RINGOSTAR – An Evolutionary Performance Upgrade of Optical Ring Networks

vorgelegt von
Diplom-Ingenieur

Martin Herzog

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
– Dr.-Ing. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr.-Ing. Klaus Petermann
Berichter:	Prof. Dr.-Ing. Adam Wolisz
Berichter:	Prof. Dr.-Ing. Jörg Eberspächer

Tag der wissenschaftlichen Aussprache: 20. April 2006

Berlin 2006

D 83

Zusammenfassung

DAS im Internet übertragene Datenvolumen vergrößert sich beständig und erfordert den kontinuierlichen Ausbau der zugrundeliegenden Infrastruktur. Ein genauerer Blick auf die Infrastruktur zeigt, daß diese hierarchisch aus Backbone-, Metro- und Access-Netzen aufgebaut ist. Die nationalen oder internationalen Backbone-Netze werden ausreichend Kapazität durch den Einsatz von Wavelength Division Multiplexing (WDM) basierten Links, welche mit optischen Add-Drop Multiplexern (OADMs) und optischen Crossconnects (OXCs) verbunden sind, zur Verfügung stellen. Metropolitan Area Networks (MANs), oder kürzer Metro-Netze, verbinden die Backbone-Netze mit den lokalen Access-Netzen, welche die Daten von und zu den einzelnen Benutzern transportieren. Durch den Einsatz von verbesserten Local Area Network (LAN) Technologien wie Gigabit Ethernet (GbE) und breitbandigem Netzzugriff mit Digital Subscriber Loop (DSL) und Cable-Modems stellen Access-Netze immer Bandbreite zur Verfügung. Die meisten bestehenden Metro-Netze basieren auf Synchronous Optical Network/Synchronous Digital Hierarchy (SONET/SDH) Technologie, welche Circuit-Switching verwendet und dadurch burst-artigen Datenverkehr nur ineffizient übertragt, wodurch ein Bandbreitenengpass im Metro-Bereich entsteht. Dieser Bandbreitenengpass, der oft als *Metro Gap* bezeichnet wird, verhindert daß High-Speed Clients und Service Provider in lokalen Access-Netzen die große Menge an im Backbone verfügbarer Bandbreite nutzen können.

In den letzen Jahren wurden verschiedene Metro-Architekturen vorgeschlagen und untersucht, wobei der Großteil dieser auf einer Ring- oder, weniger gewöhnlich, Sterntopologie basiert. Nachdem wir die vorgeschlagenen Lösungen umfassend beprochen und in Hinblick auf die Anforderungen im Metro-Bereich verglichen haben, argumentieren wir, daß eine hybride Ring-Stern Architektur, bestehend aus einem single-channel packet-switched Ring und einem single-hop WDM Stern-Netz, ein vielversprechender Ansatz zur Überwindung der Metro Gap ist. Wir schlagen eine Architektur und ein korrespondierendes Zugriffsprotokoll für ein solches hybrides Netz vor, das wir RINGOSTAR nennen. Bestehende packet-switched Ring-Netze, wie z.B. IEEE 802.17 Resilient Packet Ring (RPR), können evolutionär und dadurch kostengünstig zum RINGOSTAR Netzwerk aufgerüstet werden, indem eine *Untermenge* der Ringknoten unter Verwendung von *Dark Fiber*, welche im Metro-Bereich leicht verfügbar ist, mit einem optischen single-hop Stern-Netz verbunden wird.

RINGOSTAR basiert auf zwei Techniken zur Verbesserung der Leistung von Ring-Netzen, welche in dieser Arbeit vorgeschlagen werden. Die erste Technik heisst *Proxy Stripping* und dient dazu, die Kapazität von packetvermittelnden Ring-Netzen signifikant zu erhöhen und macht darüberhinaus die Kapazität solcher Netze *skalierbar*. Die zweite Technik, die wir *Protectoration* nennen, ermöglicht eine sowohl schnelle als auch bandbreiteneffiziente Wiederherstellung der Netzfunktionalität auch bei *mehreren* Knotenausfällen und/oder Kabelbrüchen in Ring-Netzen, welche sonst nur *einzelne* Fehler tolerieren. Desweiteren schlagen wir Mechanismen vor und untersuchen diese, welche die Unterstützung von verschiedenen Dienstgüteklassen sowie Fairnesskontrolle in RINGOSTAR ermöglichen.

Wir bewerten die Leistung der von uns vorgeschlagenden Architektur und der zugrundeliegenden Techniken zur Leistungsverbesserung umfassend für verschiedene Verkehrsszenarien, inklusive self-similar und hot-spot Verkehr, durch mathematische Analyse und verifizierende Computersimulationen. Die Ergebnisse zeigen, daß z.B. bei Verbindung von 32 von 256 Ringknoten mit einem Stern-Netz die Kapazität für uniformen Verkehr um einen Faktor von fast zehn ansteigt.

Abstract

INTERNET traffic volumes are growing and require the transmission capacity of the underlying infrastructure to be continuously extended. A closer look at this infrastructure reveals that the Internet architecturally relies on a three level hierarchy consisting of backbone networks, metropolitan area networks, and local access networks. The national or international backbone networks will provide abundant bandwidth by employing wavelength division multiplexing (WDM) links which are interconnected with reconfigurable optical add-drop multiplexers (OADMs) and optical crossconnects (OXCs). The metropolitan area networks (MANs), or metro networks for short, interconnect the backbone networks with the local access networks that carry the data from and to the individual users. By employing advanced local area network (LAN) technologies, such as Gigabit Ethernet (GbE), and broadband access, such as digital subscriber loop (DSL) and cable modems, access networks provide increasing amounts of bandwidth. Most existing metro networks are based on synchronous optical network/synchronous digital hierarchy (SONET/SDH) technology, a circuit-switched networking technology which carries bursty data traffic relatively inefficiently, thus resulting in a bandwidth bottleneck at the metro level. This bandwidth bottleneck, which is widely referred to as the *metro gap*, prevents the high-speed clients and service providers in local access networks from tapping into the vast amounts of bandwidth available in the backbone.

Numerous metro architectures have been proposed and investigated during the past few years, most of them relying on either a ring or, less common, on a star topology. After comprehensively reviewing and comparing the proposed solutions with respect to the requirements specific to the metropolitan area, we argue that a hybrid ring-star network architecture relying on a single-channel packet-switched ring and a single-hop WDM star network is a promising approach for future metro networks to overcome the metro gap. We propose an architecture and corresponding access protocol for such a hybrid network that we call *RINGOSTAR*. Existing packet-switched ring networks, such as the IEEE 802.17 Resilient Packet Ring (RPR), can be evolutionary, and therefore cost-efficiently, upgraded to the RINGOSTAR network by connecting a *subset* of the ring nodes to an optical star network using *dark fiber* which is abundantly available in metropolitan areas.

RINGOSTAR builds on two underlying performance enhancing techniques for ring networks which are proposed in this work. The first technique is called *proxy stripping* and provides a means to significantly increase the capacity of packet-switched ring networks and makes the fixed ring capacity *scalable*. The second technique, that we call *protectoration*, enables both fast and bandwidth efficient recovery from *multiple* link and/or node failures in ring networks that usually can only recover from *single* failures. Furthermore, we propose mechanisms to enable Quality of Service (QoS) support and fairness control in RINGOSTAR.

We comprehensively evaluate the performance of our proposed architecture and the underlying performance enhancing techniques for various network configurations and traffic scenarios, including self-similar and hot-spot traffic, by means of mathematical analysis and verifying computer simulations. Performance results show that, for instance, when interconnecting 32 out of 256 ring nodes via a star subnetwork, proxy striping increases the network capacity for uniform traffic by a factor almost equal to ten.

Acknowledgement

I am very grateful to *Prof. Martin Maier* who supported me in creating this work tremendously and from whom I have learned a lot. I am also indebted to my advisor *Prof. Adam Wolisz* for giving me lots of freedom in all aspects of my work while at the same time proving me on more than one occasion that in difficult situations I can rely on his advise and support. I thank *Prof. Martin Reisslein*, as well as *Prof. Hyo-Sik Yang*, *Prof. Michael Scheutzow*, and *Dr. Stefan Adams* for their important contributions to this work and for making the collaboration so trouble-less and productive.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

OVER the last ten years, the amount of traffic carried by the Internet has been growing at a rate of approximately 70 to 150% per year. The growth can be expected to continue at this rate till at least the end of this decade [2]. In analogy to Moore's Law for semiconductors, which states that the processing power and the number of transistors in a microprocessor approximately doubles every 18 months, this trend is often referred to as 'Moore's Law for Internet traffic'. Altough Moore's Law is not a natural law, but results from a complex interaction between technology, sociology, and economics, it has still held with remarkable regularity and for various technologies over many decades [2]. In case of Internet traffic growth, the basic underlying mechanism is that on one side new applications of the Internet create a demand for more transmission capacity, while at the same time advances WDM technology enable network operators to continuously increase the capacity of their networks. (With WDM multiple data channels are transmitted over a single optical fiber enabling network operators to multiply the capacity of their existing infrastructure without installing new fiber which would be very costly.) The additional capacity in turn stimulates innovation of new applications which further increase the demand for more bandwith. A point worth mentioning in this context is that in the future Internet "the typical piece of information will never be looked at by a human being" [3]. Most data will be generated from and used by machines. This is an interesting point, since otherwise the demand for more bandwidth would saturate when the Internet is able to deliver the maximum amount of data a human being is able to consume, which could be the case in near future in technically high developed countries. That also implicates that multimedia traffic, which is often used to illustrate the need for more bandwidth, will only be a small fraction of the total amount of Internet traffic in the long term. Furthermore, the majority of the worlds population still has no, or only very limited, access to the Internet. Besides technical innovation, the increasing number of Internet users alone will result in significant traffic growth.

Clearly, WDM will remain the key technology to satisfy the ever increasing demand for more bandwidth within the next years. However, a closer look at the infrastructure of today's Internet reveals that it consists of different domains that, besides the need for more bandwidth, all face different limitations and challenges. As illustrated in Fig. 1.1, the infrastructure is hierarchically composed of the long-haul (or backbone) network, metropolitan area networks (MANs), or metro networks for short, and access networks. More details on these domains and their underlying technologies are provided in the next chapter. In the following we only provide an intuitive understanding.

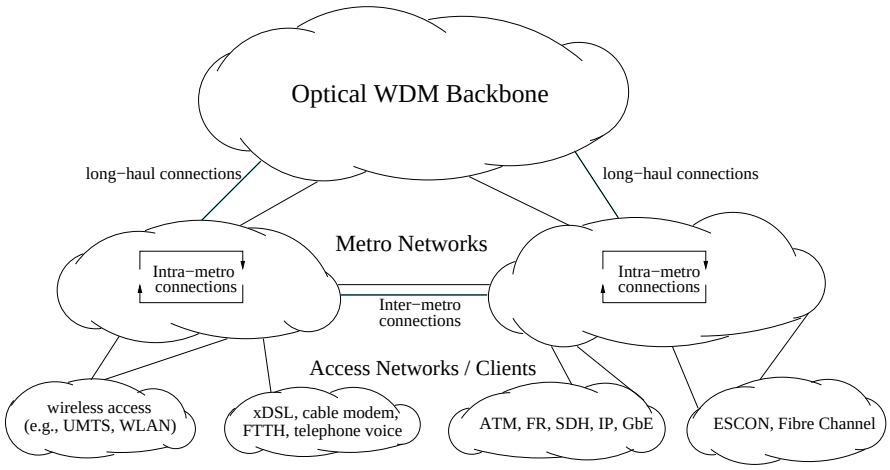


Figure 1.1: Internet infrastructure hierarchy consisting of access, metro, and long-haul networks.

At the bottom end of the hierarchy, local access networks carry the data from and to individual end users. Note that end users range from individual households up to large business premises. By employing advanced local area network (LAN) technologies, such as Gigabit Ethernet (GbE), and broadband access, such as digital subscriber loop (DSL) and cable modems, access networks provide increasing amounts of bandwidth. At the top end of the hierarchy, high capacity links interconnect different regions of a country to the national backbone network and the national backbones are connected with huge international or intercontinental bandwidth pipes to form the world wide Internet. Similar to access networks, future backbone networks will also provide increasing amounts of bandwidth. In short, this is achieved by employing WDM links which are interconnected with reconfigurable optical add-drop multiplexers (OADMs) and optical crossconnects (OXCs) [4]. In between, metro networks interconnect several local access networks within a metropolitan area with each other and with the national backbone network.

Most existing metro networks are based on synchronous optical network/synchronous digital hierarchy (SONET/SDH) technology, a circuit-switched networking technology that has originally been designed to carry telephone voice traffic which has different characteristics than Internet data traffic. A phone conversation results in two data streams, both of the same relatively small and constant bit-rate. Each stream corresponds to the speaker's voice in either direction. Such traffic is efficiently handled by setting up a bidirectional constant bit-rate data channel between the speakers which is called circuit in SONET/SDH. Internet data traffic, however, is highly asymmetric in both directions, the bit-rate is generally higher, and the data is transported in discrete packets that typically arrive in bursts. While extensions enabling SONET/SDH to carry Internet traffic exist, these still rely on constant bit-rate circuits into which the individual data packets are mapped. Due to the burstiness of the traffic the bit-rate of the circuits must be much higher than the average bit-rate of the transported data which is inefficient. This results in a bandwidth bottleneck at the metro level between high-speed access and backbone networks. Together with other limitations of SONET/SDH, which will be discussed in Chapter 2, this is widely referred to as the 'metro gap' [5, 6].

This metro gap may become more severe as proxy cache servers are more widely deployed in the metro networks. These proxy caches, which are employed to reduce the network latency,

to balance server load, and to increase the content availability [7], may result in an increase of local Internet traffic and thus exacerbate the metro gap. This trend may be further intensified by the increased use of cellular phones and handheld devices employing next generation wireless technologies, such as the Universal Mobile Telecommunication System (UMTS) and high-speed wireless local area networks (WLANs), for Internet services, which will increase the amount of locally maintained content, especially as home appliances, cars, and other electronic devices begin to utilize the metro network [8]. In addition, future peer-to-peer applications where each attached user will also operate as a server, e.g., file sharing, may dramatically increase the amount of intra-metro area traffic.

The vast majority of metro networks consist of bidirectional fiber rings running SONET/SDH. Often, multiple such rings are interconnected to form larger networks. A typical example is a metro core ring interconnecting multiple metro edge rings. Ring networks or, more precisely, bidirectional rings, are attractive because they enable simple protection against link or node failures. Recovery from failures is an important requirement for telecommunication networks in general and especially important in metropolitan areas where failures, e.g., due to cable dig-ups, occur more frequently than in the long-haul domain. In a bidirectional ring, each pair of nodes can communicate via either ring. If the ring in corresponding to the preferred direction fails the other ring can be used for the transmission, as illustrated in Fig. 1.2 (left). Another advantageous characteristic of ring networks is that, compared to other topologies such as star networks, they require less fiber to interconnect all nodes.

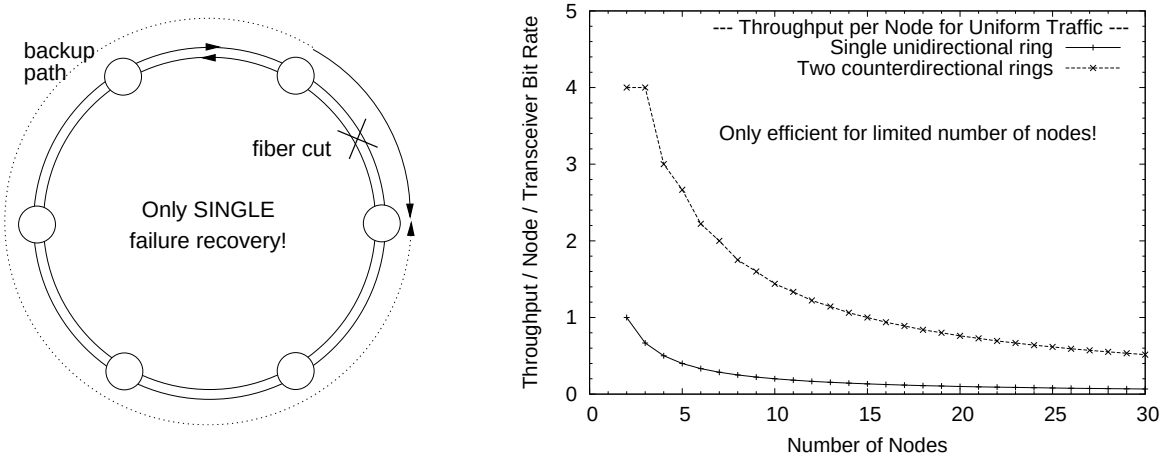


Figure 1.2: Ring networks: Topology and failure recovery in bidirectional rings (left), efficiency vs. number of nodes for uni- and bidirectional rings (right).

