Julius-Maximilians-Universität Würzburg
Institut für Informatik
Lehrstuhl für Kommunikationsnetze
Prof. Dr. P. Tran-Gia

Heuristic Design and Provisioning of Resilient Multi-Layer Networks

Michael Duelli

Würzburger Beiträge zur
Leistungsbewertung Verteilter Systeme

Bericht 02/12

Würzburger Beiträge zur Leistungsbewertung Verteilter Systeme

Herausgeber

Prof. Dr. P. Tran-Gia
Universität Würzburg
Institut für Informatik
Lehrstuhl für Kommunikationsnetze
Am Hubland
D-97074 Würzburg

Tel.: +49-931-31-86630
Fax.:+49-931-31-86632
email: trangia@informatik.uni-wuerzburg.de

Satz

Reproduktionsfähige Vorlage vom Autor.
Gesetzt in LaTeX Computer Modern 9pt.

ISSN 1432-8801

Heuristic Design and Provisioning of Resilient Multi-Layer Networks

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius–Maximilians–Universität Würzburg

vorgelegt von

Michael Duelli

aus

Schweinfurt

Würzburg 2012

Eingereicht am: 13.12.2011
bei der Fakultät für Mathematik und Informatik
1. Gutachter: Prof. Dr.-Ing. P. Tran-Gia
2. Gutachter: Prof. Dr.-Ing. J. Eberspächer
Tag der mündlichen Prüfung: 07.02.2012

Danksagung

Eine Promotion kann man getrost mit der Aufzucht einer Pflanze vergleichen. Es reicht dabei nicht, williges Saatgut in fruchtbaren Boden zu setzen. Darüber hinaus bedarf es der Wahrung der richtigen Umgebung sowie der richtigen Menge an Wasser und Licht. Ich hatte das Glück eine solche geeignete Umgebung zu erhalten, sodass meine Forschung in den letzten fünf Jahren zu einem starken Baum heranwachsen konnte, aus dem die hier vorliegenden Seiten gefertigt sind.

Mein Dank hierfür gebührt in erster Linie meinem Doktorvater Prof. Dr.-Ing. Phuoc Tran-Gia, der für mich diese kontinuierlichen Voraussetzungen schaffte und es mir ermöglichte, mich selbstständig zu entwickeln sowie an Herausforderungen zu wachsen. Darüber hinaus möchte ich mich bei meinem Zweitgutachter Prof. Dr.-Ing. Jörg Eberspächer für seine Kompetenz und seine unglaubliche Reaktionsgeschwindigkeit bedanken. Ferner gilt mein Dank meinen beiden weiteren Prüfern der Disputation, Prof. Dr. rer. pol. Andreas Hotho und Prof. Dr.-Ing. Sergio Montenegro, die mich gekonnt auf die Probe stellten.

Herzlich bedanken möchte ich mich bei Gisela Förster, für ihre gewissenhafte Mithilfe in allen Bereichen des Lehrstuhls, sowie bei Susann Schmitt, für die unkomplizierte Abwicklung aller Prüfungsangelegenheiten. Für die gute Zusammenarbeit in den Projekten 100GET, COMCON und VNREAL möchte ich allen Projektpartnern aus Industrie und Forschung danken. Diese Projekte bildeten die Wurzeln meiner Promotion und ließen mich reichlich Erfahrung sammeln.

Ein einzelner Baum kann nur schwer bestehen, daher ist es wichtig andere Gleichgesinnte um sich zu haben. Besonderer Dank gebührt unseren Gruppenleitern Tobias Hoßfeld, Michael Menth, Rastin Pries und Dirk Staehle, die mir bereitwillig mit einem offenen Ohr und mit Rat zur Seite standen. Besonders möchte ich mich auch bei den Lektoren meiner Dissertation, Rastin Pries und Robert Henjes sowie Simon Oechsner, für die Opferung ihrer Freizeit und ihre wertvollen Kommentare bedanken. Während meiner Zeit am Lehrstuhl durfte ich nicht nur mit Andreas Binzenhöfer, Valentin Burger, Steffen Gebert, Matthias Hartmann, Matthias Hirth, David Hock, Robert Henjes, Tobias Hoßfeld, Michael Jarschel, Dominik Klein, Frank Lehrieder, Andreas Mäder, Rüdiger Martin, Michael Menth, Simon Oechsner, Rastin Pries, Daniel Schlosser, Christian Schwartz, Barbara Staehle, Dirk Staehle, Kurt Tutschku, Florian Wamser und Thomas Zinner zusammenarbeiten, sondern habe sie auch jenseits des Kaffeeraums als meine Freunde lieb gewonnen. Sie waren essentiell für die gute Zusammenarbeit und die angenehme Atmosphäre am Lehrstuhl.

Die Betreuung von Studenten in Vorlesungen und das gemeinsame Entwickeln und Ausarbeiten von Ideen in ihren Abschlussarbeiten hat mir immer viel Freude bereitet. Besonders möchte ich mich für die gute Zusammenarbeit bei Kunxiang Gao, Anke Endler und Julian Ott sowie bei meinen Diplomanden Christopher Pluntke, Eduard Weber, Xiaohua Qin und Sebastian Goll bedanken.

Zu guter Letzt, möchte ich mich bei meinem Vater Martin Duelli und meiner Schwester Bianca Duelli bedanken, die immer für mich da waren, und insbesondere danke ich meiner Partnerin Stefanie Schlüsener, die meine Ablenkung durch die Arbeit am Lehrstuhl und meine Dissertation am deutlichsten zu spüren bekam und mich stets unterstützt hat.

Contents

1 Introduction

I'm a musician of sorts, and on my way to give a very special performance.
"V for Vendetta", 2006.

Accessing the Internet is as simple as plugging a cable to a given network outlet and recently gained popularity in the wireless domain by the use of mobile smart phones. There is a continuous increase in the provided capacity [39] and the consumed bandwidth [40, 41]. This race of supply and demand triggers the development of networking equipment supporting higher data rates as well as new or enhanced technologies. The complexity and structure of the *core networks*, which enable this high-speed data exchange, is hidden from the users. New and enhanced technologies are introduced to improve the utilization of the limited physical resources, e.g., by aggregation, and need to be integrated in existing network structures. This technological *evolution* and the composition of multiple technologies, like *Internet Protocol* (IP) routers and Ethernet switches, lead to *multi-layer networks* where each technology forms an individual, logical layer upon a physical topology usually consisting of fiber strands.

New equipment providing higher data rates usually comes along with non-proportionally lower cost per bit as can be concluded from the ever decreasing market prices for telecommunication in the last decades. Hence, network providers suffer from a constant cost pressure enforcing a network *design* that is aware of *Capital Expenditures* (CAPEX), i.e., the cost for the initial acquisition of equipment, and *Operational Expenditures* (OPEX), i.e., the cost for the operation and maintenance of equipment. The variety of technologies and components in a multi-layer network leads to an immense number of possible combinations and combinatorial complexity. Therein, the deployment of the right components at the right place is decisive for the economic success. Efficient approaches are needed to perform parameter studies and to consider scenarios of realistic size.

Moreover, the high data rates in core networks imply an urgent need for *protection* mechanisms since even short outages may have severe consequences with regard to *Expected Loss of Traffic* (ELT). Due to the hierarchical structure of the considered multi-layer networks, failures in one layer often cause subsequent failures in others, leading to so called *Shared Risk Groups* (SRGs), and, thus, increase the need for *resilience*. Therefore, resilience constraints are laid down in *Service Level Agreements* (SLAs) postulating amongst other characteristics the availability of a service or a network element and implying penalty fees in case of SLA violations. These resilience requirements need to be reflected in the design of multi-layer networks and increase the complexity of this task due to trade-offs between protection and cost-efficiency.

After the cost-aware design of a network in the construction stage, a network operator is interested in the efficient utilization of the installed resources in its multi-layer network. Since future traffic demands are expected to become more and more *dynamic*, today's mostly manual mechanisms to set up paths are rather insufficient to keep up with the upcoming requirements. Instead, these resources need to be *provisioned* in a fast and efficient way. The key to solve this problem is *automation* of the provisioning. This can be realized by standard-compliant implementations of *Generalized Multi-Protocol Label Switching* (GMPLS) for path signaling and set-up in combination with *Path Computation Elements* (PCEs) that calculate the paths under given constraints within seconds.

1.1 Scientific Contribution

This monograph covers three aspects of multi-layer networks. First, heuristic design principles for multi-layer networks based on a component-rich equipment and CAPEX model are presented. Second, the requirements of resilient multi-layer network design are integrated in the aforementioned approach. Third, an architecture for the efficient provisioning on a previously designed multi-layer network with dynamic traffic demands is investigated.

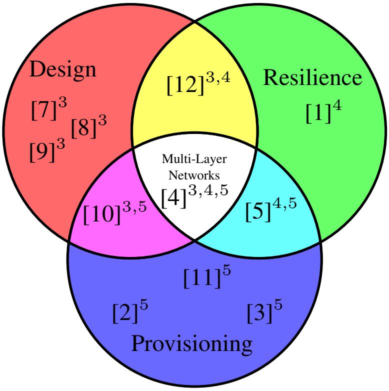


Figure 1.1: *Contribution of this work illustrated as a classification of the research studies conducted by the author. The notation $[x]^y$ indicates that the scientific publication $[x]$ is related to Chapter y of this monograph.*

Figure 1.1 gives an overview of the contribution of this work. The figure shows the three main topics of this work – design, resilience, and provisioning – which are all understood in the context of multi-layer networks and conceptually overlap to a certain extent. In addition, the figure comprises annotations of the form $[x]^y$, which denote that scientific publication $[x]$ contributes to Chapter y.

The first part of this work covers the design of multi-layer networks using a detailed CAPEX model that comprises different technologies and component types. The resulting multi-layer networks need to fulfill given traffic requirements while being CAPEX-optimal. We present and apply models to develop efficient, deterministic heuristics to deal with these requirements, the variety of input parameters, and optimization objectives. Our heuristics prove to be a useful mean to evaluate the roll-out of new technologies and to obtain valuable construction blueprints for multi-layer networks. Our findings show to what extent restrictions to certain architectures increase the overall cost. An analysis of how the heuristics themselves impact the achieved results manifests the importance of the sequential demand order of the considered heuristics.

The data rates provided in core networks and the structure implied by multi-layer networks favor the occurrence of *Shared Risk Groups* (SRGs) that mean fatal consequences even in case of a single failure. To prevent *Expected Loss of Traffic* (ELT) and to avoid penalty fees, the introduced design strategies for multi-layer networks are extended to accommodate resilience constraints and the establishment of additional backup paths. As certain failures cannot be covered by single-layer resilience, we consider protection on single as well as multiple layers. For single-layer resilience, we show that protection at different layers in combination with different protection variants has a strong impact on the performance of the design heuristics. For multi-layer resilience, we extend our study to simultaneous protection at multiple layers with different approaches and show how the immense requirements for resources and the underlying topology constrain the performance of the presented algorithms.

Beyond construction, the efficient operation of a multi-layer network is decisive for a general cost-optimality. The increasing amount of spontaneous requests for short-lived connections results in dynamic traffic patterns that network providers need to deal with. To that end, we investigate automated provisioning by means of the GMPLS protocol family and the recent PCE architecture. In particular, algorithms are developed that calculate paths within a PCE from given network constraints. We show the relevance of delayed information retrieval for PCE-based path computation with short-lived traffic demands. Furthermore, the importance of an appropriate resource utilization strategy is outlined. In addition, we find that blocking can be noticeably improved by re-optimization algorithms that require knowledge about former routing decisions. Finally, the benefit of giving feedback from operation to construction is outlined.

In sum, we address all stages of a network life-cycle within a combined optimization approach including resilience. Our heuristics allow to evaluate networks of realistic size, which is impossible with plain combinatorial approaches, like *Integer Linear Programs* (ILPs). Moreover, our approach allows to perform parameter studies to evaluate which options network providers should take. For this purpose, a software tool was developed to run the theoretical studies presented in this monograph. The software is called *Multi-Layer Network Engineering and Optimization* (MuLaNEO) [4] and is made publicly available [16].

1.2 Outline

The organization of this monograph is depicted in Figure 1.2. It shows the individual chapters together with their interdependencies and logical interrelations. After giving a broader background, a chapter is separately dedicated to the design, the resilience, and the provisioning of multi-layer networks of which each reviews the lessons learned. The remainder of this thesis is organized as follows.

In Chapter 2, the considered optimization problems are motivated by examples taken from real life and background knowledge is given that is required to understand these problems, their various parameters, and degrees of freedom. Furthermore, an overview of related work with regard to the design, resilience, and provisioning of multi-layer networks is presented. Models are developed to reflect traffic, network components, and CAPEX. The principles and models defined in this chapter are used throughout this thesis.

Chapter 3 introduces deterministic heuristics for the design of multi-layer networks from scratch using the previously defined models. Results of the heuristics are evaluated in detail on four physical topologies with a component-rich CAPEX model. On the one hand, the impact of parameters and properties of the heuristics themselves on overall cost are analyzed as well as the distribution of cost on components. On the other hand, the impact of the CAPEX model is investigated with a special focus on network architectures and cost variations.

The concepts of multi-layer network design are extended to include resilience in Chapter 4. We introduce several variants of heuristics that search for backup paths in single and multiple layers. Evaluations for single-layer resilience are performed outlining the impact of each heuristic with various protection variants as well as different protecting layers. Moreover, we investigate multi-layer resilience including sets of multiple protecting layers. An analysis of the effec-

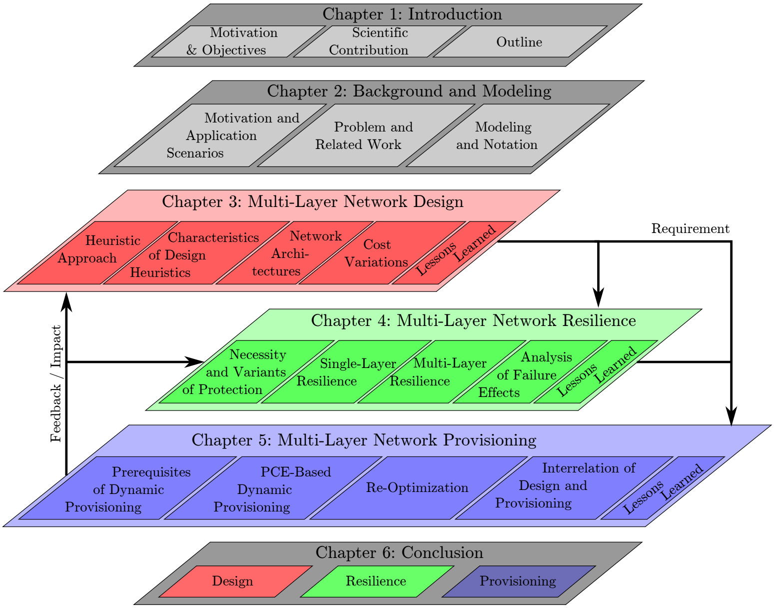


Figure 1.2: *Organization of the monograph including the interdependencies and logical interrelations of the covered topics.*

tiveness of the presented heuristics for resilient multi-layer network design is performed with regard to link, node, and equipment failures.

The automated provisioning of dynamic traffic demands covered in Chapter 5 builds on top of the designed networks. Different PCE architectures and strategies are evaluated. An efficient algorithm for PCE-based re-optimization of established routings is introduced and its ability to reduce blocking is analyzed. In addition, the possible effects of giving feedback from network operation back to network design are evaluated for resilient provisioning and dimensioning aspects.

Finally, conclusions on our findings are drawn in Chapter 6.

2 Background and Modeling

You must unlearn what you have learned.
"Star Wars: Episode V – The Empire Strikes Back", 1980.

Today's global communication is based on a variety of networks whose geographical extents reach from *Local Area Networks* (LANs) to trans-oceanic *Wide Area Networks* (WANs). These networks are built using specific equipment with different technological capabilities and requirements. However, cost-efficiency is a key issue common for all these networks, which needs to be considered from early planning stages to the regular operation mode of a network.

This chapter introduces the required background knowledge to understand the considered optimization problem and its various parameters, including related work. We show how to model the considered problems and present solution approaches, which form the basis of the strategies in the following chapters.

2.1 Motivation and Application Scenarios

Since the days of telegraphy and light houses, the technology enabling electrical and optical telecommunication (from Greek "tele", *distant*, and Latin "communicare", *to share*) evolved in many ways. We show that this evolution imposes a continuous process on networks and present applications, which outline the need for sophisticated planning methods.

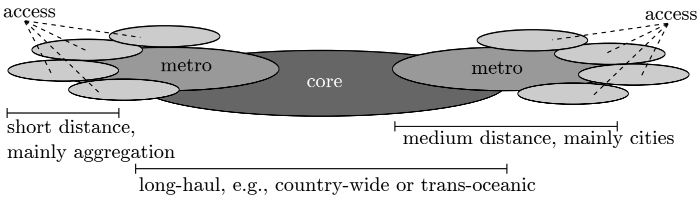


Figure 2.1: *Partitioning of networks by geographical extent and operational purpose from a top-view perspective.*

2.1.1 Network Structures

To understand the development processes and technologies in networking contexts, we first clarify the basic structure of telecommunication networks. Following [42, 43], all networks can be decomposed in two ways, which results from two different perspectives, *partitioning* and *layering*.

Partitioning

From a top-view perspective, networks form domains categorized by ownership and spatial extent. In the context of the *Internet Protocol* (IP), domains are also called *Autonomous Systems* (ASes) and tiers. Upon this categorization, a geographical *partitioning* into *access*, *metro*, and *core* networks can be performed as illustrated in Figure 2.1. The network classes differ in the prevailing technology, average data rates, reach, and main purpose, like the type of service that is common within a certain category.

Access networks are dominated by equipment that performs at comparatively low data rates, like end-user equipment in case of households or single servers in case of data centers. The purpose of access networks is aggregation. Nowadays, many access networks are still based on copper and are therefore operated electrically. In case of households, often *Asymmetric Digital Subscriber Line* (ADSL) connections are set up via *Digital Subscriber Line Access Mul-*

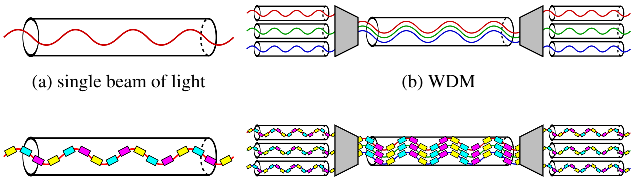
tiplexers (DSLAMs) for connectivity while *Broadband Remote Access Servers* (BRASs) are used for aggregation. In the future, optical access technologies like *Fiber To The Home* (FTTH) and *Passive Optical Network* (PON) will be introduced to keep up with the increasing demand for capacity. Metro networks or *Metropolitan Area Networks* (MANs) interconnect companies and institutions within an urban area and are used to exchange traffic between multiple access networks or to further aggregate it. Ethernet services are often used to connect distributed sites of a company. Core networks may consist of several domains or WANs and serve as the backbone between large sites, also called *Points of Presence* (PoPs). They transport immense amounts of data over large distances. Hence, topologies of core networks are typically built on optical glass fiber, which allows to transmit hundreds of gigabytes of data on distances up to several thousand of kilometers.

Layering

While the top-view perspective shows the partitioning of a network, it does not reveal what data or services are transmitted and how the transmission is realized. Each network comprises physical connections made of copper or fiber, which are piped underground and form a physical layer of interconnections between its sites. This physical layer is the common ground, which is passed by all data transmissions within a network.

However, no data could be transferred without any transmitting equipment as well as corresponding encodings and protocols. On the one hand, equipment is needed to "prepare" data for the transmission path. For example, electrical signals on copper cables need to be treated differently than optical signals on glass fiber since they are much more affected by attenuation [44], which either results in decreased capacity or limited transmission reach. More details on transmission technology are given in Section 2.2.2.

On the other hand, physical connections are rarely used directly. Instead, additional multiplexing equipment is deployed to increase the overall capacity of the physical medium. Figure 2.2 shows four examples of using an optical glass



(a) single beam of light (b) WDM

(c) TDM (d) combining WDM and TDM

Figure 2.2: *Increasing the utilization of physical resources by multiplexing.*

fiber resource. In Figure 2.2a, the fiber can only be used by a single beam of light. Thus, the resource utilization can only be increased by the capacity of the beam of light and is bound to the Shannon limit [44]. While technological progress already today allows wavelength capacities beyond 100 Gbit/s, the requirements of a single data flow lag behind this development. Hence, aggregation and de-aggregation of multiple data streams is needed to fully utilize the resources. Such mechanisms are denoted by *multiplexing* and can be applied in different ways. The main multiplexing technique in the optical domain is *Wavelength Division Multiplex* (WDM), which is a frequency-band constrained form of *Frequency Division Multiplex* (FDM), and allows to use multiple wavelengths on a single fiber as illustrated in Figure 2.2b. Another important multiplexing technique is *Time Division Multiplex* (TDM), which decomposes the carrier signal into time slots and assigns them to different user signals, as depicted in Figure 2.2c. TDM works in the electrical domain and requires *Opto-Electric-Optical* (O-E-O) conversion.

Furthermore, it is possible to combine multiplexing techniques. Figure 2.2d shows the combined use of multiple time slots via TDM on multiple wavelengths via WDM. Further multiplexing techniques as well as corresponding technologies are presented in Section 2.3. Each multiplexing technique is offered by certain technologies, which enable multiplexing by providing corresponding encodings and protocols. Each technology also provides its own *switching* or *routing* mechanisms, which determine the paths taken by multiplexed flows. This

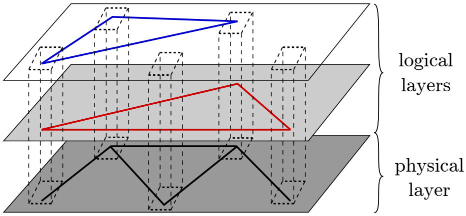


Figure 2.3: *A side-view perspective on a network reveals its multi-layer structure.*

per-technology flow-control forms individual *logical* topologies and the coexistence and combination of their technologies on top of the physical topology leads to the notion of *layers*. Figure 2.3 illustrates the resulting side-view perspective on an exemplary *multi-layer network*. As a layer with appropriate multiplexing equipment may be preferable to "transport" data of another layer, such layers are also denoted as *transport networks*.

2.1.2 Stages of a Network Life-Cycle

Due to the ongoing increase of traffic in data networks [41, 45], networks continuously underlie an evolutionary process, which keeps them in constant change. As discussed in [46], we assume that a network life-cycle repeatedly passes two main stages, *construction* and *operation*, which are illustrated in Figure 2.4.

Construction Stage

In the construction stage, a network is dimensioned and built to fit to the traffic being expected within the planning horizon. As the network is modified to suit the expected traffic, the means of this stage are also called *network engineering*.

In advance, forecasts on future traffic demands are performed and analyzed to determine the amount of required resources. When the installed resources are exceeded, the construction stage needs to be re-entered. To avoid too fre-

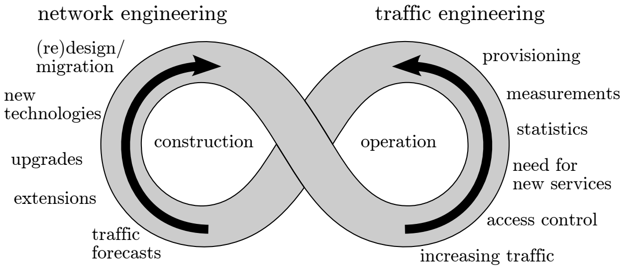


Figure 2.4: *The repeating stages of a network life-cycle.*

quent returns to the construction stage, the designers can consider capacity over-provisioning. Between two network states, upgrades and extensions might be installed or new technologies might be introduced, to support the upcoming traffic demands or to reduce cost. The transition between the current and the envisaged network state is called *migration* [47]. We consider the construction of multi-layer networks in Chapter 3 and Chapter 4. After the equipment is installed and its resources are activated, the network is ready for operation.

Operation Stage

In contrast to the construction stage, adapting the network equipment is not considered during operation. Instead, paths for traffic demands are determined and established upon the resources installed during construction. Also changing of existing paths is considered to utilize the available resources and to reduce congestion as well as blocking. This is also referred to as *traffic engineering*.

During operation, traffic demands arrive and depart dynamically. Thereby, arrivals occur at unknown, random points in time and the holding time of the traffic demands typically follows a certain characteristic. Details on such traffic is given in Section 2.3. The installed resources are dynamically provisioned to the arriving demands. With the aforementioned continuous increase in traffic, the determination of paths for newly arriving traffic demands gets more and more difficult as the load in the network increases. Hence, when the overall traffic in the network approaches its envisaged limit, either blocking occurs, which can optionally be regulated by means of access control mechanisms, or the construction stage has to be re-entered to keep up with the increasing demands.

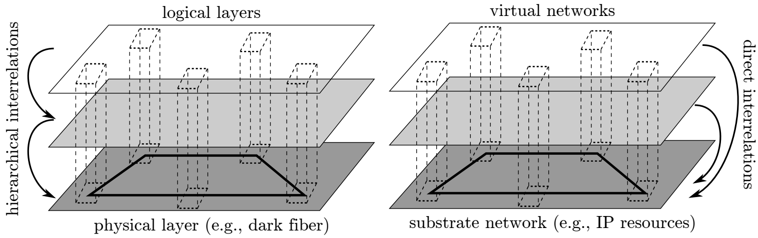
Another cause to re-enter the construction stage is the request for new services by customers that cannot be provided by the current resources and their technologies at all or very inefficiently at best. The addition of a new technology typically means the addition of a new layer to a multi-layer network. We consider the operation of multi-layer networks in Chapter 5.

2.1.3 Application Scenarios

Multi-layer structures appear in various situations with different requirements. In this work, we consider both the construction and the operation stages of multi-layer networks. The approaches presented in this work are at least applicable to the following scenarios.

Network Dimensioning

If a network is created for given traffic forecasts for multiple services (like IP or Ethernet) and corresponding technologies, the request for multiple technologies implies the construction of a multi-layer network, i.e., multiple interrelated logical layers upon a physical layer. The resulting multi-layer network needs to fit to the capacity requirements as well as to *Quality of Service* (QoS), functionality, and budget constraints. To use the resources and to fulfill the given constraints, a sufficient quantity of equipment needs to be installed. Thus, the network is *dimensioned* for the requirements. The set-up of links in logical topologies comprises hierarchical interrelations, which are illustrated in Figure 2.5a. The construction of multi-layer networks is covered in Chapter 3 and Chapter 4.



(a) network dimensioning (b) network virtualization

Figure 2.5: *Multi-layer structures in different application scenarios.*

Network Virtualization

Also network virtualization, i.e., the abstraction of physical links and nodes to virtual links and nodes forming virtual networks upon a common substrate network, comprises multi-layer structures. Hereby, the construction stage is completed and the substrate network provides the resources, which are requested by virtual networks. The direct interrelations to the substrate network by the requests of the virtual networks are illustrated in Figure 2.5b. The optimization problem to solve is *Virtual Network Embedding* (VNE), i.e., to find the optimal mapping of virtual nodes and links to substrate counterparts, which we consider in [2].

Dynamic Provisioning

The aforementioned multi-layer structures formed by construction need to be efficiently operated to react on an increasing request for flexibility not only regarding the topology but also regarding link capacities on different services of a multi-layer network. The network operators need to be able to map resources to customer requests in a given time frame and might need to set up paths in multiple layers. As resources cannot be quickly adapted in the operation stage, network operators require dynamic resource provisioning techniques. These allow to perform sophisticated path calculation based on up-to-date information and re-optimization of existing paths. We consider the *Path Computation Element* (PCE) architecture as a candidate for this purpose in Chapter 5.

2.2 Problem and Related Work

As described in the previous section, multi-layer networks are of high importance in today's telecommunication infrastructure. Hence, efficient approaches to construct and operate multi-layer networks are of great interest. In this section, we present an overview of parameters related to multi-layer network engineering and optimization. Additionally, we give a formal definition of the optimization problem. We show how the problem can be approached and what limitations each approach incorporates. Finally, we give a detailed overview of related work.

2.2.1 Parameter Space

The construction and operation of multi-layer networks involves a large number of parameters. Therefore, the definition of efficiency and optimality of a multi-layer network, i.e., the optimization objective, usually depends on multiple parameters. In the following, we go into detail of multi-layer network optimization and introduce the parameters related to it.

Equipment Parameter Space

Multi-layer networks comprise components of different technologies, which form logical layers. Hence, the available components and their properties play an important role. An overview of equipment related parameters is given in Figure 2.6.

The technologies commonly built into multi-layer networks are *Internet Protocol* (IP)/*Multi-Protocol Label Switching* (MPLS), (Carrier) Ethernet, *Asynchronous Transfer Mode* (ATM), *Synchronous Optical NETwork* (SONET)/*Synchronous Digital Hierarchy* (SDH), and *Optical Transport Network* (OTN). In general, ATM is considered outdated and only rarely used to-

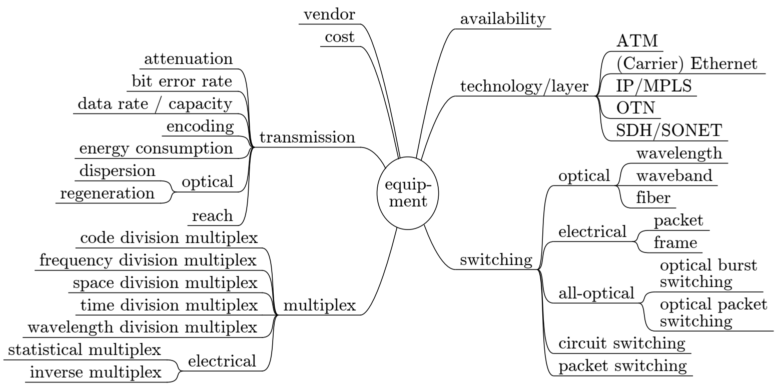


Figure 2.6: *Parameter space of equipment in multi-layer networks.*

day. We explicitly consider IP/MPLS over pure IP to avoid the inflexibility of destination-based routing, i.e., all data streams with the same destination traverse along the same path. IP/MPLS enables per-flow paths via source-based routing.

The equipment of each technology can be roughly grouped into transmission, multiplex, and switching components.

Transmission: Each technology comprises transmission equipment, which allows to set up logical connections at this layer. Due to physical impairments, e.g., expressed by the Shannon limit [44], this equipment has a certain maximum reach, at a certain data rate with a given data encoding, *Bit Error Rate* (BER), and energy consumption.

Multiplex: As mentioned in Section 2.1.1, the requirements of a single data stream usually do not fully utilize the capacity of the transmission equipment present in core networks. Hence, technologies provide multiplexing equipment to aggregate multiple data streams, cf. Figure 2.2. The type of multiplexing depends on the technology. Especially, electrical and optical equipment differs. While electrical technologies, like IP/MPLS, can access data signals bit-wise, optical technologies deal with *wavelengths* or ranges thereof, so called *wavebands*.

Electrical technologies can use *statistical multiplex*, i.e., exploiting the statistical distribution of packet inter-arrivals or perform TDM. In contrast, optical technologies are bound to optical granularities and cannot access or change the bit rate of wavelengths directly.

Switching: As mentioned earlier, the term *multi-layer* rises from the existence of switching capabilities in each technology, which leads to logical topologies and layers. As with multiplexing, the granularities that may be switched depend on whether the technology is optical or electrical. There are two basic switching techniques, *circuit switching* and *packet switching*. With circuit switching, a data flow is marked with some kind of label such that switching only considers this label. Packet switching performs switching by considering the information contained in the packet header. The latter has several disadvantages, like the aforementioned destination-based routing and the somewhat larger overhead. As we focus on core networks, packet switching is not considered in this work. Circuit switching is available in IP/MPLS, Carrier Ethernet, SDH/SONET, and OTN.

Each component of a technology is associated with a certain cost value. The split into components is covered in Section 2.3 where we introduce a specific component and cost model.

Optimization Parameter Space

The equipment, i.e., the available technologies, their components, and their properties, are the basis on which multi-layer network optimization is performed. This section describes the parameters that need to be considered in optimization techniques applied to multi-layer networks. As depicted in Figure 2.7, the field of optimization can be split into four domains.

Applicability: Certain situations make different demands and require different optimization techniques as a consequence. These situations can be reduced to the stages of a network life-cycle, which are described in Section 2.1.2.

Approach: The multi-layer network optimization problems can be approached in different ways. These approaches differ in (computational) complexity, accuracy, and flexibility regarding the optimization problem. Moreover, some ap-

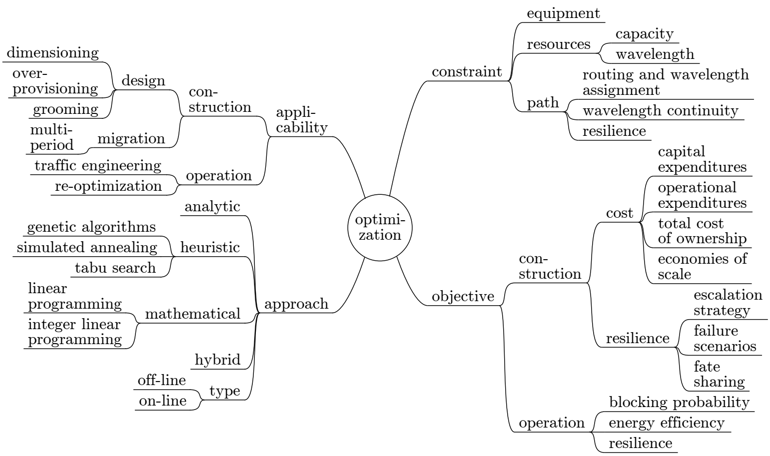


Figure 2.7: *Parameter space for multi-layer network optimization.*

proaches are more suited to certain optimization sub-problems than others. The approaches applicable to the sub-problems are covered in detail in Section 2.2.3.

