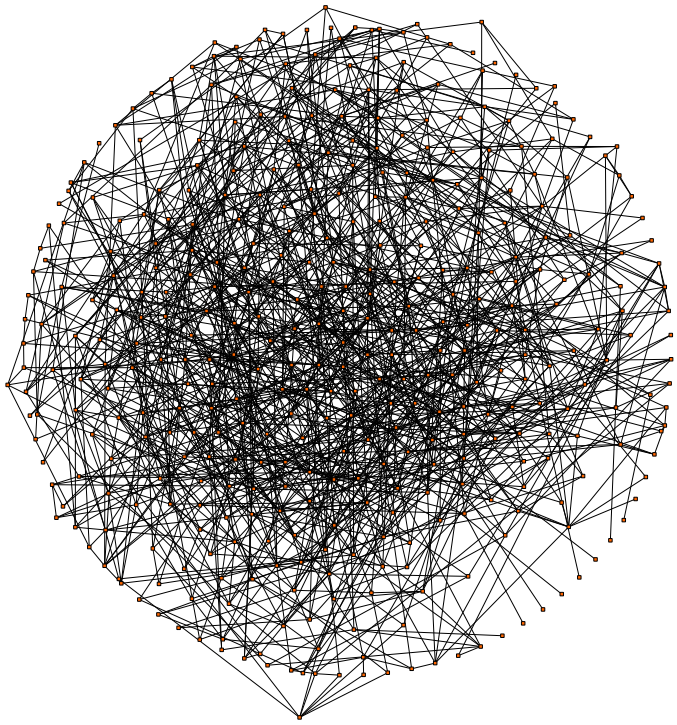


Figure B.5: *The er400d6 topology*

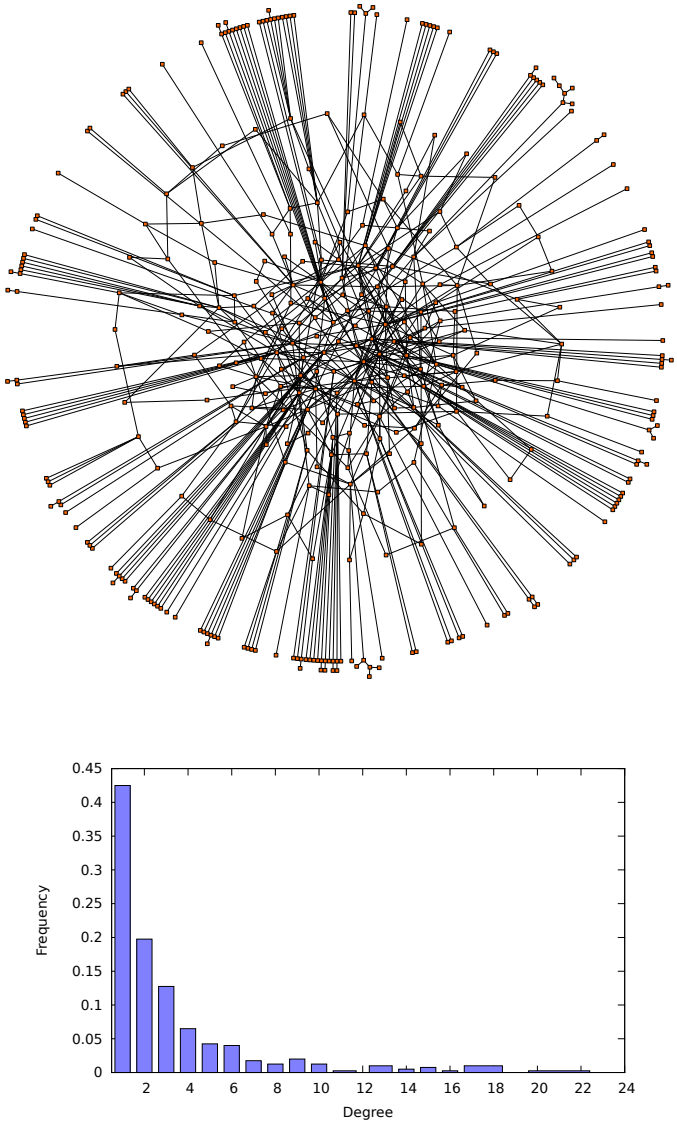


Figure B.6: *The eba400h topology*

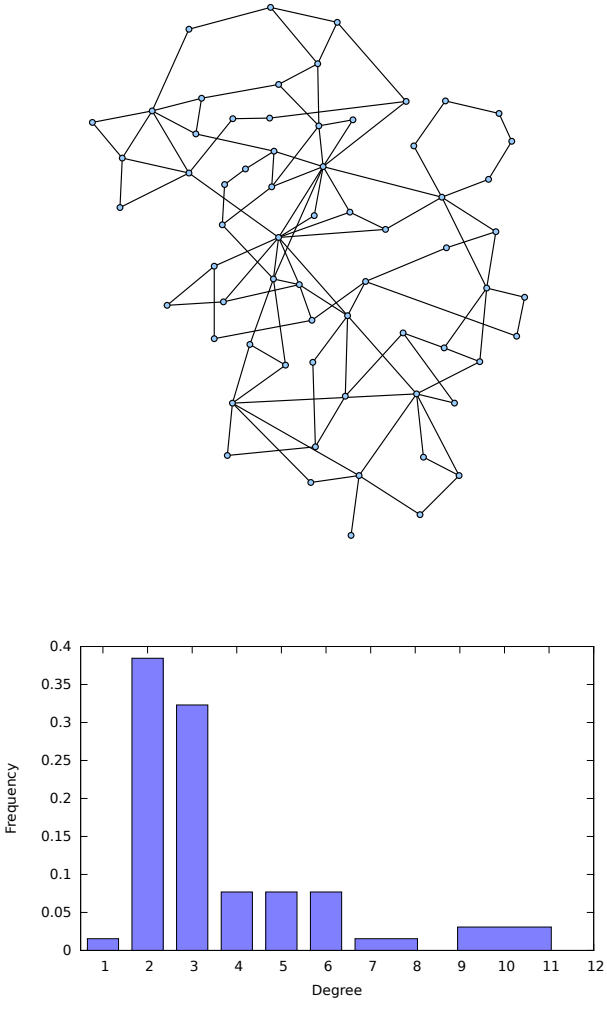


Figure B.7: *The t65 topology*

Index