Several network architectures aiming to overcome the metro gap gap have been proposed and most of them stick to the ring topology. One reason is that it is more cost efficient to reuse an existing fiber ring and only upgrade the nodes versus additionally deploying a new fiber infrastructure. A prominent example for a solution addressing the problems of SONET/SDH in the metro area is the new industry standard RPR. RPR not only allows to reuse an existing fiber ring but even offers the option to be deployed as a subsystem of a SONET/SDH ring. However, it should be noted that in the past lots of fiber has been deployed in metropolitan areas and especially unlit, so-called *dark fiber* is now abundantly available at relatively low prices. Dark fiber means that the end systems required to use the fiber for data transmission,

i.e., optical transmitters and receivers, are not included when leasing the fiber. (For a better understanding note that in the past it has been more common not to lease raw fiber but an electronic interface, to which for instance a computer could be directly attached.) While it is certainly advantageous for new metro solutions to reuse existing infrastructures, dark fiber offers an attractive opportunity to cost efficiently extend existing networks.

Furthermore, altough being very common, ring networks have two major limitations. First, while bidirectional rings can recover from *single* failures, if *multiple* failures occur, e.g., due to natural disasters or terrorism, the ring is split into two or more disjoint segments. Especially for larger metro networks, that carry the traffic of a huge number of users, recovery from multiple failure scenarios is a desirable feature. Second, and even more important, the performance of ring networks is *inherently limited* in terms of making efficient use of the available bandwidth, especially for high numbers of nodes. This also translates into a poor scalability of both the number of nodes and the network capacity. As we will see in the next chapter, bandwidth efficiency and scalability are crucial requirements for metro networks. To illustrate these limitations Fig. 1.2 (right) shows the throughput that each node in a uni- and bidirectional ring achieves for uniform traffic versus the number of nodes interconnected with the ring. (Note that the throughput per node is the reciprocal of the mean hop distance which is derived for uni- and bidirectional rings in Section 5.3.1.) Uniform traffic means that all nodes sends the same amount of data to each other node and is typical for metro core rings. In both types of ring networks the throughput per node decreases asymptotically with the number of nodes. Intuitively speaking, the higher the number of nodes, the more traffic from other nodes each node has to forward and the less capacity remains to send local traffic. Therefore, ring networks are only efficient for small numbers of nodes and the scalability of the number of nodes is limited. Furthermore, the capacity of ring networks also scales poorly since in case of an capacity upgrade *all* nodes need to be upgraded which involves high cost.

We sum up the previous discussion as follows:

- Increasing traffic volumes along with limitations in current metro technology, namely SONET/SDH, result in the metro gap.
- Most metro networks rely on a ring topology and therefore suffer inherently from limited bandwidth efficiency, scalability, and failure recovery capabilities.
- Dark fiber is abundantly available in metropolitan areas, relatively inexpensive, and could be used to cost efficienty extend existing metro infrastructures.

In this work, we propose and investigate a *performance upgrade* for optical ring networks that makes use of dark fiber to overcome the aforementioned shortcomings. The proposed architecture, which we call 'RINGOSTAR', features significantly *increased network capacity*, *scalability*, and enables *multiple failure recovery*.

1.1 Methodology and Outline

In the following we present the outline of this work to provide an overview of the structure of this document and illustrate our methodology.

- **Chapter 1 - Introduction:** We have introduced the basic problem that this thesis tackles, commonly referred to as the 'metro gap'. A discussion of inherent limitations of metro ring networks served us as motivation for our idea of providing an architectural

performance upgrade for optical ring networks.

- **Chapter 2 - Background:** This chapter starts with an introduction to optical transmission and optical networks. We then discuss currently existing optical networks with a focus on metro networks and the metro gap. Furthermore, we define the requirements that future metro solutions have to meet in order to overcome the problems in the metro area.

- **Chapter 3 - Related Work:** As the overwhelming majority of metro networks relies on a ring topology we present a comprehensive survey of previous work on metro ring systems. This work has been published in *IEEE Communications Surveys and Tutorials* [9].

- **Chapter 4 - Ring vs. Star Topology:** An alternative approach for metro systems are WDM single-hop star networks for which we conduct a detailed performance comparison with networks based on a ring topology. This work appeared in the *IEEE Journal on Selected Areas in Communications (JSAC)* [10].

- **Chapter 5 - Motivation of Our Approach:** After having gained insight about previous work on metro networks in the previous two chapters we identify and compare the major solutions with respect to the previously defined metro requirements. It turns out that providing a performance upgrade for ring networks, more precisely the combination of the IEEE 802.17 Resilient Packet Ring standard an with a WDM single-hop star network, seems to be a promising approach. Thus, our research question is the development and evaluation of such a hybrid architecture. We review previous work on ring performance enhancements.

- **Chapter 6 - RINGOSTAR:** We define the basic architecture and a corresponding access protocol for our hybrid ring-and-star network which we call *'RINGOSTAR'*. We demonstrate that RINGOSTAR features a significantly lower mean hop distance compared to a ring network what can be interpreted as an indicator for high performance, as examined in more detail in the following chapters. This work has been published in the *IEEE/OSA Journal of Lightwave Technology (JLT)* [11].

- **Chapter 7 - Proxy Stripping:** To demonstrate the potential of our approach in detail, we conduct a comprehensive performance evaluation of RINGOSTAR's basic underlying performance enhancing mechanism, which we call *'proxy stripping'*. We show that proxy stripping significantly improves the performance of packet-switched ring networks, such as RPR. The performance evaluation is performed by means of mathematical analysis with verifying computer simulations. This work appeared in the *OSA Journal of Optical Networking (JON)* [12].

- **Chapter 8 - Protectoration:** We propose the *'protectoration'* technique to provide robustness against multiple link and node failures. Using mathematical analysis and verifying computer simulations we show that protectoration results in significantly improved resilience against link or node failures compared to RPR. This work has been published in the *IEEE/OSA Journal of Lightwave Technology (JLT), Special Issue on Optical Networks* [13].

- **Chapter 9 - QoS Support & Fairness Control:** Besides resilience, Quality of Service (QoS) support is an important requirement for future metro networks. Following our strategy to combine the strengths of RPR and WDM star networks, we adapt RPR's sophisticated QoS mechanism to RINGOSTAR. In order to eliminate fairness problems related to QoS support, we extend an fairness control mechanism for RPR called Distributed Virtual-time Scheduling in Rings (DVSR) to be used with our hybrid architecture. Parts of this work appeared in *IEEE Communications Magazine* [14].
- **Chapter 10 - Conclusions:** We conclude our work and argue how we solved the research question by discussing RINGOSTAR's performance with respect to the metro requirements and in comparison with other metro solutions. Furthermore, we provide an overview on the major contributions made is this work and reflect over possible future work.

Chapter 2

Basics

THE purpose of this chapter is to introduce some basics on optical and metroplitan networks to ease the understanding of the remainder of this work. The following discussion is mostly intended for readers with no or very little background on optical networks and aims at providing an intuitive but still sufficiently detailed introduction to of the most important concepts. First, we discover the basics of fiber optic point-to-point (PtP) transmission and the components involved. We then proceed to concepts and devices used to build optical *networks* composed of many such PtP links. Finally, we take look at currently existing optical networks with a focus on metropolitan area networks and their specific limitations. As the network architecture proposed in this work is probably most relevant to future metro networks, we also discuss the specific requirements of the metropolitan area that are crucial to be addressed by future metro solutions in order to be successful. We map these relatively high-level requirements to requirements on the architectural and access control protocol level where this work concentrates on.

2.1 Optical Transmission Basics

Optical fiber is the transmission medium of choice for wired telecommunication networks. It provides huge bandwidth in the order of several Tbit/s, low attenuation of only about 0.2 dB/km, low signal distortion, and low wear. Optical transmission equipment is characterized by low power consumption and low space requirements which are important features in packed equipment rooms.

Fig. 2.1 (a) schematically shows a fiber-optic point-to-point transmission system. At the transmitter side a laser (Tx) generates light pulses corresponding to the data bits to be transmitted which propagate along the fiber and are detected by a photodiode (Rx) at the receiver side. Note that pulse code modulation (PCM) is preferred over other modulation formats because optical tranmission systems are nonlinear, i.e., the analog signals resulting from other modulation formats are distorted by the system. However, optical fiber provides much more bandwidth than the electronics involved are capable to handle. To make better use of the available bandwidth, in most systems multiple transmitter-receiver pairs use the same fiber simultaneously, as depicted in Fig. 2.1 (b). This technology is called wavelength division multiplexing (WDM) because each transmitter emits light at a different wavelength and the signals of all transmitters are multiplexed into the same fiber. At the end of the fiber the individual wavelengths are demultiplexed from the signal and fed into the corresponding

receivers. In the following, we will discuss each of the components of such a PtP transmission system in more detail. The discussion is based on [1] and [15].

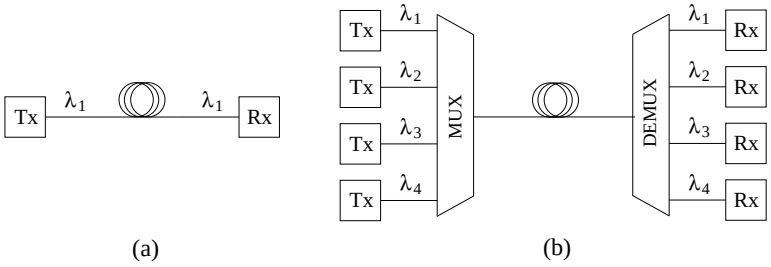


Figure 2.1: Single channel (a) vs. WDM transmission system (b).

2.1.1 Optical Fiber

Generally, fiber optic transmission systems are characterized by a tradeoff between the maximum possible fiber length and bit rate. This tradeoff results mostly from three effects that reduce the quality of the optical signal as it propagates along the fiber, namely *dispersion*, *attenuation*, and *nonlinear effects*. We will discuss each of these effects in more detail.

Chromatic Dispersion

Optical fiber comes in two flavors, single-mode fiber (SMF) and multi-mode fiber (MMF). Simply speaking, in a MMF each light pulse is composed of several overlayed components (or modes) each of which has a slightly different propagation speed. As the initially short light pulse propagates along the fiber the slower modes lag behind the faster modes and the pulse broadens, as illustrated in Fig. 2.2. This effect is called *mode dispersion*. As neighboring pulses begin to overlap they can no longer be detected properly by the receiver. Short pulses need to undergo less dispersion than longer pulses to begin to overlap. Therefore, there is a tradeoff between the maximum possible length of the fiber, i.e., the maximum tolerable amount of dispersion, and the maximum possible bit rate.

SMF has a thinner core than MMF which results in the pulse consisting of only a single mode. Altough single-mode fiber and transmission equipment is more expensive compared to multi-mode systems, in metropolitan and long-haul networks only single-mode systems are used due to their superior performance. However, single-mode systems also suffer from pulse dispersion because the propagation speed of the light pulse not only depends on the specific mode but also on the wavelength of the light. While the optical spectrum of the light emitted by a good quality laser is very narrow, it is broadened as the signal is modulated with the data to be transmitted, i.e., as pulses are sent instead of a constant power signal. A fourier transform of a light pulse shows that is it consists of different wavelength components (or frequencies) covering a spectrum with a width approximately equal to the bit rate. Each of these components propagates at slightly different speeds, again resulting in a broadening of the pulses. Because here the dispersion effect depends on the wavelength, i.e., the colour of the light, it is called *chromatic dispersion*.

As in multimode systems, this also results in a tradeoff between maximum possible bit rate B and the length of the fiber L. However, the bandwidth-length product $B\sqrt{L}$ is much larger enabling significantly higher distances at significantly higher bit rates. For standard SMF,

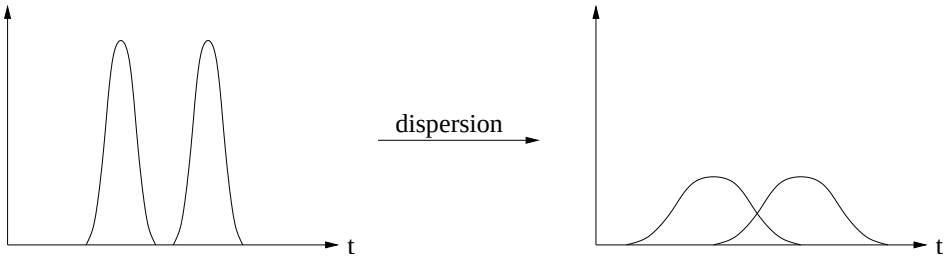


Figure 2.2: Pulse broadening due to (chromatic) dispersion.

for which the chromatic dispersion is typically of $d\tau/d\lambda = 17$ ps/(km · nm), the bandwith-length product is approximately $B\sqrt{L} = 80$ Gbit/s$\sqrt{\text{km}}$ enabling a fiber length of 64 km at 10 Gbit/s.

To enable larger transmission distances so-called dispersion compensation fiber (DCF) can be used. This fiber has a large negative dispersion, a typical value is for instance $d\tau/d\lambda = -300$ ps/(km · nm). Such fiber can be used to partly reverse signal quality degradation due to chromatic dispersion in SMF. For instance, the dispersion of 100 km SMF can be compensanted with 5.7 km DCF if the dispersion in both types of fibers is equals the aforementioned values. However, in both SMF and DCF the dispersion is not constant but depends on the wavelength and the dispersion profile of SMF and DCF cannot be matched perfectly over the whole spectrum of the pulse. Furthermore, DCF has a rather large attenuation compared to SMF which further reduces the bandwidth-length product as discussed in the next section. Other options to cope with chromatic dispersion are dispersion shifted fiber, in which the dispersion is relatively small, and electronic dispersion compensation at the receiver. Dispersion shifted fiber is attractive when new fiber is installed anyway and relatively small distances have to be covered like in metropolitan areas. Electronic dispersion compensation works best for smaller distances or in combination with other dispersion compensation techniques.

Another dispersion effect in single-mode fibers is polarisation-mode dispersion (PMD). The signal in a SMF has two perpendicular polarizations. Due to geometric asymmetries occurring in any practical fiber, these two signal components propagate at slightly different speeds which broadens the pulses. PMD is mostly relevant at data rates of 10 Gbit/s or more.

Attenuation, Noise, and Amplifiers

Pulses detected at the receiver are overlayed with noise from various sources, mostly noise introduced by the electronic circuitry that amplifies the very small photo current produced by the photodiode when a pulse arrives. The noise leads to some pulses not being detected and sometimes noise is detected as pulse, overall resulting in a certain bit error rate (BER). The smaller the signal power arriving at the receiver, the smaller the difference between a pulse and noise and the larger the BER. Typically, a BER of 10^{-9} is targeted which corresponds to a certain minimum signal-to-noise ratio (SNR) required at the receiver. If the signal is distorted by effects like chromatic dispersion or fiber nonlinearities an even larger SNR is required to achieve the same BER. The difference between the SNR required to achieve a certain BER with and without an disturbing effect is the *power penalty* resulting from that effect.

As an optical signal propagates along an optical fiber, the signal power reduces exponentially with the travelled distance, mostly due to Rayleigh scattering. Fig. 2.3 shows the wavelength dependent loss of a SMF. There are two low-loss regions, each covering a bandwidth

of 25 THz. The first is centered at approximately 1310 nm with a loss of about 0.5 dB/km while the other is centered at about 1550 nm with a loss of 0.2 dB/km. The peak at about 1400 nm results from hydroxyl ions (OH^-) in the fiber and is called water peak. Fiber optic transmission usually uses one of these two windows, to ensure a minimum attenuation of the signal. These low loss regions resulted in the standardized wavelength of SONET/SDH systems of 1310 nm and the International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) grid for WDM systems ranging from 1460-1530 nm (S-band), 1530-1560 nm (C-band), and 1560-1630 nm (L-band) [16]. Clearly, to achieve the targeted BER the fiber must not exceed a certain length. Otherwise, the SNR at the receiver would be too low due to the lost signal power. Note that it is not possible to arbitrarily increase the maximum possible length of the fiber by increasing the pulse power of the transmitter. First, the power of the laser is limited. Second, if the power in the fiber gets too high the signal is distorted by fiber nonlinearities. The latter is especially a problem in WDM systems with large channel counts.

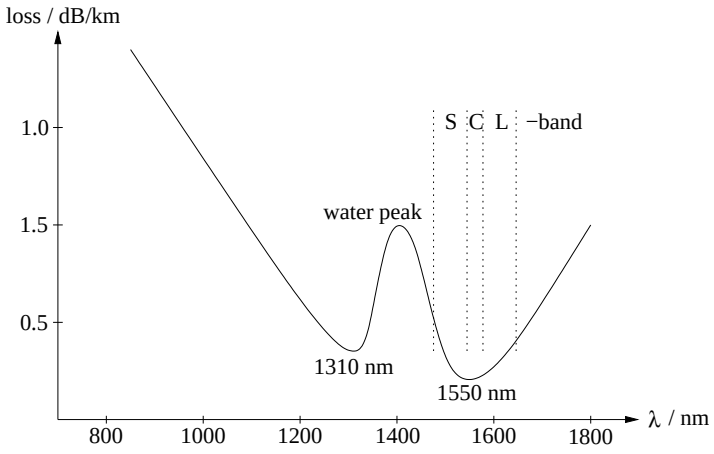


Figure 2.3: Wavelength dependent loss of SMF.