Constraint: The optimization of multi-layer networks is bound to several constraints. We already discussed various constraints given by the available equipment and illustrated them in Figure 2.6. The equipment finally provides resources of which the capacity on a link in an electrical layer and the number of wavelengths in an optical layer as well as the switching capacity in a node are the most relevant for the considered optimization. The availability of sufficient resources along paths is essential to be able to set up logical connections in multi-layer networks. Beyond simple resource availability, there are more sophisticated requirements, like the set-up of an end-to-end (so called *transparent*) light-path, which means the availability of a wavelength along the whole path, which is covered by the $\mathcal{NP}$-hard *Routing and Wavelength Assignment* (RWA) problem [42,48]. Another example of elaborate path constraints is resilience, i.e., finding additional backup paths that need to be disjoint from the primary path.

Objective: The optimization of multi-layer networks usually cannot be reduced to a single optimization objective. However, the perpetual goal in network construction and operation is cost minimization. On the one hand, there is the cost for the initial acquisition of equipment – so called *Capital Expenditures* (CAPEX). Thus, the goal of construction is to reach a minimum amount of installed equipment and highest utilization. On the other hand, there is the cost for the operation and maintenance of equipment – so called *Operational Expenditures* (OPEX) – which includes power supply and human resource, besides others. During operation, the main objective is to minimize the blocking of newly arriving demands. Also energy efficiency can be fostered by manipulating routing decisions such that links are avoided and can be (temporarily) deactivated. When resilience is considered, the minimization of affected traffic demands and *Expected Loss of Traffic* (ELT) in case of failures is aspired.

Input Parameter Space

The possibilities provided by equipment and optimization parameter space can be varied by different input parameters. As a consequence, there is a plenitude of multi-layer scenarios for which optimization can be considered. We categorize these input parameters into four groups as illustrated in Figure 2.8.

Stage: There are several input parameters that are specific for a certain stage of the network life-cycle, cf. Section 2.1.2. During construction, the order of the (static, off-line) traffic demands can be changed. This is studied in Chapter 3. Furthermore, we may decide for a parameter of the RWA problem [42, 44] in OTNs. Three types of wavelength assignments can be distinguished resulting in different optical grooming corresponding light-path topologies:

- *opaque:* wavelength assignment is done on a hop-by-hop basis, resulting in a 1-hop light-path topology,
- *transparent:* a single wavelength is used for a complete light-path, resulting in a topology consisting of end-to-end light-paths only,
- *hybrid*: a mixture of opaque and transparent, also referred to as *semi-transparent* or *translucent*.

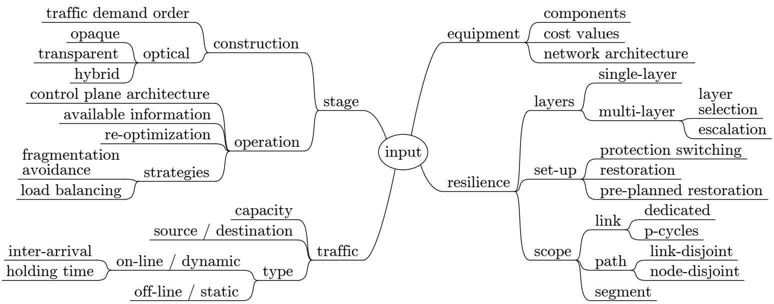


Figure 2.8: *Space of input parameters to multi-layer networks.*

During operation, the order of the (dynamic, on-line) traffic demands is not known in advance and cannot be changed. The parameters concerning operation are the control plane architecture as well as the information that is available to it, potential re-optimization strategies as well as the strategies to minimize blocking.

Traffic: There are several parameters that need to be determined for traffic demands, like distributions for capacity, source-destination pairs, as well as inter-arrival and holding times in case of dynamic provisioning. Details regarding these traffic parameters are given in Section 2.3.

Equipment: The available equipment is a very decisive parameter for the multi-layer optimization problem. The cost value assigned to a component decides whether a component is used in certain situations. The availability of certain components also affects the *architecture* of a multi-layer network, i.e., which technologies are used to operate the provided services. Details on inter-connectivity of different technologies are given in Section 2.3.

Resilience: An input parameter that branches into many aspects is resilience, i.e., the availability of disjoint primary and backup paths. During the construction stage, the *layers* performing protection need to be determined. If multiple protecting layers are determined, *escalation strategies* need to be defined to avoid interference of protection mechanisms in different layers [49]. We need to define *path set-up* strategies, which determine whether backup resources are reserved in advance (protection switching), ad-hoc (restoration), or a mixture of both (pre-planned restoration). While we are quite flexible to choose any of these schemes during operation, we carefully need to consider additional resources during network design. Equally important is the choice of the *scope* that determines whether backup paths are assigned for single links, complete paths, or path segments.

So far, we gained an insight to the parameters of the multi-layer network optimization problem. The combination of these parameters forms the actual optimization problem that is covered in the following section.

2.2.2 Problem Formulation and Notation

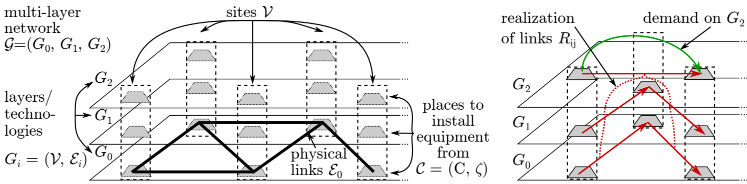
The formulation of the multi-layer optimization problem combines the presented parameters. We introduce the notation used in this work and elaborate the problem formulation on the construction and operation stage of a multi-layer network.

Notation

Before we consider the actual problem formulation for the construction and operation of multi-layer networks, we introduce some notation to describe these stages. The notation and its relation to network elements is shown in Figure 2.9a.

We denote the multi-layer network itself by $\mathcal{G}$. It consists of different layers $G_i = (\mathcal{V}, \mathcal{E}_i)$ where $\mathcal{V}$ is the set of *sites*, also called *Points of Presence* (PoPs), shared by all layers and $\mathcal{E}_i \subseteq \mathcal{V} \times \mathcal{V} \setminus \{(v, v) \in \mathcal{V}\}$ is the set of links connecting the sites in layer G_i. Each layer $G_i \in \mathcal{G}$ is a directed multi-graph. That means, each link $e \in \mathcal{E}_i$ has a source $\alpha(e) \in \mathcal{V}$ and a destination $\omega(e) \in \mathcal{V}$ and multiple links may exist between a source-destination pair. We emphasize the physical layer $G_0 = (\mathcal{V}, \mathcal{E}_0)$ since it is immutable in our optimization and consists of raw fiber ducts containing a number of fiber strands, so called *dark fiber*. The layers above the physical layer are called logical layers.

The definition of all available layers as well as their association to technologies, components, and cost is performed by the CAPEX and component model



(a) layers and equipment (b) realization of logical links

Figure 2.9: *Illustration of notation used in the context of multi-layer networks.*

$\mathcal{C} = (C, \zeta)$ comprising the available equipment C (per layer) and a cost function $\zeta : C \mapsto \mathbb{R}_0^+$ associating cost with equipment. Our studies are based on the CAPEX and component model of [50], which defines $\mathcal{G} = \{G_0, G_1, \ldots, G_5\}$ where G_0 is the physical, G_1 and G_2 the OTN, G_3 the SDH/SONET, G_4 the Ethernet, and G_5 the IP/MPLS layer as defined in Section 2.2.1. A list of the considered technologies is given in Table 2.1. While using the technologies given in [35], we are not limited to these.

Construction

In a typical business scenario for network construction, a network provider asks a network supplier to design and build a network that suits its needs. The network provider formulates its needs with regard to the expected amount of traffic among certain source-destination pairs in different services/technologies as a multi-layer traffic matrix $\mathcal{D}^{\text{stat}}$ of directed traffic demands and might also determine the network architecture that is to be used.

Furthermore, the physical topology $G_0 = (\mathcal{V}, \mathcal{E}_0)$ is usually given and immutable due to the immense cost of earthwork, which can easily raise to several USD 100.000 per kilometer in urban areas. In core networks, the physical links $\mathcal{E}_0$ typically are fiber ducts that contain a number of fiber strands on each physical link $e \in \mathcal{E}_0$ and give the amount of available fiber resources.

Table 2.1: *Overview of the considered technologies.*

$G_i \in \mathcal{G}$	name	data processing	multiplexing technique
G_0	fiber/physical	optical	SDM
G_1	OMS } OTN	optical	WDM
G_2	OCh } OTN	optical	–
G_3	SDH/SONET	electrical	TDM
G_4	Ethernet	electrical	statistical
G_5	IP/MPLS	electrical	statistical

Upon the information of the multi-layer network request, the network supplier has to construct a multi-layer network that fulfills the requirements of the network provider and is cost-minimal at the same time. To that end, the network supplier can choose from its pool of available equipment C (vendors, properties, and etc.) according to the requirements of the network provider and deploy components in the sites of the logical layers to establish connections, which form the logical layers. Thus, the logical links in a multi-layer network are defined recursively by links in lower layers. We describe the hierarchical interrelation of layers in a link-wise manner by means of functions $R_{i,j} : \mathcal{E}_i \mapsto \mathcal{P}_j$, $i > 0, j \geq 0$ where $\mathcal{P}_j$ is the set of possible paths in layer G_j. This is illustrated in Figure 2.9b.

The initial state of the construction is illustrated in Figure 2.10. As no equipment has been deployed yet, no logical connections are established and the logical layers are empty. As each technology/layer provides switching functionality and corresponding equipment, the logical layers can be arbitrarily modified by the network supplier within constraints given by the equipment. Finally, it is up to the network supplier to find cost-efficient logical topologies that allow to carry the forecast of traffic demands given by the network provider.

As only the fiber strands of the physical layer provide resources in the initial construction (empty network), network suppliers have to perform *network engineering* to mitigate this severe limitation. Each link $e \in \mathcal{E}_i$ in a logical layer G_i

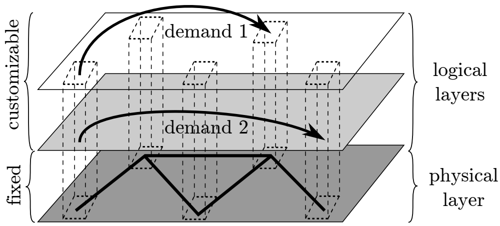


Figure 2.10: *An exemplary request for a multi-layer network of a network provider to a network supplier comprising two demands.*

is associated with equipment and provides a total amount of resources $r_{\text{total}}(e)$ (i.e., capacity or wavelengths). The amount of its free resources is denoted by $r_{\text{free}}(e)$. However, these resources cannot be sufficiently utilized or at least be cost-efficiently used by all technologies. As a consequence, technologies like OTN need to be installed to open up the optical resources for efficient usage by multiple technologies. For similar reasons, TDM equipment, like SDH/SONET, is installed to allow the aggregation of smaller data flows. The use of one technology as the base ground for another technology leads to the notion of *transport networks* [42,43]. In cost optimization, also a less expensive cost structure might be achieved by the introduction of additional equipment as we show in Chapter 3.

The use of technologies to build transport networks that carry aggregated data streams leads to hierarchical structures in multi-layer paths as illustrated in Figure 2.11. The aggregation of smaller data streams via multiplexing techniques provided by the transport network technology forms the optimization subproblem where to perform the aggregation of data streams which is called *grooming* [42,43,51]. A review of grooming techniques is given in [52].

Beyond mere design constraints, *resilience* is an important input parameter that opens up a wide range of variations for multi-layer network optimization as shown in Section 2.2.1. Especially, in combination with the various multiplexing and grooming techniques, *Shared Risk Groups* (SRGs) are created, which

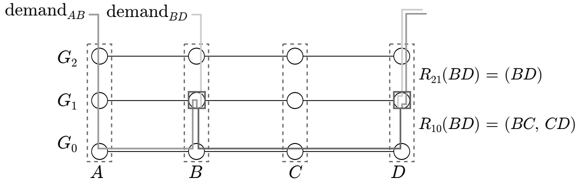


Figure 2.11: *Abstract side-view perspective on a path in a multi-layer network.*

can cause immense cost in case of failures. For instance, the failure of a physical link carrying 40 wavelengths of which each is operated at 40 Gbit/s generates 1.6 Tbit/s of *Expected Loss of Traffic* (ELT). A very important parameter for the network design regarding resilience is the choice of the layers that perform the protection. This choice affects the reaction and recovery time, the size of SRGs [53], and the CAPEX. In general, resilience adds additional constraints to the finding of paths for traffic demands resulting from the disjointness of the backup and primary paths across all layers.

Finally, the special properties of optical equipment need to be mentioned. In contrast to electrical technologies, which handle bits, optical technologies deal with wavelengths that are affected by non-linear effects in optics depending on the fiber component quality and transmission distance. Furthermore, wavelengths are a limited resource such that the RWA problem needs to be considered when wavelength conversion is not provided. Both properties impact the choice of wavelengths for light-paths and each of these is an $\mathcal{NP}$-hard problem [51, 54].

To conclude the construction problem formulation, the multi-layer network design problem can be mapped to a combinatorial problem of deploying and interconnecting equipment that has a large solution space.

Operation

Compared to network construction and multi-layer network design, the optimization of network operation seems to be easier at a first glance. However, most of the considered optimization sub-problems bear the same complexity. In the operation of multi-layer networks, the existing resources need to be efficiently used and cannot be immediately modified, cf. Section 2.1.2. The network operator is confronted with customer requests of directed dynamic traffic demands, which we summarize by the set $\mathcal{D}^{\text{dyn}}$ to whom we give details in Section 2.3.3. The added dynamics of the traffic demands bind the processing order of the demands to their arrival time. As a consequence, the network operator needs to perform *traffic engineering* to deal with newly arriving traffic demands. This is also known as *dynamic capacity provisioning*. Traffic engineering in multi-layer networks was supported by two developments in the last decade.

The first support came from the availability of *Generalized Multi-Protocol Label Switching* (GMPLS) specifications [21] and its extensions for traffic engineering in multi-layer networks [28], which is described in detail in [55,56]. GMPLS generalizes the concept of labels from IP packets to fibers, wavelengths, and time slots such that *Space Division Multiplex* (SDM), WDM [29], and TDM [25] of technologies supporting GMPLS can be integrated in a common control plane.

The second support arose from the upcoming of the PCE architecture [30], which is based on GMPLS and *Resource ReSerVation Protocol with Traffic Engineering* (RSVP-TE) [22]. The PCE architecture is a vendor-independent, unified approach to deal with path computation for on-line traffic demands across layers [34] and domain boundaries [57,58].

Therefore, we consider a GMPLS-based PCE architecture, in which the arrival of an on-line traffic demand triggers a request to a PCE for a path that fits to the requirements of the traffic demand. The algorithm that actually calculates the path within the PCE is not specified. Beyond common shortest-path algorithms, like Dijkstra [59], sophisticated *Constrained Shortest Path First* (CSPF) algorithms may be used that consider different constraints of the multi-layer network. Therefore, the CSPF algorithm must retrieve the required information and must not exceed a given calculation time limit. As discussed in Section 2.2.1, different deployments of multiple PCEs, e.g., per site, per layer, or per layer and site [60], as well as different knowledge (stateful, stateless) of a PCE [30] are possible. Furthermore, a PCE might include different strategies to reduce the blocking probability, like load balancing and fragmentation avoidance.

2.2.3 Approaches and Limitations

In the following, we briefly describe four ways to approach the presented optimization problems and the main limitations of these approaches.

Analytical: Analytical approaches are typically applied to the quantity of components and the amount of resources, like switching capacity at a node or link capacity that is required to fulfill the given traffic demands. Thus, the deployment of components during design can be roughly estimated analytically. However, the analytical approach is not feasible for selecting various equipment combinations and giving a construction blue print for a multi-layer networks. Furthermore, analytical approaches usually require a lot of expert knowledge that needs to be put into the analytical models.

Combinatorial: As stated before, the considered optimization problems of multi-layer network design can be reduced to combinatorial problems. These can usually be mapped to *Linear Programs* (LPs) or *Integer Linear Programs* (ILPs) as we deal with integral numbers of components. While real number solutions of ILPs can be efficiently found, e.g., using the *Simplex* algorithm [61], finding integral solutions is $\mathcal{NP}$-hard [62]. Thus, the usage of ILPs is limited in practice due to long computation times even though professional ILP solvers like CPLEX [36] implement sophisticated approaches, like *branch-and-bound* or *branch-and-cut* [63], to shrink the solution domain and to reduce the number of constraints. This makes ILPs infeasible for extended parameter studies on larger network instances and inflexible regarding variations of input properties. The use of combinatorial approaches in operation is possible, but the additional time constraints as well as uncertainty increase the complexity. Moreover, ILP solvers [36, 37] do not allow to derive principles for network design.

Heuristic: In contrast to combinatorial approaches, heuristic approaches imply that certain parameters, like the order of traffic demands, are not elaborated in detail. Instead, a heuristic aims to find a good solution to the optimization problem in a fraction of time. This is also the reason for the inability of heuristics to guarantee the optimality of the results. As with mathematical approaches, heuristics directly deal with the deployment of equipment to set up connections in logical layers. Therefore, construction blue prints can be created. As heuristics allow to directly implement methodologies to design networks, they open up to derive conclusions for design rules and to evaluate new rules for optimization.

Hybrid: Hybrid approaches usually combine combinatorial and heuristic approaches and aim at speeding up the solution while not diverging too much from the optimal result. The closer such hybrid approaches are to combinatorial approaches, the more they inherit the same trade-offs.

A summary of the approaches and their properties is given in Table 2.2. In the following, we give an overview of publications that are related to the considered optimization problem and are based on the presented approaches.

2.2.4 Related Work

As described in the previous sections, multi-layer network engineering and optimization are complex issues of high importance for all network providers and suppliers. In this section, we describe existing publications on network economics, network design, network resilience, and network provisioning.

Network Economics

While network traffic increases year by year [41, 45], the providers' revenues on a per bit basis decrease. Hence, cost-awareness is an important issue and so is the analysis of cost factors within the network context. There are two important types of cost, *Capital Expenditures* (CAPEX) and *Operational Expenditures* (OPEX).

Table 2.2: *Pros and cons of approaches for the considered optimization problems. (legend: $+$ = pro, $-$ = con, $\circ$ = depends)*

criteria	approaches: analytical	combinatorial	heuristic	hybrid
guaranteed optimality	$\circ$	$+$	$-$	$\circ$
scalability	$+$	$-$	$+$	$\circ$
flexibility regarding input	$-$	$-$	$+$	$\circ$
possibility to learn from	$\circ$	$-$	$+$	$\circ$

CAPEX is the cost of facilities, hardware, or digging cable trenches. That means, cost that usually comes up once during the network design and first installation [50]. OPEX [64] is the cost that has to be paid regularly, especially cost for electricity and human resources as well as cost to repair network failures and to pay penalty fees from violated *Service Level Agreements* (SLAs) [65]. While CAPEX is usually a larger cost factor, the regular maturity of OPEX may not be neglected. Studies performed in [66] state that OPEX constitute $50-80$ % of the *Total Cost of Ownership* (TCO), i.e., the sum of CAPEX and OPEX, on a ten year basis. However, reasonable values for OPEX are hard to estimate in practice as OPEX heavily depends on business models, political developments, and location factors. A general OPEX model for (multi-layer) networks is still lacking.

Thus, cost-aware network design has mostly focused on CAPEX. Studies on the structure of cost factors in transport models [67] and the development of cost [68] were performed. In recent years, the availability of detailed component and CAPEX models [50, 69] has attracted a lot of attention in research. This data is based on the European *Information Society Technologies* (IST) integrated *Next generation Optical networks for Broadband European Leadership* (NOBEL) phase 1 and phase 2 projects [35]. These projects conducted world-wide polls on cost data for networking equipment by surveying network equipment suppliers. The collected data was blinded and generalized as well as normalized to the cost of a 10 Gbit/s WDM transponder to be vendor-independent.

Design

Existing literature on multi-layer network design can be split into two groups. The first group considers multiple layers separately and, in a way, performs single-layer optimization multiple times. The second group follows an integrated approach that jointly optimizes all considered layers.

An overview of general mathematical formulations for the *discrete cost multi-commodity flow problem*, i.e., installation and deployment of equipment, and a review of single-layer studies is given in [70]. A main focus of such single-layer studies is on the optimization of grooming to mitigate routing optimization. A review of grooming techniques and the role of grooming in network design is given in [52]. In [71], a graph model is developed for traffic grooming that facilitates the change of the optimization objective and two algorithms for the optimization of these graphs are presented. The number of wavelength links and transponders is reduced to indirectly minimize network cost. The authors of [72] consider multi-layer switches in OTN, which can deal with wavelengths, wavebands, and fiber as aggregation levels. Evaluations are performed for a simple non-linear cost function that maps the utilization of the used aggregation level. The problem is approached by an ILP as well as a heuristic, which is faster, but less successful. In [73], an ILP as well as a simple heuristic is presented for a two-layer approach with IP/MPLS on OTN. However, the actual optimization only covers the placement of *Label Switching Routers* (LSRs) in the IP/MPLS layer. This work is extended by the same authors for statistical multiplex in [74]. Both papers use a simple cost model reflecting link and node cost. The authors of [75] consider an IP/MPLS layer on top of an SDH or a WDM transport network. The authors analyze the trade-off between additional cost of IP/MPLS support in the nodes, and cost savings due to traffic aggregation via statistical multiplexing and the granularity of capacity values of links and nodes. An ILP is given to optimize the number and location of LSRs in the network and the link capacity dimensioning. The authors of [76] study cost-efficient routing heuristics which are able to reduce the installation cost for the number of ports needed for an *Optical Cross Connect* (OXC). Instead of cost, the number of line cards is considered. The authors of [77] consider single-layer OTN optimization and focus on the $\mathcal{NP}$-hard RWA problem with and without availability of wavelength conversion. An ILP is given and a k-shortest path heuristic is proposed. An abstract, analytical approach for equipment deployment in single-layer scenarios from access networks to OTN core networks is presented in [78]. To that end, a distribution of equipment on a geographical area is considered. Construction blue prints for a multi-layer network cannot be extracted from such an approach. The authors of [54] use a subset of the optical equipment model from [69] with seven components, introduce an additional "grooming layer", and give an ILP for CAPEX optimization relying on pre-calculated paths to speed up the optimization. In [7], we present an ILP that does not depend on pre-calculated paths and develop heuristics for the considered optimization problem. In [9], we propose ILPs for CAPEX minimization for transparent, semi-transparent/hybrid, and opaque OTNs.

In [79], a presentation of the complexity in single-layer network design is given and extended to the complexity of integrated multi-layer network design. The authors compare global design to a two-step design approach and present an approach for iterative multi-layer network design that iterates over all layers of a multi-layer network until a convergence boundary is reached. The authors of [80] develop an ILP formulation for multi-layer network design based on a detailed, but theoretical equipment model. Evaluations are not given. In [81], the authors of [80] enhance their ILPs by a heuristic branch-and-cut algorithm and focus on a two-layer network design. Still, no evaluations are given. In [82], the same authors further tweak their ILP approach and give evaluations for networks with up to 17 sites but no details on their cost model. The authors of [83] present estimation formulas for the number of required equipment based on the multi-layer equipment model of [50]. The authors consider two-layer networks using

SDH over OTN. A construction layout or multi-layer paths for traffic demands are not given. In [84], a heuristic for planning GMPLS-based transport networks is presented. The focus is on the optimization of OTN equipment considering wavelengths, wavebands, and fiber. In [8], we use the CAPEX model from [50] and present a CAPEX-aware multi-layer network design heuristic.

Resilience

So far resilience aspects had been excluded on purpose to emphasize the mere design aspects. Now, we focus on the design of resilient multi-layer networks. A broad overview of resilience and protection schemes is given in [53]. In [42, 43], causes and consequences of failures are illustrated for different technologies and network architectures. Studies on failure statistics of equipment in different technologies of multi-layer networks are given [85]. In [1], we assign such availability and failure probabilities to links and nodes to analyze the probability of different failures. In [86], an overview of multi-layer network survivability is given. The authors consider the popular technologies of their time, ATM and SDH/SONET.

As discussed before, resilience can simply be seen as additional constraints for the combinatorial optimization problem of multi-layer network design. Therefore, many publications considering multi-layer network design also consider resilience and are only briefly listed here if they were discussed before. However, resilience can be implemented in many ways in multi-layer networks as elaborated in Section 2.2.1. Especially, we distinguish related work regarding single-layer protection and multi-layer protection.

Path protection on the IP layer is considered in [80] but no evaluations are given as aforementioned. Besides grooming layers with pre-calculated paths, the authors of [54] apply protection on the *Optical Channel* (OCh) layer. Evaluations are focused on results including protection. Comparison to the unprotected case is not presented. In [7], we use the cost model of [54] as a basis for our own cost model and present an ILP that does not depend on pre-calculated paths. We develop heuristics for the considered optimization problem with and without resiliency as well as a comparison of the corresponding results. The authors of [87] use the cost model of [54] and investigate installation cost with dedicated path protection by using simulated annealing. The authors model several layers for different capacity granularities in the OCh layer. A routing and grooming heuristic is proposed for medium-sized networks. The authors of [88] apply *Mixed Integer Programming* (MIP) to yield cost minimal solutions for a reference network provided by Swisscom with transparent, opaque, and hybrid OTN architectures. Protection is applied on the IP layer and guaranteed for single-link as well as single-node failures. The study reveals dependencies on traffic loads, available data rates, and cost ratios of interfaces with different data rates. In [9], we also introduce ILP formulation for CAPEX minimization on these OTN architectures with and without resilience but in a canonical way. In [89], a heuristic approach for two-layer network equipment planning is presented. Two probabilistic algorithms are presented and run on networks with up to 321 sites. Compared to presented ILPs, the heuristics are significantly faster and show a good approximation to the lowest CAPEX. The used CAPEX model defines equipment for edges and nodes but only contains four components. In [90], the heuristics from [89] are extended by a restrained look-ahead method. Results are shown for the heuristics with and without restraining look-aheads.

In contrast to resilience on a single layer, multi-layer resilience, i.e., simultaneous protection at multiple layers, has not been covered broadly in literature. The *Protection Across NEtwork Layers* (PANEL) project [49] considers multi-layer networks with multiple protecting layers. By means of simulations, a quantitative comparison of protection at the highest and lowest layer for ATM and SDH equipment are performed with a simple CAPEX model. These concepts are extended in [91] to consider static and dynamic multi-layer recovery strategies. During these studies also escalation strategies, i.e., the inter-working between layers, in case of failures is considered to determine which protection layer should take care of a failure. In [12], we propose a heuristic for the combined search of primary and backup paths in multi-layer networks. We analyze various protection schemes in different layers and perform evaluations on overall CAPEX as well as CAPEX for protection and the number of traffic demands that can be established.

Provisioning

Capacity provisioning means the operation of a network with dynamic traffic requests. Therein, not cost but utilization and blocking ratio are the most crucial criteria. The author of [46] considers such dynamic routing and resource allocation in WDM transport networks and postulates a feedback loop from provisioning to design as described in Section 2.1.2.

In the multi-layer context, the set-up of multi-layer paths can be supported by GMPLS, which leads to *Automatically Switched Optical Network* (ASON)/*Automatic Switched Transport Network* (ASTN) or *Intelligent Optical Network* (ION) in optical layers [92]. To perform network operation and traffic engineering, a multi-layer network must pre-exist, e.g., yield by multi-layer network design strategies, that can deal with the expected dynamic traffic demands. The main goal of the provisioning is to reduce the blocking probability for newly arriving traffic demands by means of traffic engineering. The need for integrated design of multiple layers was recognized quite early. Before the final release of the GMPLS specification in 2003, the work in this area mainly focused on single-layer approaches. The authors of [93] state a "trend in backbone networks toward dynamic provisioning of bandwidth guaranteed or wavelength paths". The authors provide ILP formulations under the assumptions that arrivals are one-by-one with no "a priori" knowledge of future arrivals and path set-up using the *Resource ReSerVation Protocol* (RSVP)/*Label Distribution Protocol* (LDP). In addition, routing algorithms are presented that effectively establish backup paths. In [93], only the IP layer is considered and the underlying OTN is neglected. Thus, combined topology and resource usage information at IP and optical layer cannot be used. In [94], the authors of [93] switch to an integrated two-layer approach in which algorithms for integrated dynamic routing of bandwidth guaranteed paths in IP over WDM networks are developed. However, the layer integration is done assuming that layers can somehow interact without providing a communication interface. In [95], the concept of *Intelligent Optical Network* (ION) is introduced to enable and to perform a comparison of static and dynamic routing in terms of cost and capacity from an operational point of view. In [96], ASTN with a dynamic photonic layer is considered. Traffic is assumed to follow a Poisson arrival process with negative-exponential holding time. In [96], network design is restricted to link capacities. Evaluations on SDH request blocking probabilities are performed. The authors of [97] perform dynamic routing in transparent optical networks. An experimental study of various greedy algorithms for a dynamic RWA problem is performed. The experiments are based on a traffic model taken from measurements.

With the availability of the GMPLS specification [24], approaches of doing integrated design of multiple layers could be justified in a much cleaner way. In [98], SDH over OTN is considered with dynamic traffic demands and an analytical model to evaluate blocking probabilities is presented including simulation results. The authors of [99] give a description of transport network technologies and a comparison of reference network architectures. Numerical evaluations are performed to show the impact of network architectures on the resource requirements, but no algorithms are given. The authors of [100] focus on a shared protection of segments, i.e., protecting paths partially or piece-wise. A peculiarity of their approach is the utilization of knowledge on holding times of dynamic demands. In [101], also protection of dynamic traffic demands is considered. The authors implemented dedicated and shared path protection in an adaptive multilayer traffic engineering framework.

After the introduction of the PCE architecture in 2006, the research focus moved towards automated provisioning of dynamic multi-layer paths. Five years before, distributed algorithms for routing with backup restoration were already investigated in [102]. However, the proposed backup load distribution matrices were never deployed. With the PCE architecture, such information could be exchanged by extensions to the *Open Shortest Path First* (OSPF) protocol and stored in a *Traffic Engineering Database* (TED). The authors of [57] and [103] focus on path set-up times in single- and multi-domain scenarios with respect to the protocols of the GMPLS architecture, especially RSVP-TE. The evaluations performed in [58] show that GMPLS is sufficiently supported in existing networking equipment, such that the PCE concept can be successfully implemented in principle. The authors of [60] give an overview of trends in standardization of the PCE architecture. Recently, a framework for multi-layer networks was introduced [92], which proposes capability planes in future multi-layer networks.

2.3 Modeling

To approach the optimization problems in multi-layer network construction and operation, we need feasible models to access all required information and to combine all envisaged parameters. However, we also need to keep our approach computationally tractable. Therefore, we introduce models to define the layers and resources provided by components including their cost, the interconnection of components and layers to set up multi-layer paths, and traffic definitions.

2.3.1 Component and Cost Model

A multi-layer network component model C defines the available technologies and their possible interconnections. In turn, a CAPEX model $\mathcal{C} = (C, \zeta)$ associates cost values with each component in C by means of a cost function $\zeta : C \mapsto \mathbb{R}_0^+$. In [50], components and corresponding CAPEX values are given for common transport network technologies including IP/MPLS, Carrier Ethernet, SDH/SONET, and OTN technologies with information based on [35].

Technologies

In the following, we present the technologies considered in [35, 50] and listed in Table 2.1, including their basic concepts and properties. As more technological and transmission-related constraints need to be considered the closer a layer comes to the physical layer, such technologies become more complicated from a modeling point of view. As we consider core networks based on optical fiber, technologies are also turning from the electrical to the optical domain. An overview of data rates and granularities is given in Table 2.3.

Table 2.3: *Common data rates and equivalents in different technologies.*

data rate (Gbit/s)	SDH/SONET	OTN	*Gigabit Ethernet* (GE)
1	–	ODU-0	1
2.5	STM-16	ODU-1	–
10	STM-64	ODU-2	10
40	STM-256	ODU-3	40
100	–	ODU-4	100

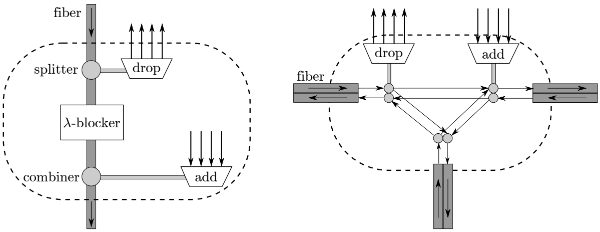
IP/MPLS: In IP/MPLS networks, a label is added to each packet of a data flow such that circuit-switching can be done instead of IP packet-switching. Therefore, *Label Switched Paths* (LSPs) are set up between *Label Edge Routers* (LERs) traversing *Label Switching Routers* (LSRs). This technology is electrical and allows for arbitrary statistical multiplex, i.e., multiplexing data packet streams. In this model, the IP/MPLS technology comprises interfaces for data rates of 2.5, 10, and 40 Gbit/s using the encapsulation formats *Packet over SONET* (PoS) and GE. Nowadays, IP technology is dominated by few, mostly US companies. Hence, IP is the most expensive technology in this model.

Carrier Ethernet: Since its beginnings in the 1970s, Ethernet has spread from LANs to almost all areas of networking [104]. *Carrier Ethernet* (CE) [105] is the logical extension of Ethernet to core and transport networks by incorporating MPLS functionalities. CE is electrical, provides statistical multiplex, and also allows for circuit-switching as in IP/MPLS. However, CE components are less expensive than IP due to its popularity and the resulting *economies of scale*, i.e., reduced production cost caused by bulk buying. However, the equipment considered in the NOBEL model is oriented towards aggregation networks and metropolitan networks, cf. Section 2.1.1, and only provides interfaces at 1 and 10 Gbit/s. Furthermore, the CE technology was quite new at the time this model was introduced. Therefore, component prices are relatively high compared to other technologies providing similar capacity granularities.

SDH/SONET: *Synchronous Digital Hierarchy* (SDH) and its US counterpart *Synchronous Optical NETwork* (SONET) were created as an electrical transport network technology that allows to aggregate several smaller flows into one larger flow using *Time Division Multiplex* (TDM) [42, 43] that is sent via optical technology. As IP/MPLS and Carrier Ethernet, SDH/SONET is also circuit-switched but uses TDM and not statistical multiplex. By means of GMPLS, we are able to set up arbitrary paths in this technology. In the NOBEL model, the SDH/SONET technology comprises interfaces for data rates of 2.5, 10, and 40 Gbit/s, which are given in units of *Synchronous Transport Modules* (STMs).

Optical Transport Networks: In contrast to the previous technologies, the *Optical Transport Network* (OTN) technology [106] is not electrical and, hence, does not support bit-wise examination of data as required for statistical multiplex. OTNs can also handle wavelengths within *Wavelength Division Multiplex* (WDM) by wrapping data in *Optical channel Data Unit* (ODU) [107]. Thereby, an *Optical Channel* (OCh) is a wavelength along a light-path and *Optical Multiplex Section* (OMS) is a combined view of OChs on the multiplexed wavelengths in the *Optical Transport Hierarchy* (OTH). The transmitting interfaces in this layer are optical transponders that create wavelengths at data rates of 2.5, 10, and 40 Gbit/s. Special equipment within this technology are *muxponder*, i.e., transponders that combine, e.g., four 10 Gbit/s data streams to a single 40 Gbit/s data stream, as well as optically switched components, *Optical Add/Drop Multiplexer* (OADM) and *Optical Cross Connect* (OXC), which are illustrated in Figure 2.12. Both, OADM and OXC, allow to add or drop wavelengths to an OMS and to switch a wavelength from an incoming fiber to an outgoing fiber without O-E-O conversion. OADMs are a special form of OXCs with a fixed degree of two that results from their origin in ring topologies. OXCs can be extended to a degree of ten in the considered model. The main cost factors in OTNs are the reach of the transponders and the degree of the optical switching equipment.

Fiber: On the fiber layer, i.e., the physical topology, data has already been encapsulated by the interfaces in higher layers. However, components need to be installed in the fiber layer to lengthen the reach of the transmitting interfaces. To



(a) OADM (b) OXC of degree 3

Figure 2.12: *Optical switching components in* OTN.

that end, a *Dispersion Compensation Filter* (DCF) is installed every 80 km to mitigate chromatic dispersion together with an *Optical Light Amplifier* (OLA) to reduce attenuation on the fiber. In addition, a *Dynamic Gain Equalizer* (DGE) that cooperates with *Erbium-Doped Fiber Amplifier* (EDFA) is installed at every fourth OLA to further reduce attenuation.

Since the release of the CAPEX model in [50], 100 Gbit/s equipment has recently made progress for Carrier Ethernet [38] and optical equipment. Furthermore, studies on technical possibilities [39] state that the capacity of a fiber will continue to increase for the next 20 years at least without exceeding the 50 THz band on optical glass fiber or the Shannon limit.

Components

Networking equipment is built in a modular fashion. This simplifies the exchange of components, increases the flexibility, and reduces cost. We seize the modular decomposition and homogeneously split the equipment of all technologies considered in [50] into four component groups [8] illustrated in Figure 2.13.

Basic nodes provide the core functionality within a technology, like power supply, cooling, and backbone switching for all incorporated components. Their price mainly depends on the backbone switching capacity. *Slot cards* can be

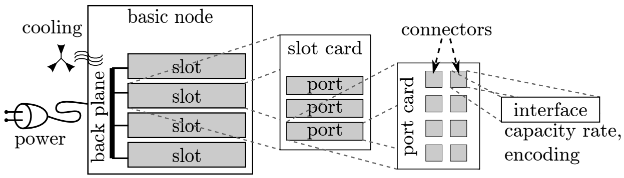


Figure 2.13: *Exemplary hierarchy of the considered four component groups.*

plugged into each slot of a basic node and provide access to its backbone switching. Each slot card has a certain share of the backbone switching capacity that is 40 Gbit/s in the considered model. *Port cards*, also called *line cards*, can be plugged into each port of a slot card. The port cards can be equipped with *interfaces*. The type and the number of interfaces depend on the type of the port card. Interfaces transmit data using a specific encoding according to their technology, data rate, and maximum transmission reach. Consequently, we provide zero-cost dummy components if component groups are not defined in the original model. This also holds for our strict split of the OTN layer into OCh and OMS layer.

Basic nodes, slot cards, and port cards provide the input, output, and exchange of data within a technology. Interfaces enable inter-layer communication.

2.3.2 Connectivity Model

Within a layer, the assembly of the four modular component groups at a site is only restricted by constraints of the components themselves. For example, the number of slots or an upper limit for the sum of requested switching capacities that cannot be exceeded. To interconnect different layers at a site or across remote sites, transmitting interfaces are required that encapsulate the data in a *compatible* format, like *Packet over SONET* (PoS), *Gigabit Ethernet* (GE), *Optical channel Data Unit* (ODU), or the more general *Generic Framing Procedure* (GFP) that can be processed by the receiving interface. An overview of encodings used between the layers covered in [50] is given in Figure 2.14.

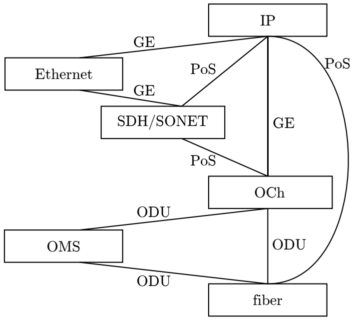


Figure 2.14: *Possible interconnections of layers due to encodings defined in [50].*

We distinguish between two types of interfaces. *Trunk* interfaces communicate to the network, i.e., downwards in the layer hierarchy while *tributary* interfaces result from client signals, i.e., injected into the network by customers, and communicate upwards in the layer hierarchy. Figure 2.15 shows an extended view of Figure 2.13 illustrating the interconnection of two layers via trunk and tributary interfaces as well as comprising components. The modulation and framing of interfaces depend on their layer and capabilities to encapsulate and de-encapsulate data. The framed data is transmitted with a specific bit rate, the interface's *capacity*. Two interfaces must use the same data encoding, i.e., modulation, framing, and capacity, to be compatible. By the interconnection of two layers, data might also be aggregated or de-aggregated via an interface, e.g., a muxponder, or during the switching process at the basic node, to increase the resource utilization.

The interconnection of two sites across multiple layers and intermediate sites may involve cascaded encapsulation and aggregation of data at each layer as well as the de-encapsulation and de-aggregation of data at intermediate sites. Figure 2.16 illustrates a *multi-layer path* between three sites on which data is encapsulated and de-encapsulated several times such that data can be processed at the corresponding layers. The figure also illustrates that not all data exchanged by

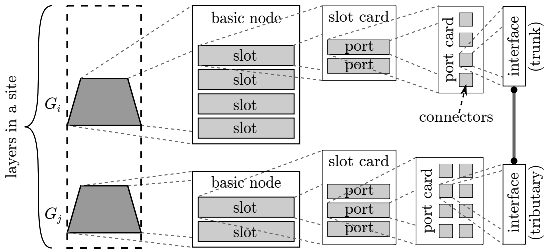


Figure 2.15: *Interconnection of two layers at a site via interfaces.*

two sites must be processed at the same layers. Instead, it is up to the optimization to find such hubs for grooming data.

2.3.3 Traffic Model

As we consider network construction and network operation, we need to distinguish between the traffic requirements of both stages. Since we consider directed multi-graphs $G \in \mathcal{G}$, we consider directed traffic demands as well.

Traffic patterns for network construction usually result from forecasts of the traffic load, like [41, 45, 108] that is expected to happen at a given layer/technology between a pair of sites. As such traffic is used for dimensioning of connections between a site pair, such traffic demands are time-invariant, i.e., *static* or *off-line*. Thus, our general definition for a static traffic matrix is

$$\mathcal{D}^{\text{stat}} = \big\{(\alpha, \omega, G, b) \in \mathcal{V} \times \mathcal{V} \times \mathcal{G} \times B^{\text{stat}} \mid \alpha \neq \omega\big\}, \tag{2.1}$$

with $\alpha, \omega \in \mathcal{V}$ being the source and destination site of a demand, $G \in \mathcal{G}$ giving the layer the demand starts at, and $b \in B^{\text{stat}}$ being the demand's requested bandwidth in Mbit/s out of the set of available bandwidths B^{stat}, which technology-dependent comprises a number of granularities as listed in Table 2.3.

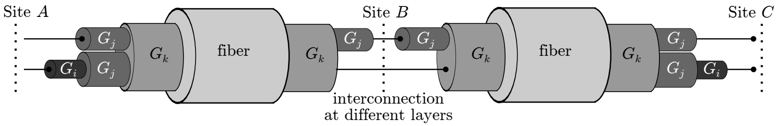


Figure 2.16: *Data encapsulations along a multi-layer path across layers and sites.*

In contrast, a network operator deals with ad-hoc traffic requests between site pairs which start at a certain layer and last for a certain holding time. Thus, traffic demands arrive and depart over a given simulation time span and are *dynamic* or *on-line*. Thus, our general definition for a dynamic traffic matrix is

$$\mathcal{D}^{\text{dyn}} = \left\{(\alpha, \omega, G, b, t_A, t_H) \in \mathcal{V} \times \mathcal{V} \times \mathcal{G} \times B^{\text{dyn}} \times \mathbb{R}^+ \times \mathbb{R}^+ \mid \alpha \neq \omega\right\}, \quad (2.2)$$

with $b \in B^{\text{dyn}}$ being the bandwidth request of a dynamic demand where B^{dyn} may be an arbitrary set, which is not given by equipment granularities, t_A being the point in time of arrival, and t_H being the holding time of the demand. The definition of α, ω, and G remains unchanged compared to $\mathcal{D}^{\text{stat}}$.

There are many ways to define traffic patterns to create traffic matrices. The authors of [109] propose traffic classes based on services and consider the distance between sites as well as their population. In [42], an introductory to different numerical models for demand creation is given while [110] uses data based on measurements. In the course of this work, we focus on two schemes for traffic matrices. On the one hand, we consider *population-based* traffic matrices in which the population ratios determine the amount of traffic transferred to another site. The population model is based on the population data $\pi(v)$ of a site $v \in \mathcal{V}$ taken from recent governmental statistic. The number of demands between a pair of sites $x, y \in \mathcal{V}$ is determined by

$$n_{x,y} = \left\lceil \frac{\pi(x) \cdot \pi(y)}{\sum_{a,b \in \mathcal{V}, a \neq b} \pi(a) \cdot \pi(b)} \right\rceil \text{ and } n_{x,x} = 0. \quad (2.3)$$

On the other hand, we study *uniform* traffic patterns which are homogeneous and more academic. The uniform model creates a given number of K demands between any two sites, i.e.,

$$n_{x,y} = K \text{ and } n_{x,x} = 0, \tag{2.4}$$

which leads to the notion of the K-*uniform* traffic pattern. For the sake of simplicity, we focus on a homogeneous bandwidth distribution for all traffic demands. However, this is not a limitation of the presented approaches.

In the following chapters, we cover the construction and operation of multi-layer networks. We focus on multi-layer network design, multi-layer network resilience, and multi-layer network provisioning while focusing on deterministic heuristics to approach the different optimization problems.

3 Multi-Layer Network Design

I love this plan! I'm excited it could work!

"Ghost Busters", 1984.

In the previous chapter, we presented the background and principles needed to understand the optimization problems considered in this work. In this chapter, we approach the design methods for the construction of multi-layer networks that suit given requirements of network operators. Our focus is on the evaluation of parameters, the minimization of cost with regard to *Capital Expenditures* (CAPEX), and the maximization of fulfillment with regard to static traffic demands requested by the network operators.

We introduce the *Auxiliary Cross Layer* (AXL) algorithm, which is a heuristic multi-layer network design approach. A from-scratch design of multi-layer networks is pursued, which maximizes the degrees of freedom but also results in a high computational complexity. In general, a slightly modified and extended variant of the CAPEX model from the IST NOBEL phase 2 project [35] is used, which was introduced in Section 2.3. However, we try to be as technology agnostic as possible except for some special requirements of optical technology. Studies are performed on the drawbacks caused by the heuristic approach and it is shown how to mitigate these drawbacks. We analyze the impact of different network architectures comprising the availability of equipment and perform studies on cost variations. The content of this chapter is the base of Chapter 4 on multi-layer network resilience as well as for Chapter 5 on dynamic provisioning of such multi-layer networks.

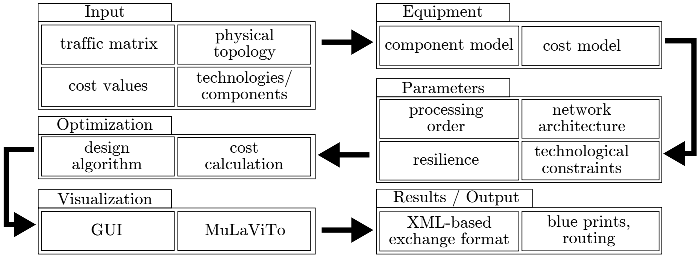


Figure 3.1: *The overall work flow with models, parameters, and algorithms [4].*

3.1 Heuristic Approach

The following sections give an overview of our approach to multi-layer network design. We present a heuristic algorithm for this purpose and introduce basic means to evaluate multi-layer networks.

3.1.1 Overview and Foundations

In Chapter 2, we described the space of parameters of multi-layer network construction regarding input, equipment, and optimization. In general, we consider the construction of a multi-layer network from scratch, i.e., only a physical topology of fiber ducts is assumed to be given. As stated in Section 2.3, we consider static traffic demands in this chapter, which are given by a traffic matrix $\mathcal{D}^{\text{stat}}$ defined by the customer – a network operator. We focus on CAPEX models and deterministic heuristics to evaluate our refined component model presented in Section 2.3.1. An overview of the overall work flow of models, parameters, and algorithms for multi-layer network design is shown in Figure 3.1.

To handle the multi-layer network design optimization problem, we integrated all steps of the work flow into a Java®-based software application. The software is called *Multi-Layer Network Engineering and Optimization* (MuLaNEO) [4,16] and is based on our *Multi-Layer Visualization Tool* (MuLaViTo) [14] to visual-

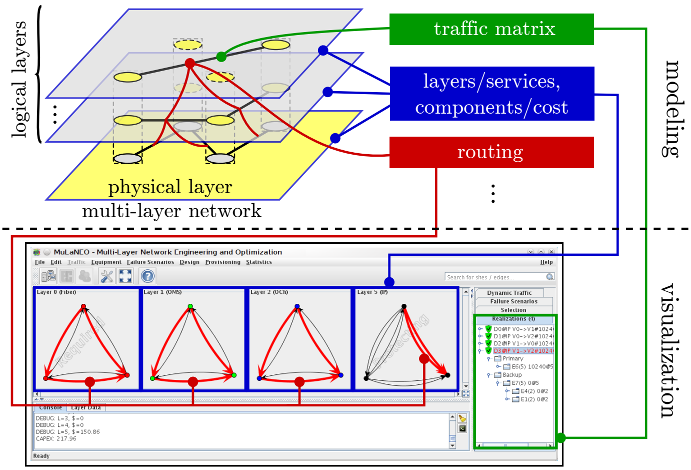


Figure 3.2: *Visual relations of modeling and tooling with* MuLaNEO *[16] on an exemplary multi-layer network.*

ize the resulting multi-layer networks. MuLaNEO provides flexibility in terms of algorithms and allows to define and to evaluate components as well as architectural models via *Extensible Markup Language* (XML), traffic patterns, and more. Some of the relations of visualization to modeling and evaluation are illustrated in Figure 3.2 for a small exemplary multi-layer network instance, which was designed using MuLaNEO.

3.1.2 Auxiliary Cross Layer Algorithm

In this section, we present our *Auxiliary Cross Layer* (AXL) algorithm [8, 12] for multi-layer network design. The algorithm is decomposed into five main steps, which are described in a canonical fashion.

Basic Principle

The work flow of the proposed heuristic algorithm is illustrated in Figure 3.3. Initially (**Step 1**), the set of interfaces $\mathcal{I} \subset C$ defined in the CAPEX model $\mathcal{C}$, cf. Section 2.3.1, is analyzed. This results in the creation of an auxiliary construct, the interface tree $\mathfrak{T}_{\mathcal{I}}$, which contains all possible interface combinations to connect any layer $G \in \mathcal{G}$ to the physical layer G_0, which contains the immutable physical connections of sites in the multi-layer network.

Following (**Step 2**), the heuristic sequentially processes all traffic demands $d \in \mathcal{D}^{\text{stat}}$. The heuristic uses the given input order of the demands to reduce the complexity of the considered optimization problem. Otherwise, a factorial number of orders ($|\mathcal{D}^{\text{stat}}|!$) needs to be considered to find the optimum. The consequence of this simplification are discussed in Section 3.2.3.

In principle, traffic demands may originate in any layer of a multi-layer network, e.g., there might be demands on *Internet Protocol* (IP) layer as well as *Synchronous Digital Hierarchy* (SDH) layer. Thus, we denote by $G(d)$ the layer a demand originates from. For each interface combination $t \in \mathfrak{T}_{\mathcal{I}}$ that connects $G(d)$ to the physical layer G_0 (**Step 3**), the heuristic gathers the free resources of each traversed layer in a free resource graph G_{free}. Thus, G_{free} consists of all links with free resources of the traversed layers. An ordinary *Constrained Shortest Path First* (CSPF) algorithm can be used to search for paths in G_{free}. The resulting paths consist of edges of different layers and are added to a candidate list for the current demand (**Step 4**).

If candidate paths (**Step 5**) exist, the equipment cost is evaluated to complement the candidate paths in G_{free} to valid multi-layer paths according to $t \in \mathfrak{T}_{\mathcal{I}}$ in $\mathcal{G}$. Greedily, the candidate path with lowest cost is chosen and the equipment required for path completion is installed. If no path with sufficient free resources exists, the demand would be blocked. In this case, we try to alter existing resource allocations such that creating the required resources allocations gets possible.

These steps are repeated for each demand $d \in \mathcal{D}^{\text{stat}}$. Next, the individual steps of the heuristic are described in detail.

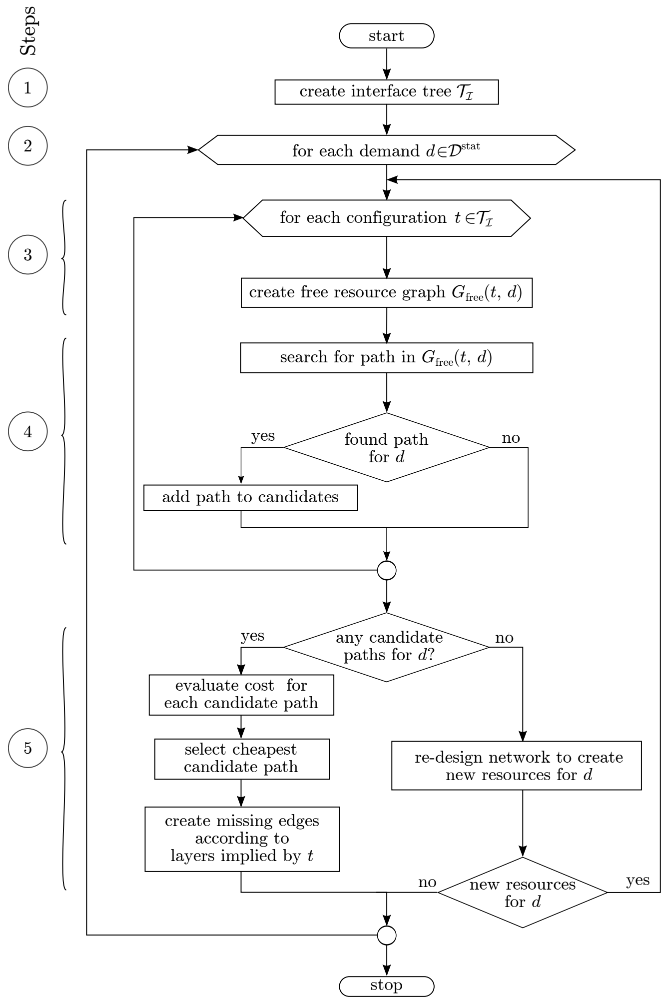


Figure 3.3: *Schematic flow chart of our multi-layer network design heuristic.*

Creation of an Interface Tree

The possibilities to hierarchically connect a layer to any other layer via trunk interfaces can be represented by an interface tree $\mathfrak{T}_{\mathcal{I}}$. An exemplary interface tree is illustrated in Figure 3.4. In Figure 3.4a, a trunk interface, which connects to a certain lower layer, is assigned to each branch in the tree. The tree lists the interfaces $i_1 \in \mathcal{I}_{xy}$, $i_2, i_3 \in \mathcal{I}_{yz}$, $i_4 \in \mathcal{I}_{xz}$, where $\mathcal{I}_{ab} \subset \mathcal{I}$ is the set of interfaces connecting layer G_a to lower layer G_b. Thus, interfaces i_1 and i_4 differ in the layer they connect to, while interfaces i_2 and i_3 connect to the same layer but differ, e.g., in the provided capacity. Each path in the tree from the root to a leaf is a possible configuration to connect layers G_x and G_z. The highlighted path in Figure 3.4a is (i_1, i_3).

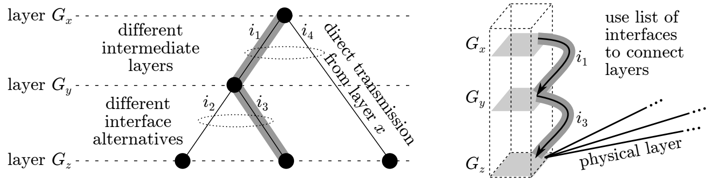
We introduce a formalism $\mathfrak{T}_{\mathcal{I}}^m$ to describe the interface tree to connect layer G_m to the physical layer. To that end, we use a set of tuples of trunk interfaces, which is given by

$$\begin{aligned}\mathfrak{T}_{\mathcal{I}}^m := \Big\{ \, & (i_1, i_2, \ldots, i_k) \in \mathcal{I}_{x_1 y_1} \times \ldots \times \mathcal{I}_{x_k y_k} \mid \\ & x_1 = m \wedge \ldots \wedge y_{i_j} = x_{i_{j+1}} \wedge \ldots \wedge x_k = y_k = 0 \, \Big\}.\end{aligned} \tag{3.1}$$

This set of tuples of interfaces is mapped to a set $\mathfrak{T}_{\mathcal{G}}^m$ of tuples of layers that are traversed when a certain interface configuration in $\mathfrak{T}_{\mathcal{I}}^m$ is used. The mapping is given by

$$\mathfrak{T}_{\mathcal{G}}^m := \Big\{ \, \big(G(i_1), \ldots, G(i_k)\big) \subseteq \mathcal{G}^k \mid (i_1, i_2, \ldots, i_k) \in \mathfrak{T}_{\mathcal{I}}^m \, \Big\}. \tag{3.2}$$

Thus, each tuple of interfaces in $\mathfrak{T}_{\mathcal{I}}^m$ corresponds to a tuple of layers in $\mathfrak{T}_{\mathcal{G}}^m$. The combination of $\mathfrak{T}_{\mathcal{I}}^m$ and $\mathfrak{T}_{\mathcal{G}}^m$ defines at which layers data is processed and how data is encapsulated using interfaces and vice versa, as depicted in Figure 3.4b. For instance, the highlighted path in Figure 3.4a is $(G_x, G_y, G_z) \in \mathfrak{T}_{\mathcal{G}}^x$. The layer configurations given by $\mathfrak{T}_{\mathcal{G}}^m$ are used to create free resource graphs, which are defined next.



(a) an interface tree listing possible interfaces to connect layers G_x and G_z

(b) connecting layers G_x and G_z using the highlighted branch

Figure 3.4: *Possibilities to connect multiple layers via interfaces.*

Selection of Traversed Sites

The selection of the traversed sites for a multi-layer path depends on the existence of paths with sufficient resources. Considering a single layer at a time cannot always yield an optimal solution. For example, there might be no path with sufficient resources between two sites on any layer in $\mathcal{G}$. Thus, the set-up of a completely new path would be required.

Therefore, we developed the AXL algorithm [8, 12]. Instead of always setting up a completely new path, the AXL approach allows to use resources that cannot make up a valid multi-layer path by their own, but for which it might be less expensive to complete these existing resources to a valid path than setting up a new one. Therefore, we evaluate all layer configurations given by $\mathfrak{T}_{\mathcal{G}}^{G(d)}$ for the processing of a demand $d \in \mathcal{D}^{\text{stat}}$ originating at layer $G(d)$. For each tuple $t = (l_1, \dots, l_k) \in \mathfrak{T}_{\mathcal{G}}^{G(d)}$, we define a *free resource graph* $G_{\text{free}}(t, d) = \big(\mathcal{V}, \mathcal{E}_{\text{free}}(t, d)\big)$, which consists of the edges $\mathcal{E}_{\text{free}}(t, d)$ of all layers listed in t with sufficient free resources for the bandwidth request of d. The construction of $\mathcal{E}_{\text{free}}$ can be formulated as

$$\begin{aligned} \mathcal{E}_{\text{free}}(t, d) := \big\{ e \in \mathcal{E}_l \mid l \in (l_1, \dots, l_k) = t \in \mathfrak{T}_{\mathcal{G}}^{G(d)} \\ \wedge\, r_{\text{free}}(e) \geq b(d) \big\}. \end{aligned} \tag{3.3}$$

The layer of the edges implies a rating, which is based on two rules of thumb derived from the CAPEX model $\mathcal{C}$. On the one hand, using edges in lower layers is usually less expensive, e.g., IP equipment is more expensive than *Optical Channel* (OCh) equipment on a per-bit basis. On the other hand, using an edge in the physical layer, a fiber, for the first time implies high cost as this means that additional equipment has to be installed in all envisaged layers at the source and destination of this fiber. We represent this rating by applying a weight function $w_{\text{axl}} : \mathcal{E}_{\text{free}} \mapsto \mathbb{R}^+$ on the edges of G_{free}, which we define as

$$w_{\text{axl}}(e) := \begin{cases} \gamma\big(G(e)\big) \cdot \dfrac{r_{\text{free}}(e)}{r_{\text{total}}(e)} & \text{if } G(e) > 0, \\ 1 & \text{if } G(e) = 0 \land e \text{ is unused}, \\ \infty & \text{otherwise}, \end{cases} \tag{3.4}$$

where the rating function $\gamma : \mathcal{G} \mapsto \mathbb{R}_0^+$ is defined as $\gamma(l) := l/|\mathcal{G}|$ in this work. Considering the utilization of the edges [12] proved a significant enhancement to the AXL algorithm presented in [8]. To find paths in the weighted free resource graph, any k-shortest path algorithm [111] as well as Dijkstra [59] can be used.

Completing AXL Paths to Multi-Layer Paths

With the previous steps of the algorithm, a weighted free resource graph $G_{\text{free}}(t, d)$ was created for each layer configuration $t \in \mathfrak{T}_{\mathcal{G}}^{G(d)}$ and CSPF algorithms were used to find candidate paths providing sufficient resources for the current demand $d \in \mathcal{D}^{\text{stat}}$.

As mentioned before, the candidate paths found in $G_{\text{free}}(t, d)$ may contain edges from any layer given by the layer configuration $t \in \mathfrak{T}_{\mathcal{G}}^{G(d)}$. Hence, the candidate paths have to be complemented with further equipment to make up a valid multi-layer path for the given layer configuration t. In particular, we have to ensure connectivity of the source and destination of demand d at $G(d)$. Figure 3.5 shows an exemplary AXL path with three layers. The dashed edges need to be installed in addition to create a valid multi-layer path from the AXL path.

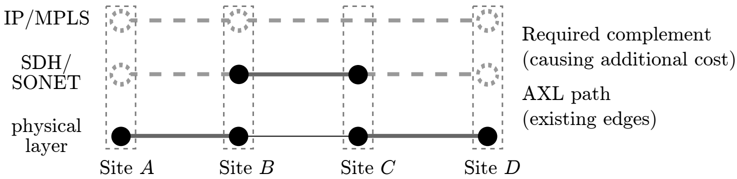


Figure 3.5: *Required complement of an* AXL *path.*

We calculate the components that need to be installed to complement the AXL candidate paths according to the corresponding interface configuration $\mathfrak{T}_{\mathcal{I}}^{G(d)}$ in a bottom-up fashion and determine the additional CAPEX caused by this completion. In this step, we can apply additional constraints and influence design decisions. We can even reject path candidates by assigning them infinite cost. Especially, this flexibility is used to consider different kinds of grooming in Section 3.3.1. Finally, we select the cheapest of the candidates and install the required equipment complement.

Re-Design

The steps of the algorithm presented up to here mostly work in a *top-down* fashion by handing the requirements of one layer down to the next lower layer. If no path with sufficient free resources can be found, a demand would be blocked. To avoid blocking, we consider *re-design* approaches to alter existing resource allocations such that new resources are established to carry further demands. As described in Section 2.1.1, different multiplexing techniques in different layers can be combined to increase the utilization of the physical resources but the additional equipment also increases CAPEX. An optimal network-wide re-design is a complex task as it requires that all established traffic demands are considered for de- and reallocation of resources. Hence, our heuristic focuses on the re-design of single edges in logical layers and applies the following two strategies to increase the amount of available resources where appropriate.

Interface upgrades: As long as few demands are in the network, inexpensive interfaces providing few resources are installed to minimize cost. If the resources of these interfaces are exceeded, we can replace them by more expensive interfaces providing more resources to be able to carry further demands. The new interface is selected according to its cost to keep CAPEX minimal. With this modification, additional resources are created without changing existing topologies. Additional cost occurs for exchanging the interfaces.

Aggregation at intermediate layers: If possible, additional transport layers are avoided for cost reasons. However, an additional transport layer can provide additional resources as an intermediate indirection between two layers that are already in use. The most common case in our studies is the use of the *Optical Multiplex Section* (OMS) layer and its *Wavelength Division Multiplex* (WDM) capabilities when the number of demands increases and more resources are needed. The routing of existing logical edges is changed to make use of the intermediate transport layer. The equipment associated with these edges is changed accordingly and the additional cost is considered.

Both re-design approaches create new free resources, which are available in the free resource graph and can be used to aggregate further edges to mitigate resource blocking and increase resource utilization. Hereby, resources are handed from lower to higher layers in a *bottom-up* fashion. If new resources were created, the path finding is repeated for demand d. As the heuristic is deterministic, the inability to create new resources would prevent infinite re-design loops.

Wavelength Continuity

While we pursue the routing part of the *Routing and Wavelength Assignment* (RWA) problem, we neither perform an explicit *wavelength assignment* nor ensure *wavelength continuity*, i.e., the availability of the same wavelength along a path in the optical layer. This can be justified by two reasons. First, today's technology can provide up to 160 wavelengths per fiber [44]. However, we only consider 40 wavelengths per fiber strand which significantly stays behind the possibilities of today's technology to avoid these considerations. Second, wavelength converters can be used, which can increase wavelength reuse by $10 - 40\ \%$ [51]. Thus, the RWA is assumed to be solved in a separate post-processing step, as done in [88].

With these steps, we are able to create a multi-layer network for given traffic matrices. Next, we consider means to evaluate the solution found by our heuristic.

3.1.3 Metrics and Objectives

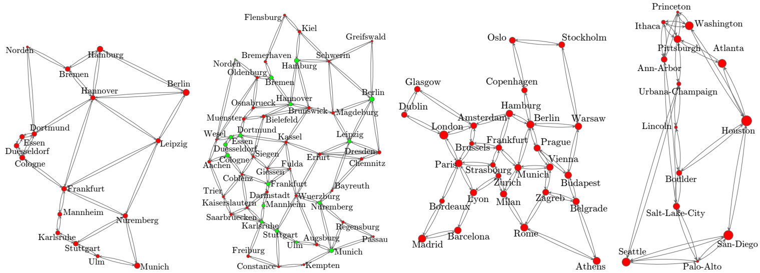
Multi-layer network design is a multi-objective optimization problem, i.e., several objectives have to be considered to find the optimum that fulfills all requirements. In our approach, we consider at least two optimization objectives, the minimization of CAPEX and the maximization of the number of demands carried by the multi-layer network. Without the second objective, the apparent result would be an empty network. To evaluate the optimality of a multi-layer network, we introduce a couple of metrics to compare the results for different parameters.

Overall cost: The design algorithm installs components at sites in different layers of the resulting multi-layer network. As discussed in Section 2.3, each component $c \in C$ is associated with a CAPEX value given by the cost function $\zeta : C \mapsto \mathbb{R}_0^+$. Thus, we define the overall CAPEX $\zeta(\mathcal{G})$ of a multi-layer network $\mathcal{G}$ by

$$\zeta(\mathcal{G}) = \sum_{v \in \mathcal{V}} \sum_{G \in \mathcal{G}} \sum_{c \in C} \zeta(c) \cdot N(v, G, c), \tag{3.5}$$

where $N : \mathcal{V} \times \mathcal{G} \times C \mapsto \mathbb{N}_0$ denotes the quantity to be installed of a component $c \in C$ at site $v \in \mathcal{V}$ in layer $G \in \mathcal{G}$.

Routed and blocked demands: For several reasons, the proposed network design approach might not be able to set up a multi-layer network that is able to carry all requested traffic demands $\mathcal{D}^{\text{stat}}$, cf. Section 3.2. Therefore, we introduce the number of *routed demands* n_{r} holding $0 \leq n_{\text{r}} \leq |\mathcal{D}^{\text{stat}}|$. Consequently, we define the number of *blocked demands* n_{b} by $n_{\text{b}} = |\mathcal{D}^{\text{stat}}| - n_{\text{r}} \geq 0$.



(a) Nobel Germany (b) Germany50 (c) Nobel Europe (d) Nobel U.S.