To enable transmission over larger distances, fiber amplifiers, mostly erbium doped fiber amplifiers (EDFAs) are frequently used. Besides WDM technology, the introduction of EDFAs was the second key enabler of todays Internet. An EDFA amplifies all WDM channels in a fiber simultaneously and in the optical domain, i.e., without processing the signal electronically, and is therefore very cost efficient. Fiber amplifiers enable all-optical links spanning several thousand kilometers, e.g., transatlantic cables.

A fiber amplifier basically consists of several meters fiber doped with a certain seldom earths, like erbium (Eb) in case of an EDFA. The fiber is inserted in the optical link. At the input side, a laser pumps light inside the doped fiber. The wavelength of the laser is chosen according to the material used to dope the fiber and results in population inversion like inside a regular laser. When a light pulse travels through the doped fiber it stimulates emission of light at the pulses wavelength, effectively resulting in an amplification of the pulse. The output power of an EDFA is typically 20 to 50 mW. Note that this is the aggregate power of all WDM channels and that this value is independent of the WDM channel count.

However, a fiber amplifier also produces a certain amount of so-called amplified spontaneous emission (ASE) noise and therefore degrades the SNR of the optical signal. The noise figure F is the ratio of the SNR at the input and the output of the amplifier, a typical value

for an EDFA is 5 dB. To compensate for the power loss in longer links several several fiber amplifiers are required. Fig. 2.4 shows a typical setting where the link is divided segments of equal size. At the end of each segment DCF and an EDFA compensate for the chromatic dispersion and the power loss in the individual segment. As the signal propagates along the fiber the ASE noise power reduces by the same factor as the signal power and is amplified by the same factor as the optical signal when passing an EDFA, i.e., the SNR remains the same. However, each fiber amplifier produces additional ASE noise itself which adds to the amplified noise from the amplifier input and decreases the SNR with each passed segment. As already mentioned above, a certain minimum SNR is required to achieve the targeted BER, putting a limit on the maximum possible link length. Still, transoceanic cables exceeding a length of 10000 km have been successfully deployed.

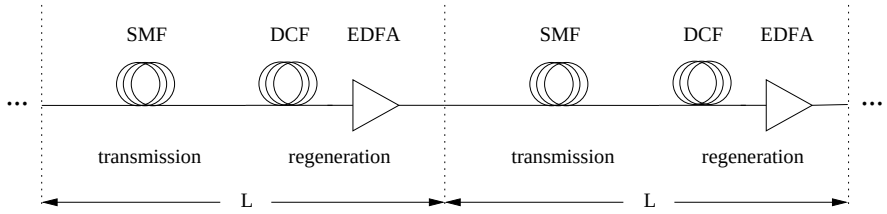


Figure 2.4: Segmented optical link with all-optical signal regeneration.

Nonlinear Effects

Besides attenuation and chromatic dispersion so-called nonlinear effects are the third major performance limiting factor of fiber optic transmission systems. Nonlinear effects can be divided into nonlinear scattering effects and effects resulting from changes of the refraction index of the fiber at high signal powers. Let us first consider scattering effects. Stimulated Raman Scattering (SRS) results from interaction of the optical signal with molecular vibrations in the fiber. The scattered light is called stokes wave and has an up to 40 GHz lower frequency than the original signal. Depending on the frequency, scattered light might interfere with lower frequency WDM channels and reduces the power of the original signal. The intensity of the effect depends on the power of the optical signal. However, the power must be very large to result in significant scattering.

Stimulated Brillouin Scattering (SBS) occurs due to interactions of the signal with acoustic waves. Here, the Stokes wave always propagates in the other direction than the optical signal. The power of the wave is much greater than with SRS while the frequency shift is smaller, i.e., only up to 10 GHz lower than to original signal. The intensity of SBS is also power dependent and significant interference with the optical signal occurs if the signal power is greater than several mW within a spectrum of 100 MHz. Very short pulses generally suffer less from SBS than longer pulses.

The second type of nonlinear effects can be further classified into self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM) and result from the power dependency of the refraction index known as Kerr effect. In both SPM and XPM the power of the optical signal modulates the refraction index of the fiber which in turn modulates the phase of the signal. In SPM the signal that causes the effect modulates itself. The modulation broadens the optical spectrum of the signal intensifying the negative effects of chromatic dispersion. In XPM the signal that causes the effect modulates the phase of other WDM

channels and leads to crosstalk between the individual channels. In FWM signals of different frequencies mix to new signals at other frequencies resulting from addition or subtraction of the frequencies of the original signals. E.g., three WDM channels with a frequency of f_1, f_2, and f_3 can mix into a fourth signal at $f_1 \pm (f_2 - f_3)$. This results into crosstalk if the signals mix into frequencies occupied by other WDM channels.

Overall, nonlinear effects are mostly relevant to WDM systems where the optical power in the fiber is relatively high. Furthermore, the exact impact of these effects on the shape of the pulses on the individual WDM channels is hard to describe as it results from a complex interaction between different nonlinear effects, multiple channels, as well as chromatic dispersion.

2.1.2 Optical Transmitters

In telecommunication networks the optical signal is generated using semiconductor lasers. Note that the term *laser* is an abbreviation for 'light amplification by stimulated emission of radiation'. In the following we discuss briefly how a semiconductor lasers work and how tunable lasers can be implemented. Tunable lasers are useful for all-optical wavelength-switched networks, as we will see in Section 2.2.2.

Semiconductor Basics

Let us first discuss some semiconductor basics. Semiconducting materials are characterized by the fact that electrons can have two different energy levels the low level corresponding to a the so-called *valence band* and the high level to the *conducting band*, as shown in Fig. 2.5. Electrons in the valence band are attached to an atom and cannot move freely while electrons in the conducting band can move relatively freely in the material and, as the name suggests, produce an electric current flow. The energy difference between the two bands is called *band gap*. Note that electrons cannot have any energy levels between the two bands. If an electron moves from the valence band to the conduction band it absorbs energy and leaves a so-called *hole* in the valence band. The energy absorbed is proportional to the width of the band gap. The reverse process, i.e., an electron moving from the conduction band to the valence band, is called recombination and releases the same amount of energy as required for lifting up the photon. The released energy may lift up a different electron to the conduction band or result in a photon of being emitted. The wavelength of a photon is proportional to its energy, therefore the band gap determines the wavelength of the emitted light.

Semiconduction materials can be *doped* or *impurified* with certain other materials to increase the number of holes or electrons in the conduction band. An n-type doped semiconductor has an increased number of holes while a p-type semiconductor has an increased number of electrons in the conduction band. If a p-type semiconductor is layered over an n-type semiconductor this results in a *p-n junction*. In the *active region* around the p-n junction almost all electrons from the n-layer recombine with holes in the p-layer. The active region electrically isolates both layers since there are no more mobile carriers in this region. When a voltage is applied to the p-n junction the size of the active region either increases or decreases, depending in the polarity of the voltage. The voltage either pulls more electrons from to the n-layer to the p-layer or pushes them back. If the voltage is high enough the active region vanishes completely and a current starts flowing. Note that this how semiconductor diodes works that conduct in one direction and isolate in the other. While the current is flowing the

electrons in the p-n junction frequently recombine with holes and produce photons. If the p-layer in top of the n-layer is thin enough the photons can pass through and the diode emits light. This kind of diode is called light emitting diode (LED) and can be used as cost efficient transmitters for multimode systems operating at bit rates around 100 Mbit/s.

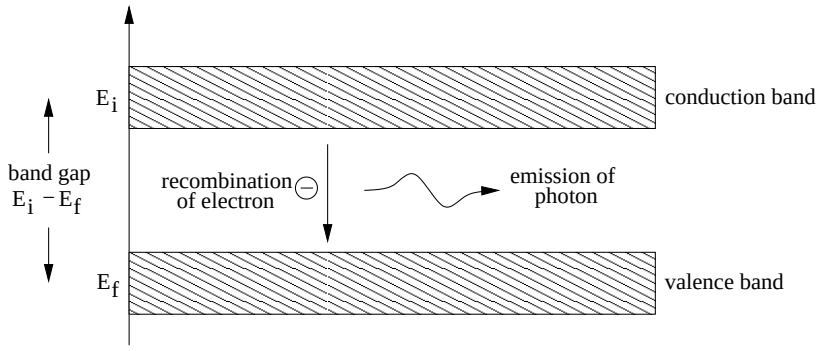


Figure 2.5: Energy bands of a semiconductor and light emission due to recombination.

Semiconductor Laser

If two faces of the p-n junction are mirrored, with one of the mirrors being partly transmitting as illustrated in Fig. 2.6, the structure represents a semiconductor diode laser, or semiconductor laser for short. Photons produced from recombination reflect back from the surfaces and oscillate in laser cavity. Part of the light leaves at the partly transmitting side of the structure. Note that only photons whose wavelength λ matches the cavity length build up in the cavity. More precisely, the length L of the cavity must be a multiple of half of the wavelength λ, i.e., $L = m\lambda/2$, with m being integer. The point of a laser is that photons oscillating in the cavity *stimulate* recombination of electrons in the p-n layer which in turn produces new photons so that a strong lightwave builds up. Note that the wavelength corresponding to the band gap should be matched to the cavity length that determines the emission wavelength. Photon produced by stimulated emission have the same wavelength as the photons stimulated the recombination process. Therefore, light that leaves the laser structure has a very narrow optical spectrum or *linewidth*.

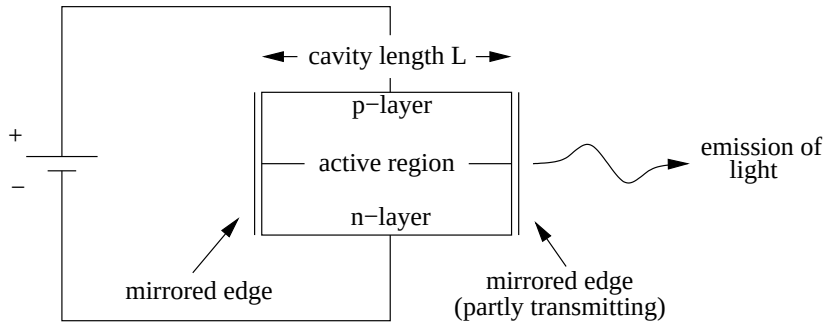


Figure 2.6: Basic structure of semiconductor laser.

Lasers used in high speed WDM telecommunication networks are usually multiple-quantum-well (MQW) lasers that consist of several very thin layers and feature an even narrower spec-

trum minimizing the effects of chromatic dispersion. The wavelength is chosen to fall in the low loss regions of the fiber, i.e., typically 1310 or 1550 nm. Semiconducting materials used in the laser are often based on indium gallium arsenide (InGaAs).

Due to a optical power dependency of the refraction index of the cavity material, the wavelength of a laser depends on the emission power. This effect is called *chirping*. When the power of a laser is directly modulated with an electrical current to produce light pulses, the emission frequency is also modulated and the beginning of the pulse has a different wavelength than the end. Therefore, *direct modulation* may limit the channel spacing in a WDM system. Furthermore, this chirping also increases the impact of chromatic dispersion. In SMF the chirp of a directly modulated laser leads the pulses to broaden. Therefore, special fibers exist whose dispersion is matched to the chirp of a directly modulated laser. Such fibers can be used to reduce, or even reverse, pulse dispersion. While this approach clearly increases the range of systems based on directly modulated lasers, e.g., in the cost-sensitive metro market, it is unfortulately not suitable for long-haul networks. Experience has shown that for covering larger distances the best strategy is to reduce the chirping to a minimum. For this purpose, in long-range systems the laser is run at constant power and pulses are generated using an external modulator. However, the external modulator, which is usually a Mach-Zehnder inferometer (see below), adds cost to the system.

Tunable Lasers

Tunable lasers, i.e., lasers whose emission frequency can be tuned to different WDM channels, are useful in all-optical wavelength-switched networks which are discussed below. The two important characteristics of such a laser are *tuning range* and the *tuning time* from channel to channel. A large tuning range enables to tune to a large number of different wavelength while the tuning time should be as small as possible for fast switching. Ideally, the tuning range would be wide enough to cover the whole spectrum used for WDM systems which is approximately 40 nm [16]. The tuning time would be small enough to switch individual packets. Assuming a packet size of 1500 byte and a bit rate of 10 Gbit/s the packet duration is 1.2 μs. To avoid tuning time overhead the tuning time should be significantly smaller. Another characteristic of tunable lasers is whether the laser is continuosly tunable or only to discrete wavelengths.

In *mechanically-tuned lasers* an external Fabry-Perot cavity adjacent to the lasing medium filters out unwanted wavelengths. The Fabry-Perot cavity consists of two parallel mirrors whose distance can be mechanically adjusted. Light is reflected back and forth between the mirrors and wavelengths which do not match the distance between the mirrors interfere destructively. The mirror at the side of the Fabry-Perot cavity opposite to the laser is partly transmitting and passes the tuned in wavelength which interferes constructively between the mirrors. As the tuning involves mechanical movement the tuning time is relatively slow, typically 1-10 ms. The tuning rage is very wide, about 550 nm, and covers the whole full gain spectrum of a semiconductor laser.

Acoustooptically- and electrooptically-tuned lasers use an external tunable filter that consists of materials whose refraction index can be changed either using soundwaves or with an electrical current. The refraction index controls which wavelength the filter passes. This tunable laser type provides a good compromise between low tuning time and wide tuning range. In acoustooptically-tuned lasers, the tuning time is about 9 μs with a tuning rage of 750 nm. Electrooptically tuned lasers provide a tuning time of 1-10 ns over a range of 7 nm.

Type of Tunable Laser	Tuning Range	Tuning Time
Mechanically-Tuned	550 nm	1-10 ms
Acoustooptically-Tuned	750 nm	9 μs
Electrooptically-Tuned	7 nm	1-10 ns (estimated)
Injection-Current-Tuned	45 nm	1-10 ns

Table 2.1: Tunable lasers: Comparison of the approximate tuning range and tuning time of different implementation types (from [1]).

Both types are not continuously tunable to arbitrary wavelengths.

Finally, in *injection-current-tuned lasers* a Bragg diffraction grating is placed either inside or outside the lasing medium. The diffraction grating consists of alternating thin layers of two materials with different refraction indices and passes only a certain wavelength. If a current is applied the index of refraction changes and different wavelength is passed thereby enabling tunability. The tuning time is 1-10 ns over a spectrum of about 40 nm.

An alternative to tunable lasers are *laser arrays* where several lasers, each emitting at a different wavelength, are integrated into a single component. These can be implemented relatively cost-efficient using vertical-cavity surface-emitting lasers (VCSELs), a type of semiconductor laser that is rotated by 90 degrees and emits light in the direction vertical, instead of parallel, to the wafer on which it is produced.

Table 2.1 provides an overview of the tuning time and tuning range of the different implementations of tunable lasers. Note that in current tunable laser types there is a tradeoff between these two parameters.

2.1.3 Optical Receivers

The most popular receiver concept in optical telecommunication networks is *direct detection* which is illustrated in Fig. 2.7. First the incoming signal is bandpass filtered to select the appropriate WDM channel and to remove some of the ASE resulting from optical amplifiers on the link. The filtered signal is fed into a photodiode which produces a small photocurrent when an optical pulse arrives. The photocurrent is amplified and analyzed by a threshold device which decides during each bit period whether the signal received during the bit period was strong enough to be interpreted as pulse. Due to noise and signal detoriation, e.g., by chromatic dispersion, this decision is not always correct resulting in a certain BER. Optionally, to improve the BER and enable larger transmission distances, complex signal processing operations can be performed, e.g., to partially compensate the effects of chromatic dispersion.

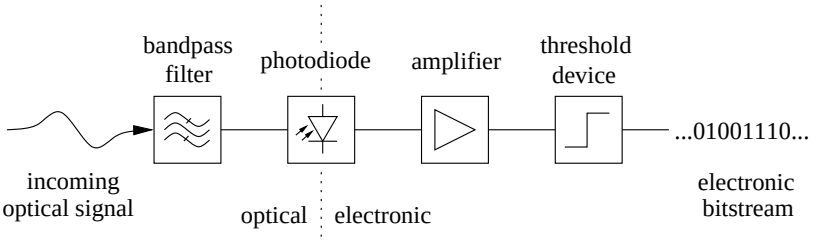


Figure 2.7: Basic structure of a direct detection optical receiver.

The photodiode is usually a PIN-photodiode or an avalanche photodiode (APD). The structure of a PIN-photodiode is a p-n junction with an intrinsic or only weakly doted semiconductor layer, the i-layer, in between the p- and n-layer. Note that intrinsic means not doted at all. The photodiode is reversed biased resulting in more electrons and holes to recombine than without external current which increases the width of the active region without mobile carriers. The width of the intrinsic region is chosen equal to the width of the active region and the band gap of the intrinsic material is chosen to match the wavelength of the incoming signal. For the p- or n-layer in between the intrinsic region and the fiber, a material that is transparent to the signal's wavelength is used. Photons arriving from the fiber pass the transparent region and are injected in the intrinsic region. Since the band gap of the intrinsic material matches the wavelength of the photons electron-hole pairs are produced due to the photoelectric effect. Due to the external voltage, the electrons move in the direction of the n-region and holes move in the direction of the p-region resulting in a small current flow.