Figure 3.6: *The four considered physical topologies [110, 112]. (Remark, the Nobel U.S. topology is rotated counter-clockwise by 90 degrees.)*

Normalized cost: The overall CAPEX $\zeta(\mathcal{G})$ of a multi-layer network is not a suitable criterion to compare cost on different algorithms or parameters. Therefore, we normalize the overall CAPEX by the number of routed demands and the average length of the paths in the physical layer $\overline{h}_0 = \frac{1}{n_\mathrm{r}} \sum_{d \in \mathcal{D}^{\mathrm{stat}}} |R_{G(d),0}|$, i.e.,

$$\overline{\zeta}(\mathcal{G}) = \frac{\zeta(\mathcal{G})}{n_\mathrm{r} \cdot \overline{h}_0}. \tag{3.6}$$

3.1.4 Topologies and Traffic Patterns

Besides a component and CAPEX model $\mathcal{C}$, we rely on two basic input parameters, the physical topology and the envisaged traffic matrix for the network design. In the course of this work, we focus on four physical topologies depicted in Figure 3.6, which comprise the geographic area of Germany, Europe, and the United States. These are the *Nobel Germany*, *Germany50*, *Nobel Europe*, and *Nobel U.S.* topologies, which are available from the *Survivable Network Design library* (SNDlib) project [112] and mostly contributed via cooperations with large telecommunication providers [110]. Table 3.1 lists all key parameters of the four topologies. The sites of the *Nobel Germany* topology are a sub-set of the *Ger-*

Table 3.1: *Details on the considered network topologies [112].*

Topology (short name)	$\vert\mathcal{V}\vert$	$\vert\mathcal{E}_0\vert$	$\vert\mathcal{D}^{\text{stat}}\vert$
Nobel Germany (G17)	17	52	121
Germany 50 (G50)	50	176	662
Nobel Europe (EU)	28	82	378
Nobel U.S. (US)	14	42	91

many50 topology. Thus, the *Nobel Germany* topology can rather be thought of as a backbone topology.

Regarding the traffic matrices that are to be carried by multi-layer networks on top of these physical topologies, we pursue the two approaches presented in Section 2.3.3. On the one hand, we consider population-based traffic matrices, in which the population ratios determine the amount of traffic transferred to another site. These traffic matrices are based on the population data $\pi(v)$ of a site $v \in \mathcal{V}$, which is taken from recent statistics and is indicated by different sizes of sites in Figure 3.6. On the other hand, we study homogeneous traffic patterns (K-uniform), which are more academic and create a given number of K demands between any two sites. For the sake of simplicity, we focus on a homogeneous bandwidth distribution for all traffic demands. However, this is not a limitation of the presented approaches.

In the following sections, we study different variants of the algorithm presented in Section 3.1.2 in combination with different parameters and use the introduced metrics to evaluate the resulting multi-layer networks.

3.2 Characteristics of Design Heuristics

In this section, we consider different parameters of our heuristics to examine their impact on the results of the heuristics and analyze the characteristics of the proposed heuristic optimization approach. Moreover, we investigate means to mitigate the impact of the heuristic and continue our analysis with these parameters. In this section, we use the CAPEX model defined in Section 2.3 with all available technologies and without any modifications.

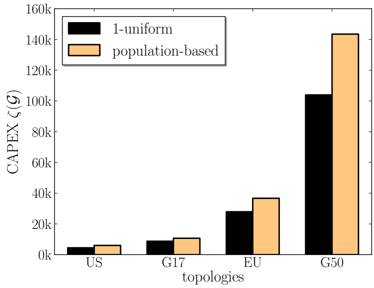
3.2.1 Topologies and Traffic Patterns

In the following, we outline the impact of the underlying physical network topology as well as the envisaged traffic matrix on the resulting multi-layer network.

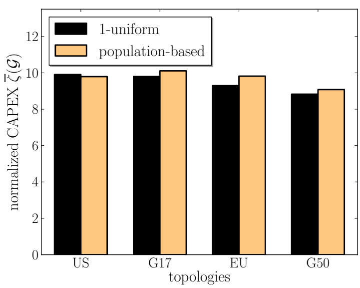
To that end, we set up homogeneous as well as population-based traffic matrices for the IP layer. For simplicity, we use a 1-uniform homogeneous traffic matrix ($K = 1$). In a first study, we consider an average bandwidth $\mathrm{E}[B^{\text{stat}}] = 10\,\mathrm{Gbit/s}$ and perform evaluations for all four physical network topologies introduced in Section 3.1.4. The results of this study are depicted in Figure 3.7. Figure 3.7a shows the resulting overall CAPEX $\zeta(\mathcal{G})$ for each pair of topology and traffic pattern. Thereby, all requested traffic demands can be routed in all cases. However, the number of traffic demands drastically differs for the different topologies as $|\mathcal{D}^{\text{stat}}| = |\mathcal{V}| \cdot (|\mathcal{V}| - 1)$ for the homogeneous traffic pattern. Thus, the overall CAPEX is not suited to compare multi-layer networks with different physical topologies and traffic matrices.

For this purpose, we consider the normalized CAPEX $\overline{\zeta}(\mathcal{G})$ of a multi-layer network, which is illustrated in Figure 3.7b for the same test cases. As defined in Equation 3.6, this metric respects the number of routed demands n_{r} and the average path length $\overline{h}_0$ in the physical layer of all demands. In the considered evaluations, normalized cost proves to be a suitable metric to compare multi-layer networks on different physical topologies and traffic patterns, like population-based and 1-uniform, with a slight variation of a few percent.

As homogeneous and population-based traffic patterns make up a difference of a few percent with regard to normalized CAPEX $\overline{\zeta}(\mathcal{G})$, we restrict the following studies to homogeneous traffic patterns.



(a) overall CAPEX $\zeta(\mathcal{G})$.



(b) normalized CAPEX $\overline{\zeta}(\mathcal{G})$.

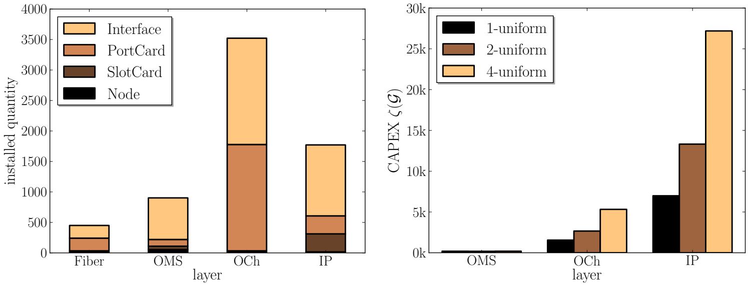
Figure 3.7: *The four considered topologies with different traffic patterns.*

3.2.2 Cost Structures and Increasing Traffic

Besides the overall CAPEX of a multi-layer network, we want to outline the composition of multi-layer networks into components and component groups separated by layers with increasing traffic.

First, we analyze the quantity of installed components per technology split up into the four component groups (nodes, slot cards, port cards, interfaces) defined by the CAPEX model in Section 2.3. Figure 3.8a shows the installed quantity of components caused by a homogeneous, 1-uniform IP traffic matrix on the *Nobel Germany* topology. The figure shows several CAPEX model specific properties, like the one-to-one relation of port cards to interfaces in the OCh layer and our zero-cost dummy component extensions. These relations are important when considering cost variations as done in Section 3.4.

While the overall amount of traffic was kept constant in the last study, we now investigate the impact of increasing traffic. We use an average bandwidth $\mathrm{E}[B^{\mathrm{stat}}] = 10\,\mathrm{Gbit/s}$ and scale the traffic matrix by increasing the parameter K of the homogeneous traffic matrix. Since an increase in overall CAPEX is quite apparent, we also consider the impact on cost in different technologies of



(a) cost structure with 1-uniform traffic

(b) increasing traffic, K-uniform

Figure 3.8: *Cost structure split by component groups and layers as well as impact of increasing traffic on cost distribution to different layers.*

the multi-layer networks. Therefore, we separate the cost by individual layers as depicted in Figure 3.8b.

We choose the *Nobel Germany* topology, which is well suited for increasing traffic. The case for $K = 1$ is identical to the ones depicted in Figure 3.7. The results from Figure 3.8b are twofold. On the one hand, we see that IP equipment is the dominant cost factor in all three evaluations. It is followed by OCh equipment, which mostly comprises optical transponders and muxponders. Last, the WDM components of the OMS equipment make up a fraction of the overall cost. On the other hand, the evolution of cost for the different technologies is illustrated with increasing traffic. It shows that the OMS cost more or less stagnates as almost all sites have been equipped with OMS components already for the smallest traffic matrix. In contrast, the OCh cost shows an almost linear increase with the additional number of demands. Finally, the IP cost rather increases exponentially which is for two reasons. First, the CAPEX model considers a dramatic cost increase from 16.67 to 111.67 cost units when the smallest available IP node needs to be upgraded to the next higher one which is a multi-chassis node. Second, the price of an IP interface at 40 Gbit/s is more than four times higher then the price of one at 10 Gbit/s. However, multi-chassis nodes as well as 40 Gbit/s interface are required to carry the highest considered traffic scaling ($K = 4$).

3.2.3 Impact of Demand Order

As mentioned in Section 3.1.2, the proposed heuristic processes demands sequentially and uses the given input order of the traffic demands, which reduces the complexity of the considered optimization problem. In contrast to *Integer Linear Programs* (ILPs), this creates a new optimization sub-problem to find an optimal order of demands.

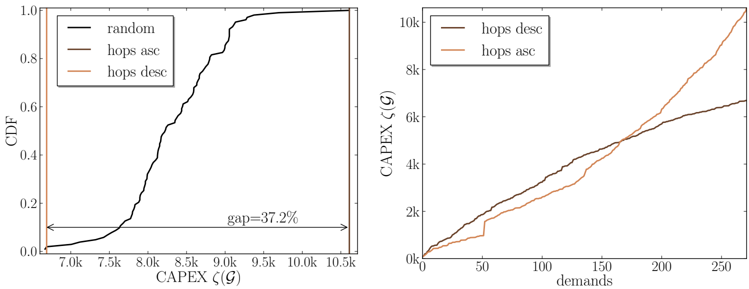
The demand order has a strong impact on CAPEX and an inappropriate demand order might cause drastically increased CAPEX as well as blocking of demands. Therefore, we consider different demand order strategies that we compare in the following. We evaluate (a) the number of hops of the shortest path in the physical layer, as also done in [84], and (b) the bandwidth of a demand. Both strategies can be applied in ascending or descending order. In addition, we hypothetically consider two more elaborated strategies:

- Determining the demand order by evaluating which demand would be the cheapest to route next (greedy look-ahead strategy). However, this approach requires $|\mathcal{D}^{\text{stat}}| \cdot \big(|\mathcal{D}^{\text{stat}}| + 1\big)/2$ runs.

- Considering all $|\mathcal{D}^{\text{stat}}|!$ demand orders which yields the optimal order.

In general, look-ahead strategies cannot guarantee an optimal order as demands chosen at the beginning – due to their originally low cost – can lead to more expensive results for following demands.

We demonstrate the performance of the proposed strategies and compare them with 100 random demand orders. The results for the *Nobel Germany* topology are shown in Figure 3.9. For the sake of completeness, Table 3.2 lists the main results for all considered topologies and shows that the values are in the same order of magnitude.



(a) distribution of $\zeta(\mathcal{G})$ (b) evolution of $\zeta(\mathcal{G})$

Figure 3.9: *Different demand order strategies upon the Nobel Germany topology.*

The *Cumulative Distribution Functions* (CDFs) for CAPEX shown in Figure 3.9a indicates a gap of almost 40 % in CAPEX between best and worst demand order strategy. This stresses the importance of the demand order. Due to the consideration of a homogeneous bandwidth distribution, the results for the strategies with sorting by bandwidth are not illustrated. However, additional studies on bandwidth ascending and descending demand order were performed, which allow to draw the conclusion that sorting by bandwidth has some impact, which is rather negligible. Furthermore, we see that sorting by hops in descending order (hops desc) performs very good while sorting by hops in ascending order (hops asc) performs worst in our studies. This finding is confirmed for the other topologies as well.

Figure 3.9b depicts the evolution of cost for the best and worst case, i.e., hops in descending and ascending order, respectively. The figure shows that the hops-descending strategy leads to higher cost up to the 160th demand compared to the hops-ascending strategy. However, the additional cost spend in the beginning opens up the possibilities to realize further demands at lower cost compared to hops-ascending. This also emphasizes once more the inappropriateness of look-ahead strategies for cost-aware multi-layer network design.

Table 3.2: *Results of* 100 *runs with random demand order.*

Topology	min. CAPEX	avg. CAPEX	max. CAPEX	std. dev. (σ)
US	4059.19	4937.71	5975.53	392.61
G17	6666.19	8318.22	10610.29	614.03
EU	24532.75	27331.06	38551.34	1638.72
G50	98496.73	105776.97	148353.89	5211.52

As the hops-descending strategy proved to be the best while still being efficient, we use this strategy in all following studies.

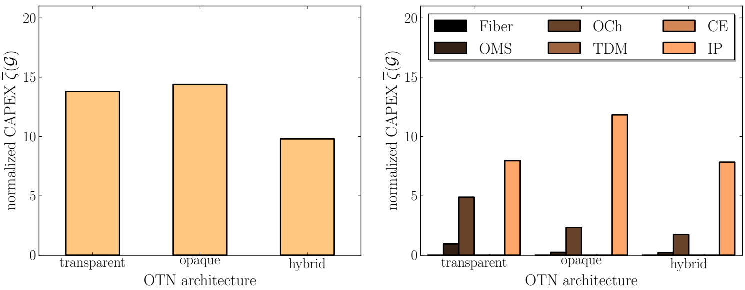
3.3 Network Architectures

The network architecture is a decisive factor in multi-layer network design, which is indirectly given by customer constraints to the available equipment. Thus, we can achieve different network architectures by varying the CAPEX model $\mathcal{C}$. Such variations include the removal of certain equipment to prevent that a technology is used or bypassed. We explicitly do not change any cost values of the CAPEX model in this section.

3.3.1 Optical Network Architectures

As discussed in Section 2.2.1, there are different ways to restrict grooming in the *Optical Transport Network* (OTN), which we split up into OMS and OCh layer. These OTN architectures are motivated by different kinds to approach the RWA problem and the existence or absence of wavelength converters.

To explain the following results, we delve into the effects of the different OTN architectures. An OTN architecture is called *opaque* if wavelength assignment is done on a hop-by-hop basis, resulting in a 1-hop OCh topology of light-paths and all layers above. In this case, wavelength conversion is not necessary as wavelength continuity on a single hop is always guaranteed and data must be processed



(a) normalized CAPEX $\overline{\zeta}(\mathcal{G})$ (b) normalized CAPEX $\overline{\zeta}(\mathcal{G})$ by layer

Figure 3.10: *Different* OTN *architectures with the Nobel Germany topology.*

in a higher layer at each hop. An OTN architecture is called *transparent* if a single wavelength is used for a complete (multi-hop) light-path. The resulting OCh topology consists of end-to-end light-paths only. This strict requirement is hardly solvable if no wavelength conversion is considered in the optical layer. A mixture of opaque OTN and transparent OTN is called *hybrid*, which we considered in all previous studies. In our case, the hybrid architecture arbitrarily combines transparent and opaque light-paths to achieve a cost-minimal network design.

As in [88], we consider the three OTN approaches and illustrate the impact on normalized CAPEX $\overline{\zeta}(\mathcal{G})$ as well as the distribution of cost on individual layers for the *Nobel Germany* topology in Figure 3.10. Herein, cost of fiber, *Time Division Multiplex* (TDM), and *Carrier Ethernet* (CE) are zero and not visible since none of this equipment was used. Transparent OTN seems to be second to best of the OTN architectures. In fact, transparent OTN is the most expensive of the considered variants since it is only able to route $n_{\mathrm{r}} = 102$ of 272 traffic demands while it is possible to route all demands with the other two OTN architectures. However, this breaks the comparability of the architectures as setting up further demands actually means infinite cost for transparent OTN in this study.

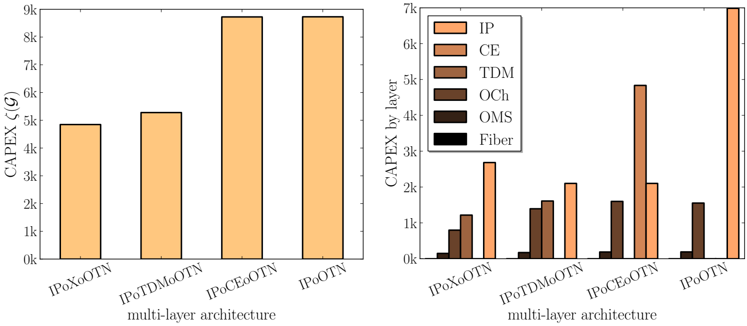
Of special interest are the cost of the individual layers. Transparent OTN is most expensive mainly caused by the amount of IP equipment due to the direct light-paths from source to destination of a demand since no intermediate aggregation at IP layer can be performed to use synergies, e.g., at OCh layer. The opposite approach, opaque OTN, results in lower cost than transparent OTN, but still enforces the installation of several IP interfaces at each site. Only the hybrid OTN architecture, as used by our AXL approach, is able to further reduce the amount of IP interfaces by performing arbitrary grooming.

3.3.2 Multi-Layer Network Architectures

So far, we mostly considered multi-layer networks that contain an OTN transport layer in addition to the IP layer. However, there are reasons for operators to have multiple layers. On the one hand, operators need to provide technologies beyond IP, like Carrier Ethernet or SDH, if there is a sufficient request from customers for this additional service. In this case, the multi-layer network design needs to take care of multiple technologies to be able to find an optimal solution. We consider this case in Section 3.3.3. On the other hand, the consideration of more than just OTN as a transport technology might be a mean to further reduce cost.

In either case, certain layers are obligatory or forbidden in the multi-layer network design. The available layers form a *multi-layer architecture*. A typical example of such an architecture is IP over OTN (in short, *IPoOTN*), which determines that only IP and OTN technology may be used. Analogously, we define *IPoTDMoOTN*, which uses SDH/SONET as a TDM-based transport layer, and *IPoCEoOTN*, which does the same using CE. In addition to these three architectures, we let our algorithm decide which layers to use to transport the data. We refer to this scenario by *IPoXoOTN* where *X* is arbitrarily chosen from OTN, TDM, or CE on a per-demand basis.

The results of this comparison on the *Nobel Germany* topology are depicted in Figure 3.11. For each multi-layer architecture, Figure 3.11a illustrates the overall CAPEX of the calculated solution and Figure 3.11b shows the CAPEX split up by



(a) overall CAPEX $\zeta(\mathcal{G})$ (b) CAPEX by layer

Figure 3.11: *Comparison of four different network architectures.*

layer. From Figure 3.11a, we can see that IPoXoOTN and IPoTDMoOTN yield the most cost-efficient solutions with regard to the considered CAPEX model. Especially, IPoOTN, which is seen as the simplest architecture from a management perspective, is comparatively expensive. Figure 3.11b shows the shift of IP to CE equipment from IPoOTN to IPoCEoOTN architecture. However, the overall cost for these architectures is almost identical in the considered scenario. Figure 3.11b allows to gain an insight how our algorithm is able to achieve a more cost-efficient solution if the freedom to choose layers per demand is given. Whilst more cost is spent on IP equipment in the IPoXoOTN architecture, this can be more than compensated by cost savings in OCh and TDM layer.

The previous two studies covered different network architectures while traffic only originated from IP layer and further layers were only used for transport. In the next study, we also use several transport layers but consider traffic on multiple layers and investigate whether it is reasonable to simplify such an architecture with regard to CAPEX.

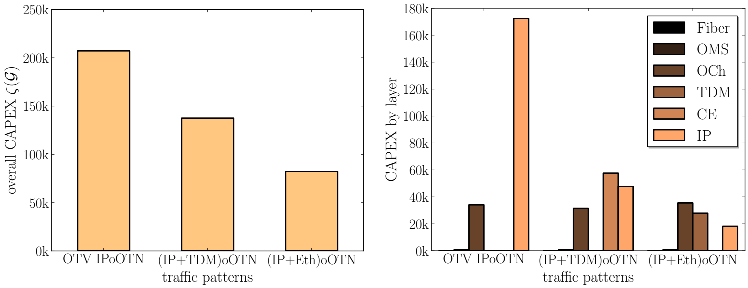
3.3.3 Tunneling Legacy Services

We showed that having multiple transport technologies within a multi-layer network is reasonable from a CAPEX point of view. However, any additional technology causes maintenance and management overhead during operation. Therefore, an ongoing trend for future network deployments is the reduction to the technologies that are essentially required.

In our case, this is IP/MPLS since we consider this kind of traffic as well as OTN as a powerful transport network with the potential to be extensible in the future [39]. As a network operator cannot simply quit contracts for legacy services, different solutions have come up to transport legacy services like Ethernet or SDH via the IP layer. That means, legacy services are encapsulated and transmitted through IP and therefore just are seen as additional IP traffic from the planning point of view. Recent approaches to realize such piggy-backing of services are Cisco's *Overlay Transport Virtualization* (OTV) [113] or Alcatel-Lucent's *Virtual Private LAN Service* (VPLS) [114].

While such approaches lead to reduced management and maintenance overhead, the consequences regarding CAPEX are not clear. Therefore, we perform a study on piggy-backing legacy service on the IP layer and compare such a solution to the original multi-layer network providing two services in their original form. Besides the existence of IP, we consider Ethernet and SDH services as a second service provided by a multi-layer network. Thereby, all services cause an identical amount of traffic such that piggy-backing the other service via IP simply doubles the IP traffic. The results of this study are shown in Figure 3.12. In Figure 3.12a, we can easily see that this way of simplifying the management and maintenance raises the overall CAPEX by $50 - 130\ \%$. The reason of this dramatic cost increase can be identified by the drastically increased quantity of IP interfaces, which is illustrated in Figure 3.12b.

As a consequence of the obtained results, a simplification to IPoOTN cannot be justified from a mere CAPEX point of view. However, the main motivation to perform such transitions is in reduced *Operational Expenditures* (OPEX) over



(a) overall CAPEX $\zeta(\mathcal{G})$

(b) CAPEX by layer

Figure 3.12: *Cost for piggy-backing services on* IP*/MPLS.*

a long time span, which can compensate such initial additional charge. Furthermore, cost is in a continuous process of change. Thus, studies like these need to be repeated with updated cost values in case of cost changes and their results can only represent a snapshot of an ongoing development.

To that end, we consider the variability of cost in the following two studies.

3.4 Cost Variations

As stated before, cost studies are snapshots of a certain point in time since the market of networking equipment is moving continuously. While equipment cost typically decreases over time, the introduction of new, more powerful equipment, e.g., 100 Gbit/s interfaces, can break up the current market situation and price structures. To that end, we perform two studies on equipment cost changes. The first considers a price discount on a whole technology, while the second investigates the introduction of and the trend towards new more powerful equipment.

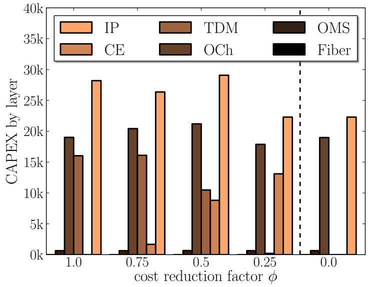
3.4.1 Carrier Ethernet Discounts

As shown in Section 3.3.2, especially Figure 3.11b, Carrier Ethernet is not used if our algorithm is given the choice to select other layers. This is due to the rather expensive cost structure of the CAPEX model for Carrier Ethernet equipment, which was still a new concept at the time this cost model was constituted. We already stated that a market is continuously changing and the cost of newly available equipment decreases with time as it gets more and more common.

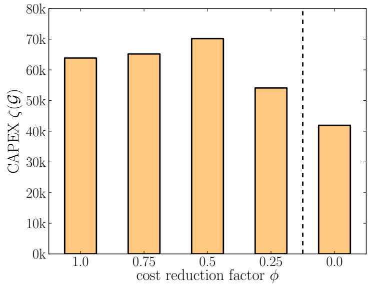
Therefore, we pick up our former studies but apply a cost reduction factor $\phi \in \{1, 0.75, 0.5, 0.25, 0\}$ on all Ethernet equipment defined in the CAPEX model $\mathcal{C}$. The results of this study are shown in Figure 3.13. As we can see in Figure 3.13a, the usage of Ethernet equipment increases with decreasing ϕ. Hereby, $\phi = 0$ serves as a theoretical threshold since cost for Ethernet equipment cannot be further reduced and also shows whether a solution that uses Ethernet is feasible at all. So in this case, the overall CAPEX is exactly the cost of all equipment except for Ethernet as Ethernet equipment is for free in this case. Figure 3.13b shows the impact of the cost reduction on the overall CAPEX $\zeta(\mathcal{G})$. Counterintuitively, the overall CAPEX increases at first when applying the cost reduction factor $\phi < 1$. In this case, it is preferable for the greedy heuristic to route a demand using Ethernet layer but this liaison to the Ethernet layer comes at the price of higher cost for all demands. Thus, the break-even for Ethernet equipment in this CAPEX model with regard to the cost reduction factor ϕ is with $\phi < 0.5$ as is shown by the reduction of overall CAPEX for $\phi = 0.25$.

3.4.2 Increasing Line Rates

The ongoing increase of IP traffic forces network providers to continuously increase the data rates in their IP networks. At the time the CAPEX model [50] was created, 40 Gbit/s IP interfaces were quite new and their cost-benefit ratio was higher compared to 10 Gbit/s IP interfaces. In particular, having four times more capacity comes at a cost increase of factor 5.82 when choosing one 40 Gbit/s interface over four 10 Gbit/s IP interfaces.



(a) CAPEX by layer



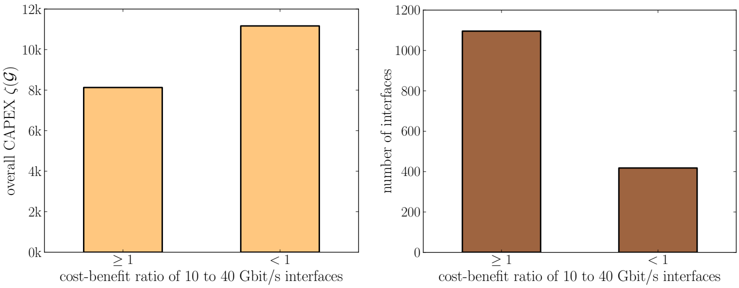
(b) overall CAPEX $\zeta(\mathcal{G})$

Figure 3.13: *Varying Ethernet cost using a cost reduction factor* ϕ*.*

Several studies are performed that gradually introduce new line rates. We start with merely 10 Gbit/s IP interfaces and increase the ratio of 40 Gbit/s IP interfaces until it is the only type of interface. For the gradual introduction, cost increase factors of 5.5 and 3.5 are used. The latter seizes a rule of thumb for the introduction of new equipment and prefers the installation of one 40 Gbit/s IP interface over four 10 Gbit/s IP interfaces. We use the IPoOTN architecture to minimize the impact of other technologies.

The results of these studies are depicted in Figure 3.14. The gradual introduction showed that our algorithms refuse to install 40 Gbit/s IP interfaces as long as the cost increase factor is greater than four. This behavior would be different in cases where 10 Gbit/s IP interfaces do not provide sufficient resources but can be easily replaced with 40 Gbit/s IP interfaces by the re-design mechanisms presented in Section 3.1.2. So in general, the acceptance of new more powerful equipment relies on clear cost benefits if decisions are merely based on CAPEX.

As shown in Figure 3.14a, the overall CAPEX of the 40 Gbit/s IP interfaces is higher than that of the 10 Gbit/s IP interfaces even though the cost-benefit ratio is below one. The reasons for this increase are twofold and need to be carefully considered. On the one hand, the higher granularity of the 40 Gbit/s IP interfaces



(a) overall CAPEX $\zeta(\mathcal{G})$

(b) number of interfaces

Figure 3.14: *Cost modifications causing a shift of 10 to 40 Gbit/s interfaces.*

cannot be fully utilized in all cases. Thus, additional capacity is provided but wasted. On the other hand, the number of interfaces cannot be decreased to the aspired quarter of the number of 10 Gbit/s IP interfaces, cf. Figure 3.14b.

3.5 Lessons Learned

This chapter proved the versatility of our heuristic approaches in multi-layer network design with regard to the huge number of parameters and degrees of freedom, cf. Chapter 2. The efficient implementation of the heuristics in a software tool [16] that comprises visualization was a prerequisite to get deeper insights into the complex interrelations of multi-layer networks with different parameters.

We saw that the heuristics introduced in this chapter are able to find solutions for realistic scenarios and to fulfill their traffic demands. In addition, the sequential processing of demands allows to easily append additional demands at no extra complexity. We found that the impact of physical topologies and traffic patterns can be compensated by means of a specific cost metric that allows for a simple comparison across these parameters. Our studies on the distribution of cost in different layers and increasing traffic volumes revealed that the WDM technology in the OMS layer was almost fully deployed in all considered scenarios such that this could be seen as a constant and rather negligible part. As a result, it seems eligible to assume OTN equipment, like *Optical Cross Connects* (OXCs), at each site. An analysis on the impact of the sequential demand order reveals its importance for the results of the multi-layer network design heuristics. From this analysis, we recommend to use the hops-descending strategy.

Studies on the consequences of determining network architectures manifested a trade-off between cost and complexity with the considered CAPEX model. The results reveal that neither opaque nor transparent OTN architectures are able to yield optimal results although these might be preferable for management and strategic reasons. Given all degrees of freedom, our heuristics showed to find hybrid OTN solutions that outperform the latter two OTN architectures with regard to CAPEX. The trade-off was confirmed in studies on the impact of general multi-layer architectures. We found that complex architectures are preferred over simple ones with regard to the underlying CAPEX model. Again, the best results are achieved when our heuristics are given the freedom to choose from all layer interconnections on their own. In our studies, we identified the trade-off between complexity and CAPEX. Studies on tunneling legacy services reveal that architectures being simple from a management perspective can be by far more expensive in terms of CAPEX than more complex architectures. Therefore, the consideration of CAPEX and OPEX is essential to decide for a network architecture or tunneling legacy services.

Motivated by these findings, we performed studies on cost variations, i.e., changing the cost values of the given CAPEX model, to explore its implications. We showed that CE technology is too expensive in this CAPEX model and, thus, is not used by our heuristics until cost of CE are reduced by at least 50 %. Another study on the gradual introduction of interfaces with higher bit rates reveals that the cost-benefit ratio of these new interfaces must distinctly drop below 1 to foster the deployment of such equipment when only CAPEX is considered.

The heuristic and analysis presented in this chapter are the foundation for resilience and provisioning in multi-layer networks in the remainder of this work.

4 Multi-Layer Network Resilience

Remember, with great power comes great responsibility.
"Spider-Man", 2002.

In Chapter 3, we introduced heuristics to design multi-layer networks. The focus of the previous chapter is on general design principles and their optimization. In this chapter, we shift our focus to resilient multi-layer network design to adequately cover the large parameter space of resilience in multi-layer networks, as described in Section 2.2.1.

Protection mechanisms enable to guarantee a certain *Quality of Service* (QoS) also in case of failures, i.e., to *survive* and to recover from certain failures and, thus, increase a network's *survivability* and make it *resilient*. As we consider circuit-switching technologies, traffic requires well-defined backup paths. Similar to Chapter 3, we focus on the minimization of *Capital Expenditures* (CAPEX). However, we strive for routing and protecting as many traffic demands as possible. Evaluations are again performed using the *Multi-Layer Network Engineering and Optimization* (MuLaNEO) software [16] briefly described in Section 3.1.1. The protection heuristics developed in this chapter are based on the design heuristics presented in Chapter 3. We perform an additional evaluation of where, how, and how much backup capacity is installed to fulfill a traffic demand's requirements. In this chapter, we evaluate the performance of the developed heuristics in different protection scenarios with respect to CAPEX and fulfillment of requirements. Additionally, we study the consequences of failures and different failure types in multi-layer networks.

4.1 Necessity and Variants of Protection

Networks comprise a plethora of components. We introduce definitions and structures of possible failures of these components as well as corresponding protection mechanisms in multi-layer networks.

4.1.1 Availability and Failures

Contracts between customers and providers formulate *Service Level Agreements* (SLAs), which determine, amongst others, the postulated *availability* of a specific service. Requirements for availability range up to $99.999\ \%$ (so called "five nines") or even higher. Consequently, there might be a maximum overall outage of roughly five minutes per year. Violating such requirements can lead to penalty fees, which can also be part of an SLA. A statistical, more formal definition of availability of a network element is the probability p_A of a network element to be operational. This probability is derived from the *Mean Time Between Failures* (MTBF), i.e., the average time between the occurrence of two failures affecting the same component, and the *Mean Time To Recover* (MTTR), i.e., the average time that a component is not operable after a failure occurred. Thus, the availability of a network is defined by

$$p_A = 1 - \frac{\text{MTTR}}{\text{MTBF}}. \tag{4.1}$$

Failures, e.g., being caused by hardware, software, or human error as well as natural disaster, decrease the availability.

Thus, network providers need to be able to estimate the frequency of such failures and their consequences. To that end, studies on the availability of network nodes and links, e.g., [42, 43], as well as studies on the availability of different equipment in multi-layer networks, e.g., [85], have been conducted. These studies provide statistical data for the failure of certain network components and allow to give a probability for certain failure scenarios [1]. Due to studies [115, 116] on the frequency and extent of typical failures, we focus in this work on three failure types out of the set of all possible failure scenarios $\mathcal{F}$.

We consider the *set of single physical link failures* $\mathcal{F}_{\mathcal{E}_0}$ with

$$\mathcal{F}_{\mathcal{E}_0} = \big\{\{e\} \mid e \in \mathcal{E}_0\big\}, \tag{4.2}$$

the *set of single layer-node failures* $\mathcal{F}_{\mathcal{V}\cap G}$ comprising all incident links of a site, i.e., all links starting (α) or ending (ω) at a site, at a layer $G \in \mathcal{G}$, given by

$$\mathcal{F}_{\mathcal{V}\cap G} = \bigcup_{v\in\mathcal{V}} \big\{\{e \in \mathcal{E}_G \mid \alpha(e) = v \vee \omega(e) = v\}\big\}, \tag{4.3}$$

and the *set of single total-node failures* $\mathcal{F}_{\mathcal{V}}$ with

$$\mathcal{F}_{\mathcal{V}} = \bigcup_{G\in\mathcal{G}} \mathcal{F}_{\mathcal{V}\cap G} \tag{4.4}$$

Notably, any of these failure scenarios may cause multiple link failures in a multi-layer network, which is covered in the following.

4.1.2 Definitions and Metrics

Analogous to Section 3.1.3, we define several terms and metrics for the evaluation of resilient multi-layer networks.

Protecting layers: In principle, each considered technology can provide a resilience mechanism. We denote the set of layers with active protection mechanism, the *protecting layers*, by $\mathcal{G}_P \subseteq \mathcal{G}$. For *single-layer resilience* as considered in Section 4.2, $|\mathcal{G}_P| = 1$ holds and $|\mathcal{G}_P| > 1$ holds for *multi-layer resilience* as examined in Section 4.3.

Protected layers: A layer above a protecting layer that is protected by this protecting layer is denoted as a *protected layer*.

Protected demands: In addition to the number of routed demands n_r, cf. Section 3.1.3, we introduce the number of demands with backup routings denoted by $n_{\mathrm{p},G}$, where $G \in \mathcal{G}_P$. Hereby, $0 \leq n_{\mathrm{p},G} \leq n_\mathrm{r} \leq |\mathcal{D}^{\mathrm{stat}}|$ always holds since it is possible that no backup path can be found for a routed demand.

Normalized CAPEX: For resilient networks, the definition of normalized CAPEX $\hat{\overline{\zeta}}(\mathcal{G})$ differs from $\overline{\zeta}(\mathcal{G})$ defined in Equation 3.6 to consider the definition of primary and backup paths for a demand in different protecting layers. It is given by

$$\hat{\overline{\zeta}}(\mathcal{G}) = \zeta(\mathcal{G}) \cdot \frac{\overline{h}_0}{n_\mathrm{r}} \cdot \prod_{G \in \mathcal{G}_P} \frac{\hat{\overline{h}}_{G,0}}{n_{\mathrm{p},G}}, \tag{4.5}$$

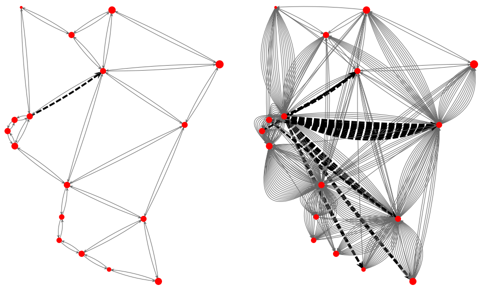
where $\hat{\overline{h}}_{G,0} = \frac{1}{n_{\mathrm{p},G}} \sum_{d \in \mathcal{D}^\mathrm{stat}} |\hat{R}_{G(d),0}|$ is the average path length of backup paths $\hat{R}_{G(d),0}$ of demand $d \in \mathcal{D}^\mathrm{stat}$ on protecting layer $G \in \mathcal{G}_P$ in the physical layer.

4.1.3 Shared Risk Groups and Failure Discovery

The use of a hierarchy of transport layers as comprised by multi-layer networks bears the risk of two characteristic issues caused by the inherent structure of a multi-layer network.

The first of these characteristics are *Shared Risk Groups* (SRGs). An SRG is a set of components whose operability is bound to the operability of another, common network entity. For example, all links using a common resource form a group sharing the risk of this resource's failure. An SRG caused by a single physical link failure and its consequences in higher layers of a multi-layer network are illustrated in Figure 4.1. Herein, a single failure propagates a cascade of further failures [53]. Therefore, we distinguish *original* and *subsequent* failures – also called *primary* and *secondary* failures [43]. For each failure scenario $\mathcal{F}_x \in \mathcal{F}$, we define the set of subsequent failures $\mathcal{F}_x^\mathrm{sub}$ where $\mathcal{F}_x \subseteq \mathcal{F}_x^\mathrm{sub} \in \mathcal{F}$ holds. In general, SRGs increase the risk of contract penalty fees due to violations of SLAs and, hence, intensify the importance of resilience mechanisms in multi-layer networks.

A second characteristic regarding resilience in multi-layer networks is the inability to discover certain failures. Any failure in a protecting layer is detected and covered by the protection mechanism itself, e.g., by switching from the primary



(a) fiber, original failure (b) OCh, subsequent failures (c) IP, subsequent failures

Figure 4.1: *Subsequent failures of a single link failure in the physical layer.*

to the backup path. As described above for SRGs, failures in lower layers propagate to higher layers and, thus, cross any protecting layer, which cares about the failure. However, failures originating above the highest protecting layer cannot be detected since protection mechanisms only check for connectivity on the protecting layer but not whether any meaningful data is transmitted. Hence, there is no malfunction from the point of view of lower layers. Such a situation is illustrated in Figure 4.2, which might be caused by a failure of an interface or port card. As a consequence, the logical connection in the protected layer is broken, but no failure is detected by or signaled to the lower protecting layer. In Section 4.3, we discuss how such failures can be handled.

In any case, special care must be taken regarding these two characteristics of multi-layer networks as well as resilience when developing design heuristics for resilient multi-layer networks.

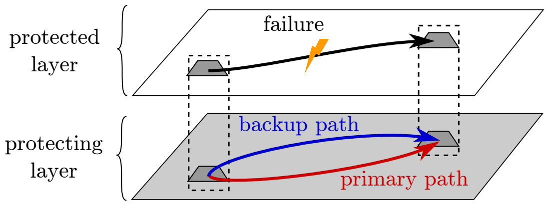


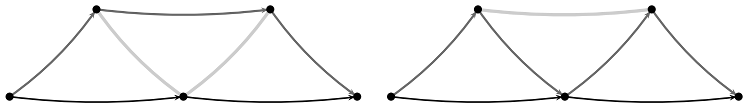
Figure 4.2: *A failure that cannot be discovered by the protection mechanism.*

4.1.4 Characteristics of Protection

Due to the numerous degrees of freedom in network design, there is a variety of possibilities to realize protection in multi-layer networks. A survey on resilience mechanisms and methodologies as well as a classification of protection variants is given in [53]. The concepts of resilience in multi-layer networks are also explained in [43] while the *Protection Across NEtwork Layers* (PANEL) project [49] focused on multi-layer resilience. We distinguish protection variants by the following criteria.

Assignment: Backup paths can be assigned in two ways. On the one hand, a backup path can be assigned *link-wise*, e.g., to apply protection on a link carrying important traffic. On the other hand, backup paths can be assigned *demand-wise*, i.e., protection is provided on the basis of business models and customer requests.

Disjointness: As mentioned earlier, primary and backup paths need to be disjoint to be useful. Hereby, we distinguish disjointness with respect to links, i.e., primary and backup paths must not share any common link, and disjointness with respect to nodes, i.e., primary and backup paths must not pass any common node except for the source and sink of a demand. In general, node-disjointness implies link-disjointness since links connect nodes. If the node failure probability is rather negligible, link-disjoint protection can be sufficient to guarantee the desired resiliency.



(a) path protection (b) link protection

Figure 4.3: *Different protection mechanisms for a demand.*

Layers: We defined earlier the set of protecting layers $\mathcal{G}_P \subseteq \mathcal{G}$. For technological reasons, the physical layer, G_0, and the *Optical Multiplex Section* (OMS) layer, G_1, cannot be chosen for protection since there is no equipment to process and to duplicate data streams. Depending on the number of protecting layers, we refer to *single-layer resilience* and *multi-layer resilience*.

Path set-up: There are two alternatives with regard to when backup paths are established. With *protection switching*, backup paths are set up "a priori", i.e., in advance of failures. Traffic is simultaneously sent on primary and backup paths, i.e., traffic is duplicated. With *restoration*, backup paths are set up right after a failure appeared, i.e., "ad-hoc". There are also combinations of protection switching and restoration, e.g., to cover multiple failures and set up additional backup paths when the primary or backup paths are broken due to a failure.

Resource usage: There are two possibilities to use resources for backup paths. With *dedicated* usage, backup resources are exclusively reserved for a single demand and are always available for this demand. With *shared* usage, backup resources can be assigned to multiple demands. In contrast to dedicated usage, backup resources might not be available in case multiple of the assigned demands are affected by failures. A general notation for the resource usage defines N primary and M backup paths, $N \geq M$. The syntax M+N is used for dedicated and M:N for shared protection, cf. [42, 43, 53]. In this work, we especially use 1+1 dedicated protection.

Scope: There are two commonly considered variants in the scope of protection, *path protection* and *link protection*, which are illustrated in Figure 4.3. While path protection sets up a disjoint path from the source to the sink of a traffic demand, link protection sets up a backup path for each link of the primary path. Thus, traffic sticks to the primary path as much as possible with link protection, while it can be completely detoured in case of path protection.

In this work, we focus on the commonly used demand-wise, dedicated 1+1 path and link protection with link or node disjointness in single-layer and multi-layer resilience scenarios. However, the proposed methodologies, which we introduce and evaluate in the next sections, are not limited to these protection variants.

4.2 Single-Layer Resilience

In this section, we introduce and evaluate algorithms for the design of multi-layer networks that comprise a single protecting layer.

4.2.1 Algorithms for Single-Layer Protection

The algorithms to be presented for single-layer resilience are based on the heuristics for non-resilient multi-layer network design introduced and evaluated in Chapter 3. Hence, the following algorithms comprise all parameters used in non-resilient network design and additionally respect the constraints for resilience described in Section 4.1.4.

Compared to [8], our "protected" *Auxiliary Cross Layer* (AXL) algorithm [12] realizes dedicated 1+1 protection for each traffic demand on a given layer $\mathcal{G}_P$ with different resilience requirements. The heuristic sequentially processes demands in the given input order. In general, this demand-wise processing allows to apply a different resilience mechanism for each traffic demand. For the sake of simplicity, we apply the same resilience mechanism to each traffic demand. If it is not possible to set up a backup path for a demand, only the primary path is established. We developed three heuristics for single-layer resilience: *Separate* (SEP),

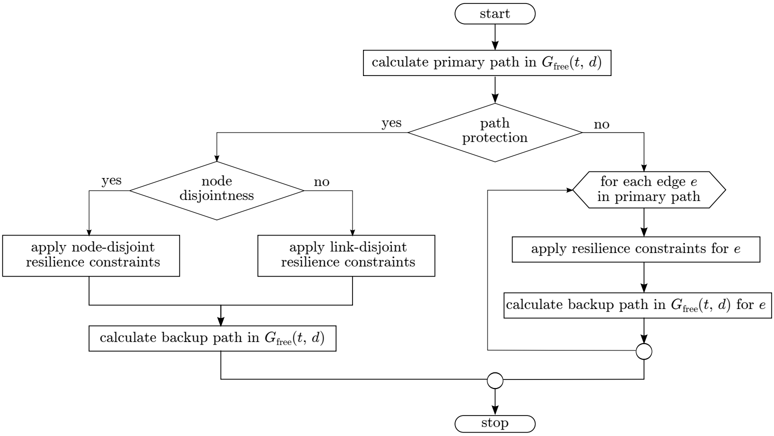
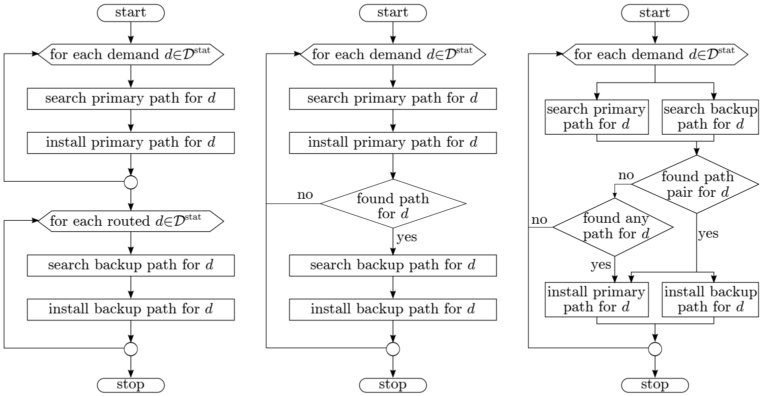


Figure 4.4: *Schematic flow chart of resilient path finding extending Figure 3.3.*

Alternating (ALT), and *Combined* (COM) search for primary and backup paths and corresponding protection mechanisms.

All three heuristics have a common basis to calculate backup paths for traffic demands, which is illustrated in Figure 4.4 and, in addition to Figure 3.3, covers the scope and disjointness parameters of protection, cf. Section 4.1.4. The hidden challenges in the calculation of paths for resilient multi-layer network design are the disjointness constraints of primary and backup paths, which not only affect the protecting layers $\mathcal{G}_P$ but all layers below, especially the physical layer that is passed by all traffic. The same holds for the re-design mechanisms presented in Section 3.1.2. In particular, the re-design technique *"aggregation at intermediate layers"* must not aggregate edges whose realization must be disjoint in the physical layer. To that end, we apply the disjointness criteria on the free resource graph G_{free}, introduced in Section 3.1.2, by forbidding those edges from solutions that would break the disjointness.

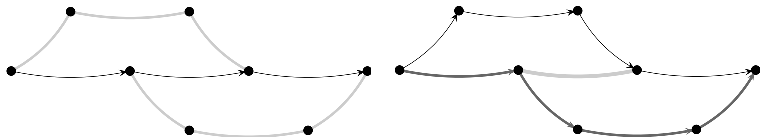


(a) SEP (b) ALT (c) COM

Figure 4.5: *Principle of* SEP, ALT, *and* COM *protection.*

The difference of the three heuristics is in the order in which they calculate and assign primary and backup paths. The principle of the calculation and assignment is depicted in Figure 4.5. In case of SEP protection, primary and backup paths are calculated separately, cf. Figure 4.5a. Thus, primary paths are searched for all traffic demands first, and backup paths are searched for all routed traffic demands afterwards. Hence, as many traffic demands as possible can be routed before the backup paths of traffic demands occupy resources. In addition, SEP protection can be used for restoration, i.e., ad-hoc set-up of backup paths.

The other two heuristics, ALT and COM protection more and more break up this separate establishment of primary and backup paths. With ALT protection, primary and backup paths are searched and assigned alternatingly. Thus, primary paths of other traffic demands cannot block the search for the backup paths of a demand. As illustrated in Figure 4.5b, the principles used with ALT protection are mostly the same as with SEP protection.



(a) shortest path, no further disjoint path possible

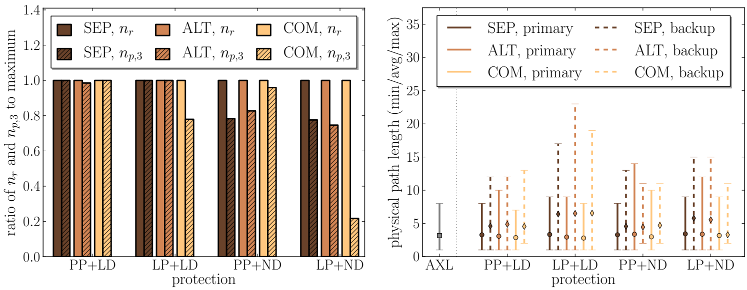
(b) disjoint shortest path pair

Figure 4.6: *Shortcomings of a shortest path algorithm with the "trap topology".*

The COM protection is the most complex of the three presented algorithms. It combines and simultaneously runs the search for primary and backup paths for a traffic demand, cf. Figure 4.5c. Similar to [117], we developed the COM algorithm to avoid situations, in which an unmodified *Constrained Shortest Path First* (CSPF) algorithm cannot find two disjoint paths as illustrated in Figure 4.6. However, an unmodified shortest disjoint path pair algorithm, like the *Suurballe-Tarjan* algorithm [118, 119], proved to be not suitable to find cost-efficient solutions since we need to guarantee disjointness not only in a single layer but in multiple layers. Therefore, we use k-shortest path algorithms, especially *Eppstein* [120] and *Yen* [121], to calculate up to k candidates for primary paths. Then, we evaluate the possibilities to set up corresponding backup paths. In Chapter 5, we consider resilience in network provisioning, i.e., routing of dynamic traffic demands on an existing multi-layer network, and use our own variation of the Suurballe-Tarjan algorithm to guarantee disjointness in the physical layer.

4.2.2 Impact of Protection Algorithms

In the following, we evaluate the impact of the SEP, ALT, and COM protection algorithms on the resulting resilient multi-layer network. For evaluation, we re-use the *Nobel Germany* (G17) topology presented in Table 3.1 and apply a homogeneous 1-uniform traffic pattern as in Chapter 3. For the moment, we restrict our evaluations to using the *Time Division Multiplex* (TDM) layer as the protecting



(a) ratio of routed and protected demands (b) path lengths in physical layer

Figure 4.7: *Performance of* SEP, ALT, *and* COM *single-layer protection.*

layer, i.e., $\mathcal{G}_P = \{G_3\}$. We consider four protection variants, i.e., combinations of resource usage and scope, cf. Section 4.1.4. These are *Path Protection* (PP) and *Link Protection* (LP) each with *Node Disjointness* (ND) and *Link Disjointness* (LD). In Figure 4.7, we compare the performance of the three single-layer protection algorithms with respect to the number of routed (n_r) and protected ($n_{\text{p},3}$) traffic demands as well as the path lengths of primary and backup paths compared to the unprotected case using the plain AXL algorithm, cf. Chapter 3.

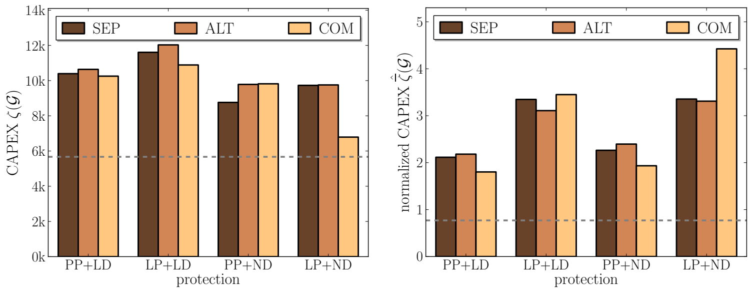
Figure 4.7a illustrates the ratio of routed and protected traffic demands to the overall number of traffic demands, which is $|\mathcal{D}^\text{stat}| = 272$ for the G17 topology with 1-uniform traffic pattern. With path protection and link disjointness (PP+LD), all three algorithms are able to route all traffic demands. Except for ALT protection, which cannot protect 3 traffic demands, all traffic demands can be protected as well. Regarding the routed demands, the situation is the same for link protection and link disjointness (LP+LD). Contrarily, COM fails to protect all demands in this protection variant. The other two evaluations comprise node disjointness, which is harder to realize than link disjointness. The increased complexity can be seen for all three algorithms.

While the three algorithms are able to still route all traffic demands for path protection and node disjointness (PP+ND), none of them is able to protect all demands. In this protection variant, SEP protection performs worst and COM protection performs best. Thus, the separated calculation of primary and backup paths leads to reduced ability of protection in this case. Finally, with link protection and node disjointness (LP+ND), SEP and ALT protection nearly keep their level of protection compared to PP+ND, but COM protection shows a drastic decline in protected demands.

Concluding, all algorithms are able to route all demands while the effectiveness in the protection of demands depends on the considered protection variant. The combined search for primary and backup paths of COM protection performs best with path protection.

Figure 4.7b depicts the lengths of the primary and backup paths in the physical layer. For the sake of clarity, we only show minimum, average, and maximum of the path length distribution over all routed and protected traffic demands. Demands that were not routed or protected are not considered in the distribution. We consider the three algorithms and four protection variants as in Figure 4.7a. For a clear comparison, we additionally plotted the unprotected case using the AXL algorithm, which can only show the routed demands and is identical to the values for primary paths with SEP protection. In general, the average and maximum length of backup paths are at least twice of the average and maximum length of primary paths. Due to the per-link assignments, backup paths with link protection are always longer than with path protection. This seems to be not that dramatic with LP+ND, but here rather few demands are protected at all.

Additional resources need to be installed to provide protection and path lengths increase as a consequence of disjointness constraints. Apparently, the overall CAPEX is increased by protection. Hence, we also evaluate the overall CAPEX resulting from the single-layer protection algorithms and apply the metric for normalized CAPEX in resilient multi-layer networks $\hat{\bar{\zeta}}(\mathcal{G})$ defined in Equation 4.5. The results are shown in Figure 4.8.



(a) overall CAPEX $\zeta(\mathcal{G})$

(b) normalized CAPEX $\hat{\bar{\zeta}}(\mathcal{G})$

Figure 4.8: *Resulting* CAPEX *with different, single-layer resilience algorithms.*

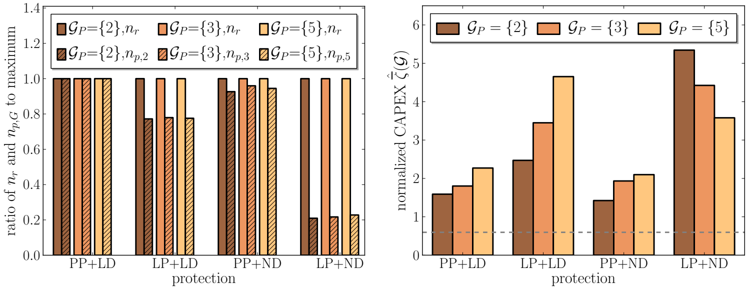
In Figure 4.8a, the overall CAPEX $\zeta(\mathcal{G})$ is depicted for the three single-layer protection algorithms. For comparison, the CAPEX value without protection (AXL) is plotted as a horizontal dashed line. A major finding from this figure is that the CAPEX is not doubled although additional backup paths are set up with lengths that are at least twice the path length of primary paths. Comparing PP+LD and LP+LD shows that link protection is more expensive than path protection. This is expected as backup paths in link protection are longer since they have to pass each node of the primary path. For LD+ND, COM achieves lowest cost, but also cannot protect many demands as aforementioned. Due to different numbers of routed and protected demands in combination with the algorithms and protection variants, further conclusions cannot be easily drawn from this figure.

To that end, the normalized CAPEX $\hat{\bar{\zeta}}(\mathcal{G})$, which considers routed and protected demands as well as path lengths, cf. Equation 4.5, is depicted in Figure 4.8b. The figure shows that COM protection is the best choice for path protection. However, COM protection yields slightly higher CAPEX with LP+LD and, in case of LP+ND, a drastic increase. The latter indicates that COM protection cannot protect as many demands as the other protection algorithms. In summary, $\hat{\bar{\zeta}}(\mathcal{G})$ is better suited to compare the algorithms and protection variants.

4.2.3 Variation of Protecting Layer

As described in Section 4.1.4, the set of protecting layer $\mathcal{G}_P$ can be almost arbitrarily chosen from the layers in a multi-layer network. To study the effect of different protecting layers with single-layer resilience, we focus on the COM protection algorithm as it shows best the impact of varying the protecting layer $\mathcal{G}_P \in \big\{\{G_2\}, \{G_3\}, \{G_5\}\big\}$, i.e., OCh, SDH/SONET, and IP/MPLS, respectively. We do not consider $\mathcal{G}_P = \{G_4\}$, i.e., Ethernet, due to the performance drawbacks caused by inappropriate equipment described in Chapter 3.

Since the different layers do not provide identical equipment with respect to cost and capabilities, having different protecting layers inherently leads to different results and CAPEX. Thus, protecting layers cannot easily be compared as the different capabilities of the equipment lead to individual routing solutions and especially different numbers of routed (n_r) and protected ($n_{\mathrm{p},G}$) demands. Figure 4.9 shows the results of the studies on the protecting layers. In Figure 4.9a, the number of routed and protected demands is depicted. Apparently, the protecting layer has a rather low impact on the ability of the protection algorithm to route and protect traffic demands. This emphasizes our findings in Chapter 3 that the main cause for such limitations is in the restriction to a single fiber strand in the physical layer and the degrees of freedom of the underlying network topology. In Figure 4.9b, we focus on the impact of the protecting layer on CAPEX. It shows the normalized CAPEX $\hat{\bar{\zeta}}(\mathcal{G})$ and outlines the importance of the choice of the protecting layer, which enforces the use of a certain technology and increases the corresponding cost ratio of this technology. Furthermore, it confirms the findings on network architectures given in Section 3.3.2 that IP is the most expensive technology in the considered cost model. Compared to IP, the set-up of paths in the optical layer is less expensive. This advantage of protection in the OCh layer is only violated for the LP+ND protection variant. With this variant, traffic can be switched between primary and backup path at each node along the primary path, i.e., the ability to process data at each node is required. However, OCh cannot process data itself. Thus, the data has to be sent to higher layers, i.e., not only



(a) ratio of routed and protected demands

(b) normalized CAPEX $\hat{\bar{\zeta}}(\mathcal{G})$

Figure 4.9: *Different protecting layers and single-layer resilience.*

optical technology is to be installed at each node but also electrical equipment, like IP. As a consequence, the LP+ND protection variant is not recommended on the optical layer.

In the next section, the variation of protecting layers is generalized and $|\mathcal{G}_P| > 1$ is considered.

4.3 Multi-Layer Resilience

It might seem sufficient to activate protection on a single layer to cover all considered failures. However, failures in higher layers cannot be discovered in lower layers, cf. Section 4.1.3. To that end, we introduce and evaluate heuristics for CAPEX-aware multi-layer network design with multiple protecting layers.

4.3.1 Algorithms for Multi-Layer Protection

Applying protection in the highest layer (G_5 in our model, i.e., IP/MPLS) might seem to avoid the problems to discover failures in higher layers. For technical reasons, this is neither always sufficient nor cost-optimal.

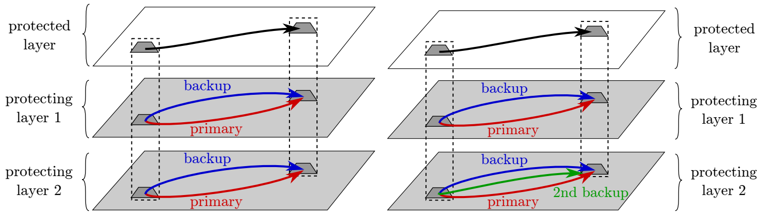
First, traffic in multiple layers, cf. Section 3.3.3, also requires protection in multiple layers. Second, the *Mean Time To Recover* (MTTR) usually increases the higher a layer is in the multi-layer hierarchy, e.g., MTTR of IP takes up to seconds while SDH can fulfill 50 ms requirements, cf. [42]. Third, there is a trade-off between protection and corresponding additional cost that also typically increases for upper layers. Fourth, every protecting layer also requires additional capacity in all lower transport layers.

Hence, we developed algorithms for CAPEX-aware multi-layer network design with multi-layer resilience. Herein, interrelations of layers, like considering the capacity of backup paths in lower layers, require cooperation amongst protection mechanisms in different layers. To that end, we presume corresponding control plane and management plane mechanisms as defined in [20]. We developed three heuristics for multi-layer resilience: *Identical Multi-Layer* (IML), *Disjoint Multi-Layer* (DML), and *Recursive Multi-Layer* (RML) protection.

IML starts by applying one of the single-layer resilience algorithms presented in Section 4.2 on the lowest protecting layer and tries to *identically* replicate the structures of primary and backup paths in higher protecting layers. Thereby, existing link resources for backup paths are re-used in higher layers and leave further paths unused to increase the degree of freedom in finding paths for other demands. The principle of IML protection is illustrated in Figure 4.10a. However, a failure on a certain layer breaks all backup paths in higher protecting layers as well. Thus, IML only provides additional protection for failures in higher layers.

This handicap of IML is eliminated by DML. While DML also starts by applying one of the single-layer resilience algorithms, DML searches new backup paths for each protecting layer that are disjoint in the physical layer to all lower primary and backup paths of the current traffic demand. As the degrees of freedom for paths in the physical layer are limited, it is more likely with DML that not all protecting layers can provide a backup path compared to IML. The principle of DML protection is illustrated in Figure 4.10b.

With RML, primary and backup paths are *recursively* searched for each lower primary and backup path in a top-down fashion and all primary and backup paths



(a) IML protection (b) DML protection

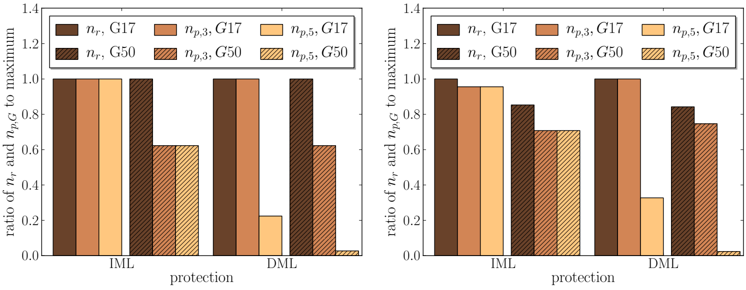
Figure 4.10: *Principles of* IML *and* DML *multi-layer protection.*

need to be disjoint in the physical layer to avoid *Single Point Of Failures* (SPOFs). Due to the recursiveness, RML requires a lot of resources. For example, three protecting layers already require eight-fold protection in the physical layer where all eight paths need to be disjoint from each other. Consequently, RML is only reasonable for tightly meshed topologies with many degrees of freedom in path finding. Since none of our considered topologies is able to sufficiently fulfill these requirements, we do not further focus on RML nor provide evaluations for it.

In conclusion, IML is efficient with resources but does not provide a lot of additional protection, while RML provides a lot of additional protection but is very expensive and therefore only applicable in special cases. Thus, DML is a reasonable compromise.

4.3.2 Impact of Algorithms and Topologies

In this section, we evaluate the performance of multi-layer resilience algorithms and the impact of physical topologies with respect to resulting CAPEX and to the number of protected demands. For this study, we evaluate IML and DML protection on the *Nobel Germany* (G17) and *Germany50* (G50) fiber topologies. Both multi-layer protection algorithms are evaluated with underlying SEP and COM single-layer protection, which showed different behavior in Section 4.2. As in the previous evaluations, we use the homogeneous, 1-uniform traffic pattern on



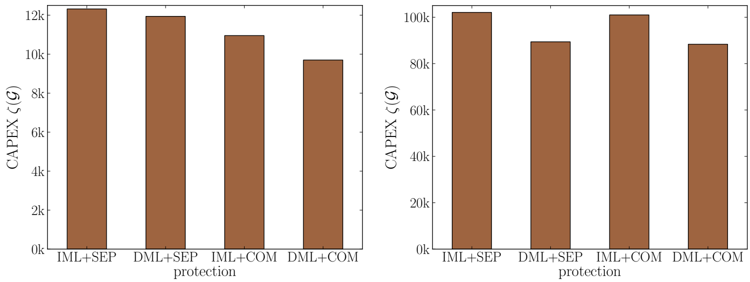
(a) SEP

(b) COM

Figure 4.11: *Ratio of routed and protected demands with different underlying single-layer resilience algorithms and physical topologies.*

the IP layer, which results in $|\mathcal{D}^{\text{stat}}| = 272$ for the G17 topology and $|\mathcal{D}^{\text{stat}}| = 2450$ for the G50 topology. For multi-layer resilience, we use $\mathcal{G}_P = \{G_3, G_5\}$ in this study since both layers proved to work well with SEP and COM single-layer protection. Thereby, we focus on *path protection* with *link disjointness* (PP+LD) as it proved to be the least restrictive protection variant in Section 4.2. The results of this study are shown in Figure 4.11 and Figure 4.12.

Figure 4.11 depicts a comparison of IML and DML protection with respect to the number of routed and protected demands on each protecting layer defined in $\mathcal{G}_P$. The evaluation in Figure 4.11a shows the results when using underlying SEP single-layer protection, while COM single-layer protection was applied for Figure 4.11b. While SEP and COM protection performed equally well on the PP+LD scenario with single-layer resilience in Section 4.2, both algorithms show a different behavior with IML and DML multi-layer resilience. In addition, we perform the studies with G17 and G50 topologies, cf. Table 3.1. These topologies do not only differ in the number of nodes, but also in the number of edges and the resulting node degrees that is important for the degrees of freedom in finding disjoint paths in the physical topology.



(a) G17

(b) G50

Figure 4.12: *Overall* CAPEX $\zeta(\mathcal{G})$ *with different underlying single-layer resilience algorithms and physical topologies.*

On the G17 topology, IML performs slightly better with SEP than with COM protection. With DML protection, neither SEP nor COM protection are able to protect all demands on both layers. However, COM protection is able to protect more demands than SEP protection. The differences become clearer on the G50 topology. Due to the separate search for primary and backup paths, SEP protection is still able to route all demands in contrast to COM protection. However, the underlying COM protection is able to protect more demands than SEP since the combined search for primary and backup paths deals more efficiently with resources in the physical layer that can be used for further disjoint backup paths needed for DML. In general, DML protection is only able to protect up to 30 % on layer G_5. This already shows the immense constraints put by DML protection to find three disjoint paths in case of two protecting layers.

Figure 4.12 shows the overall CAPEX of the protection algorithms in the considered scenarios split by the underlying topology. With underlying SEP protection, the cost for IML protection is only about 20 % higher than the cost for single-layer protection depicted in Figure 4.8a for the G17 topology. For an underlying COM protection, the difference in cost is even lower, at about 10 % but again not all demands are protected. In general, this outlines that IML protection can be cost-efficiently realized. All results for DML protection have lower cost than IML. However, the constraints regarding path disjointness of DML protection yield incomplete multi-layer resilience as we saw before.

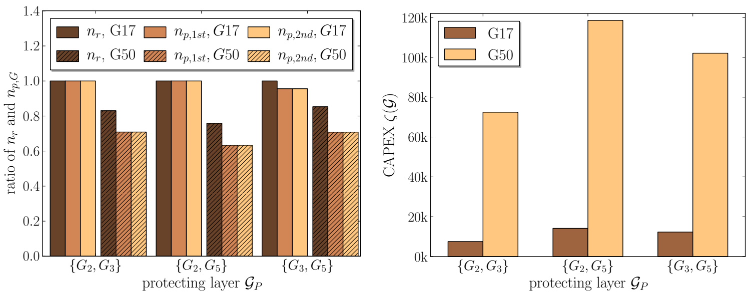
4.3.3 Selection of Protecting Layers

In this section, we generalize the variation of a single protecting layer performed in Section 4.2.3 to multiple protecting layers. In general, there are $\binom{m}{n}$ possibilities to choose the layers in $\mathcal{G}_P$ when considering m available and n protecting layers. For this study, we restrict ourselves to scenarios with two protecting layers, i.e., $n = |\mathcal{G}_P| = 2$. Since fiber and OMS cannot be used as protecting layers and we excluded the Ethernet layer in Section 4.2.3, i.e., $m = 3$, this leaves us with three possibilities, $\mathcal{G}_P \in \big\{\{2,3\},\{2,5\},\{3,5\}\big\}$.