In an APD geometry and semiconductor material is chosen in a way that makes the electric field in the p-n junction much stronger than in a PIN-photodiode. When a photon produces an electron-hole pair the two carriers are accelerated very rapidly producing additional electron-hole pairs. Overall, this results in an avalanche effect multiplying the photocurrent.

Tunable Filters

Tunable receivers are required for many proposed WDM architectures and are generally implemented by replacing the fixed tuned filter of an optical receiver by a tunable device. In addition to Fabry-Perot filters and acousto- or electrooptically tuned filters that have already been discussed in the context of tunable transmitters, liquid crystal (LC) Fabry-Perot filters and Mach-Zehnder chains can be used to implement tunable filters. Again, the important parameters are tuning time and tuning range.

A *LC Fabry-Perot filter* works similar to the previously discussed mechanically tunable Fabry-Perot filter. However, instead of mechanically varying the distance between the two mirrors, tunability is achieved by using a LC cavity between the mirrors. The refraction index of the LC can be modulated by an externally applied current used to select which wavelength to be filtered. The tuning time is about 0.5-10 μs with a tuning range of about 50 nm.

As the name suggests, a *Mach-Zehnder chain* consists of several Mach-Zehnder inferometers each of which filters out a certain wavelength so that only one selected wavelength arrives at the end of the chain. A Mach-Zehnder works as follows. The incoming signal is split and one of the components is routed through an adjustable delay before both components of the signal are recombined at the output if the device. The amount of delay is adjusted to phase-shift the wavelength to be removed by 180 degrees resulting in the wavelength to destructively interfere at the combiner. Since thermal elements are used in the delay component the tuning time is relatively high, i.e., in the order of milliseconds. Another disadvantage is that a chain of such inferometers is hard to control as the amount of delay in each stage depends on the amount of delay in the previous stages. The tuning range is about 16 nm.

Table 2.2 provides an overview of tuning time and tuning range of the discussed tunable filters. Similar to tunable lasers there is a tradeoff between the two parameters.

Type of Tunable Filter	Tuning Range	Tuning Time
Mechanically-Tuned	500 nm	1-10 ms
Acoustooptically-Tuned	250 nm	≈10 μs
Electrooptically-Tuned	16 nm	1-10 ns
LC Fabry-Perot	50 nm	0.5-10 μs
Mach-Zehnder Chain	around 16 nm	ms range

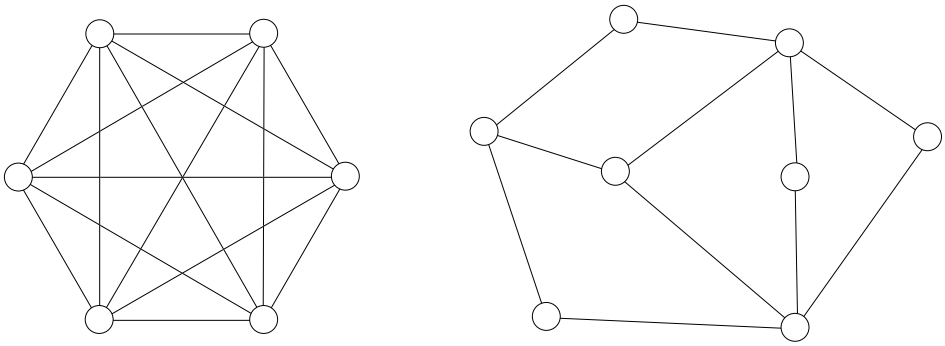
Table 2.2: Tunable filters: Comparison of the approximate tuning range and tuning time of different implementation types (from [1]).

2.2 Optical Network Basics

So far, our discussion only focused on a PtP WDM link connecting two nodes. However, real telecommunication networks consist of a potentially large number of nodes. Directly interconnecting each node with each other node would not be feasible because too many links would be required. As illustrated in Fig. 2.8 (a), to implement such a *full mesh* between N nodes, $N(N-1)/2 \in O(N^2)$ links must be deployed. In real networks only some nodes, usually those nodes geographically close to each other, are directly connected with a physical link, as depicted in Fig. 2.8 (b). In such networks, communication between two nodes which are not direct neighbors involves one or more intermediate nodes that have to *switch* the data to the right output link in the direction of the destination node. Switching can either be done *opaque* or optically *transparent*, i.e., with or without converting the optical signal to the electronic domain. In the former case, so-called optical-electronic-optical (OEO) conversion is performed which means that the signal is converted to the electronic domain using an optical receiver, electronically processed, and output on the appropriate link with an optical transmitter. In the latter case, the node performs all-optical (OOO) switching. Current networks almost exclusively rely on opaque switching. The advantage of opaque node architectures is that the signal is fully recovered from transmission impairments as OEO conversion results in reamplifying, reshaping, retiming (3R) regeneration. Furthermore, the electronically converted data can be used for performance monitoring, modified, and, most important, be electronically stored and processed to determine the appropriate output link. On the downside, OEO conversion requires costly transceivers and high-speed electronics. These disadvantages have pushed the development of optically transparent networks and future networks are expected to perform all-optical switching. Furthermore, all-optical node architectures are modulation format and bit rate independent and can therefore be considered scalable and future proof. In the following we review the optical components used to implement all-optical WDM networks based on [1].

2.2.1 Optical Switching Basics

The two switching concepts underlying most telecommunication networks are *circuit switching* and *packet switching*. The term circuit originates from telephone systems and means that a connection at a fixed data rate is setup between two nodes, just like during a phone call. Note that the bandwidth reserved for the circuit cannot be claimed by other nodes, even if the circuit is currently idle. In contrast, when packet switching is performed the data to be transmitted is divided into discrete packets of fixed or variable size. The data in each packet is preceeded by a header that contains the address of the destination node. When a packet



(a) full mesh (b) real network

Figure 2.8: Fully interconnected network vs. 'real' network.

arrives at an intermediate node, the destination address is used to determine to which output link the packet must be forwarded in order to reach the destination. Here, the full capacity of each link is shared among all nodes, i.e., idle nodes leave the available bandwidth to the other nodes. The characteristics of circuit and packet-switching are somewhat contrary in terms of technical complexity, efficiency for bursty traffic, and QoS as discussed in the following.

Concerning technical complexity, circuit switching is relatively simple compared to packet switching. Intermediate nodes just cut-through the circuit which is technically is relatively simple task. Packet switching requires more complex node structures, namely *routers*, that processes each single packet and at least determine the appropriate output link from the destination address. Since more than one million packets per second can arrive from a high-speed link operating at 10 Gbit/s demanding this demands for huge electronic processing capacities.

Another factor that distinguishes both switching technologies is their performance for bursty traffic, i.e., as opposed to a continuous flow the data to be transmitted arrives in bursts. Note that bursty traffic is typical for metro and access networks which are close to the end users, as detailed in Section 2.3.1. To see the difference circuit and packet switching make for bursty traffic consider the simple four node network depicted in Fig. 2.9 (a). Node 1 generates bursty, packet based traffic at a rate of λ per node to node 3 and 4. Each of the three links has a capacity of μ. Fig. 2.9 (b) and 2.9 (c) show the queuing systems corresponding to this network for circuit switching and packet switching, respectively. In case of circuit switching, each circuit has a capacity of $\mu/2$, i.e., both circuits share the capacity of the link connecting node 1 and 2 equally. Each of the circuits corresponds to a server operating at a rate of $\mu/2$ at which packets arrive at a rate of λ. In case of packet switching, the link between node 1 and 2 corresponds to a single server operating at a rate of μ at which packets arrive at a rate of 2λ. Note that the other two links do not need to be included in the model as the traffic is sufficiently smoothened by the first link for no queuing to occur at these links. It can be shown that the mean waiting time of the packets in a system with a single big server is smaller than in a system with multiple smaller servers whose aggregate capacity is the same as that of the big server. In average, the data is delayed less in the packet-switched system. Furthermore, if the maximum queue length is limited, as it is the case in a real networks, the packet loss is smaller in the system with the single big server, i.e., the packet-switched network. Intuitively speaking, this is due to the fact that in the dual server system one of the

servers might be idle while the queue of the other server is not empty or even overflowing. In a system with a single big server, the traffic would be processed with twice the rate of the small servers. Overall, delay and packet loss are smaller when packet switching is employed than with circuit switching. In other words, the available link capacity is utilized more efficiently. This is not only the case in our specific example but holds for the packet switched networks in general. Since packet switching implements *statistical multiplexing* this improvement is often referred to as *statistical multiplexing gain*.

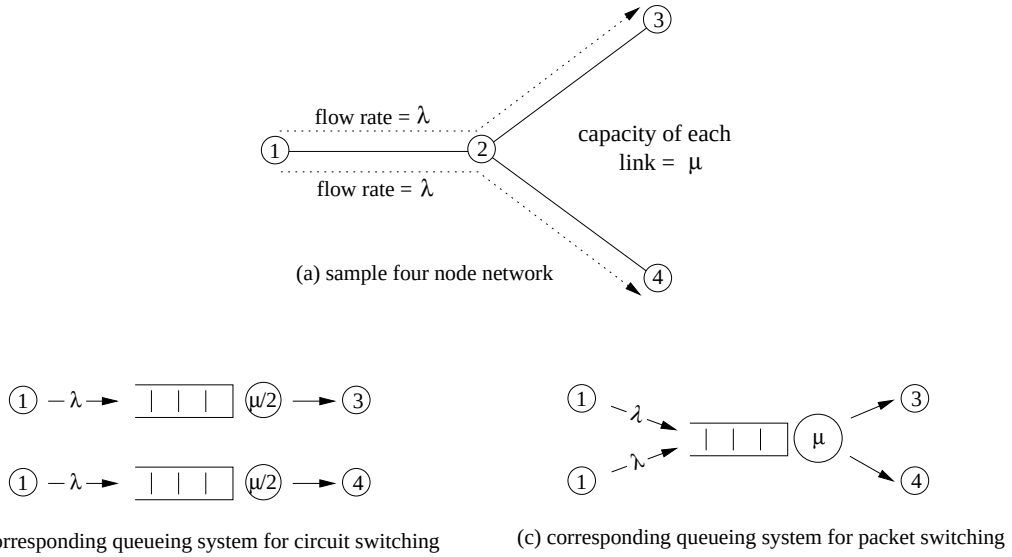


Figure 2.9: Sample four node network and corresponding queuing systems for circuit-switching and packet-switching.

Note that the statistical multiplexing gain relates to mean values. In *average* the network performs better. However, at different time instances the delay a packet experiences between the same pair of nodes as well as packet loss and the maximum transmission rate may differ significantly when packet switching is employed. As discussed in more detail in Section 2.3.2 some network applications require connections with a guaranteed transmission rate and/or approximately constant delay over the whole lifetime of the connection, i.e., the application has certain QoS requirements. Such requirements can hardly be met in packet switched networks where the load at each router on the path to the destination is unpredictable and changes continuously. With circuit switching, on the other hand, the full bandwidth of the connection is continuously available and due to the cut-through principle the data is rarely delayed at intermediate nodes.

In optical WDM networks, circuit switching corresponds to *wavelength routing* and packet switching to *optical packet switching (OPS)* and *optical burst switching (OBS)*. Wavelength routing means that nodes communicate via WDM channels and intermediate nodes perform optically transparent switching at the granularity of individual wavelengths. OPS can with current technology only be implemented for very simple topologies such as ring and star networks which are discussed in detail later in this work. More general topologies require the optical packet to be temporarily stored to evaluate the destination address and determine the right output link which is not feasible with current optical technology due to the lack of optical random access memory (RAM) and optical processing capabilities. A compromise

between wavelength switching and OPS is OBS where each source node aggregates packets to the same destination to *optical bursts*. Because the burtst are longer than individual packets, the switching times can be longer than with packet switching without introducing too much overhead. To transmit the burst usually one of the following two approaches is taken. One option is to setup the optical switches along the path to the destination prior to transmitting the burst, i.e., a reservation of the required wavelength channel on all links to the destination is made. The advantage is that collisions of bursts at intermediate nodes are completely avoided. On the downside the pretransmission coordination introduces an additional delay to the transmission of the burst. The other approach is to delay the burst at intermediate nodes sufficiently long with fiber delay lines (FDLs) to evaluate the destination address and to setup the switch. A FDL is basically wound up fiber of a certain length proportional to the required delay. Due to the propagation delay though the fiber the signal is delayed. Some of the power of the signal is tapped from the signal before the input of the FDL to read the destination address. The advantage of this implementation of burst switching is that there is no additional delay due to pretransmission coordination. The disadvantage is that collisions occur at intermediate nodes when two bursts that need to be forwarded to the same output fiber at the same wavelength arrive simultaneously. Some approaches use more complex FDLs structures with adjustable delays to be able to resolve some of the collisions. While there is no delay due to pretransmission coordination, this approach also introduces additional delays as collided bursts need to be retransmitted. Overall, OBS does not achieve the same degree of statistical multiplexing as OPS would but reduces the switching time to the minimum possible with current technology.

From a more general perspective, each technology switches the optical signal at a certain timescale, the so-called λ-timescale. In traditional SONET/SDH based networks, the λ-timescale is about several weeks since this is the time required to manually setup a new circuit. The other extreme would be future OPS networks operating on λ-timescales of less than milliseconds. However, any improvement on the λ-timescale results in an increased degree of statistical multiplexing. For instance, if wavelength can be dynamically setup or teared down within minutes or seconds, the network can be adapted to time-of-the-day changes in the traffic which is, although far apart from OPS, already a significant improvement over legacy SONET/SDH.

2.2.2 Passive Switching Devices

In the following we discuss passive optical devices used to construct optical WDM networks. Note that some of the components discussed in this section, namely combiners, splitters, multiplexers, and demultiplexers, are usually not considered switching devices. However, they are still discussed here as they are functionally similar to switching components and used as building blocks for those. One important characteristic of passive optical devices is the *insertion loss*. When an optical signal passes through a passive component some signal power is generally lost, e.g., due to reflections. The insertion loss is ratio of the power at the output of the device and the power the signal would have without inserting the device. Remember from Section 2.1.1 that reducing the power of the signal decreases the bandwidth-length product. Therefore, the insertion loss should be as small as possible.

Wavelength Insensitive Passive Devices

Fig. 2.10 (a) shows a 1×2 *optical splitter* which broadcasts the incoming signal to both output ports. The power of the incoming signal is usually evenly distributed among all output ports. In case of a 1×2 splitter the splitting loss is therefore 50%, or approximately 3 dB. Note that all other power distributions can be implemented as well, e.g., a 10:90 splitter which is fequently used to tap 10% of the power of an optical signal from a fiber. The insertion loss of a splitter is mostly determined by the splitting loss. However, as in any component, there is also some *coupling loss* of about 50-60 dB because the input fiber coupled to the device perfectly and some power is reflected back at the interconnection point. Furthermore, flaws in the production process result in the relatively small *excess loss* in any passive optical device.

An *optical combiner*, which is illustrated in Fig. 2.10 (b) combines the signal from the individual input ports at the single output. The power at the output is the aggregate power of all signals minus coupling and excess loss which are the same as in a splitter. Note that a combiner and a splitter are basically the same device operated in opposite directions. Further note that if there are signals at the same wavelength at two or more inputs of a combiner the wavelengths collide at the output and the information carried on all collided channels is corrupted.

A coupler is a combination of a combiner and a splitter, a 2×2 *coupler* with two input and outputs is depicted in Fig. 2.10 (c). The signal from any input is broadcasted to all outputs.

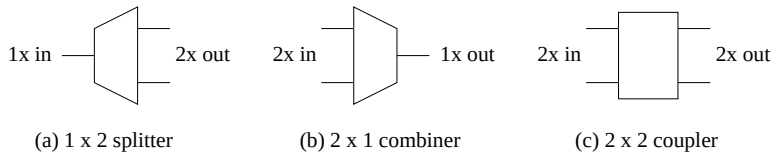


Figure 2.10: Basic passive optical coupling devices.

Splitters and combiners with more than 2 ports can be implemented by building a binary tree of 1×2 splitters or 2×1 combiners, respectively. The number of ports is then always a power of two. From a $N\times1$ splitter and a $1\times N$ combiner a so-called *passive star coupler (PSC)* can be implemented, as illustrated in Fig. 2.11. However, there are more power efficient ways to implement a PSC. In an ideal implementation, the input power would be evenly distributed among all outputs with no loss. In real devices the insertion loss can be as small as 1-2 dB. Many proposed all-optical LANs rely on an $N \times N$ PSC to interconnect N nodes. In such a network, each node is attached to one input an one output.

Wavelength Selective Passive Devices

In wavelength selective devices it depends on the wavelength of the signal to which output it is routed. The simplest form of such devices are optical multiplexers and demultiplexers, or MUX and DEMUX for short. An optical multiplexer is similar to a combiner but at each input only a certain wavelength is passed through. Analogously, a demultiplexer filters the individual wavelengths of a signal and passes each wavelength to a different output. The most common use of these devices is multiplexing and demultiplexing the individual channels in a WDM system to and from the fiber, as illustrated in Fig. 2.1 (b).