We use underlying COM protection with PP+LD protection variant as it proved to perform well, both with single-layer and multi-layer resilience, cf. Section 4.2 and Section 4.3.2. As the results presented in Section 4.3.2 showed that DML protection heavily depends on the degrees of freedom in the physical layer, we focus on IML protection for the evaluation of different protecting layers on multi-layer resilience. The results of this study are depicted in Figure 4.13.

The number of routed and protected demands is shown in Figure 4.13a. It can be noticed that the number of routed as well as protected demands distinguish with different $\mathcal{G}_P$. While $\mathcal{G}_P = \{2,3\}$ and $\mathcal{G}_P = \{3,5\}$ only differ marginally with respect to the number of routed and the number of protected demands in the first and second protecting layer, the performance of $\mathcal{G}_P = \{2,5\}$ is about 10 % lower. This difference results from the equipment of the protecting layers.

The consequences of $\mathcal{G}_P$ on the overall CAPEX are illustrated in Figure 4.13b. Herein, $\mathcal{G}_P = \{2,5\}$ proves to not only be a bad choice with regard to numbers but also is almost twice the cost of $\mathcal{G}_P = \{2,3\}$ for an underlying G17 as well as G50 topology. While $\mathcal{G}_P = \{3,5\}$ is slightly better than $\mathcal{G}_P = \{2,3\}$ with respect to the number of routed and the number of protected demands in the



(a) ratio of routed and protected demands (b) overall CAPEX $\zeta(\mathcal{G})$

Figure 4.13: IML *protection with different protecting layers and underlying physical topologies.*

first and second protecting layer, this comes at an enormous increase of cost. Therefore, $\mathcal{G}_P = \{2, 3\}$ seems to be the best choice for multi-layer resilience in the considered CAPEX model.

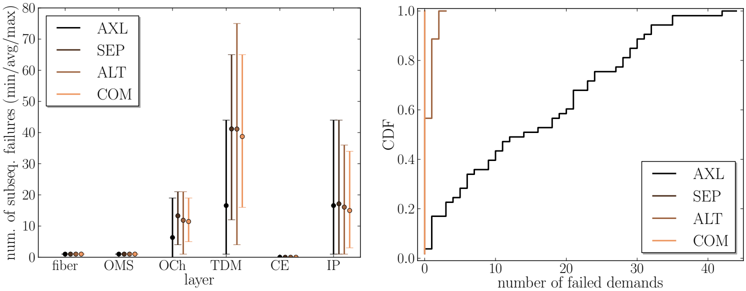
In the next section, we evaluate how well the presented algorithms for single-layer and multi-layer resilience are able to cover different failure scenarios.

4.4 Analysis of Failure Effects

So far, algorithms to create resilient multi-layer networks were presented that prevent the consequences of catastrophic failures in multi-layer networks. On the basis of our analysis of failure effects in [12], we now evaluate the consequences of different failure types and the effectiveness of our algorithms.

4.4.1 Impact of Shared Risk Groups

As mentioned in Section 4.1.3, the hierarchical structure of multi-layer networks implies SRGs. Besides the structure of a multi-layer network, the effects of an



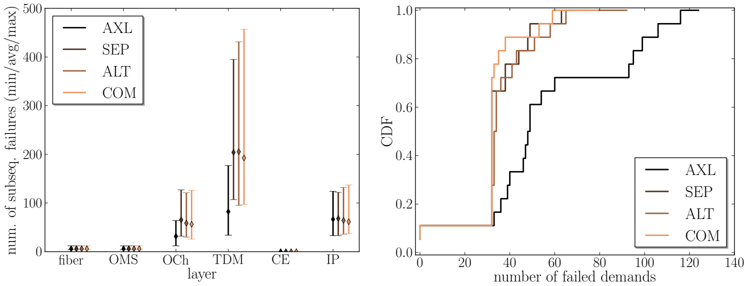
(a) SRG sizes

(b) reliability of protection

Figure 4.14: *Impact of single-link failures* $\mathcal{F}_{\mathcal{E}_0}$ *on the Nobel Germany topology.*

SRG, i.e., the number of subsequent failures $|\mathcal{F}_x^{\text{sub}}|$ caused by a failure $\mathcal{F}_x \in \mathcal{F}$, depend on the specific failure scenario $\mathcal{F}_x$. To that end, we perform an analysis of two failure scenario types defined in Section 4.1.1 We consider all single-link failures in the physical layer $\mathcal{F}_{\mathcal{E}_0}$ as well as all single total-node failures $\mathcal{F}_{\mathcal{V}}$. The basis for our evaluations is a multi-layer network that was designed upon the *Nobel Germany* (G17) topology using one of the single-layer resilience algorithms presented in Section 4.2. For now, we focus on a single protecting layer and choose $\mathcal{G}_P = \{G_3\}$ with PP+LD protection variant.

The results for single-link failures are illustrated in Figure 4.14. Figure 4.14a shows the minimum, average, and maximum number of subsequently failed links in each layer of the multi-layer network. Noticeably, *Carrier Ethernet* (CE) shows no failures as this technology was not used by any of the considered design algorithm in this study. The fiber layer lists exactly one failure since we consider single-link failures in the physical layer. The OMS layer also shows exactly one failure due to the one-to-one relation of fiber and OMS stated in Chapter 2. The number of subsequent failures is given for SEP, ALT, and COM single-layer protection in addition to the unprotected case using plain AXL algorithms, cf. Chapter 3. Counter-intuitively, the size of SRGs with resilience can be larger than



(a) SRG sizes

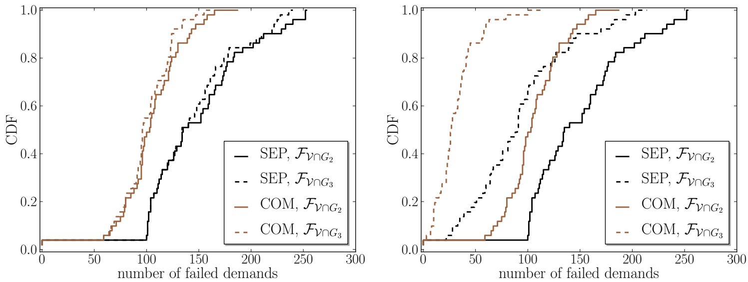
(b) reliability of protection

Figure 4.15: *Impact of single total-node failures $\mathcal{F}_{\mathcal{V}}$ on the Nobel Germany topology.*

without resilience as shown for TDM in Figure 4.14a. This increase is due to the higher number of links being installed in total when providing backup paths.

In Figure 4.14b, the consequences of these SRGs on the survival of demands are illustrated using the same design algorithms. Therefore, we plot the *Cumulative Distribution Functions* (CDFs) of demands that failed in spite of different protection mechanisms due to single-link failures. The results for unprotected AXL emphasize the importance of resilience in multi-layer networks. Furthermore, the presented single-layer resilience algorithms prove to successfully protect demands from single-link failures, except for ALT which could not set up backup paths for three demands as shown in Figure 4.7a.

Although most common and important, the consequences of single-link failures are a lot less drastic than single total-node failures, for which the results are given in Figure 4.15. The number of subsequent failures is depicted in Figure 4.15a, which shows that single total-node failures can cause almost five times more subsequent failures than the single-link failure scenarios. In this figure, the size of the SRGs even more increases with resilience than in Figure 4.14a and clearly dominates the unprotected AXL case. The CDFs of failed demands illus-



(a) IML protection

(b) DML protection

Figure 4.16: *Impact of single-layer node failures $\mathcal{F}_{\mathcal{V}\cap G}$ in multi-layer resilient scenarios on the Germany50 topology.*

trated in Figure 4.15b make clear that our single-layer protection algorithms are able to provide a certain level of protection but none of the algorithms is able to protect against all single total-node failures. Additional constraints for network design would be needed that explicitly consider such failures.

4.4.2 Dealing with Equipment Failures

Besides the previously considered link failures and the failure of a whole site that affects all layers, single layer-node equipment failures $\mathcal{F}_{\mathcal{V}\cap G}$, cf. Section 4.1.1, are another type of failure scenarios relevant to multi-layer networks. Such single layer-node failures only affect the node equipment of a single layer, e.g., the failure of a router in the IP layer. Nevertheless, such failures have immense SRGs.

In the following, we analyze multi-layer resilience as one possibility to mitigate the impact of single layer-node failures. We focus on multi-layer resilience with $\mathcal{G}_P = \{2, 3\}$ and the PP+LD protection variant. To allow a better comparison, we consider the *Germany50* (G50) topology as some single-layer node failures would affect all links in the *Nobel Germany* (G17) topology. Due to the different requirements and properties, we evaluate the performance of both IML and DML protection with underlying SEP and COM single-layer protection. The results of these evaluations are given in Figure 4.16. Figure 4.16a shows the behavior of IML protection while Figure 4.16b depicts the results of DML protection. In all combinations, underlying COM protection leads to less failed demands than SEP protection. In addition, the number of failed demands due to single layer-node failures on the TDM layer, G_3, can be drastically decreased when using DML protection in favor or IML protection both with underlying SEP and COM single-layer protection. Although we showed in Section 4.3 that the requirements of DML protection make it impossible to protect all demands on the second protecting layer, the combination of DML with COM is able to nearly protect against all considered failures scenarios in $\mathcal{F}_{\mathcal{V}\cap G_3}$.

Thus, certain multi-layer resilience strategies proved that mitigating the impact of single layer-node equipment failures is possible.

4.5 Lessons Learned

The introduction of resiliency constraints in multi-layer network design outlined the advantage of our heuristic approach. Since the heuristics were built from several algorithmic steps, great parts could easily be re-used. Thus, we adhered to the general design concepts introduced in Chapter 3.

The development of heuristics for the design of resilient multi-layer networks showed the increased complexity of resilience and the need for further parameter studies regarding resilience. We explored that resilience can be motivated in several ways and introduced three heuristics for the design of resilient multi-layer networks with a single protecting layer, i.e., single-layer resilience. As expected, the average path length increased when compared to unprotected scenarios and the increase can be up to several orders of magnitude for some protection variants. However, the overall CAPEX does not even double due to synergies in the transport layers. A general recommendation for any of the heuristics cannot be given since we found that the heuristics behave differently with certain protection variants. Furthermore, the heuristics differ in the number of routed and protected demands, so that the choice of the heuristic depends on the objective of the network provider. Besides the performance of the heuristics, the impact of the choice of the single protecting layer was evaluated. While the results showed that the protecting layer is rather negligible with regard to the number of routed and protected demands, it is decisive with regard to CAPEX. An analysis of different failure types and their resulting SRGs by the number of subsequently failed links reveal the immense consequences of failures in hierarchical structures that are inherent to multi-layer networks. Evaluations showed that our single-layer protection algorithms are able to protect against all considered single-link failures. However, single total-node failures and their resulting SRGs cannot be fully covered by these algorithms. Thus, we recommend to consider such failures within the design algorithm itself to provide adequate protection.

Extending the concepts of single-layer resilience, algorithms for the design of resilient multi-layer networks with multiple protecting layers, i.e., multi-layer resilience, were developed in this chapter. Again, the heuristic approach allows to build these new heuristics on top of the single-layer resilience heuristics. We evaluated the performance of two multi-layer resilience algorithms with two protecting layers. We found that multi-layer resilience is hard to realize when strict disjointness is required in all layers. In addition, the results heavily depend on the properties of this physical layer. This was worked out by comparing different physical topologies. The variation of the protecting layers on all possible layer combinations reveal nearly identical performance with different protecting layers and outline the difference in cost for combining certain layers. Thus, we recommend to protect at OCh and SDH/SONET layer with regard to this CAPEX model. Furthermore, the effectiveness of multi-layer resilience algorithms was analyzed with single layer-node failures types. The results thereof manifested the advantages of multi-layer resilience even when strict disjointness in all layers is not postulated.

Upon the design of multi-layer networks, the provisioning of dynamic traffic demands is considered in the next chapter.

5 Multi-Layer Network Provisioning

Marty. You're not thinking fourth dimensionally!
"Back to the Future – Part III", 1990.

In the previous chapters, we studied optimization problems that deal with the cost-aware construction of multi-layer networks. In this chapter, we consider the construction to be finished. Thus, equipment can neither be altered nor added and we enter the next stage of the network life-cycle – the *operation* – as described in Section 2.1.2. Network operation runs on top of the layers of a constructed multi-layer network. Hence, the focus is on the efficient utilization of the deployed equipment in order to maximize the revenue of the network operator.

In contrast to the previous two chapters, this chapter deals with *dynamic*, time-restricted traffic demands in multi-layer networks. Thus, the time domain and probabilistic uncertainty need to be considered. The *provisioning*, i.e., the set up and tear down of dynamic traffic demands including the occupation and freeing of resources in a multi-layer network, underlies given time restrictions that cannot be fulfilled by legacy business processes, like sending a fax to a provider. Hence, we consider an automated control plane architecture that enables dynamic provisioning in multi-layer networks.

5.1 Prerequisites of Dynamic Provisioning

In the following, we introduce a model for dynamic traffic in multi-layer networks as well as the fundamentals to evaluate and set up automated provisioning for such traffic demands.

5.1.1 Dynamic Traffic Modeling

In Section 2.3.3, we gave a general notation for a set of dynamic multi-layer traffic demands $\mathcal{D}^{\text{dyn}}$. As considering dynamic traffic demands is still a future scenario, no information on parameters of dynamic traffic is available. The *Survivable Network Design library* (SNDlib) project [112] recognized this lack and released data on time-dependent usage of different network topologies in May 2011. However, this data currently contains only time-dependent utilization of links but no information on contiguous flows of dynamic traffic demands. Therefore, we rely on an analytic model whose assumptions are based on measurements of real traffic to yield realistic dynamic traffic characteristics.

In Figure 5.1, we illustrate several arrivals, causing an increase in load, and holding times of dynamic traffic demands, ending up in departures and decreasing load, during a considered time span – the simulation time T. In the following, we introduce stochastic processes to describe the arrivals, departures, and corresponding changes in load by means of random variables. All time-dependent parameters are given in units of seconds if not explicitly stated otherwise.

Arrival Process

A key property of a dynamic traffic demand is its unforeseen and random point in time of its arrival. This makes it impossible to take the order of dynamic demands into account during optimization as done in Section 3.2.3. Arrivals of dynamic traffic demands can be described by stochastic processes. As stated in [110], it is eligible for circuit-switched connections in core networks to assume statistical independence of arrivals of traffic demands. This leads to a *Poisson* process

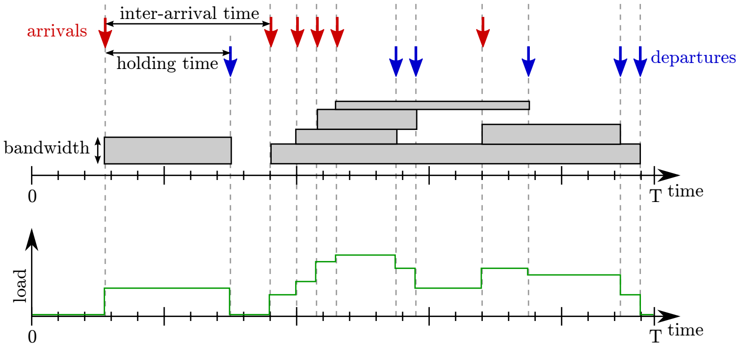


Figure 5.1: *Arrivals and departures of dynamic traffic demands as well as the network load resulting from the corresponding bandwidth requests.*

model, which is *memoryless*, i.e., future arrivals do not depend on past arrivals.

A Poisson arrival process with rate λ (arrivals/second) is based on a Poisson distribution of the number of arrivals in a given time span. The Poisson arrival process results in a negative-exponentially distributed inter-arrival time A given by its *Cumulative Distribution Function* (CDF), which is

$$A(t) = 1 - e^{-\lambda t} \text{ with mean } \mathrm{E}[A] = \frac{1}{\lambda}. \tag{5.1}$$

The definition of the arrival rate λ controls the network load. We consider a population-based traffic model [11], cf. Equation 2.3, wherein the amount of traffic between two sites $x, y \in \mathcal{V}$ is determined by the ratio of the population $\pi(v)$ ($v \in \mathcal{V}$) of these two sites to all pairs of sites in the network. In the course of this work, we always use the population distribution that was used as a basis for the design of the underlying multi-layer network. To distribute the traffic demands among the sites, we define the arrival rate $\lambda_{x,y}$ of dynamic traffic demands per site pair $(x, y) \in \mathcal{V} \times \mathcal{V}$. Thus, we define

$$\lambda_{x,y} = \lambda \cdot \frac{\pi(x) \cdot \pi(y)}{\sum_{a,b \in \mathcal{V}, a \neq b} \pi(a) \cdot \pi(b)} \text{ and } \lambda_{x,x} = 0, \tag{5.2}$$

where the latter means that there is no traffic from a site $x \in \mathcal{V}$ to itself. This is valid, as a superposition of Poisson processes with rate $\lambda_{x,y}$ again results in a Poisson process with rate $\lambda = \sum_{x,y \in \mathcal{V}} \lambda_{x,y}$.

Holding Times

The *holding time*, also called life time or sojourn time, is the duration, a dynamic traffic demand stays active in the network, i.e., the time from its arrival to its departure. As we consider dynamic traffic with unforeseen properties, the holding time also is described by a stochastic process. However, we assume that the average duration of the demands does not change over time. Thus, the remaining time of a "conversation" is independent from the amount of elapsed time as was already shown by A. K. Erlang [122] and is confirmed for core networks in [123]. As with the arrival process, this leads to a negative-exponential distribution of the holding times H with rate μ given by its CDF, which is

$$H(t) = 1 - e^{-\mu t} \text{ with mean } \mathrm{E}[H] = \frac{1}{\mu}. \tag{5.3}$$

Bandwidth Distribution

In core networks, highly aggregated traffic of many access networks is transmitted using circuit-switching technologies. As described in Chapter 2, the equipment only provides certain bandwidth granularities. Like the authors of [110], we assume that dynamic traffic in core networks occurs in certain granularities.

Especially, we assume dynamic traffic demands of 1 Gbit/s and 10 Gbit/s for our studies on core networks as we assume large companies with medium demands and data centers with high demands to be the first profitable customers for such an automated provisioning. However, the principles and algorithms presented in this chapter are not limited to such coarse granularities and can be applied for any bandwidth distribution B^{dyn}. In this chapter, we focus on homogeneous bandwidth distributions and vary only the mean $\mathrm{E}[B^{\mathrm{dyn}}]$.

5.1.2 Evaluation Metrics

During the evaluation of dynamic traffic upon a constructed multi-layer network, we consider the following metrics.

Average number of active demands: The average number of active demands a is given by the mean of the random variable L of active demands in the network. By means of Little's law this is

$$a = \mathrm{E}[L] = \lambda \cdot \mathrm{E}[H] = \frac{\mathrm{E}[H]}{\mathrm{E}[A]}. \tag{5.4}$$

Offered load: The (average) offered load ρ_o is given by the product of the mean of the random variables L of active demands in the network and the mean of the random variable B^{dyn} of bandwidths, i.e.,

$$\rho_o = \mathrm{E}[L] \cdot \mathrm{E}[B^{\mathrm{dyn}}] = \lambda \cdot \mathrm{E}[H] \cdot \mathrm{E}[B^{\mathrm{dyn}}]. \tag{5.5}$$

Blocking probability: The blocking probability p_b, which gives the ratio of blocked demands to the overall number of demands.

$$p_b = \frac{\text{number of blocked demands}}{|\mathcal{D}^{\mathrm{dyn}}|} \tag{5.6}$$

Carried load: The (average) carried load ρ_c, i.e., the fraction of offered load ρ_o that is not blocked and carried by the network, is defined by

$$\rho_c = (1 - p_b) \cdot \rho_o. \tag{5.7}$$

5.1.3 Control Plane Architectures

As we deal with dynamic traffic demands and their automated provisioning during network operation, we require automated control mechanisms that allow to access equipment in different layers as well as to reserve and free resources on command. We consider the *Generalized Multi-Protocol Label Switching* (GMPLS) and *Path Computation Element* (PCE) architectures as the basis

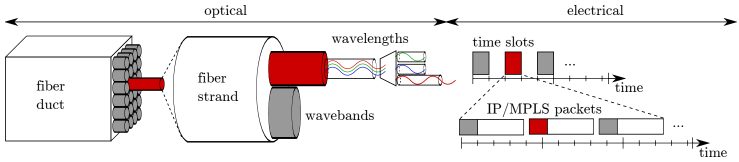


Figure 5.2: *Generalized label concept in* GMPLS *to address multiplexing mechanisms in different technologies of a multi-layer network.*

for our dynamic provisioning considerations as GMPLS-enabled equipment has reached a certain market penetration and GMPLS supports access to multiple technologies. Via GMPLS signaling, the upcoming PCE architecture allows to utilize the multi-layer concepts of GMPLS and to perform efficient path finding.

GMPLS Architecture

The GMPLS architecture as defined in [24] extends the concept of *Multi-Protocol Label Switching* (MPLS) [18] to perform switching decisions for flows based on a *label* put in front of an *Internet Protocol* (IP) packet. In the context of GMPLS, labels are generalized in order to fit not only to IP packets. Thus, a GMPLS label can be assigned to a fiber strand in a fiber duct, a wavelength on a fiber or waveband, as well as a time-slot in a *Time Division Multiplex* (TDM) technology. This is illustrated in Figure 5.2 where the generalization of labels is supported by GMPLS-enabled equipment. This generalization allows to cover further technologies that are based on *Packet Switch Capable* (PSC), *Time Division Multiplex* (TDM) *Capable*, *Lambda Switch Capable* (LSC), or *Fiber Switch Capable* (FSC) interfaces.

Especially in *Optical Transport Networks* (OTNs), GMPLS is often used as a synonym for *Automatic Switched Transport Network* (ASTN) (or former *Automatically Switched Optical Network* (ASON)). However, GMPLS is not a single protocol but rather a “family of protocols” that comprises the so called GMPLS *stack* [55], which consists of three main protocols.

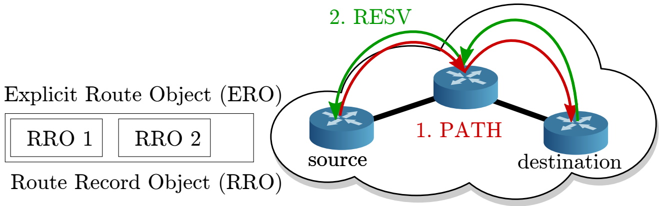


Figure 5.3: *Path signaling using the* RSVP *protocol with a given path definition.*

Management: The *Link Management Protocol* (LMP) [27] is used to combine multiple data links into a single *Traffic Engineering* (TE) link, e.g., for scalability, and to define out-bound control channels, e.g., for optical links to whom control data cannot simply be added or sent in-bound together with ordinary data.

Routing: The *Open Shortest Path First with Traffic Engineering* (OSPF-TE) routing protocol [23] and its adaption to GMPLS [26] allow to exchange the TE topology within an OSPF area. Besides others, the conveyed information includes overall, unused, and maximum reservable bandwidth on links and TE metric data [23].

Signaling: The *Resource ReSerVation Protocol with Traffic Engineering* (RSVP-TE) [19] and its adaption to GMPLS [22] is used to establish connections. Therefore, RSVP-TE sends reservation messages (`PATH`) along a desired path to check for available resources. The request is encoded in *Route Record Objects* (RROs) defined in [19], which comprise an *Explicit Route Object* (ERO) defining the path that should be reserved. At the destination of the path, a notification on the success of the reservation (`RESV`) is sent back on the same path and resources are reserved. This principle is illustrated in Figure 5.3. While RSVP-TE provides a soft-state mechanism to remember reservations, we use the explicit but optional tear-down of demands to avoid RSVP timing issues.

Given a valid ERO that contains the desired path, the mechanisms provided by the GMPLS protocol family allow to signal a path across multiple layers [28]. The calculation of such a path, however, is not part of GMPLS itself.

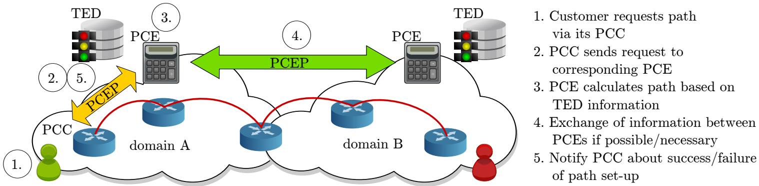


Figure 5.4: *Overview of the* PCE *architecture illustrating the steps of information exchange for on-demand path computation.*

PCE Architecture

The *Path Computation Element* (PCE) architecture [30] builds upon an underlying GMPLS network. The sole purpose of a PCE is to compute paths through the network such that *Label Switched Paths* (LSPs) can be set up by means of GMPLS signaling. As illustrated in Figure 5.4, a *Path Computation Client* (PCC) being accessible to customers of the automated provisioning requests a PCE for a path that fulfills a set of given requirements. The general requirements for communication between PCC and PCE are specified in the *Path Computation Element Communication Protocol* (PCECP) defined in [31]. A request from the PCC consists of at least one ERO defined by RSVP-TE and contains, e.g., source, destination, requested bandwidth, resilience, and other *Quality of Service* (QoS) constraints. A specific realization of PCECP is given by the *Path Computation Element Communication Protocol* (PCEP) [33].

For the path computation, a PCE needs information on the current utilization of resources. To that end, the concept of a *Traffic Engineering Database* (TED) was already proposed in [23] to perform global TE. The TEDs can be queried by the PCE. The knowledge gathering of the TED can be done in different ways as described in [30], e.g., based on the information from OSPF-TE by means of the GMPLS architecture, cf. Section 5.1.3.

The concept of the PCE architecture is very powerful. It includes communication between PCEs in different *Autonomous Systems* (ASes), as illustrated in

Figure 5.4, and enables multi-domain path set-up [32] as considered in [57, 58]. Of special interest for our studies is the possibility to perform inter-layer TE by defining adequate requests as proposed in [34]. Furthermore, several capability levels for PCEs to perform TE and *re-optimization* exist.

5.1.4 Constrained Path Computation

Path computation for dynamic provisioning takes place on a layer $G_i = (\mathcal{V}, \mathcal{E}_i)$ of the multi-layer network $\mathcal{G}$ as introduced in Section 2.2.2. The directed links $\mathcal{E}_i$ of layer G_i provide resources that can be utilized for routing dynamic traffic demands by using the GMPLS architecture. In the following studies, we always use the IP layer $G_{\text{IP}} = (\mathcal{V}, \mathcal{E}_{\text{IP}})$.

To compute paths on such a topology, a PCE takes information from TEDs that serve as input and give constraints for the computation of a path. To enable path computation using such *Constrained Shortest Path First* (CSPF) algorithms, we associate a weight $w(e)$ with each link $e \in \mathcal{E}_i$ that represents the constraints given by the TED. Thus, we yield a weighted graph (G_i, w) with a weight function $w : \mathcal{E}_i \mapsto \mathbb{R}^+$ that yields a real-valued, positive weight for each link. A CSPF computes a minimum-weight path in this weighted graph, i.e., links with lower weight are preferred. If not explicitly stated otherwise, we use the *Dijkstra* shortest-path algorithm [59] for our studies.

The algorithm for the actual path computation inside the PCE is not part of its specification. Hence, we focus on the development and analysis of efficient PCE algorithms and strategies in the following sections.

5.2 PCE-Based Dynamic Provisioning

In this section, we consider different parameters of the PCE architecture and evaluate their impact on the blocking probability of dynamic traffic demands under different PCE algorithms, capabilities, and strategies. The studies are performed on the network layers resulting from multi-layer network construction, discussed in the previous chapters.

5.2.1 Network Topologies and Traffic Patterns

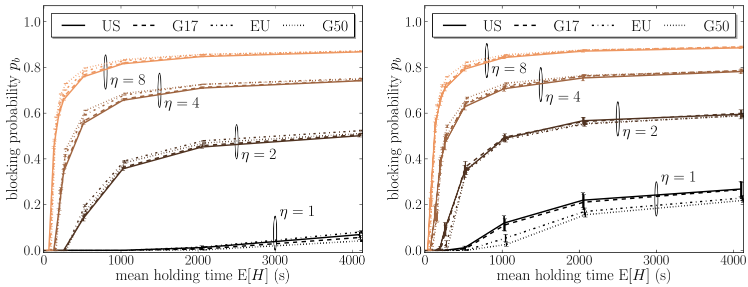
Building on the studies on network construction performed in Chapter 3 (design) and Chapter 4 (resilience), we perform our provisioning studies on the same four network instances. The load of a network mainly depends on three input parameters. The inter-arrival rate A directly affects the average number of active demands a, i.e., the number of concurrent demands in the system, cf. Equation 5.4. Similar but indirectly, the average holding time $\mathrm{E}[H]$ of a traffic demand affects the average number of active demands a. Finally, the average bandwidth $\mathrm{E}[B^{\mathrm{dyn}}]$ of a traffic demand increases the offered load ρ_o, cf. Equation 5.5. The impact of $\mathrm{E}[B^{\mathrm{dyn}}]$ depends on the capacity B^{stat} provided by the underlying network topology. An increase in any of these parameters also increases the network load.

To be able to compare results for different network topologies, we define a way to create a common load on different network topologies. To achieve a comparable network load in different network topologies, each dynamic traffic parameter needs to be selected on the basis of the topology and the static traffic matrix $\mathcal{D}^{\mathrm{stat}}$ for whom the underlying topology was designed. Therefore, we define the per-topology arrival rate λ by

$$\lambda = \left|\mathcal{D}^{\mathrm{stat}}\right| \cdot \frac{\mathrm{E}[B^{\mathrm{stat}}]}{\mathrm{E}[B^{\mathrm{dyn}}]} \cdot \frac{\eta}{T}, \tag{5.8}$$

where T is the length of the considered simulation time span in which arrivals may occur. We use $T = 1000$ if not explicitly stated otherwise. In this equation, the scalar $\eta > 0$ is used to yield different network loads. The ratio of mean bandwidths is multiplied to balance the available bandwidth (B^{stat}) and requested bandwidth (B^{dyn}) in a network.

Figure 5.5 shows the results of a study to validate Equation 5.8 and to analyze the impact of network topologies and traffic patterns. The studies regarding the blocking probability p_b are performed on the IP layer G_{IP} of the four reference topologies used in Chapter 3, which are created with a homogeneous traffic pattern and $\mathrm{E}[B^{\mathrm{stat}}] = 10\,\mathrm{Gbit/s}$. The results of this study with $\eta \in \{1, 2, 4, 8\}$ and $\mathrm{E}[B^{\mathrm{dyn}}] = 1\,\mathrm{Gbit/s}$, cf. Figure 5.5a as well as $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$, cf.



(a) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 1\,\mathrm{Gbit/s}$ (b) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$

Figure 5.5: *Impact of network topologies and traffic characteristics on blocking probabilities.*

Figure 5.5b are shown. Due to the random nature of the dynamic traffic, we performed ten runs per topology and traffic pattern and plotted confidence intervals at a significance level of 95 %.

The figures show several outcomes. First, the arrival rate definition given in Equation 5.8 results in a comparable behavior of different network topologies grouped by the load-control factor η. Small differences between network topologies are caused by different nodal degrees. Therefore, it seems to be eligible to present results only for a single network topology in the following studies to evaluate PCE algorithms and parameters. We use the *Germany50* topology from now on, cf. Table 3.1. Second, the blocking probability p_b rapidly increases with increasing mean holding time $\mathrm{E}[H]$. Assuming that at most ten percent blocking is acceptable for a network provider, $\eta = 1$ is a good choice for small traffic demands and $\eta < 1$ should be used when considering large traffic demands. It is up to the network provider to perform *access control* to hold these limits of η. Further increase of holding times or bandwidths only has a minor impact, as already many demands are blocked if a certain blocking probability is exceeded. Third, the confidence intervals of the evaluations typically range within a few

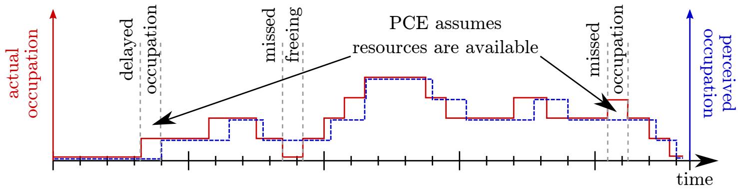


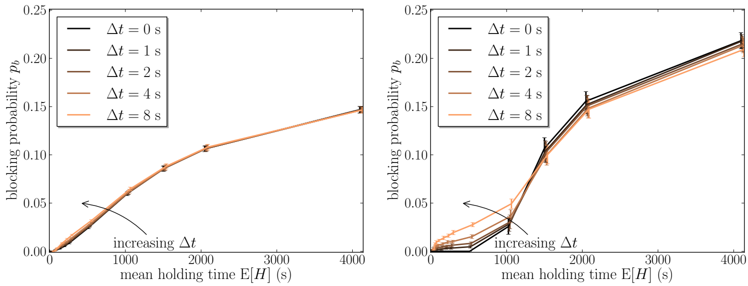
Figure 5.6: *Illustration of mismatches due to synchronization issues.*

percent. However, the confidence intervals increase for larger mean bandwidth and increasing holding times. This is due to the longer lasting effect with increased holding times. The impact of varying traffic demands (η) increases with their holding time.

5.2.2 Delayed Information Retrieval

Apart from the general availability of information, the timeliness of information is decisive to make the right and valid decisions in automated provisioning. As mentioned earlier, a PCE relies on the information it obtains from its *Traffic Engineering Database* (TED) as declared in [30]. Hereby, the TED typically retrieves its information from the routing protocol OSPF-TE.

The synchronization between PCE and TED may be disturbed, e.g., by delayed retrieval of information from OSPF-TE to the TED or by fluctuating transmission times that delay the TED information perceived by a PCE. Both effects may result in synchronization issues between the real resource utilization in the network and the utilization perceived at the PCE. Such synchronization issues are illustrated in Figure 5.6, which may comprise delayed occupation, missed occupation, or missed freeing of resources. While the latter is simply a waste of resources, delayed and missed occupation may cause the computation of paths that are valid from the PCE's point of view but are rejected due to unfulfillable requests for resources and lead to blocking.



(a) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 1\,\mathrm{Gbit/s}$ (b) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$

Figure 5.7: *Impact of synchronization offset Δt with increasing mean holding time and load factor $\eta = 1$.*

We analyzed the impact of the TED synchronization offset Δt on the overall blocking probability p_b with different mean holding times $\mathrm{E}[H]$. For the sake of simplicity, we assume a constant synchronization offset and performed studies for $\Delta t = \{0\,\mathrm{s}, 1\,\mathrm{s}, 2\,\mathrm{s}, 4\,\mathrm{s}, 8\,\mathrm{s}\}$ on the *Germany50* topology. We consider all reservations to reach their destination regardless of Δt. Especially, loss of RESV/PATH messages due to packet loss is not considered. According to the results in Section 5.2.1, we use $\eta = 1$ to keep the blocking probability within a realistic bound and investigate scenarios for $\mathrm{E}[B^{\mathrm{dyn}}] = 1\,\mathrm{Gbit/s}$ as well as $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$. The results of these studies are shown in Figure 5.7. The synchronization offset leads to a rather low increase of p_b in general. The lines for $\Delta t = 0\,\mathrm{s}$ are identical to the ones given in Figure 5.5.

This increase of p_b is due to mismatches of perceived and real resource occupation. For small $\mathrm{E}[H]$, i.e., a system with a higher fluctuation of resource occupation, as shown in Figure 5.7a, the impact of Δt is more obvious as more demands occur in the considered simulation time T. With increasing holding times, the effect of Δt decreases due to the decreased fluctuation in the network and the remaining impact gets lost in the overall increasing blocking probability.

Figure 5.7b shows that the impact of Δt with high fluctuation drastically increases with higher mean bandwidth requests. Nevertheless, the impact of Δt gets lost in the increasing $\mathrm{E}[H]$. In case of larger $\mathrm{E}[H]$ in combination with larger $\mathrm{E}[B^{\text{dyn}}]$, also the confidence intervals enlarge such that the impression of inverted impact of Δt with larger $\mathrm{E}[H]$ in Figure 5.7b is not valid.

5.2.3 Resource Utilization Strategy

A typical approach to reduce the blocking probability is the use of "a priori" strategies, i.e., acting before we are forced to. Such strategies affect the evaluation of the utilization of available resources and do not require additional knowledge or capabilities from the PCE. As described in Section 5.1.4, such evaluation of resources can be expressed by a weight function $w : \mathcal{E}_i \mapsto \mathbb{R}^+$. The resulting weighted graph reflects the utilization strategy by giving lower weights to links that should be preferred. We introduce three common resource utilization strategies for "a priori" routing optimization that can be used with any shortest path algorithm within a PCE.

Minimum Hop-Count: The minimum hop-count strategy aims at using as few resources as possible. To that end, all links are assigned a common, positive weight, resulting in a weight function $w_{\text{mh}}(e) = 1$ for $e \in \mathcal{E}_i$.

Load Balancing: To achieve a homogeneous network utilization and keep an equal amount of resources available throughout the network, load balancing can be applied. Load balancing can be realized by taking the current utilization of a link into account in the weighting function $w_{\text{lb}}(e) = 1 + r_{\text{free}}(e)/r_{\text{total}}(e)$. Note, that we do not split up demands as done in *Equal Cost Multipath* (ECMP).

Resource Fragmentation Avoidance: Choosing the shortest path or load balancing is not always the best strategy. As we do not consider the split up of bandwidth requests, situations may occur in which small amounts of available bandwidth are scattered on many links but no demand exceeding this small bandwidths can be routed. This is called external *fragmentation* in the context of hard disks. To avoid such fragmentation, we define a weight function w_{fa} that fosters

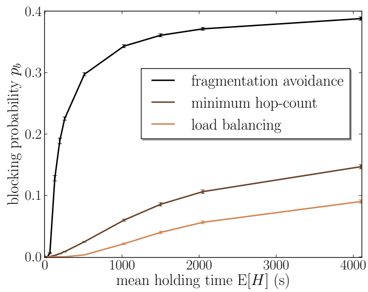


Figure 5.8: *Comparison of resource utilization strategies with increasing mean holding time,* $\mathrm{E}[B^{dyn}] = 1\,\mathrm{Gbit/s}$*, and* $\eta = 1$.

close matches of requests to resources by

$$w_{\text{fa}}(e) = 1 + \begin{cases} r_{\text{free}}(e) - b(d) & r_{\text{free}}(e) \geq b(d), \\ \infty & \text{otherwise,} \end{cases} \tag{5.9}$$

i.e., the closer the bandwidth request $b(d)$ of a demand $d \in \mathcal{D}^{\text{dyn}}$ is to the amount of available capacity $r_{\text{free}}(e)$ on a link $e \in \mathcal{E}_i$, the more preferable is link e to carry demand d. In case a resource is not sufficient for the demand, it is assigned an infinite weight to exclude it from path finding. However, this is likely to choose a longer path for a demand.

We analyzed these three strategies in a PCE on the *Germany50* network topology for $\eta = 1$ and $\mathrm{E}[B^{\text{dyn}}] = 1\,\mathrm{Gbit/s}$. The results are shown in Figure 5.8. The line representing minimum hop-count is identical to the ones shown in Figure 5.5a and Figure 5.7a for $\Delta t = 0\,\mathrm{s}$. The minimum hop-count strategy performs quite well which emphasizes this simple strategy. However, load balancing outperforms the other two strategies as it is better suited to distribute the small bandwidth requests and achieves a reduction of up to $5\ \%$ in blocking probability for large mean holding times. Load balancing also helps to avoid single points of failures and reduces the impact of network failures.

In our studies, fragmentation avoidance performs worst for two reasons. First, this strategy prefers longer paths and occupies more resources, which leads to higher blocking. Second, the performance of fragmentation avoidance must be seen in relation to the very good performance of the other strategies, which deal very well with the homogeneous bandwidth distribution used in our studies. A non-homogeneous bandwidth distribution would outline the situations, in which this strategy performs better. This is fostered by studies, in which the average bandwidth request $\mathrm{E}[B^{\mathrm{dyn}}]$ comes closer to the capacity provided by the links. Then, all three strategies perform identically. Hence, the effect of the strategies can only be seen for request to capacity ratios less than one.

5.3 Re-Optimization

As outlined in the previous section, we analyzed three PCE algorithms using "a priori" strategies to reduce the overall blocking probability. However, the capabilities of such "a priori" strategies are limited. To overcome such situations, we consider the re-optimization concept, proposed in [30], i.e., the re-routing of already established demands to resolve the blocking state and to minimize the blocking rate "a posteriori", i.e., acting on demand.

This leads to a new optimization problem, the PCE *re-optimization* problem. To solve this problem, a (minimal) sub-set of established dynamic traffic demands is to be found such that these demands need to be altered in order to block fewer demands while not breaking any established demands. This "make before break" concept is supported by RSVP-TE [19] using the "`SENDER_TEMPLATE`" object and enables a seamless transition of established connections.

In this section, we present the different types of PCE architectures as well as their capabilities. Furthermore, we introduce an efficient algorithm for re-optimization, which takes a PCE's capabilities into account, and we analyze the impact of PCE parameters in combination with this algorithm.

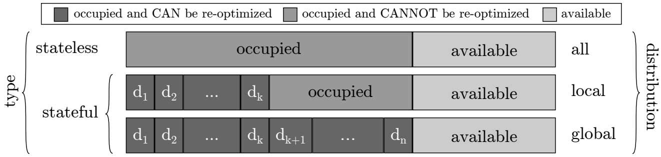


Figure 5.9: *Information available with different* PCE *architectures.*

5.3.1 Distributed and Stateful Path Computation

Up to now, we considered a PCE to be a mere dumb path computing device that only makes use of information it retrieves from its TED. Whilst this is the simplest approach, the original specification of the PCE architecture [30] already proposes additional capabilities to store routing decisions made in the past. Such a PCE with a memory is denoted as *stateful*, and the original type is therefore called *stateless*. Figure 5.9 illustrates, which information can be accessed by the two types of PCEs. With the knowledge of this memory, a PCE can take actions to re-route established demands to release resources required to route a demand that would get blocked otherwise.

Additionally, the definition of the PCE architecture [30] fosters distributed path computation approaches, i.e., multiple, communicating PCEs instead of a single “all-seeing oracle in the sky” [30] for scalability reasons as well as for administrative domains. Different kinds of distributions with different administrative domains are possible and as PCEs are distributed, each PCE only might have local knowledge, i.e., locally established demands, instead of global knowledge, i.e., all established demands. This radically influences the possibilities for re-optimization at a given PCE. The four commonly considered distributions [60] are a single, global PCE, one PCE per site, one PCE per layer, and one PCE per site and layer. These distributions are illustrated in Figure 5.10.

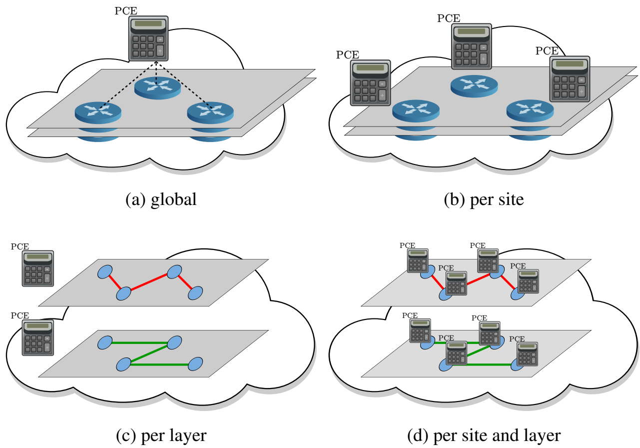


Figure 5.10: *Possible* PCE *distributions in a multi-layer network.*

5.3.2 Stateful Re-Optimization Algorithms

As aforementioned, stateful PCEs store the routing decision of their administrative domain, e.g., local or global. For an algorithmic description, we introduce the set of dynamic traffic demands $\mathcal{M}$, a PCE stores in its memory. As a PCE gets information from a TED, we also introduce the term $\mathcal{T}$ for a PCE's TED.

A naive re-optimization approach that finds the optimal re-routing solution is to try all sub-sets of $\mathcal{M}$, i.e., iterating the full power-set $\wp(\mathcal{M})$. Apparently, it can easily get very time consuming to check $2^{|\mathcal{M}|}$ sub-sets. Thus, this approach does not scale well with the number of stored routings. Another strategy is to try to re-route all single element sub-sets of $\mathcal{M}$. This has linear complexity but is not likely to find a solution as multiple demands might need to be involved to resolve the blocking state.

Therefore, we propose a systematic, efficient, and deterministic algorithm, `MINBW`, to resolve the blocking of a dynamic traffic demand $d \in \mathcal{D}^{\text{dyn}}$ that chooses the path for a blocked demand such that the amount of affected demands is minimal with regard to their cumulated bandwidth. The algorithm is structured as follows:

1. Compute a path for a demand $d \in \mathcal{D}^{\text{dyn}}$ on layer G_i, on which the PCE operates. On success, no re-optimization is required, STOP.
2. If no path was found for d, apply weight function w_{minbw} as defined in Algorithm 5.1 that prefers links with minimum bandwidth mismatch on G_i. In this algorithm, `realized(`$\mathcal{T}$`, `e`)` gives the demands realized by a link e and `freeCapacity(`$\mathcal{T}$`, `e`)` gives the available capacity, both according to the current information of the TED $\mathcal{T}$.
3. Calculate candidate path p using a CSPF algorithm in the PCE.
4. Identify set of affected demands $\mathcal{D}' \subseteq \mathcal{D}^{\text{dyn}}$ using Algorithm 5.2.
5. Tear down all demands in $\mathcal{D}'$.
6. Set up d with path p, now that sufficient resources should exist.
7. If no success for d on path p, restore demands in $\mathcal{D}'$, STOP. This may happen with $\Delta t > 0\,\text{s}$.
8. If the set-up of d was successful, try to set up all removed demands in $\mathcal{D}'$ again.
    a) Calculate new path p' for c.
    b) If the set-up of c was not successful, tear down d and restore demands in $\mathcal{D}'$, STOP.
    c) If the set-up of c was successful, continue.

Algorithm 5.1: Weight function w_{minbw}.

input : the considered link $e \in \mathcal{E}_i$ of layer G_i, on which the PCE operates, the PCE's TED $\mathcal{T}$, the PCE's memory $\mathcal{M}$, the blocked demand $d \in \mathcal{D}^{\text{dyn}}$

output: w_{minbw}

sumBw $\leftarrow \sum_{c \in \texttt{realized}(\mathcal{T},\ e) \cap \mathcal{M}} b(c)$ `// all known demands`
if *sumBw* + *freeCapacity*($\mathcal{T}$, e) $< b(d)$ **then**
  return ∞ `// there is nothing we can do`
else
  `// prefer links with minimum bandwidth mismatch`
  return $1 + \max(0, b(d) - \texttt{freeCapacity}(\mathcal{T},\ e))$

Algorithm 5.2: Demand identification for re-optimization.

input : the proposed path p in layer G_i, on which the PCE operates, the PCE's TED $\mathcal{T}$, the PCE's memory $\mathcal{M}$, the blocked demand $d \in \mathcal{D}^{\text{dyn}}$

output: The set of affected demands $\mathcal{D}'$ or $\emptyset$

$\mathcal{D}' \leftarrow \emptyset$
foreach *link* $e \in p$ **do** `// get affected demands per link`
  if *freeCapacity*($\mathcal{T}$, e) $< b(d)$ **then**
    cs $\leftarrow$ `realized`($\mathcal{T}$, e) $\cap\, \mathcal{M}$ `// all known demands`
    if *cs* $\neq \emptyset$ **then** `// select sub-set`
      sort cs $= \{c_1, \ldots, c_{|\text{cs}|}\}$, such that $b(c_i) \leq b(c_{i+1})$
      $\text{cs}_{\text{minbw}} = \{c_1, \ldots, c_k\}$ such that $\sum_{i=1}^{k} b(c_i) \geq b(d)$
      $\mathcal{D}' \leftarrow \mathcal{D}' \cup \text{cs}_{\text{minbw}}$
    else `// no success, abort!`
      return $\emptyset$
return $\mathcal{D}'$

Quite similar to MINBW would be an algorithm that chooses the candidate path for d such that the number of affected demands is minimal (MINNUM). However, the results of both algorithms would be the same in our case as we consider homogeneous bandwidth distributions in our studies.

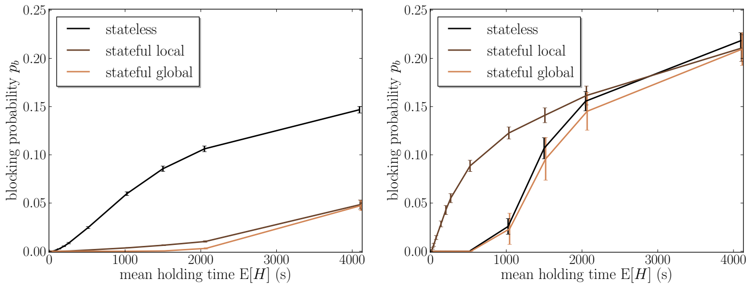
5.3.3 Impact of Re-Optimization

In the following, we analyze the impact of the proposed PCE re-optimization algorithm MINBW and the interaction of PCE parameters with re-optimization.

Blocking Probability

Analogous to Section 5.2.1, we analyze the impact of re-optimization on the resulting blocking probability p_b by performing studies on the *Germany50* topology with load-control factor $\eta = 1$ and $\mathrm{E}[B^{\text{dyn}}] = 1\,\mathrm{Gbit/s}$ as well as $\mathrm{E}[B^{\text{dyn}}] = 10\,\mathrm{Gbit/s}$. Whilst Section 5.2.1 only covered stateless PCE, cf. Section 5.3.1, we also include stateful PCE types in this study with global knowledge as well as local, per-site distribution of the PCEs.

The results of this study are shown in Figure 5.11. Figure 5.11a impressively shows the positive impact the MINBW re-optimization algorithm has with small bandwidths. Both global and local re-optimization achieve an absolute reduction of the blocking probability by $10\,\%$ with large mean holding times. As expected, the local re-optimization is outperformed by the global re-optimization, which may re-route any established demand. The results are not as clear for large bandwidths as is shown in Figure 5.11b. While the global re-optimization is still able to slightly outperform the stateless PCE, the local re-optimization performs worse than no re-optimization at all. On the one hand, this is caused by the bandwidth requirements of the demands since a single demand of 10 Gbit/s fills up the capacity of all used links. On the other hand, this negative effect is issued by the local re-optimization strategy, which comprises that a re-routing decision can block many upcoming demands that do not start at this PCE/PCC as no other PCE can re-route this demand with the local strategy.



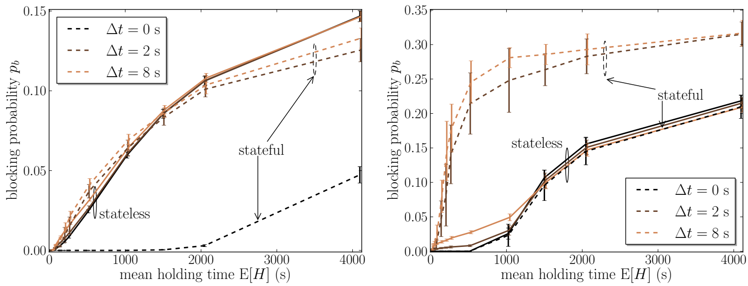
(a) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 1\,\mathrm{Gbit/s}$ (b) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$

Figure 5.11: *Comparison of central and distributed* PCE *architectures with and without state information.*

Synchronization

In Section 5.2.2, we showed that delayed information retrieval from the TED has a negative impact on the blocking probability in case of stateless PCE path computation. As re-optimization relies on valid occupation data during the calculation of new paths for the re-routed demands, it is likely that the synchronization offset Δt has a larger impact in this case. To analyze this impact, we repeat the study of Section 5.2.2 with global, i.e., the most powerful, information availability and plot the results of stateless path calculation for comparison.

As shown in Figure 5.12, mismatches due to delayed information retrieval heavily increase the blocking probability of a stateful PCE. In case of small traffic demands as shown in Figure 5.12a, the performance of the stateful PCE decreases to the level of the stateless PCE for $\Delta t > 0\,\mathrm{s}$ as decisions to re-route traffic are based on expired information and lead to re-routing decisions that cannot be fulfilled. Again, the results are different and even worse for large traffic demands, which could have been expected from the results in Figure 5.11b. The increased size of the confidence intervals in Figure 5.12b results from the impact every



(a) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 1\,\mathrm{Gbit/s}$ (b) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$

Figure 5.12: *Impact of* TED *synchronization on* PCE *re-optimization on the Germany50 topology with load factor* $\eta = 1$.

single routing decision made by a PCE has with $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$ and the randomness of demand arrivals and consecutive TED mismatches. Thus, PCE re-optimization heavily suffers from the TED synchronization offset.

Resource Utilization Strategy

In analogy to Section 5.2.3, we study the impact of re-optimization on p_b in combination with different resource utilization strategies. We revisit the studies performed in Section 5.2.3 on the *Germany50* topology with load-control factor $\eta = 1$ and $\mathrm{E}[B^{\mathrm{dyn}}] = 1\,\mathrm{Gbit/s}$. The results are shown in Figure 5.13 together with the those of stateless PCEs from Figure 5.8. Again, the `MINBW` PCE re-optimization is able to absolutely decrease the blocking probabilities by more than 10 %. The re-optimization proves to be especially effective for the fragmentation avoidance strategy, which partly sees improvements beyond 20 % but is still the worst strategy due to the homogeneous bandwidth distribution.

Finally, extra knowledge can be used in re-optimization to achieve better results in case of blocking. However, the basic requirement for acceptable blocking

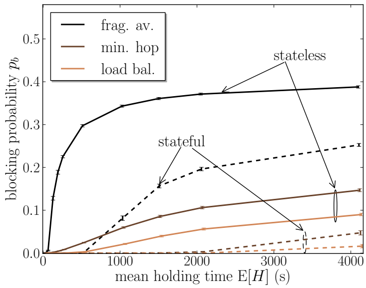


Figure 5.13: *Impact of resource utilization strategy on* PCE *re-optimization.*

rates is the match of design and provisioning, regardless of the goodness of the re-optimization. Thus, we focus on the interrelation of design and provisioning in the following.

5.4 Interrelation of Design and Provisioning

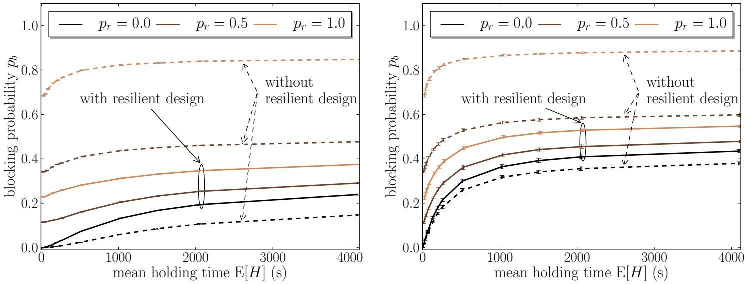
During the network construction stage, a certain amount of data, specified in $\mathcal{D}^{\text{stat}}$, was assumed to be transferred among a pair of sites/*Points of Presence* (PoPs) at a certain layer. Thus, the network has been dimensioned to carry this amount of data, which should be respected during network operation and which we control, e.g., by means of Equation 5.8. Otherwise, a mismatch between traffic forecasts and real traffic either results in waste of *Capital Expenditures* (CAPEX) due to *over-provisioning* or potentially increased blocking of dynamic traffic demands caused by *over-booking*. In this section, we consider such mismatches and their consequences.

5.4.1 Resilient Provisioning

In Chapter 4, we considered the design of resilient multi-layer networks by deploying resources not only for a primary but also for a backup path of a traffic demand. Next, we consider the introduction of resilient provisioning, i.e., providing primary and backup paths to dynamic traffic demands. Especially, we consider an over-booking-like business model to introduce dynamic resilience, i.e., the underlying multi-layer network remains unchanged while more and more dynamic traffic demands request for resilience. To that end, we introduce the probability p_r with that a dynamic traffic demand requests for dedicated 1+1 path protection, i.e., a primary and a disjoint backup path is set up and exclusively used for this dynamic traffic demand. Thus, $p_r = 0$ means no resilience at all as we considered so far and $p_r = 1$ means all dynamic demands request for resilience. However, if the resilience request cannot be fulfilled, the demand is completely rejected.

For the computation of the disjoint primary and backup path pair, we use the *Suurballe-Tarjan* algorithm [118, 119] as it is intended to work on directed weighted graphs and re-uses the *Dijkstra* shortest path algorithm that we already assumed for our PCE algorithms. However, we generalize the way the algorithm guarantees disjointness by checking for equity of links not in the IP layer, but more importantly in the physical layer. Therefore, we assume that the PCE has knowledge about the physical layer topology as described in [34].

Figure 5.14 shows the results for an extended study on resilience in dynamic provisioning. We consider three levels of resilience requests, represented by $p_r \in \{0, 0.5, 1\}$, and we compare two underlying scenarios, a network designed without resilience and a network designed with resilience, on top of the *Germany50* topology, which provides dedicated 1+1 protection for each static traffic demand in $\mathcal{D}^{\text{stat}}$. In Figure 5.14a, the results for this study with mean bandwidth $\mathrm{E}[B^{\text{dyn}}] = 1\,\mathrm{Gbit/s}$ are shown. It illustrates that the occurrence of any demand requesting resilience ($p_r > 0$) tremendously increases the blocking probability. The increased blocking is mainly caused by rejection of demands for that no disjoint path pair can be found. The use of an underlying topology, which was



(a) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 1\,\mathrm{Gbit/s}$ (b) mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$

Figure 5.14: *Impact of underlying resilience mechanism on blocking probability of traffic with increasing resilience ratio* p_r.

designed for resilient demands, is required to serve dynamic resilience as soon as p_r increases to a certain level. However, an underlying resilient design is not able to fully compensate the additional requirements of dynamic resilience. This is due to the way the dynamic traffic demands arriving during provisioning use the resources for dedicated backup that were installed during design. This usage can differ from the disjoint path pair that was intended during design and can especially use longer paths if resources are used by other demands. As in former studies, we also consider mean bandwidth $\mathrm{E}[B^{\mathrm{dyn}}] = 10\,\mathrm{Gbit/s}$. The results of this study are shown in Figure 5.14b. As with former results, the impact of dynamic resilience is immanent for small holding times, i.e., high fluctuation, and gets lost in the increase of p_b for higher holding times.

In this section, we focused on the consequences of mismatches of design and provisioning requests. In the next section, we show how provisioning studies can give feedback to the design decisions.

5.4.2 Design Feedback-Loop

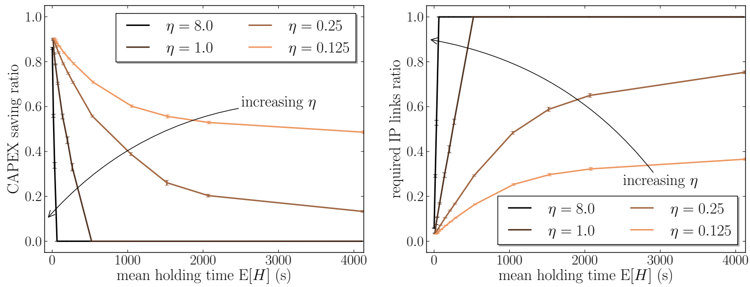
In a second study on interrelation of design and provisioning, we want to analyze a feedback mechanism from operation to construction, i.e., information about actual utilization during provisioning are fed back to the design stage. We already postulated such a mechanism [5] and presented a first algorithm using simple strategies to down-size a network designed for a forecast of static traffic requirements ($\mathcal{D}^{\text{stat}}$) to a set of dynamic traffic demands ($\mathcal{D}^{\text{dyn}}$) using a probabilistic IP routing algorithm [11].

For the current study, we extend the principle of the previous work and transferred them to the PCE concept while also avoiding the non-determinism. For this purpose, we introduce a weight function w_{au} that aims at avoiding to use resources that have not been used before. This weight function is defined for a link $e \in \mathcal{E}_i$ as

$$w_{\text{au}}(e) = \begin{cases} 1 & N(e) > 0, \\ W_{\text{max}} & \text{otherwise}, \end{cases} \tag{5.10}$$

where $N(e)$ is the number of times link e was used before and $W_{\text{max}} = 1000$ is the penalty for using a link that was not used before. The function $N(e)$ requires historical knowledge of the demand routings, which can in principle be provided by a global, stateful PCE.

With this new criterion, we performed a study on the utilization of an IP topology that is the result of our network design. We assume this topology to be suited for a certain amount of dynamic traffic such that no blocking occurs for all considered holding times. For the *Germany50* topology used in this study, this is reached for $\eta_0 = 0.25$. We varied the value of η between 0.125 and 8. The results are shown in Figure 5.15. In this study, the blocking probability p_b is not of interest as the weight function w_{au} itself does not cause additional blocking since the usage of links is not prevented but simply not preferred.



(a) relative CAPEX savings

(b) ratio of required IP links

Figure 5.15: *Evaluation on feedback from the provisioning to the design stage.*

Instead, we focus on CAPEX savings shown in Figure 5.15a and the ratio of links that need to be deployed shown in Figure 5.15b. These two figures indicate that for values of η lower than the intended design level η_0, a huge amount of links (up to 70 %) and CAPEX (up to 50 %) can be saved. This again shows that CAPEX does directly correlate with the number of IP links. When the load-control factor η exceeds the intended level η_0, the CAPEX savings as well as the unused links rapidly decrease to zero.

Furthermore, the weight function w_{au} can be used as a basis for first studies on energy-aware network operation with temporarily de-activated links, presuming that adequate equipment exists.

5.5 Lessons Learned

Compared to the design of multi-layer networks, their operation with dynamic traffic demands posed new optimization problems and objectives. The exploration of the possibilities of our approach for automated provisioning, which is based on GMPLS and determines paths via PCEs, showed a notable range in the results with different parameters.

For studies on the general impact of the PCE architecture, we developed algorithms and analyzed the impact of different parameters, like network topologies and traffic patterns, on *stateless* PCEs and the resulting blocking probability p_b. We found that the impact of load increases with the holding time and is especially crucial if dynamic demands request bandwidths near to the capacity of the links. Furthermore, we analyzed delayed retrieval of information from the TED. The results show the importance of TED synchronization, which is the more important the higher traffic is fluctuating. Different resource utilization strategies proved to reduce the impact of load. The proposed minimum hop count and load balancing strategies always performed best in our studies. However, we found that their choice depends on traffic and network characteristics.

Following the studies on stateless PCEs, we introduced the *stateful* PCE concept that allows to perform re-optimization of established demands. We presented algorithms that use information on established demands to reduce blocking. The availability of global knowledge, e.g., by information exchange, proved to outperform local knowledge, which even could lead to results worse than without re-optimization. An efficient re-optimization heuristic outlined the impact of re-optimization, which ranges from 10 to 50 % with different resource utilization strategies. We also showed that the PCE re-optimization causes an increased dependency on up-to-date TED information. We even found that inappropriate TED synchronization might lead to worse results than without re-optimization.

Finally, we focused on interrelations of design and provisioning. Our studies dealt with mismatches in the planned and the actual utilization. We proved that resilience constraints in operation should always be considered during network design as otherwise limited degrees of freedom cause severe blocking even with a rather low ratio of resilience requests. In addition, the possibility to give feedback of resource utilization to the design process was considered. This proved to be a feasible way to not only plan networks for a given traffic matrix for future requirements, but also to allow intermediate steps to be considered. Our results stated notable CAPEX savings and equipment reduction by the introduction of such steps. These principles can also be used for energy-aware network operation.

6 Conclusion

But then again, all good things must come to an end.

"Star Trek – The Next Generation", episode "All Good Things...", 1994.

Core networks always have to deal with growing bandwidth requests and increasingly more dynamic traffic as well as with technological evolutions with respect to data rates, energy efficiency, and cost-benefit ratios. Along with a general cost pressure, this steady change in conditions and possibilities forces network providers to integrate new technologies that extend and improve their networks. This evolution of technologies paired with continued support for legacy services leads to a hierarchy of technologies, the *multi-layer networks*.

In this monograph, we considered the two main stages in the life cycle of a multi-layer network: the *construction* and the *provisioning*. The variety of available networking components makes the construction and operation of multi-layer networks an important and complex optimization problem. We identified the input and optimization parameters and developed efficient heuristics for the design of resilient multi-layer networks as well as their operation. Thus, we succeeded in finding general principles for a technology-agnostic optimization of multi-layer networks and efficient operation that significantly extend existing knowledge.

For the construction of multi-layer networks, most approaches use combinatorial procedures that neither scale with the problem size nor give rules how the actual solution was found. Thus, we focused on heuristic approaches that integrate comprehensible rules whose consequences can be evaluated in parameter studies. We created a component model that is able to map all kinds of current networking equipment and allows for a mostly technology agnostic network design, which brings a lot of flexibility not present in existing approaches.

We analyzed the impact of characteristics of our heuristics on the resulting multi-layer network. It was shown that the order, in which traffic demands are given for sequential processing, is decisive. Further heuristics were developed that determine the demand order in advance and decrease the overall *Capital Expenditures* (CAPEX) by up to 40 % in the performed studies. Furthermore, the impact of input parameters was evaluated. We analyzed different multi-layer architectures and found that those being simple to manage can be far more expensive in terms of CAPEX than more complex architectures. An analysis of the resulting distribution of cost on the different layers showed that the cost for optical switching and enabling of *Wavelength Division Multiplex* (WDM) is constant and almost negligible if a certain load level is exceeded in the network. If given the freedom to arbitrarily select amongst all available layers, our algorithms proved to create networks that outperform all considered candidates.

The construction of *resilient* multi-layer networks was considered after exploring the general design principles of unprotected multi-layer networks. In this step, we focused on the numerous parameters of resilience. In contrast to existing literature, we considered different as well as multiple layers taking care of the actual protection and analyzed the impact on the resulting multi-layer networks.

The performance of the introduced heuristics was evaluated with different protecting layers. For single as well as multiple protecting layers, this choice was found to be decisive with respect to the resulting CAPEX as well as the ability to protect a certain scenario. While our single-layer protection algorithms showed to be able to protect against all considered single-link failures, single total-node failures cannot be fully covered. Thus, such failures need to be considered within the design algorithm itself to provide adequate protection. We revealed that effective multi-layer resilience is hard to realize if special constraints for disjointness in the physical layer cannot be met. As well, a relation of the properties of the physical layer and the performance of our algorithms with multi-layer resilience was worked out. We showed that the algorithm as well as the protection variant impact the consequences of failures. Thus, algorithms need to be trimmed for reduction of subsequent failures instead of mere cost optimality.

In current networks, the set-up of new traffic demands typically goes through well-defined business processes. These often include legally binding communication between partners, e.g., by fax, which can take days. Dynamic traffic patterns with on-demand provisioning of resources are not considered yet and the impact of such dynamic operation is not sufficiently known. To improve the knowledge in this field, we considered *automated* provisioning based on *Generalized Multi-Protocol Label Switching* (GMPLS) and *Path Computation Elements* (PCEs).

We discussed several PCE architectures and introduced algorithms for PCE path computation, whose definition is not part of the PCE specification. The algorithms reveal the necessity to carefully monitor the network load and to perform access control to limit the blocking probability of demands. We discussed the general concept of PCE re-optimization and provided two realistic implementations that proved to leverage blocking of demands. We evaluated the impact of delayed and unreliable information used for path computation and outlined that its negative impact especially affects scenarios with very short-lived demands and is especially important when performing re-optimization. Lastly, we investigated the impact of mismatches of design and provisioning, i.e., under- and over-booking. While the latter results in blocking, the first can be expressed by possible CAPEX savings as well as equipment reduction and can also be considered for energy-aware network operation. We found that the possibility to give feedback from network operation to construction is able to reduce the cost.

In the course of this monograph, a comprehensive overview of resilient multi-layer network design and provisioning was given, taking into account new and extended aspects compared to existing literature. On the basis of this work and the evaluation software *Multi-Layer Network Engineering and Optimization* (MuLaNEO) [16] developed along with it, further elaborated evaluations can be realized. Amongst others, shared protection would be a worthwhile extension of the construction and provisioning. This would open up improved resilience in case of single failures and reduced resource requirements. In addition, the acceptance at a lower data rate instead of fully blocking a demand would lead to trade-offs with regard to a business model reflecting such service downgrades.