One demultiplexer and one multiplexer can be combined to an *optical add-drop multiplexer (OADM)*. As Fig. 2.12 (a) illustrates, in an OADM some wavelengths are passed through while

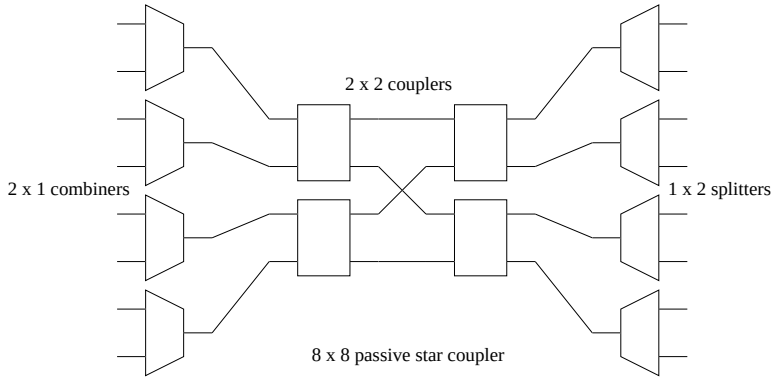
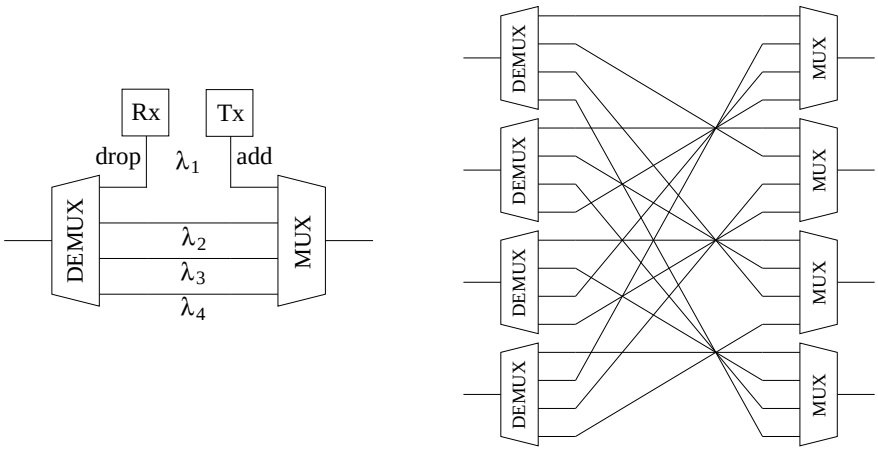


Figure 2.11: Implementation of PSC by hierarchical composition of smaller devices.

others are locally *dropped* and *added*. OADMs are commonly used in optical node structures where the dropped wavelengths are converted to the electronic domain and can therefore be processed. E.g., part of the data carried by the dropped wavelength may be destined to the node itself and is removed from the signal while the remaining data is forwarded along with additional traffic originating from the node itself on the added wavelength. Incoming wavelengths which do not carry any data for the node itself can be bypassed without electronic processing reducing complexity and cost of the system.



(a) optical add−drop multiplexer (b) nonreconfigurable wavelength router

Figure 2.12: Illustration of OADM and nonreconfigurable wavelength router.

Furthermore, multiple demultiplexers and multiplexers can be combined to a nonreconfigurable wavelength router, as depicted in Fig 2.12 (b). A $N \times N$ wavelength router with N input and output ports designed for W wavelengths consits of N demultiplexers and multiplexers with W ports. Consider a simple N node network where each node is connected to one input/output pair. By choosing the appropriate wavelength, e.g., with a transmitter array, each node can send data to a specific destination node. Note that multiple source-destination pairs can communicate simultaneously without any collisions. If the wavelength router is configured in a way that the same wavelength at different inputs is routed to a different outputs, which is the case in Fig. 2.12 (b), *all* nodes can communicate at *all* wavelengths simulta-

neously resulting in $N \cdot W$ simultaneously usable channels. This is called *spatial wavelength resuse* and can be considered as an improvement over connecting the nodes via a PSC where each wavelength can only be used once, i.e., only W channels are available simultaneously.

A wavelength router that features the previously described routing scheme can be implemented efficiently as a single component, namely an *arrayed-waveguide grating (AWG)*. The basic structure of an AWG is shown in Fig. 2.13. Two star couplers are interconnected with a waveguide grating. The numbers of waveguides between the couplers N' is much larger than the number of input and output waveguides N. Furthermore, the length l_n of each waveguide is by Δl larger than the length l_{n-1} of the previous waveguide between the couplers. The star couplers are free propagation regions to which the input/output waveguides are connected at an angle α and α' between the grating waveguides, respectively. Wavelength dependent routing is achieved as follows. Due to the geometric setup of the free propagation regions, light arriving from a certain input waveguide focuses at certain grating waveguides. Since these waveguides have different lengths, the light arrives at a different phases at the second free propagation region resulting in the light from some waveguides to interfere constructively and destructively between others. The geometric position of the waveguides whose light interferes constructively results in the light focusing at a certain output waveguide. Note that wavelength of the next higher or lower WDM channel inserted at the same AWG input constructively interferes at other waveguides. The geometric design of the AWG corresponds to the WDM channel spacing, resulting in the light to focus exactly at the neighboring output waveguide. For each wavelength the light focuses at a different output, and after one cycle including all waveguides, at the same output again. The spectrum covered by one such cycle is called free spectral range (FSR). Furthermore, note that the same wavelength from a different input waveguide focuses at different grating waveguide inputs and interferes constructively at different waveguide positions at the grating output. Therefore, the same wavelength from different input waveguides will focus at different output waveguides enabling spatial wavelength reuse. The geometric design of an AWG must be manufactured very precise to in order achieve a low insertion loss and low crosstalk between the individual channels. Also, the performance of a AWG depends on its temperature which changes the geometric dimensions. However, recent AWG designs are relatively insensitive to temperature changes and feature a low insertion loss around 5 dB with about -40 dB crosstalk. Finally, note that an AWG can also be used to implement a multiplexer or demultiplexer if one of the star couplers has only one input waveguide.

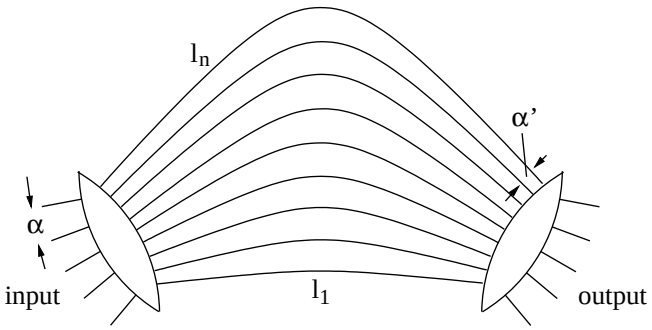


Figure 2.13: Structure of an AWG nonconfigurable wavelength router.

2.2.3 Reconfigurable Switching Devices

More flexibility is gained with reconfigurable switching devices. Such devices can either be wavelength insensitive, i.e., switch all WDM channels of an incoming fiber collectively, or wavelength selective and switch individual wavelengths between input and output fibers. In the former case the switch is called *optical crossconnect (OXC)*, in the latter reconfigurable *wavelength-routing switch (WRS)* or *wavelength selective crossconnect (WSXC)*. If individual wavelengths are switched is it further distinguished whether the switch is able to perform *wavelength conversions*, i.e., switching a WDM channel from an input fiber to another WDM channel on the output fiber, or not.

Optical Crossconnects

Typically, OXCs are either be based on arrangements of 2×2 cross-bar switches or implemented as micro-electro-mechanical system (MEMS). While implementations based on electrooptical cross-bar switches feature low reconfiguration times as required for OPS and OBS, MEMS are the technology of choice for larger-scale OXCs. Besides configuration time, an important characteristic of OXCs is if the switch is fully nonblocking which is the case if all inputs and outputs can be connected in any permutation simultaneously.

OXCs Based on Cross-Bar Switches Fig. 2.14 (a) shows the basic principle of a 2×2 cross-bar switch. As the name suggests, the switch can be either in cross or in bar state. In cross state, input 1 is connected with output 2 and input 2 is connected with output 1. In bar state input 1 is connected with output 1 and input 2 is connected with output 2. The two most common implementations of a cross-bar switch are the *directive switch* and the *gate switch*.

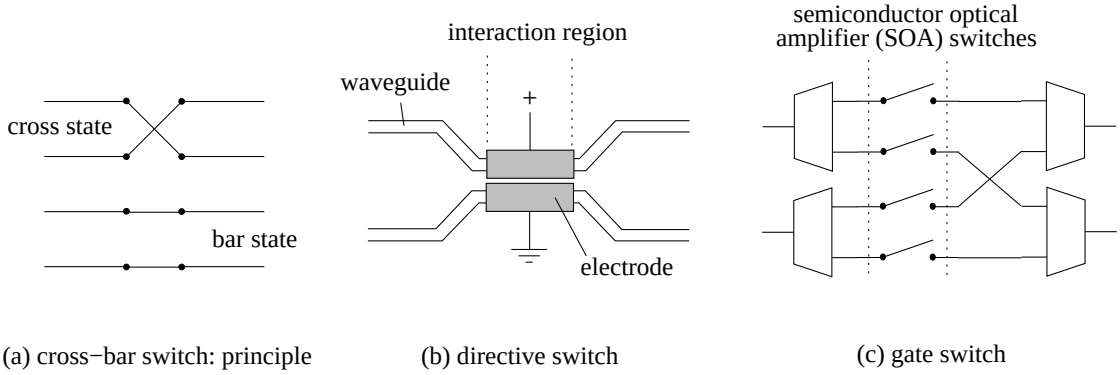


Figure 2.14: Cross-bar switch: principle and implementations.

Fig. 2.14 (b) depicts a simple implementation of a directive switch. It consists of two parallel waveguides between which the distance is reduced in an interaction region of a certain length. In the interaction region the waveguides are covered by electrodes. The switch works as follows. Part of the optical signal, the so-called evanescent wave, propagates outside the waveguide. If the geometric dimensions are chosen properly, the evanescent wave results in the signal completely coupling into the other waveguide in the interaction region. This effect is called evanescent coupling and effectively puts the switch in cross state. For switching to the bar state, a voltage is applied to the electrodes which changes the propagation constant of

the underlying waveguides and avoiding evanescent coupling between the waveguides. Note that this implementation of a directive switch is wavelength specific and that even small deviations from the right interaction length result in significant crosstalk between the two channels. These problems can be overcome with improved directive switch implementations such as the reverse delta-beta coupler, which is functionally similar to the device discussed here, the balanced bridge interferometric switch, and the intersecting waveguide switch. Other types include mechanic and thermooptic directive switches [17].

The basic structure of a 2×2 gate switch is depicted in Fig. 2.14 (c). Generally, a gate switch can have N inputs and N outputs. Each input is connected to a $1 \times N$ splitter. Each output of each splitter is connected to an optical switch, i.e., overall N^2 switches are required. Ideally, these switches are semiconductor optical amplifiers (SOAs) which has the advantage that splitting losses can be compensated for. Finally, N combiners, one for each output, are attached to the outputs of the central switching array. Each input of each of the $N \times 1$ combiners is attached to a switch corresponding to a different input port. This structure provides a path between each input and each output port and each path can be enabled or disabled individually using the switching array, i.e., the structure implements an $N \times N$ fully nonblocking switch. A shortcoming of the gate switch is the limited scalability due to the splitting loss which linearly increases with N and at some point exceeds the gain of the amplifiers.

Larger nonblocking switches can be constructed by arranging multiple cross-bar switches to a *switching matrix*, as depicted in Fig. 2.15. In this structure the individual cross-bar switches are called *crosspoints*. To build an $n \times k$ switching matrix with n inputs and k outputs $n \cdot k$ crosspoints are required. However, the maximum number of crosspoints and therefore the size of the switch is limited due to power consumption and space constraints. A more efficient way to build larger switches is the *three stage clos architecture*. The clos architecture has been shown to require the minimum number of crosspoints for a switch of given dimension $N \times N$. As illustrated in Fig. 2.16, the clos architecture consists of three stages and each stage consists of several switching matrices. The first and last stage each consist of N/n switching matrices of dimension $n \times k$ and $k \times n$, respectively. The central stage consists of k switching matrices of dimension $N/n \times N/n$. Each switch in the central stage has one connection to each switch in both the first and the last stage. The minimum number of crosspoints is required for $n = \sqrt{N/2}$ and $k = 2n - 1$ which results in $C = 4N\sqrt{2N-1} \in O(N^{3/2})$ crosspoints as opposed to $C = N^2 \in O(N^2)$ crosspoints required for a plain $N \times N$ switching matrix.

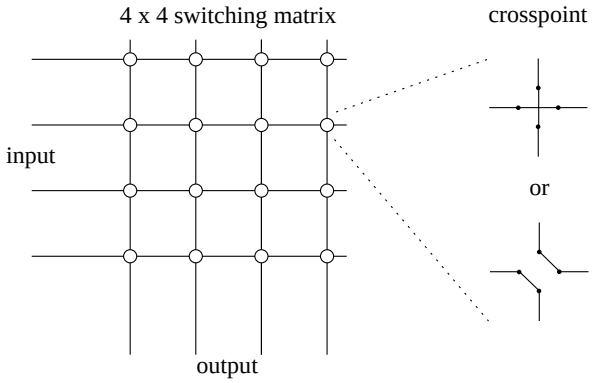


Figure 2.15: Switching matrix with cross-bar switches as crosspoints.

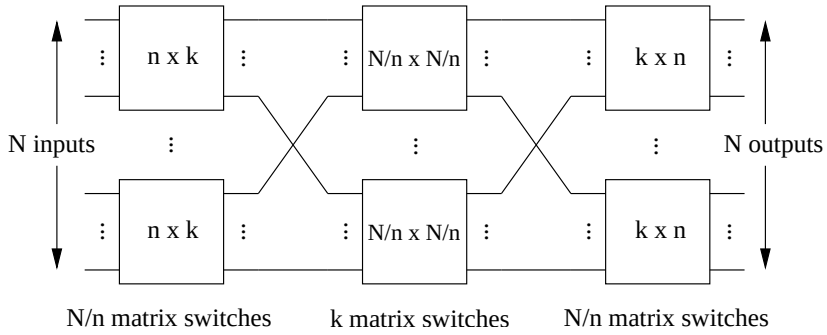


Figure 2.16: Clos architecture composed from matrix switches.

Micro-Electro-Mechanical Systems

Currently, the most promising technology for large scale OXCs are MEMS switches which can be categorized into free-space optics and waveguide based switches. The former category can further be divided into switches based on either mirrors or membranes. MEMS switches can operate in two or three dimensions. The most popular type is the three-dimensional (3D) mirror-based MEMS switch which is illustrated in Fig. 2.17. The incoming fibers are bundled to a two-dimensional (2D) fiber array. The signal from each fiber is focused with a lens on a mirror corresponding to this fiber in a 2D tilt mirror array opposite to the fiber array. The tilt of each mirror can be adjusted in two directions. By choosing the right tilt, the signal from each input fiber can be focused on any mirror of a second mirror array. Analogous to the input side, each of the mirrors in the second mirror array corresponds to one output fiber. If the tilt of the mirror at the output side is adjusted towards the mirror at the input side the signal focuses on the corresponding output fiber and couples in through the lens. As each mirror in one array can focus on each other mirror in the other array, any input fiber can be connected to any (unused) output fiber, i.e., and the switch is fully nonblocking.

Based on this architecture, optically transparent crossconnects with more than 1000 ports, low insertion loss, and low crosstalk between the channels can manufactured using batch process techniques.

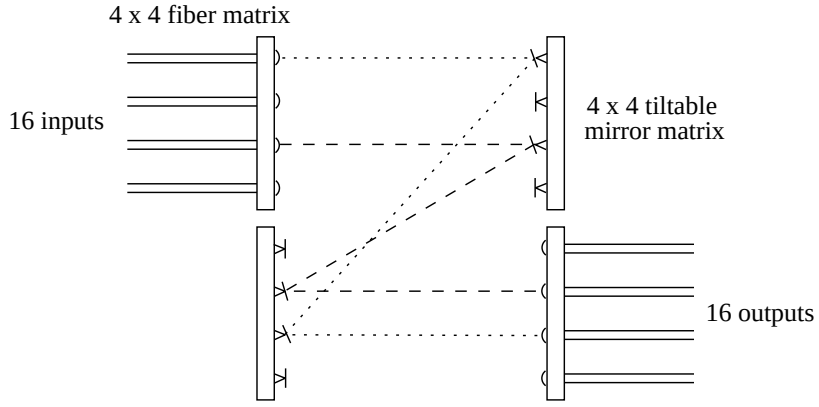


Figure 2.17: Structure of three dimensional MEMS switch.

Wavelength Selective OXCs

In order for two nodes to communicate in a wavelength-switched network, a transparent wavelength path must be setup between source and destination. However, in a network relying on OXCs as switching elements, the circuit consists of a number of fibers between source and destination, i.e., using OXCs the switching is performed at a granularity of fibers, not individual wavelengths. Note that, as long as that circuit exists, none of the fibers used by the circuit can be used for communication between other pairs of nodes. This significantly reduces the flexibility of the network and generally leads to many communication attempts to be blocked. Furthermore, the huge amount of bandwidth of all WDM channels on such a path of fibers is rarely required for communication between a single pair of nodes. It is much more efficient to perform the switching at a per-wavelength granularity using a WSXC or WRS. As illustrated in Fig. 2.18, an $N \times N$ WSXC capable of switching W wavelengths individually, can for instance be constructed from W OXCs of the same dimensions, where each OXC is associated with one wavelength. At each input fiber the wavelengths channels are demultiplexed from the incoming signal, switched individually, and multiplexed into the signal of the outgoing fiber.

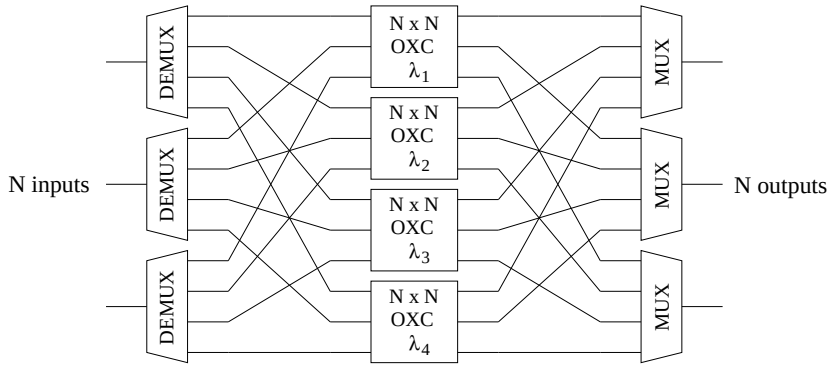


Figure 2.18: Structure of reconfigurable wavelength routing switch.

An important aspect of wavelength switched networks based on WSXCs is the *wavelength-continuity constraint*. A wavelength path between a pair of nodes can only be setup if there is a set of links between source and destination where the intended wavelength is available on each link. Otherwise, the communication attempt is blocked. To increase the flexibility of the network the switches can be extended by *wavelength conversion* capabilities. One way to achieve that would be to perform OEO conversion at each node and switch the wavelengths electronically. However, as mentioned earlier, optically transparent systems are currently pushed as such systems lack the complexity of electronic processing and enable scalability due to modulation format and bit-rate transparency. Therefore, lots of research efforts has been put into the development of all-optical wavelength converters. The approaches taken are mostly based on exploiting nonlinear effects like FWM or XPM (see Section 2.1.1).

For completeness we also mention reconfigurable add-drop multiplexers (ROADMs) which are OADMs that can be reconfigured to any wavelengths to be added or dropped. A ROADM is basically a 2×2 WSXC where the fiber is attached to one input/output pair while the other input/output pair is used to locally add/drop the desired wavelengths. Another popular implementation of an ROADM is depicted in Fig. 2.19. Here, some power is tapped from the fiber at the input of the ROADM and fed into a tunable receiver tuned to the dropped

wavelength. Similarly, the signal of a tunable transmitter tuned to the added wavelength is inserted in the fiber at the output of the ROADM using a combiner. To be able to remove the added/dropped wavelength channel from the incoming signal all channels are demultiplexed and routed through a switch array, e.g., consisting of SOAs.

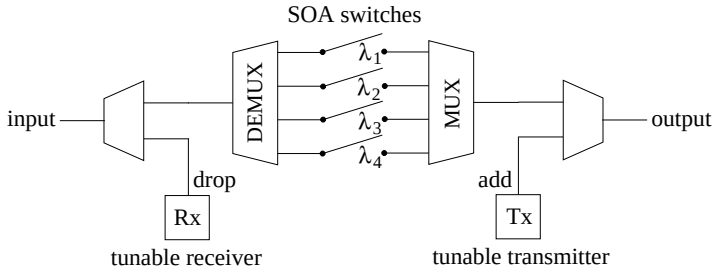


Figure 2.19: Structure of ROADM.

2.3 Metro Network Basics

In the following we briefly discuss the 'real' optical infrastructure of the current Internet with a focus on the metropolitan area. More specifically, we discuss the role of metro networks within their context, i.e., long-haul and access networks, and the limitations that result in the metro gap. This discussion leads to a discussion of the requirements future metro solutions have to meet in order to overcome the bandwidth bottleneck at the metro level.

2.3.1 The Metro Gap

As discussed in Chapter 1, Internet traffic volumes will keep increasing during the next decade. Clearly, WDM will remain the key technology to satisfy the ever increasing demand for more bandwidth within the next years. However, remember from Chapter 1 that the Internet's infrastructure consists of different domains, namely long-haul networks, access networks, and metro networks that interconnect the former two. Besides the need for more bandwidth, each of these types of networks face different future limitations and challenges. In the following discussion, which closely follows [18], we will have a look at each of the domains in more detail and elaborate how the current setting result in the metro gap.

Long-haul Networks

Long-haul networks are traditionally the domain of relatively few large trans-national and global telecom carriers. They span distances ranging from national regions up to thousands of kilometers and connect to metropolitan area networks and amongst each other to provide global connectivity between regional domains. The primary concern in these networks has been to improve the underlying WDM technology to increase the transport capacity of these networks. Long-haul networks are traditionally based on SONET/SDH technology, where each fiber only carries a single data channel. As the traffic volume increased during the 90's, fiber-exhaust became a problem and carriers begun to deploy WDM technology on a wide scale. The introduction of WDM is the key enabler of todays Internet, since it made it possible for carriers to multiply the capacity of their networks without costly installing new fiber or even

building new conduits. However, since severe impairments can arise for large WDM channel counts careful analog engineering provisions are required to maintain good signal quality over large distances. In many cases OEO conversion along with 3R regeneration is required to maintain the required channel quality over long distances. A key development in this area was the introduction of optical amplifiers, mostly EDFAs, which significantly extended the distance the optical signal can travel without costly 3R regeneration and was especially advantageous for transoceanic cables. Long-haul solutions are generally very expensive and represent longter-term, strategic investments. But, with improving technology as amplifiers, filters, isolators, dispersion compensation, and fiber, long-haul networks continue to evolve with every generation of WDM enhancements making them more cost efficient and robust. Since MANs are the source of the long-haul traffic this translates into the need of also improving the robustness and capacity of metropolitan networks.

Access Networks

Access networks connect the end users to the regional metropolitan area network and are characterized by a diverse variety of protocols and infrastructures. Access rates span over a large bandwidth spectrum ranging from 10 Mbit/s Ethernet or even less to full wavelength capacities of 2.5 or 10 Gbit/s (OC-48 or OC-192). Similarly, the customer base ranges from residential Internet users up to large private, government, or educational corporations. Therefore, access networks are confronted with a diverse variety of protocols resulting from many different applications which they all need to handle efficiently. Among others, these protocols include Internet Protocol (IP), Asynchronous Transfer Mode (ATM), SONET/SDH, (Fast/Gigabit) Ethernet, multiplexed time division multiplexing (TDM) voice, digital video, and other more specific protocols such as Fibre Distributed Data Interface (FDDI), Fiber Distributed Data Interface (ESCON), and Fibre Channel. The evolution of access networks is very dynamic and unpredictable. It is driven by new end-user applications and improved, higher speed access technologies such as DSL, cable modems, and emerging next-generation wireless services. For instance, the introduction of any of the new end-user applications which are on the horizon, such as Internet video, telemedicine, or videoconferencing, can cause an abrupt rise in bandwidth demand. Especially the amount if IP traffic will continue to increase and future access networks need to handle this this bursty, asymmetric, and unpredictable kind of traffic efficiently. Overall, the development of access networks is driven by two key requirements, support for a plethora of protocols used by many different applications and flexible architectures with support for a wide range of data rates and number of users. Currently, Ethernet passive optical networks (EPONs), which are discussed in the next chapter, seem to be the most likely solution for future access networks.

Metropolitan Area Networks

MANs both carry traffic within the metro domain, i.e., inter-business and inter-office traffic, as well as from and to large long-haul point of presence (POP)s. Since metro networks are fed by the regional access networks they have to cope with the same diversity of protocols and wide range of data-rates. Today, most MANs rely on bidirectional SONET/SDH fiber rings. In SONET/SDH, so-called circuits (permanent connections operating at fixed data rates) are established between pairs of ring nodes at data rates usually ranging from 155 Mbit/s to 2.5 Gbit/s (OC-3 to OC-48). As illustrated in Fig. 2.20, elecronic circuits are added to the

ring by the source node and dropped from the ring at the destination node using add-drop multiplexers (ADMs). Note that these are electronic ADMs, rather than OADMs as discussed in Section 2.2.2. SONET/SDH based metro networks suffer from a number of shortcomings:

- Capacity scaling limitations: Upgrading the ring capacity to adapt to traffic growth normally requires expensive 'forklift upgrades' where a large fraction of the equipment needs to be replaced which involves high costs and interruption of normal operation.
- Poor bandwidth utilization: Bursty, asymmetric IP traffic is handled only inefficiently due to SONET/SDH's lack of statistical multiplexing and responsiveness.
- High provisioning time: Provisioning of additional circuits for new customers usually takes several weeks to months which is unacceptable in the highly competitive metro market.
- High system complexity: All circuits need to be groomed (multiplexed) into SONET/SDH's rigid TDM structure which requires lots of electronic processing and results in high equipment cost, inflexibility, and complex operation and maintenance.

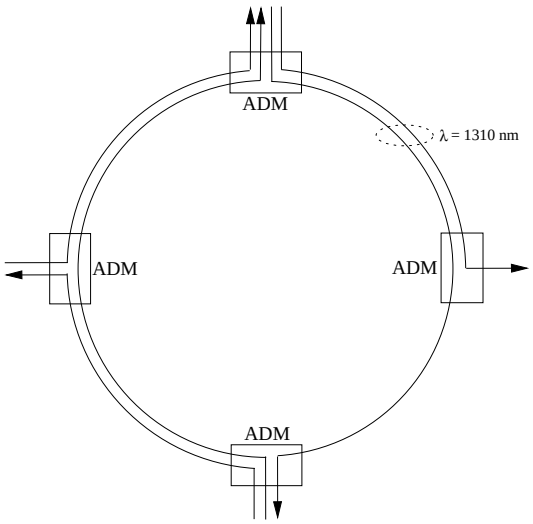


Figure 2.20: Add-drop multiplexing of circuits in SONET/SDH.

In order to address these concerns, future metro solutions must be highly scalable, provide huge transmission capacity, and be flexible enough to deal with a large variety of traffic and protocols. To achieve these goals, the deployment of WDM which offers bandwidth scalability, optical transparency to support arbitrary protocols, flexibility, and manageability in the metro area is a promising approach. In fact, many metro service providers are already deploying WDM technology in their networks. On the other hand, other properties of WDM, namely circuit-switching and large bandwidth granularities, are not very well suited to efficiently handle the wide spectrum of protocols including rather small data rates originating from the access networks. This makes the introduction of WDM to the metro area a challenging task. Collectively, these problems are known as the *'metro gap'*.

2.3.2 High-Level Metro Requirements

To close the metro gap, future metro networks must meet a number of requirements which are discussed below. MANs are not just smaller, scaled down reincarnations of long-haul networks. Instead, network operators are in need of solutions specifically addressing the characteristics of the metro domain to survive in the competitive market. We first discuss these requirements from a relatively high-level point of view, like from that of a network operator, and break down how these relate to the architecture and access protocol of metro networks in the next section. In this section we discuss the metro requirements from a rather high level-point of view similar to [18] which this section largely follows.

Multi-Protocol Support: Metro networks are facing a large variety of protocols with different requirements and traffic patterns which they all need to handle efficiently. For instance, data might be transported in SONET/SDH TDM streams, ATM cells, or Ethernet packets. Legacy voice traffic is symmetric, has a constant bit rate, and is sensitive to delay jitter and data loss. Internet traffic, on the other hand, is usually asymmetric, bursty and generally tolerates some data loss. Furthermore, new applications relying on unforeseen communication protocols and data formats may arise in future. For metro operators, it is important to handle all these different kinds of traffic with a single, common platform and fiber infrastructure. Deploying multiple, distinct metro infrastructures for different protocols clearly is an unrealistic proposition. This would involve high maintenance and deployment costs along with lengthy right-of-way concerns and service providers would not be able to provision services based on new protocols quickly. In contrast, a single multi-protocol infrastructure provides backward compatibility, significant cost reductions (by eliminating equipment and eased maintenance), simplified network management, and reduced collocation issues in packed central offices.

Optical Transparency: Optical transparency means that an optical signal is not converted to the electrical domain on its way from source to destination, i.e., bypassed optically transparent at all intermediate nodes. Switching functionality can be implemented at the optical layer using MEMS. Due to the diverse mix of data signaling formats at the access side, optical transparency is a big advantage for metro operators as it eases the burden of constantly adapting the network to the ever evolving data-formatting standards or of increasing the networks capacity since only the end systems need to be upgraded. This results in cost-efficiency and scalability. Furthermore, optical transparency allows network operators to extend their product portfolio by offering a customer wavelength channels between his sites. These can be run with arbitrary data formats and feature a high level of security against tapping or modifying the transmitted data. The channel latency and latency jitter are minimized due to the lack of electronic processing on the transmission path. Compared to long-haul networks, where optical transparency is also considered advantageous, metro area distances are much shorter and transparent wavelength channels are less susceptible to signal degradation, making its deployment in the metro area less challenging.

Differentiated SLAs and QoS Levels: The competitive metro market requires to offer more differentiated services than only SONET/SDH TDM circuits to the customer. Specifically, the deployed networking technology should enable service providers to enrich their product portfolio by offering connections with different QoS parameters, specified in a so-called service level agreement (SLA) as part of the contract with the customer. Such QoS parameters are for instance bandwidth, delay, priority, survivability, or BER. The quality of the service

determines its price. For instance, in case of fiber cut, part of the available bandwidth is lost and inexpensive, lower priority connections are dropped first while the remaining bandwidth is used for more expensive, higher priority connections.

Fast Provisioning: Provisioning of new circuits in current SONET/SDH based metro networks usually takes several weeks due to the technical complexity of rolling-out new circuits. For being successful in the competitive metro market, operators need to provision new connections or additional bandwidth quickly. The ability of the deployed technology to instantly provision bandwith also is an opportunity to offer new services like 'dialing for bandwidth' where customers are enabled to temporally set up additional connections, for instance via a web-interface, to satisfy additional bandwidth requirements like those resulting from a high-quality video conference.

Sub-Rate Bandwidth Provisioning: While there is a demand for high-capacity connections at full wavelength capacities, usually 2.5 to 10 Gbit/s, many customers require only a small fraction of this bandwidth. In an optical metro network this results in a granularity gap between the relatively low-rate electronic customer interfaces and the huge capacity of the optical transmission channels. As the number of available wavelength channels is limited, assigning each low-rate connection a full wavelength channel quickly exhausts the available number of wavelengths. This 'protocol-per-wavelength' approach severely limits the scalability of the network and clearly is not feasible. To utilize the available transmission capacity efficiently, multiple low-rate connections should be aggregated into a single, higher capacity wavelength. In this 'multiple-protocols-per-wavelength' approach, all data protocols such as Transmission Control Protocol/Internet Protocol (TCP/IP) and TDM streams such as voice traffic are multiplexed into the same wavelength channel at the ingress side and demultiplexed at the egress side. This technique is commonly termed 'sub-rate' or 'sub-wavelength' multiplexing. The wavelength channel itself can either be packet based, for instance GbE, or carry TDM frames like in SONET/SDH. Sub-rate multiplexing enables service providers to utilize the available wavelength resources efficiently, independent of the mixture of bandwidth granularities currently demanded.

High Bandwidth Utilization: Due to the proximity to the end user and the resulting low degree of traffic aggregation of the bursty IP traffic, metro traffic profiles are generally more dynamic than the highly aggregated traffic carried in long-haul systems. Furthermore, many customers are interested in inexpensive GbE service with relaxed QoS requirements. These two propositions currently make packet based architectures an attractive solution for the metro area. Packet switched networks feature statistical multiplexing and therefore utilize the available bandwidth more efficiently than their circuit based counterparts. While such networks may suffer from packet loss and relatively large delay variations, this can be often be tolerated for data connections. Circuit based architectures, like wavelength-switching with OXCs are most efficient for less bursty, aggregated IP traffic in conjunction with next-generation IP traffic engineering protocols like multi-protocol wavelength switching (MPλS). In that case, short-reach optical interfaces on terabit IP routers can connect directly to with WDM cross-connects and will allow higher-layer traffic engineering protocols to request/release bandwidth within milliseconds in an automated manner and thus achieve a certain degree of statistical multiplexing. The interface between the router and OXC is commonly termed optical User Network Interface (UNI). Another aspect of making efficient use of the available bandwidth resources is to reduce the forwarding burden for intermediate nodes between source

and destination so that a larger fraction of their capacity can be used for sending and receiving the intermediate nodes' own traffic.