Nomenclature

General:

$\wp(X)$	Power-set of a set X, i.e., the set of all sub-sets of X
$\|X\|$	Cardinality of a set X, i.e., the number of elements in X
$\sigma(X)$	Standard deviation of a random variable X
$\mathrm{E}[X]$	Expected value of a random variable X

Chapter 2:

$\mathcal{V}$	Sites of a multi-layer network, also called *Points of Presence* (PoPs)
$\mathcal{E}_i$	Links at layer G_i with $\mathcal{E}_i \subseteq \mathcal{V} \times \mathcal{V} \setminus \{(v, v) \in \mathcal{V}\}$
G_i	i-th layer of a multi-layer network, $G_i = (\mathcal{V}, \mathcal{E}_i)$, a directed multi-graph without self-loops
$\mathcal{G}$	Set of layers $\mathcal{G} = \{G_0, \dots, G_k\}$ comprising a multi-layer network
$\alpha(e)$	Source of a link, $\alpha : \mathcal{E}_i \mapsto \mathcal{V}$
$\omega(e)$	Destination of a link, $\omega : \mathcal{E}_i \mapsto \mathcal{V}$
$r_{\text{total}}(e)$	Total resources of a link
$r_{\text{free}}(e)$	Available, free resources of a link
$\mathcal{D}^{\text{stat}}$	Set of static traffic demands defined in Equation 2.1
$\mathcal{D}^{\text{dyn}}$	Set of dynamic traffic demands defined in Equation 2.2
$b(d)$	Bandwidth of a demand

Chapter 3:

$\mathcal{C}$	CAPEX model $\mathcal{C} = (C, \zeta)$ with C being the available equipment and $\zeta : C \mapsto \mathbb{R}_0^+$ associating cost with it
$\mathcal{I}$	Set of interfaces defined in C
$G(x)$	Layer of a link or component x
B^{stat}	Set of bandwidth granularities for static demands derived from the available equipment
$\zeta(\mathcal{G})$	Overall cost of a multi-layer network, cf. Equation 3.5
$\overline{\zeta}(\mathcal{G})$	Normalized cost of a multi-layer network, cf. Equation 3.6
n_{r}	Number of routed demands
n_{b}	Number of blocked demands

Chapter 4:

p_A	Probability of a service or network element to be operational at one particular point in time
$\mathcal{F}$	Set of all possible failure scenarios
$\mathcal{F}_{\mathcal{E}_0}$	Set of single-link failures in the fiber layer
$\mathcal{F}_{\mathcal{V} \cap G}$	Set of single layer-node failures in layer $G \in \mathcal{G}$
$\mathcal{F}_{\mathcal{V}}$	Set of single total-node failures
$\mathcal{F}_x^{\text{sub}}$	Set of subsequent failures caused by $\mathcal{F}_x \in \mathcal{F}$
$\mathcal{G}_P$	Sub-set of $\mathcal{G}$ providing protection
$n_{\text{p},G}$	Number of demands protected by protecting layer $G \in \mathcal{G}_P$
$\hat{\overline{\zeta}}(\mathcal{G})$	Normalized cost of a resilient multi-layer network, cf. Equation 4.5

Chapter 5:

a	The average number of active demands in the system according to Little's law, defined in Equation 5.4
ρ_o	Offered load defined in Equation 5.5
p_b	Blocking probability, i.e., the ratio of blocked to overall number of demands in $\mathcal{D}^{\text{dyn}}$, as defined in Equation 5.6
ρ_c	Carried load defined in Equation 5.7
$\pi(v)$	Population of a site $v \in \mathcal{V}$, $\pi : \mathcal{V} \mapsto \mathbb{N}^+$
T	Simulation run time in which dynamic traffic demands arrive
η	Load-control factor $\eta > 0$
$\mathcal{D}'$	Set of affected demands, used in re-optimization in Section 5.3
A	Random variable for inter-arrival times
H	Random variable for holding times
B^{dyn}	Set of bandwidth granularities for dynamic demands
L	Random variable for the number of active demands
$\mathcal{T}$	*Traffic Engineering Database* (TED)
Δt	TED synchronization offset

Acronyms

ADSL	Asymmetric Digital Subscriber Line
ALT	Alternating single-layer protection, introduced in Section 4.2
AS	Autonomous System
ASON	Automatically Switched Optical Network
ASTN	Automatic Switched Transport Network
ATM	Asynchronous Transfer Mode
AXL	Auxiliary Cross Layer algorithm, introduced in Section 3.1.2
BER	Bit Error Rate
BRAS	Broadband Remote Access Server
CAPEX	Capital Expenditures
CDF	Cumulative Distribution Function
CE	Carrier Ethernet
COM	Combined single-layer protection, introduced in Section 4.2
CSPF	Constrained Shortest Path First
DCF	Dispersion Compensation Filter
DGE	Dynamic Gain Equalizer
DML	Disjoint Multi-Layer protection, introduced in Section 4.3
DSLAM	Digital Subscriber Line Access Multiplexer
ECMP	Equal Cost Multipath
EDFA	Erbium-Doped Fiber Amplifier
ELT	Expected Loss of Traffic
ERO	Explicit Route Object, cf. [19]

FDM	Frequency Division Multiplex
FSC	Fiber Switch Capable
FTTH	Fiber To The Home
GE	Gigabit Ethernet
GFP	Generic Framing Procedure
GMPLS	Generalized Multi-Protocol Label Switching, cf. [24]
ILP	Integer Linear Program
IML	Identical Multi-Layer protection, introduced in Section 4.3
ION	Intelligent Optical Network
IP	Internet Protocol
IST	Information Society Technologies
LAN	Local Area Network
LDP	Label Distribution Protocol
LER	Label Edge Router
LMP	Link Management Protocol, cf. [27]
LP	Linear Program
LSC	Lambda Switch Capable
LSP	Label Switched Path
LSR	Label Switching Router
MAN	Metropolitan Area Network
MIP	Mixed Integer Programming
MPLS	Multi-Protocol Label Switching
MTBF	Mean Time Between Failures
MTTR	Mean Time To Recover
MuLaNEO	Multi-Layer Network Engineering and Optimization, cf. [16]
MuLaViTo	Multi-Layer Visualization Tool, cf. [14]
NOBEL	Next generation Optical networks for Broadband European Leadership, cf. [35]
OADM	Optical Add/Drop Multiplexer

OCh	Optical Channel
ODU	Optical channel Data Unit
O-E-O	Opto-Electric-Optical
OLA	Optical Light Amplifier
OMS	Optical Multiplex Section
OPEX	Operational Expenditures
OSPF	Open Shortest Path First
OSPF-TE	Open Shortest Path First with Traffic Engineering, cf. [23]
OTH	Optical Transport Hierarchy, cf. [107]
OTN	Optical Transport Network
OTV	Overlay Transport Virtualization
OXC	Optical Cross Connect
PANEL	Protection Across NEtwork Layers, cf. [49]
PCC	Path Computation Client
PCE	Path Computation Element
PCECP	Path Computation Element Communication Protocol, cf. [32]
PCEP	Path Computation Element Communication Protocol, cf. [33]
PON	Passive Optical Network
PoP	Point of Presence, also called *site*
PoS	Packet over SONET
PSC	Packet Switch Capable
QoS	Quality of Service
RML	Recursive Multi-Layer protection, introduced in Section 4.3
RRO	Route Record Object, cf. [33]
RSVP	Resource ReSerVation Protocol, cf. [17]
RSVP-TE	Resource ReSerVation Protocol with Traffic Engineering, cf. [19]
RWA	Routing and Wavelength Assignment
SDH	Synchronous Digital Hierarchy
SDM	Space Division Multiplex

SEP	Separate single-layer protection, introduced in Section 4.2
SLA	Service Level Agreement
SNDlib	Survivable Network Design library
SONET	Synchronous Optical NETwork
SPOF	Single Point Of Failure
SRG	Shared Risk Group
STM	Synchronous Transport Module
TCO	Total Cost of Ownership
TDM	Time Division Multiplex
TE	Traffic Engineering
TED	Traffic Engineering Database
VNE	Virtual Network Embedding
VPLS	Virtual Private LAN Service
WAN	Wide Area Network
WDM	Wavelength Division Multiplex
XML	Extensible Markup Language

Bibliography and References

Bibliography of the Author

— Journals —

[1] M. Menth, M. Duelli, R. Martin, and J. Milbrandt, "Resilience Analysis for Packet-Switched Communication Networks," *IEEE/ACM Transactions on Networking*, vol. 17, Dec. 2009.

[2] A. Fischer, J. F. Botero, M. Duelli, D. Schlosser, X. Hesselbach, and H. de Meer, "ALEVIN – A Framework to Develop, Compare, and Analyze Virtual Network Embedding Algorithms," *Electronic Communications of the EASST, Kommunikation in Verteilten Systemen 2011*, vol. 37, Mar. 2011.

[3] S. Meier, M. Barisch, A. Kirstädter, D. Schlosser, M. Duelli, M. Jarschel, T. Hoßfeld, K. Hoffmann, M. Hoffmann, W. Kellerer, A. Khan, D. Jurca, and K. Kozu, "Provisioning and Operation of Virtual Networks," *Electronic Communications of the EASST, Kommunikation in Verteilten Systemen 2011*, vol. 37, Mar. 2011.

[4] M. Duelli, X. Qin, J. Ott, and E. Weber, "MuLaNEO – Eine Software zur Planung und Optimierung mehrschichtiger Netze," *Praxis der Informationsverarbeitung und Kommunikation (PIK)*, vol. 34, no. 3, Sep. 2011.

— Conference Papers —

[5] M. Duelli, M. Hartmann, M. Menth, R. Hülsermann, and M. Düser, "Performance Evaluation of IP over Cost-Optimized Optical Multilayer Networks with SRLGs," in *9th ITG-Fachtagung Photonische Netze*, Leipzig, Germany, Apr. 2008.

[6] M. Dzida, M. Zagozdzon, M. Zotkiewicz, M. Pióro, M. P. Petterson, M. Duelli, and M. Menth, "Three Methods for Optimizing Single-Shortest Path Routing," in *4th Conference on Next Generation Internet Design and Engineering (NGI)*, Krakow, Poland, Apr. 2008.

[7] M. Duelli, C. Pluntke, and M. Menth, "Minimizing Installation Costs of Survivable DWDM-Mesh Networks: A Heuristic Approach," in *4th Conference on Next Generation Internet Design and Engineering (NGI)*, Krakow, Poland, Apr. 2008.

[8] M. Duelli, E. Weber, and M. Menth, "A Generic Algorithm for CAPEX-Aware Multi-Layer Network Design," in *10th ITG-Fachtagung Photonische Netze*, Leipzig, Germany, May 2009.

[9] C. Pluntke, M. Menth, and M. Duelli, "CAPEX-Aware Design of Survivable DWDM Mesh Networks," in *IEEE International Conference on Communications (ICC)*, Dresden, Germany, Jun. 2009.

[10] T. Michaelis, M. Duelli, M. Chamania, B. Lichtinger, F. Rambach, and S. Türk, "Network Planning, Control and Management Perspectives on Dynamic Networking," in *European Conference and Exhibition on Optical Communication (ECOC)*, Vienna, Austria, Sep. 2009.

[11] M. Duelli, A. Endler, and M. Menth, "Deferring CAPEX Investments in Multi-Layer Networks through IP Traffic-Dependent Expansion Stages," in *11th ITG-Fachtagung Photonische Netze*, Leipzig, Germany, May 2010.

[12] M. Duelli, X. Qin, and M. Menth, "Greedy Design of Resilient Multi-Layer Networks," in *6th Conference on Next Generation Internet Design and Engineering (NGI)*, Paris, France, Jun. 2010.

[13] D. Schlosser, M. Duelli, and S. Goll, "Performance Comparison of Hardware Virtualization Platforms," in *IFIP/TC6 NETWORKING 2011*, Valencia, Spain, May 2011.

— Software —

[14] M. Duelli, J. Ott, and T. Müller, "MuLaViTo – Multi-Layer Visualization Tool," Apr. 2011, http://mulavito.sf.net.

[15] M. Duelli, V. Singeorzan, D. Schlosser, A. Fischer, M. T. Beck, J. F. Botero, L. D. Cervantes, and X. Hesselbach, "ALEVIN – ALgorithms for Embedding VIrtual Networks," May 2011, http://alevin.sf.net.

[16] M. Duelli, X. Qin, J. Ott, and E. Weber, "MuLaNEO – Multi-Layer Network Engineering and Optimization," Dec. 2011, http://mulaneo.sf.net.

Internet Documents

— Request For Comments —

[17] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource ReSerVation Protocol (RSVP)," Sep. 1997, http://tools.ietf.org/html/rfc2205.

[18] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," Jan. 2001, http://tools.ietf.org/html/rfc3031.

[19] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels," Dec. 2001, http://tools.ietf.org/html/rfc3209.

[20] W. Lai and D. McDysan, "Network Hierarchy and Multilayer Survivability," Nov. 2002, http://tools.ietf.org/html/rfc3386.

[21] L. Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description," Jan. 2003, http://tools.ietf.org/html/rfc3471.

[22] ——, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions," Jan. 2003, http://tools.ietf.org/html/rfc3473.

[23] D. Katz, K. Kompella, and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2," Sep. 2003, http://tools.ietf.org/html/rfc3630.

[24] E. Mannie, "Generalized Multi-Protocol Label Switching (GMPLS) Architecture," Oct. 2004, http://tools.ietf.org/html/rfc3945.

[25] E. Mannie and D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control," Oct. 2004, http://tools.ietf.org/html/rfc3946.

[26] K. Kompella and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)," Oct. 2005, http://tools.ietf.org/html/rfc4203.

[27] J. Lang, "Link Management Protocol (LMP)," Oct. 2005, http://tools.ietf.org/html/rfc4204.

[28] K. Kompella and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)," Oct. 2005, http://tools.ietf.org/html/rfc4206.

[29] D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control," Jan. 2006, http://tools.ietf.org/html/rfc4328.

[30] A. Farrel, J.-P. Vasseur, and J. Ash, "A Path Computation Element (PCE)-Based Architecture," Aug. 2006, http://tools.ietf.org/html/rfc4655.

[31] J. Ash and J.-L. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements," Sep. 2006, http://tools.ietf.org/html/rfc4657.

[32] J.-L. Le Roux, "Path Computation Element Communication Protocol (PCECP) Specific Requirements for Inter-Area MPLS and GMPLS Traffic Engineering," Jun. 2007, http://tools.ietf.org/html/rfc4927.

[33] J.-P. Vasseur and J.-L. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)," Mar. 2009, http://tools.ietf.org/html/rfc5440.

[34] E. Oki, T. Takeda, J.-L. Le Roux, and A. Farrel, "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering," Sep. 2009, http://tools.ietf.org/html/rfc5623.

— Others —

[35] "IST NOBEL and IST NOBEL-2 Project Website," Sep. 2008, http://www.ist-nobel.org.

[36] *CPLEX*, ILOG, Inc., http://www.cplex.com.

[37] *SCIP: Solving Constraint Integer Programs*, Zuse Institute Berlin (ZIB), http://scip.zib.de.

[38] IEEE 802.3ba – Standard for Local and Metropolitan Area Networks, "40Gb/s and 100Gb/s Ethernet Task Force," Jun. 2010.

General References

[39] E. B. Desurvire, "Capacity Demand and Technology Challenges for Lightwave Systems in the Next Two Decades," *Journal of Lightwave Technology*, vol. 24, no. 12, Dec. 2006.

[40] A. M. Odlyzko, "Internet Traffic Growth: Sources and Implications," in *Optical Transmission Systems and Equipment for WDM Networking II*, vol. 5247, Aug. 2003.

[41] Cisco Systems, Inc., "Cisco Visual Networking Index – Forecast and Methodology, 2010–2015," San Jose, CA, USA, Jun. 2011.

[42] W. D. Grover, *Mesh-Based Survivable Networks: Options and Strategies for Optical, MPLS, SONET, and ATM Networking.* Upper Saddle River, NJ, USA: Prentice Hall PTR, 2004.

[43] J.-P. Vasseur, M. Pickavet, and P. Demeester, *Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS.* Waltham, MA, USA: Morgan Kaufmann, 2004.

[44] B. Mukherjee, *Optical WDM Networks*, 1st ed. New York, NY, USA: Springer, 2006.

[45] Cisco Systems, Inc., "Entering the Zettabyte Era," San Jose, CA, USA, Jun. 2011.

[46] J. Späth, "Dynamic Routing and Resource Allocation in WDM Transport Networks," *Computer Network*, vol. 32, May 2000.

[47] C. Meusburger, D. A. Schupke, and J. Eberspächer, "Multiperiod Planning for Optical Networks - Approaches Based on Cost Optimization and Limited Budget," Beijing, China, May 2008.

[48] A. M. C. A. Koster and A. Zymolka, "Provably Good Solutions for Wavelength Assignment in Optical Networks," in *IFIP Working Conference on Optical Network Design and Modeling (ONDM)*, Milan, Italy, Feb. 2005.

[49] P. Demeester, M. Gryseels, A. Autenrieth, C. Brianza, L. Castagna, G. Signorelli, R. Clemente, M. Ravera, A. Jajszczyk, D. Janukowicz, K. V. Doorselaere, and Y. Harada, "Resilience in Multilayer Networks," *IEEE Communications Magazine*, vol. 37, no. 8, Aug. 1999.

[50] R. Hülsermann, M. Gunkel, C. Meusburger, and D. A. Schupke, "Cost Modeling and Evaluation of Capital Expenditures in Optical Multilayer Networks," *Journal of Optical Networking*, vol. 7, no. 9, Aug. 2008.

[51] A. K. Somani, *Survivability and Traffic Grooming in WDM Optical Networks.* Cambridge, England: Cambridge University Press, 2006.

[52] K. Zhu and B. Mukherjee, "A Review of Traffic Grooming in WDM Optical Networks: Architecture and Challenges," *Optical Networks Magazine*, Mar./Apr. 2003.

[53] P. Cholda, A. Mykkeltveit, B. E. Helvik, O. J. Wittner, and A. Jajszczyk, "A Survey of Resilience Differentiation Frameworks in Communication Networks," *IEEE Communications Surveys & Tutorials*, vol. 9, no. 4, 2007.

[54] M. Scheffel, R. G. Prinz, C. G. Gruber, A. Autenrieth, and D. A. Schupke, "Optimal Routing and Grooming for Multilayer Networks with Transponders and Muxponders," in *IEEE Globecom*, San Francisco, CA, USA, Nov. 2006.

[55] A. Farrel and I. Bryskin, *GMPLS: Architecture and Applications.* Waltham, MA, USA: Morgan Kaufmann, 2005.

[56] S. Tomic, B. Statovci-Halimi, A. Halimi, W. Muellner, and J. Fruehwirth, "ASON and GMPLS - Overview and Comparison," *Photonic Network Communications*, vol. 7, no. 2, 2004.

[57] S. Dasgupta, J. Cavalcante de Oliveira, and J.-P. Vasseur, "Path-Computation-Element-Based Architecture for Interdomain MPLS/GMPLS Traffic Engineering: Overview and Performance," *IEEE Network Magazine*, vol. 21, no. 4, 2007.

[58] W. Imajuku, E. Oki, R. Papneja, S. Morishita, K. Ogaki, M. Miyazawa, K. Miyazaki, H. Nakazato, H. Sugiyama, J. Allen, S. Hasegawa, N. Sakuraba, I. Nishioka, S. Seno, Y. Nakahira, D. Ishii, S. Okamoto, T. Unen, M. Blumhardt, H. Rakotoranto, and V. Pandian, "A Multi-Area MPLS/GMPLS Interoperability Trial over ROADM/OXC Network," *IEEE Communications Magazine*, vol. 47, no. 2, Feb. 2009.

[59] E. W. Dijkstra, "A Note on Two Problems in Connexion with Graphs," *Numerische Mathematik*, vol. 1, 1959.

[60] V. López, B. Huiszoon, J. Fernández-Palacios, O. G. de Diosx, and J. Aracil, "Path Computation Element in Telecom Networks: Recent Developments and Standardization Activities," in *IFIP Working Conference on Optical Network Design and Modeling (ONDM)*, Kyoto, Japan, Feb. 2010.

[61] G. B. Dantzig, *Linear Programming and Extensions*, 1st ed. Princeton, NJ, USA: Princeton University Press, 1963.

[62] S. Even, A. Itai, and A. Shamir, "On the Complexity of Timetable and Multicommodity Flow Problems," *SIAM Journal on Computing*, vol. 5, no. 4, 1976.

[63] M. Pióro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks.* Waltham, MA, USA: Morgan Kaufman, 2004.

[64] S. Verbrugge, D. Colle, M. Pickavet, P. Demeester, S. Pasqualini, A. Iselt, A. Kirstädter, R. Hülsermann, F. Westphal, and M. Jäger, "Methodology and Input Availability Parameters for Calculating OpEx and CapEx Costs for Realistic Network Scenarios," *Journal of Optical Networking*, vol. 5, no. 6, May 2006.

[65] C. Meusburger and D. A. Schupke, "Method to Estimate the Break-Even Point Between SLA Penalty Expenses and Protection Costs," in *Optical Fiber Communication Conference (OFC)*, San Diego, CA, USA, Feb. 2008.

[66] C. Mas Machuca, J. Eberspächer, M. Jäger, and A. Gladisch, "Service Migration Cost Modeling," in *8th ITG-Fachtagung Photonische Netze*, Leipzig, Germany, 2007.

[67] R. Prinz, D. A. Schupke, A. Kirstädter, A. Iselt, A. Autenrieth, and B. Edmaier, "Cost Structures of Transport Networks," in *Beiträge zum ITG VDE NetWorkshop*, Sep. 2003.

[68] M. Doms and C. Forman, "Prices for Local Area Network Equipment," *Information Economics and Policy*, vol. 17, no. 3, Jan. 2005.

[69] M. Gunkel, R. Leppla, M. Wade, A. Lord, D. A. Schupke, G. Lehmann, C. Fürst, S. Bodamer, B. Bollenz, H. Haunstein, H. Nakajima, and J. Martensson, "A Cost Model for the WDM Layer," in *Photonics in Switching*, Crete, Greece, Oct. 2006.

[70] M. Minoux, "Discrete Cost Multicommodity Network Optimization Problems and Exact Solution Methods," *Annals of Operations Research*, vol. 106, 2001.

[71] H. Zhu, H. Zang, K. Zhu, and B. Mukherjee, "A Novel Generic Graph Model for Traffic Grooming in Heterogeneous WDM Mesh Networks," *IEEE/ACM Transactions on Networking*, vol. 11, no. 2, 2003.

[72] G. Huiban, S. Pérennes, and M. Syska, "Traffic Grooming in WDM Networks with Multi-Layer Switches," vol. 5, 2002.

[73] P. Belotti, A. Capone, G. Carello, F. Malucelli, F. Senaldi, and A. Totaro, "MPLS over Transport Network: Two Layers Approach to Network Design with Statistical Multiplexing," in *2nd*, Valencia, Spain, Apr. 2006.

[74] P. Belotti, A. Capone, G. Carello, and F. Malucelli, "Multi-Layer MPLS Network Design: The Impact of Statistical Multiplexing," *Computer Networks*, vol. 52, no. 6, 2008.

[75] A. Capone, G. Carello, and R. Matera, "Multi-Layer Network Design with Multicast Traffic and Statistical Multiplexing," in *IEEE Globecom*, Washington, DC, USA, Nov. 2007.

[76] M. Batayneh, H. Zhu, L. Song, and B. Mukherjee, "Cost-Efficient WDM Mesh Network Design with Line Cards of Multiple Ports," in *IEEE Globecom*, San Francisco, CA, USA, Nov. 2006.

[77] K. Assis, W. Giozza, and H. Waldman, "WDM Optical Networks: A Complete Design," *Journal of Communications and Information Systems (JCIS) IEEE ComSoc/SBrT Joint Issue*, vol. 20, no. 3, Dec. 2005.

[78] M. Düser and A. Gladisch, "An Analytical Approach to Cost Optimal Network Design," in *10th ITG-Fachtagung Photonische Netze*, Leipzig, Germany, 2009.

[79] A. Knippel and B. Lardeux, "The Multi-Layered Network Design Problem," *European Journal of Operational Research*, vol. 183, no. 1, Nov. 2007.

[80] S. Orlowski and R. Wessäly, "An Integer Programming Model for Multi-Layer Network Design," Konrad-Zuse-Zentrum für Informationstechnik Berlin, ZIB Preprint ZR-04-49, Dec. 2004. [Online]. Available: http://www.zib.de/PaperWeb/abstracts/ZR-04-49

[81] S. Orlowski, A. M. C. A. Koster, C. Raack, and R. Wessäly, “Two-Layer Network Design by Branch-and-Cut featuring MIP-based Heuristics,” in *International Network Optimization Conference (INOC)*, Spa, Belgium, Apr. 2007.

[82] A. M. C. A. Koster, S. Orlowski, C. Raack, G. Baier, and T. Engel, “Single-Layer Cuts for Multi-Layer Network Design Problems,” Seattle, WA, USA, Nov. 2007.

[83] C. Pavan, R. M. Morais, and A. N. Pinto, “Estimating CapEx in Optical Multilayer Networks,” in *Conference on Telecommunications (ConfTel)*, Santa Maria da Feira, Portugal, May 2009.

[84] N. Naas and H. T. Mouftah, “Cost-Effective Heuristics for Planning GMPLS-Based Transport Networks,” *Optical Switching and Networking: A Computer Networks Journal*, vol. 7, no. 1, 2010.

[85] S. Verbrugge, D. Colle, P. Demeester, R. Hülsermann, and M. Jäger, “General Availability Model for Multilayer Transport Networks,” in *International Workshop on the Design of Reliable Communication Networks (DRCN)*, Island of Ischia, Italy, Oct. 2005.

[86] D. Medhi and D. Tipper, “Multi-Layered Network Survivability – Models, Analysis, Architecture, Framework and Implementation: An Overview,” in *DARPA Information Survivability Conference & Exposition (DISCEX)*, Hilton Head Island, SC, USA, Jan. 2000.

[87] M. B. Kiese and A. Autenrieth, “A Fast Heuristic for DWDM Routing and Grooming,” in *International Workshop on the Design of Reliable Communication Networks (DRCN)*, La Rochelle, France, Oct. 2007.

[88] T. Engel, A. Autenrieth, and J.-C. Bischoff, “Packet Layer Topologies of Cost Optimized Transport Networks,” in *IFIP Working Conference on Optical Network Design and Modeling (ONDM)*, Brunswick, Germany, Feb. 2009.

[89] H. Höller and S. Voß, "A Heuristic Approach for Combined Equipment-Planning and Routing in Multi-Layer SDH/WDM Networks," *European Journal of Operational Research*, vol. 171, no. 3, Jun. 2006.

[90] H. Höller, B. Melián, and S. Voß, "Applying the Pilot Method to Improve VNS and GRASP Metaheuristics for the Design of SDH/WDM Networks," *European Journal of Operational Research*, vol. 191, no. 3, Jun. 2008.

[91] M. Pickavet, P. Demeester, D. Colle, D. Staessens, B. Puype, L. Depré, and I. Lievens, "Recovery in Multilayer Optical Networks," *Journal of Lightwave Technology*, vol. 24, no. 1, 2006.

[92] T. Lehman, X. Yang, N. Ghani, F. Gu, C. Guok, I. Monga, and B. Tierney, "Multilayer Networks: An Architecture Framework," *IEEE Communications Magazine*, vol. 49, no. 5, 2011.

[93] M. S. Kodialam and T. V. Lakshman, "Dynamic Routing of Bandwidth Guaranteed Tunnels with Restoration," in *IEEE Infocom*, Tel-Aviv, Israel, Mar. 2000.

[94] ——, "Integrated Dynamic IP and Wavelength Routing in IP over WDM Networks," in *IEEE Infocom*, Anchorage, AK, USA, Apr. 2001.

[95] S. De Maesschalck, D. Colle, A. Groebbens, C. Develder, A. Lievens, P. Lagasse, M. Pickavet, P. Demeester, F. Saluta, and M. Quagliatti, "Intelligent Optical Networking for Multilayer Survivability," *IEEE Communications Magazine*, vol. 40, no. 1, 2002.

[96] C. M. Gauger and M. Köhn, "Dimensioning of SDH/WDM Multilayer Networks," in *4th ITG-Fachtagung Photonische Netze*, Leipzig, Germany, May 2003.

[97] R. Hülsermann, M. Jäger, S. O. Krumke, D. Poensgen, J. Rambau, and A. Tuchscherer, "Dynamic Routing Algorithms in Transparent Optical

Networks," in *IFIP Working Conference on Optical Network Design and Modeling (ONDM)*, Budapest, Hungary, Feb. 2003.

[98] R. Srinivasan and A. K. Somani, "Analysis of Optical Networks with Heterogeneous Grooming Architectures," *IEEE/ACM Transactions on Networking*, vol. 12, Oct. 2004.

[99] M. Köhn, S. Bodamer, C. M. Gauger, S. Gunreben, G. Hu, and D. Saß, "Comparison of IP/WDM Transport Network Architectures for Dynamic Data Traffic," in *11^{th} European Conference on Networks and Optical Communications (NOC2006)*, Berlin, Germany, 2006.

[100] M. Tornatore, M. Carcagni, C. S. Ou, B. Mukherjee, and A. Pattavina, "Efficient Shared-Segment Protection Exploiting the Knowledge of Connection Holding Time," in *IEEE Globecom*, San Francisco, CA, USA, Nov. 2006.

[101] T. Cinkler, P. Hegyi, G. Geleji, and J. Szigeti, "Protection for Adaptive Multi-Layer Traffic Engineering," in *International Workshop on the Design of Reliable Communication Networks (DRCN)*, La Rochelle, France, Oct. 2007.

[102] S. Norden, M. M. Buddhikot, M. Waldvogel, and S. Suri, "Routing Bandwidth Guaranteed Paths with Restoration in Label Switched Networks," *IEEE International Conference on Network Protocols*, vol. 0, 2001.

[103] S. Gunreben and F. Rambach, "Assessment and Performance Evaluation of PCE-based Inter-Layer Traffic Engineering," in *IFIP Working Conference on Optical Network Design and Modeling (ONDM)*, Vilanova i la Geltrú, Spain, Mar. 2008.

[104] J. Sommer, S. Gunreben, F. Feller, M. Kohn, A. Mifdaoui, D. Sass, and J. Scharf, "Ethernet – A Survey on its Fields of Application," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 2, Apr. 2010.

[105] A. Kasim, P. Adhikari, N. Chen, N. Finn, N. Ghani, M. Hajduczenia, P. Havala, G. Heron, M. Howard, and L. Martini, *Delivering Carrier Ethernet: Extending Ethernet Beyond the LAN.* New York, NY, USA: McGraw-Hill Osborne Media, 2007.

[106] G. Barlow, "A G.709 Optical Transport Network Tutorial," Jan. 2009, JDSU White Paper.

[107] ITU-T Recommendation G.709, "Interfaces for the Optical Transport Network (OTN)," Dec. 2009.

[108] E. Palcopolou, D. A. Schupke, C. G. Gruber, A. Kirstädter, and C. Merkle, "Service-Oriented Traffic Models for Future Backbone Networks," in *8th ITG-Fachtagung Photonische Netze*, Leipzig, Germany, May 2007.

[109] A. Dwivedi and R. E. Wagner, "Traffic Model for USA Long-Distance Optical Network," in *Optical Fiber Communication Conference (OFC)*, Baltimore, MD, USA, Mar. 2000.

[110] R. Hülsermann, S. Bodamer, M. Barry, A. Betker, M. Jäger, J. Späth, C. M. Gauger, and M. Köhn, "A Set of Typical Transport Network Scenarios for Network Modelling," in *5th ITG-Fachtagung Photonische Netze*, Leipzig, Germany, Apr. 2004.

[111] A. W. Brander and M. C. Sinclair, "A Comparative Study of k-Shortest Path Algorithms," in *11th UK Performance Engineering Workshop for Computer and Telecommunications Systems*, Sep. 1995.

[112] S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessäly, "SNDlib 1.0–Survivable Network Design Library," *NETWORKS*, vol. 55, no. 3, 2010. [Online]. Available: http://sndlib.zib.de

[113] Cisco Systems, Inc., *Cisco Overlay Transport Virtualization*, San Jose, CA, USA, Jun. 2011.

[114] Alcatel-Lucent., *Managed Carrier Ethernet VPNs Enabled by VPLS*, Paris, France, Sep. 2007.

[115] G. Iannaccone, C.-N. Chuah, R. Mortier, S. Bhattacharyya, and C. Diot, "Analysis of Link Failures in an IP Backbone," in *ACM SIGCOMM Internet Measurement Workshop*, Marseille, France, Nov. 2002.

[116] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, and C.-N. Chuah, "Characterization of Failures in an IP Backbone," in *IEEE Infocom*, Hongkong, Mar. 2004.

[117] S. D. Patre, G. Maier, M. Martinelli, and A. Pattavina, "Design of Static WDM Mesh Networks with Dedicated Path-Protection," in *IFIP Working Conference on Optical Network Design and Modeling (ONDM)*, Torino, Italy, Feb. 2002.

[118] J. W. Suurballe, "Disjoint Paths in a Network," *Networks*, vol. 4, no. 2, 1974.

[119] J. W. Suurballe and R. E. Tarjan, "A Quick Method for Finding Shortest Pairs of Disjoint Paths," *Networks*, vol. 14, no. 2, 1984.

[120] D. Eppstein, "Finding the k Shortest Paths," *SIAM Journal on Computing*, vol. 28, no. 2, 1998.

[121] J. Y. Yen, "Another Algorithm for Finding the k Shortest Loopless Network Paths," *Management Science*, vol. 17, no. 11, Jul. 1971.

[122] A. K. Erlang, "Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges," *Elektroteknikeren*, vol. 13, 1917.

[123] J. Späth, "Impact of Traffic Behaviour on the Performance of Dynamic WDM Transport Networks," in *28th European Conference on Optical Communications*, Copenhagen, Denmark, Sep. 2002.

ISSN 1432-8801