Scalability: The number of users and the traffic volume in metro networks steadily increases. However, details about the future development of the traffic and geographic dispersion of the network are difficult to predict, primarily due to changing customer requirements driven by new applications. Therefore, future metro solutions must be enable the network providers to adapt their network infrastructure to changes in the demanded services, traffic volumes, and even the geometric dispersion of their customers. Considering this, topological flexibility is a strong advantage for future metro solutions and any kind of topological constraints might turn out to be too restrictive. Besides this, the capacity of the network and the number of nodes and users must also be cost-efficiently scalable to enable continuous adaption to increasing demands. Current metro networks are mostly based on individual or multiple interconnected SONET/SDH rings. This approach requires careful network pre-planning, is not flexible for geometrically dispersed, unpredictable traffic demands, and will therefore face serious limitations in future. In the long term, a physical topology independent framework, namely a mesh based approach, will be the most valuable solution. Mesh networks require little up-front planning and allow to progressively grow the network as demand increases. Such a network will most likely be based on the wavelength switching paradigm using reconfigurable OXCs. However, currently used SONET/SDH equipment and especially the existing fiber infrastructure are huge investments and the operators are not likely to switch to a mesh topology in one step. Instead, the operator's upgrade strategy is to perform cautious upgrades for an evolution towards more efficient architectures. Therefore, new metro solutions should provide a smooth, future proof migration path which allows the operators to upgrade their network in a 'pay-as-you-grow' manner.

Efficiency for Different Traffic Patterns: A network architecture designed for the metro area requires to efficiently handle a large variety traffic patterns. Traffic demands differ depending on where the network is deployed and even within an individual metro area. For instance, current metro networks often consist of interconnected SONET/SDH rings, where several metro edge rings are connected to a central metro core ring. Note that one of the edge rings usually represents the interconnection point to the global Internet. As the edge rings are the gateway between the customers and the core ring, the traffic is highly asymmetric. Most of the traffic is coming from or destined to the node interconnection the edge and the core ring. This traffic pattern is also called 'hot-spot' traffic with the node interconnecting the two rings being the hot-spot. In the core ring the traffic is widely symmetric, as the core distributes the traffic between the individual edge rings. This traffic pattern, where each node sends the same amount of traffic to all other nodes, is called 'uniform' traffic. Note that it does not suffice if the network can be statically configured or the fiber infrastructure be layed out to support a specific traffic pattern. The network needs to support different traffic inherently. For instance, in the daytime business areas are more active while in the evening the traffic moves to residential areas. Due to this periodic change the the network architecture cannot fully be optimized for a specific traffic pattern. Additionally, as the longer term development of the traffic is difficult to predict, a tolerance for for changes in the traffic results in longer capacity upgrade intervals which makes the network more cost-efficient. Note that since hot-spot and uniform traffic are complementary they are generally well suited to evaluate the efficiency of a network for different traffic patterns and not only relevant in the

context of the aforementioned interconnected rings.

Survivability: Survivability is the capability of a system to remain functional in the presence of failures. Network operators generally try to deliver their services 24 hours a day, 365 days a year, with as few as interruptions as possible. The most prominent type of failure in optical networks are fiber cuts, node failures are relatively rare. In 2002, the Federal Communications Commission (FCC) published statistics that in metro networks the rate at which fiber cuts occur is approximately 13 cuts per 1000 miles of fiber per year, and 3 cuts per 1000 miles per year in long-haul networks. Almost 60% of all cuts are caused by cable dig-ups ('backhoe fades'). Recovery timescales of networking solutions are traditionally gauged against SONET/SDH, namely sub-50 ms Automatic Protection Switching (APS). The 50 ms benchmark results from the fact that in a SONET/SDH system interruptions of more than 50 ms cause undesirable complications. For instance, voice connections start being disconnected and ATM cell re-routing may begin. If the interruption lasts 2 s or longer, all switched circuits are disconnected and data connections may be dropped. Generally, the longer the interruption the more problems occur. For more details on the previous discussion refer to [19]. Note that for the user perspective, interruptions within the rage of several 100 ms in a video or voice stream could probably be tolerated. Concerning data connections, interruptions in this range can be easily handled by most data protocols including TCP/IP. While interruptions may cause data retransmission, the major part of current Internet traffic results from applications like web-browsing, e-mail, file-sharing, or FTP and is relatively insensitive to retransmissions. Therefore, the 50 ms should be regarded as technical constraint to ensure that new metro solutions seamlessly integrate in existing SONET/SDH environments, rather than a requirement resulting from the applications and services run on the network. Whether 50 ms recovery is really a hard requirement or not, especially in new solutions which might not be based on SONET/SDH, has been argued without resolution for over a decade. Note that the reason why the debate still persists is that it is "not entirely based on technical considerations which could resolve it, but has roots in historical practices and past capabilities and also has been used as a tool of certain marketing strategies" [19]. Overall, it is most reasonable to assume that future metro solutions should offer at least one survivability option with recovery times in the order of several tens to few hundreds of milliseconds for mission-critical traffic while sub-50 ms recovery is a desirable feature for compatibility with legacy SONET/SDH systems and from a marketing perspective. Advantageous are additional survivability options with relaxed recovery timescales or even no recovery guarantees to offer inexpensive best-effort service for the increasing amount of data applications with relaxed survivability requirements. These add value to the system as they enable the service provider to offer SLAs specifically addressing the needs of individual customers. Another aspect of survivability is that many deployed technologies already provide recovery capabilities, such as APS or IP re-routing. New solution must prevent any possible destructive interference between multiple recovery mechanisms, vital to a smoother migration from today's networks.

Manageability: Multi-protocol support and interoperability with existing technologies increase the complexity of new metro solutions and make their management a challenging task. Therefore, a standards-based, bit-rate independent network management solution that provides detailed network and equipment observability is a crucial component of the overall system and may even be the primary concern for a network operator to choose which solution to deploy. The network management solution should provide a graphical user interface (GUI)

to cope with the increasing complexity and dimensionality in a user friendly way and support the operator with all important operation, administration, and management (OA&M) activities. Most importantly, these include integrated configuration, performance monitoring, fault localization, and accounting.

Cost Efficiency: Metro network operators are facing serious competition emphasising the importance of new metro solutions to be cost-efficient. Generally, the cost for a new solution in the metro area is expected to be significantly lower than for a long-haul system. Furthermore, existing SONET/SDH solutions represent a huge investment and have reached economies of scale. Therefore, network operators are rather hesitant to revolutionize their working and revenue generating network in favor of a new unproven technology. In order to gain market acceptance, new solutions must provide significant benefits at the same or lower price levels compared to the deployed technology. More specifically, features like 'low-first-cost', 'pay-as-you-grow' and future proofness are crucial to success.

Reliability and Modularity: Resulting from the high degree of service multiplexing on fibers and wavelengths, a node or subsystem failure potentially disconnects many customers. The same holds for downtimes due to maintenance or system upgrades. To reduce service interruptions in these cases to a minimum, critical subsystems should be fully redundant and capable if in-service upgrades. Furthermore, the intense competition for plant space makes modular designs with compact footprints a requirement.

Table 2.3 summarizes these requirements for future metro networks and serves as a reference throughout this work.

2.3.3 Mapping to The Architectural/MAC Level

The remainder of this work deals at the level of network topologies, node architectures, and medium access control (MAC) protocols. On the other hand, the metro requirements in the previous section are described from a relatively high-level point of view. To provide a better understanding how which features of a metro architecture and corresponding access protocol relate to these requirements we here provide mapping between the two. The discussion distinguishes between relevant features and properties of the network architecture and topology, the node architecture, and the MAC protocol. Note that it is generally advantageous to keep the complexity of the network as low as possible and that the MAC protocol builds on top of the architecture. I.e., a good MAC protocol can efficiently use a good network architecture but cannot make a poor architecture efficient. Also note that OA&M aspects like *manageability*, *reliability*, and *modularity* are not included in the discussion because their relation to the aspects investigated in this work is too vague. The high-level requirements to which each paragraph refers are printed in *italics*.

Network Architecture & Topology

Number of nodes independent from number of wavelengths: This is an important requirement to provide *scalability of the number of nodes*. For instance, some proposed optical ring architectures use a dedicated wavelength for each ring node, i.e., the number of wavelengths equals the number of nodes. As the number of wavelengths is usually limited due to cost or practical reasons such architectures do not scale well.

Geographic extendability: This feature also corresponds to the *scalability* requirement. Consider for instance two networks with a mesh and ring topology. The meshed network can be extended to additional nodes in the surrounding area relatively easy by deploying additional links while it is difficult to extend the ring while sticking to its topology.

Efficiency for different traffic patters: This high level requirement translates directly into a architectural requirement. E.g., bidirectional rings have been shown to perform relatively poorly for generic traffic distributions [20].

Small mean hop distance and spatial wavelength reuse: The mean hop distance is the average number of nodes on the path between source and destination. The smaller the mean hop distance, the fewer bandwidth resources, i.e., wavelength channels on links, are required per transmission, and the more transmissions can take place simultaneously. Spatial wavelength reuse means that the same wavelength can be used simultaneously multiple times in different spatial locations. A more detailed discussion and examples for these features, that correspond the the high level requirements *high bandwidth utilization* and *cost efficiency*, can be found in Section 5.3.

High number of paths between nodes: The more different paths exists between each source-destination pair, the higher the *survivability* of the networks. For instance, a bidirectional ring provides two paths between each source-destination pair, one per ring direction, and maintains full connectivity between all nodes if the ring is interrupted in one place.

Node Architecture

Optically transparent bypassing and switching: The concept of optical switching and the corresponding equipment like OXCs and ROADMs has been discussed in Section 2.2.3. Besides the general advantages resulting from *optical transparency*, this feature also simplifies the node structure, i.e., fewer transceivers and electronic processing are required, and therefore increases the *cost efficiency* of the system. Furthermore, transparent channels do not introduce any variable delays (jitter) or packet loss and thereby support *high quality SLAs and high QoS*.

Low switching times: As discussed in the context of optical switching in Section 2.2.1, a low switching time result in a *high bandwidth utilization* due to statistical multiplexing and thereby improve the *cost efficiency* of the system. Ideally, the switching time is small enough to enable packet-switching.

MAC Protocol

Statistical multiplexing and efficient bandwidth usage: This basically means that the MAC protocol should use packet switching to achieve a *high bandwidth utilization* if supported by the node architecture. Packet-switching also *sub-rate bandwidth provisioning*, i.e., provisioning of communication channels with less then the bandwidth of full wavelength channel. Furthermore, packet-switched networks automatically adapt to changes in the traffic distribution resulting in *efficiency for different traffic-patterns*. Generally, the MAC protocol should utilize the available bandwidth resources as good as possible, e.g., with features like shortest path routing and spatial wavelength reuse.

Efficient support for variable size packets: This is an important requirement to enable *multi-protocol support* because different protocols use different and variable packet sizes. If the MAC layer does not naturally support variable packet sizes, processing power consuming segmentation/reassembly or aggregation procedures are required to convert the data to the packet size supported by the MAC. This is for instance n problem in many all-optical ring architectures where data is transmitted in fixed size time slots.

Mechanisms implementing different QoS classes: First, such mechanisms directly relate to the high-level requirement of providing *differentiated SLAs & QoS levels*. However, in a packet switched network a high quality service class is also required to emulate circuit-like service which is an important aspect of providing *multi-protocol support*.

Mechanisms implementing survivability: To enable *survivability*, the MAC protocol should be able to recover the network from failures and route the traffic in a way circumventing the failed component.

Fairness control: Fairness control does not directly relate to any of the high-level requirements. However, as for instance discussed in Section 9.2, packet switched networks often suffer from fairness problems that need to be resolved in order for the network to operate as expected.

Requirement	Summary
Multi-Protocol Support	Single platform for all current and future protocols, especially voice, IP traffic, and GbE.
Optical Transparency	Optical bypassing of intermediate nodes, modulation format transparency, provisioning of transparent wavelength channels to customers.
Differentiated SLAs & QoS Levels	SLAs with QoS specifications, e.g., for guaranteed bandwidth, delay, priority, or survivability.
Fast Provisioning	Short provisioning intervals, 'dialing for bandwidth'.
Sub-Rate Provisioning	Overcome granularity-gap by multiplexing multiple lower rate client channels into high bit-rate wavelengths.
High Bandwidth Utilization	High utilization the available bandwidth resources for bursty data traffic, low forwarding overhead.
Scalability	Scalability of network capacity, number of users, geographical dispersion. Smooth migration from ring to more flexible mesh topology.
Eff. for Different Traffic Patterns	Efficient support for uniform and hot-spot traffic, tolerance for dynamic changes.
Survivability	Recovery from fiber cuts and equipment failures, 50 ms benchmark for high-priority traffic, relaxed survivability options for best-effort traffic.
Manageability	Integrated, standardized and GUI-driven OA&M: Configuration, performance monitoring, fault localization, accounting, network and equipment observability.
Cost Efficiency	'Low-first-cost' solutions which provide a smooth migration path to future optical metro networks in a 'pay-as-you-grow manner'.
Reliability & Modularity	Equipment which features redundancy and in-service upgrades. Limited central office space requires small footprints and modularity.

Table 2.3: Requirements for future metro networks.

Chapter 3

Related Work

ALMOST all currently deployed metro networks (as well as long-haul systems) rely on SONET/SDH. To address the aforementioned shortcomings of this technology in the metro area, three main developments are underway:

- Enhancing and adapting legacy SONET/SDH systems to more efficiently support bursty packet traffic.
- The development of a standard for a optical packet-switched ring networks, namely IEEE 802.17 Resilient Packet Ring (RPR).
- The design of packet-switched optical WDM ring networks for the metro area [6, 21].

In this chapter, we survey these developments with a focus on the last category in which lots of work has been done during past few years. Note that while the overwhelming majority of metro systems relies on a ring topology, there is also work on the design of packet-switched WDM systems with a star topology. We discuss this approach in Chapter 4. For completeness we also mention that there are networks with a bus topology, e.g., AMTRAC [22], which have received relatively less interest.

The chapter is organized as follows. In Section 3.1, we first give a brief historical overview of optical networking in the metropolitan area. This historical overview culminates in a survey of the current standardization activities on optical metro WDM networks, which includes a discussion of the aforementioned enhancements to SONET/SDH and a short overview of RPR (the latter is part of our proposed network architecture and discussed in more detail in in Section 6.1). In Section 3.2, we provide a survey of a few selected experimental metro ring WDM testbed systems. The purpose of the discussion of these testbeds is twofold. First, it gives an illustration of the network architectures that are feasible with current optical equipment. Second, we introduce and explain several key photonic hardware components used in optical networks. In Section 3.3, which is the main section of this survey, we provide a comprehensive survey of the packet-switched ring metro WDM networks that have been studied to date. In this section, we first introduce a categorization for ring networks. Our categorization is based on the MAC protocol, or *access protocol* for short, employed in the network. We then comprehensively survey the ring WDM networks within the structure provided by our categorization. In Section 3.4, we comprehensively survey fairness control and QoS support for packet-switched metro WDM ring networks. In Section 3.5, we summarize the research and development work on packet-switched ring metro WDM networks to date and outline directions for future research and development.

3.1 History and Standardization

Below we provide an historical overview on developments in the metro area after which we discuss current standardization activies related to the metro networks.

3.1.1 Historical Overview

Optical fiber is widely considered the medium of choice to provide enough bandwidth in the metro area to the ever increasing number of users and bandwidth-hungry applications, e.g., video conferences, distributed games, visualization, supercomputer interconnection, or medical imaging applications which do not allow for image compression [23].

Remember from Section 2.2 that there are two generations of optical metro networks. In first-generation optical metro networks, copper links are replaced with fiber links while the nodes at either end of the fiber remain electronic. In such *opaque* optical networks OEO conversions of the signal take place at each node [24]. Initially, each fiber carried only one wavelength such as in FDDI and IEEE 802.6 Distributed Queue Dual Bus (DQDB) networks. To cope with the increasing amount of data traffic and to fully exploit the gain bandwidth of EDFAs, WDMs was introduced in the 90's. With WDM, each fiber carries multiple wavelength channels, each operating at any arbitrary line rate, e.g., electronic peak rate. After providing these huge pipes, attention turned from optical transmission to optical networking [25]. In second-generation optical networks, OEO conversions occur only at the source and destination nodes while all of the intermediate nodes are *optically bypassed* by means of OADMs. As discussed in Section 2.2.2 an OADMs allows nodes to locally drop and add one or more wavelengths from or to an incoming or outgoing fiber link. By optically bypassing nodes, the electro-optic bottleneck is alleviated, and the number of electronic port cards can be reduced at each node, resulting in OOO node structures and significantly reduced network costs, which is one of the most important drivers for optics [26].

Optical bypasses can be used in ring WDM networks to build cost-effective node architectures [27] and to reduce the number of logical intermediate nodes between source-destination pairs, leading to a decreased logical mean hop distance [28]. The resultant all-optical lightpaths are able to provide *transparent* channels to users who are free to choose bit rate, modulation format, and protocol. This transparency enables the support of various legacy as well as future services, which may include ATM, Frame Relay, SONET/SDH, IP, ESCON, and Fibre Channel, as illustrated in Fig. 3.1. We note that there are also hybrid forms of optical networks where not all intermediate nodes are optically bypassed and OEO conversion takes place not only at the source and destination nodes but also at a few selected intermediate nodes. This type of optical network is known as *translucent* network.

3.1.2 SONET/SDH

Here, we briefly review legacy and future SONET/SDH systems. For more details on this technology the interested reader is referred to [29].

Legacy SONET/SDH

Today's metropolitan area networks are mostly SONET/SDH ring networks. These networks are circuit-switched networks. The individual network nodes access the network bandwidth in a time-division multiplex fashion, i.e., each node is periodically allocated a specific number

of slots. SONET/SDHmay be combined with WDM to establish multiple SONET/SDH rings on one fiber. Also, SONET/SDH WDM rings may employ optical bypassing and traffic grooming to alleviate the computational burden and reduce the number of electronic port cards at bypassed nodes [30]. (Traffic grooming refers here to the routing of traffic destined to a node on the wavelengths that are not bypassed at the node. In general, traffic grooming in WDM networks aims at collecting lower rate traffic and sending it on high-speed wavelength channels such that a smaller number of wavelengths is required and fewer wavelengths have to be dropped and electronically processed at each node.) The main drawback of SONET/SDH networks is that due to their TDM operation in conjunction with a circuit set-up time on the order of several weeks or months [31], they accommodate packet traffic only inefficiently [32], especially when the traffic is highly variable. In conjunction with the additional drawbacks discussed in Chapter 1 this results the aforementioned metro gap.

Data over SONET/SDH

The inefficiencies of SONET/SDH networks are addressed by three new technologies, collectively knows as Data over SONET/SDH (DoS). These technologies are the Generic Framing Procedure (GFP) [33], Virtual Concatenation (VC) [34], and the Link Capacity Adjustment Scheme (LCAS) [35] currently being standardized by the ITU-T and T1X1.5. The GFP technology allows for the transport of data packets in SONET/SDH frames. Until now many network operators use proprietary technologies based on Packet over SONET/SDH (PoS) for this purpose. With PoS the boundaries of the variable size data packets are marked with control characters which requires the receivers to have lots of processing capacity since each incoming byte has to be monitored to recognize the boundaries. In addition, occurrences of the control character in the data packet have to be masked with byte stuffing, resulting in a fluctuating data rate depending on the content of the packet. In contrast, with Frame-Mapped GFP (GFP-F) [36] each data packet is preceded by a short header providing the length of the packet so that the receiver knows the beginning of the next packet in advance and no byte stuffing is required. The header is protected with a checksum which corrects single-bit errors. For storage networks, Transparent GFP (GFP-T) provides a method to transparently transport block coded data, such as 8B/10B coded bytes, which is bandwidth efficient and introduces only small delays. 8B/10B coding, in which ten bits are transmitted for each byte, is common in storage networks and is also used in GbE. The two additional bits are used to balance the numbers of ones and zeros and to transmit link control information.

The SONET/SDH technology offers data transmission only at specific rates from a prescribed set of rates. A GbE connection with a data rate of 1 Gbit/s, for instance, would have to be transported via SONET/SDH at a data rate of 2.5 Gbit/s (OC-48), resulting in an overhead of 1.4 Gbit/s. With VC data rates of a much finer granularity are provided to reduce the overhead. This is achieved by virtually combining (concatenating) multiple SONET/SDH low data rate connections into an aggregate connection close to the desired data rate. The individual connections making up an aggregate connection can operate at different data rates and can travel on different paths through the network. Alignment is performed at the receiver. When LCAS is added, further flexibility is obtained in that the aggregate data rate can be adapted to the data rate currently required. For instance, the amount of data transported over the GbE connection might differ significantly at different times of day. To adapt the data rate, low-rate tributaries can be added or removed from the virtually concatenated connection. To add or remove connections, control packets are exchanged between the sender and

the receiver. Note that both virtual concatenation and LCAS do not require any changes inside an existing SONET/SDH network, only the sender and the receiver are affected. In conjunction with control plane protocols such as Generalized Multiprotocol Label Switching (GMPLS) or Automatic Switched Transport Networks (ASTNs), DoS enables SONET/SDH based networks to automatically adapt to the current traffic situation within minutes or even seconds. This may be sufficient to achieve a high utilization in backbone networks where the traffic flows are aggregates of many individual flows and are thus relatively smooth. In metro networks, however, the traffic is more bursty and it is desired to efficiently share the available capacity between the nodes at the time scale of individual packets (packet switching) or bursts of packets (burst switching).

3.1.3 Resilient Packet Ring

While the standardization efforts in the area of SONET/SDH are not specific to metro networks, the importance of the metro gap is reflected by the large number of recently initiated standardization activities and industry fora such as the Internet Engineering Task Force (IETF) working group (WG) IP over Resilient Packet Ring (IPoRPR), the RPR Working Group (RPRWG), the Metro Ethernet Forum (MEF), and the Resilient Packet Ring Alliance which comprises more than 70 companies. Efforts by the IETF WG IPoRPR and the RPRWG finally yielded in the IEEE 802.17 Resilient Packet Ring (RPR) standard [37] for packet switched metro ring networks which has been released in 2004.

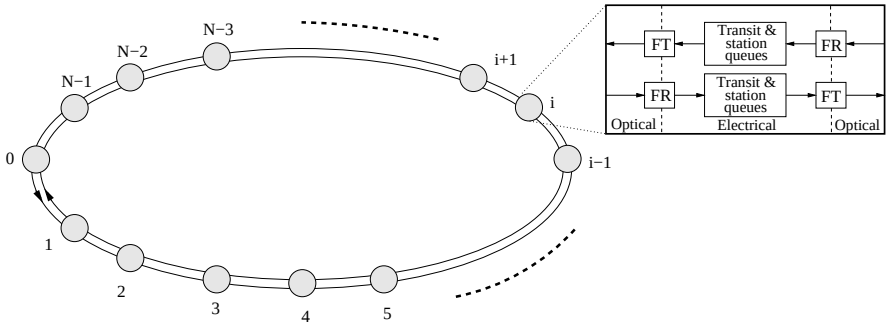


Figure 3.1: RPR network and node architecture.

The RPR network consists of a bidirectional dual-fiber ring using one wavelength for each direction which is OEO converted at each node, i.e., RPR does not implement WDM. The counter-rotating rings provide protection against any single link or node failure. The network and node architecture is illustrated in Fig. 3.1. Every node is equipped with two fixed-tuned transmitters (FTs) and two fixed-tuned receivers (FRs), one for each fiber ring. RPR is an example of a *buffer insertion ring* where each node features three different types of electronic first-in-first-out (FIFO) buffers or queues: reception, transmission, and insertion [38]. In general, the reception and transmission buffers store packets that are destined to or originate from the corresponding node. The insertion buffer temporarily stores the incoming ring traffic in the electrical domain in order to allow the local node to transmit a packet onto the ring. Details of RPR's architecture and MAC protocol are discussed in Section 6.1. RPR network design makes use of the following four underlying principles:

- *Source stripping*: With source stripping the source node removes the transmitted packet

from the ring.

- *Destination stripping*: With destination stripping, packets are removed from the ring by the receiving node rather than the transmitting node.
- *Spatial reuse*: As opposed to source stripping, destination stripping enables the destination stripping node and its downstream neighbor nodes to spatially reuse bandwidth on the ring, resulting in a higher degree of concurrency and an increased network capacity.
- *Shortest path routing*: With shortest path routing a given source node transmits packets to a destination node via the shortest path (e.g., given in terms of number of hops or distance) by using the appropriate ring.

Spatial reuse and shortest path steering are well understood, and it was shown within the MetaRing project that their use increases the network capacity significantly [39, 40]. The RPR standard also defines mechanisms to provide three different QoS classes. A fairness control algorithms allows a congested downstream node to throttle the transmission rate of upstream nodes by sending fairness control packets upstream. For more details RPR's implementation of QoS and fairness control please refer to Chapter 9. The two main limitations of RPR are (*i*) the use of only one wavelength in each fiber, and (*ii*) the OEO conversion of all traffic at each node, i.e., the fact that RPR belongs to the family of first generation (opaque) networks. WDM ring networks overcome these limitations by using multiple wavelengths in a fiber and optically bypassing transit traffic.

3.1.4 Ethernet Passive Optical Networks

For completeness we also mention EPONs, which recently have attracted considerable research and standardization activities [41]. EPONs fall into the category of access networks, i.e., they connect multiple end users to one node of a metro network. Architecturally, an EPON is a point-to-multipoint optical network with no active elements in the signal path from source to destination. The only interior elements used in an EPON are passive optical components, such as optical fiber, splices (which connect two fibers), and splitters, which fan out to multiple optical drop fibers connected to subscriber nodes. An EPON is an optical broadcast network, possibly augmented with a wavelength-routing WDM overlay. There are several EPON topologies suitable for the access networks. Typically, EPONs have a tree topology, but also other topologies such as ring, tree-and-branch, and bus are possible. An EPON carries all data encapsulated in Ethernet frames. In addition to the standardization efforts, research on the design and evaluation of efficient multiple access schemes for EPONs have begun recently [42, 43, 44].

Newly adopted QoS techniques have made Ethernet networks capable of supporting voice, data, and video. These techniques include full-duplex support, prioritization (IEEE P802.1p), and virtual LAN (VLAN) tagging (IEEE P802.1q). The standards work for Ethernet in the local subscriber access network is currently being done in the IEEE P802.3ah Ethernet in the First Mile (EFM) Task Force. Ultimately, the optical Ethernet has the potential to evolve from a pure LAN technology to a MAN technology that some predict will replace SONET/SDH, ATM, and Frame Relay [45].

3.2 Experimental Systems

In this section, we survey three of the most recent experimental testbed systems for packet-switched ring metro WDM networks: KomNet, RINGO, and HORNET. The surveyed testbed

systems illustrate the capabilities of currently readily available photonic hardware components. By the way of explaining the functioning of these testbeds we also explain the functionalities of several key photonic networking components. Most of the experimental ring metro WDM networks surveyed in this section operate at a line rate of 2.5 Gbit/s. Depending on the used technology, the systems are suited either for circuit or packet switching. While transmitters have been demonstrated to be tunable across adjacent wavelengths in a few nanoseconds, fast tunable receivers (TRs) are not yet mature. Therefore, most of the experimental packet switched WDM ring networks use FRs rather than TRs receivers.

3.2.1 KomNet

The KomNet metro WDM field trial network consists of three OADMs interconnected in a bidirectional fiber ring topology [46, 47]. The structure of an OADM is shown in detail in Fig. 3.2. On each fiber, multiple wavelengths can be dropped by deploying tunable fiber Bragg gratings (FBGs). The FBGs reflect the desired wavelengths back to the circulator, which takes them off the ring and forwards them to the demultiplexer. By using wavelength-insensitive combiners multiple wavelengths can be added to each fiber. Each FBG has a relatively small insertion loss of 0.1 dB. The FBGs can be mechanically tuned within the millisecond range. Therefore, KomNet is well suited for (lambda) circuit switching, but is inefficient for packet switching due to the relatively large tuning time of each FBG.

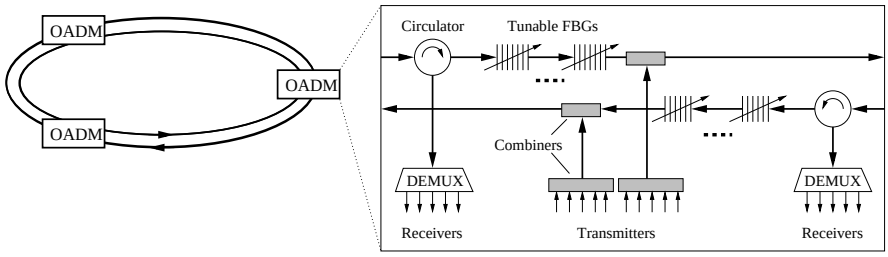


Figure 3.2: The KomNet metro WDM network.

3.2.2 RINGO

The packet-switched RING Optical network (RINGO) has a unidirectional fiber ring network architecture [48, 49]. It features N nodes, where N equals also the number of wavelengths. Each node is equipped with an array of FTs and one FR operating on a given wavelength that identifies the node. That is, node j drops wavelength λ_j from the ring. Thus, in order to communicate with node j, a given node i has to transmit data by using the laser operating on wavelength λ_j, as illustrated in Fig. 3.3. All wavelengths are slotted with the slot length equal to the transmission time of a fixed-size data packet plus guard time. Each node performs λ-monitoring, i.e., checks the state of the wavelength occupation, on a slot-by-slot basis to avoid channel collisions. This approach is a multichannel generalization of the *empty-slot approach*. In the empty-slot approach one bit at the beginning of each slot indicates the state of the corresponding slot, i.e., whether the slot is free (empty) or occupied. A monitoring node is only allowed to use empty slots for its transmissions.

Fig. 3.4 depicts the node structure in greater detail. At each node all wavelengths are demultiplexed. The drop wavelength is routed to a burst mode receiver while the status of

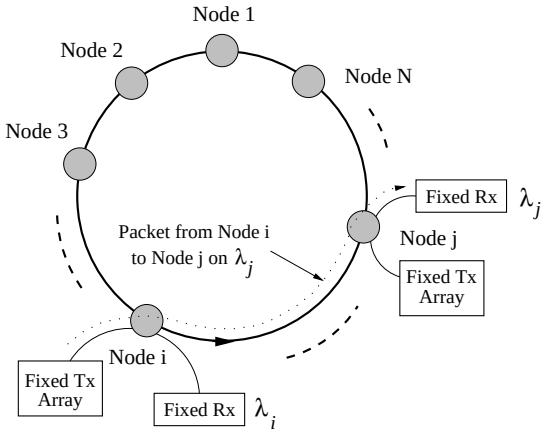


Figure 3.3: RINGO metro WDM network.

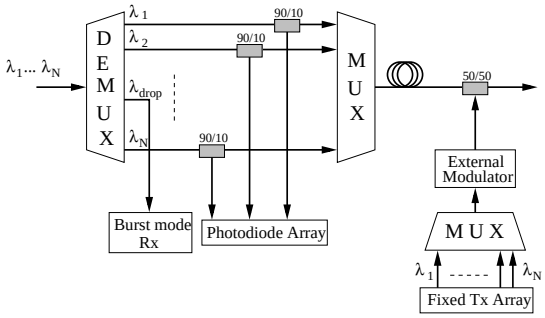


Figure 3.4: RINGO node structure.

the remaining wavelengths is monitored by using 90/10 taps and an array of photodiodes. A burst mode receiver recovers the clock for each optical burst (packet) very quickly and does not need to receive a continuous signal. A 90/10 tap splits off 10% of the optical power from the fiber. Subsequently, the wavelengths are multiplexed on the outgoing ring fiber. With a 50/50 combiner and an external modulator the node is able to send data packets by activating one or more fixed-tuned transmitters. A 50/50 combiner collects signals from two input ports and equally combines them onto one output port. Both input signals experience thereby a combining loss of 3 dB.

3.2.3 HORNET

The Hybrid Optoelectronic Ring NETwork (HORNET) is a unidirectional WDM ring network [50, 51]. All wavelengths are slotted with the slot length equal to the transmission time of a fixed-size packet (plus guard time). Each wavelength is shared by several nodes for data reception. Every node is equipped with one fast tunable transmitter (TT) [52, 53] and one fixed-tuned burst mode receiver [54]. As shown in Fig. 3.5, the node structure consists of a slot manager, a smart drop, and a smart add module [55].

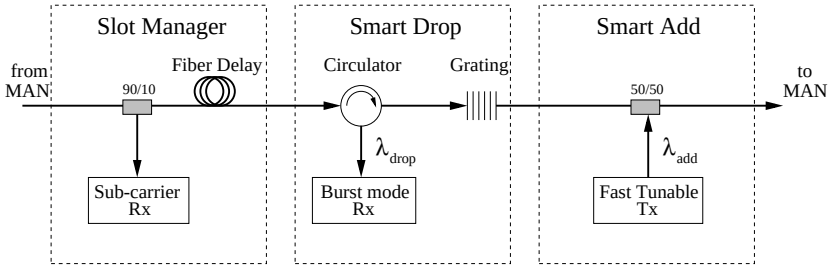


Figure 3.5: HORNET node structure.

Access to all wavelengths is governed by means of a *carrier sense multiple access with collision avoidance (CSMA/CA)* MAC protocol [56, 57]. When a node transmits a packet it multiplexes a sub-carrier tone onto the packet at a sub-carrier frequency that corresponds to the wavelength on which the packet is sent. The destination address of the packet is

modulated onto the sub-carrier multiplexing (SCM) tone using a combination of amplitude shift keying (ASK) and frequency shift keying (FSK). For carrier sensing, the slot manager taps off a small amount of optical power and detects it with one photodiode, as illustrated in Figs. 3.6 and 3.7. The payload data from all wavelengths collide at baseband while the SCM tones remain intact. The composite SCM signal is demultiplexed into the individual SCM tones using a collection of bandpass filters. The SCM tone corresponding to the drop wavelength of the node is FSK demodulated while the other SCM tones are ASK demodulated. The outcome of the ASK demodulation indicates the absence or presence of a packet on the corresponding wavelength. This allows the node to determine whether a wavelength is free for a packet transmission, which is conducted with the smart add module. The outcome of the FSK demodulation indicates whether there is a packet on the node's drop wavelength. If there is a packet, it is taken off the ring with the node's burst mode receiver. The outcome of the FSK demodulation also gives the destination address of the packet. If the destination address does not match the node's address, then the node forwards the packet using it's smart add module.

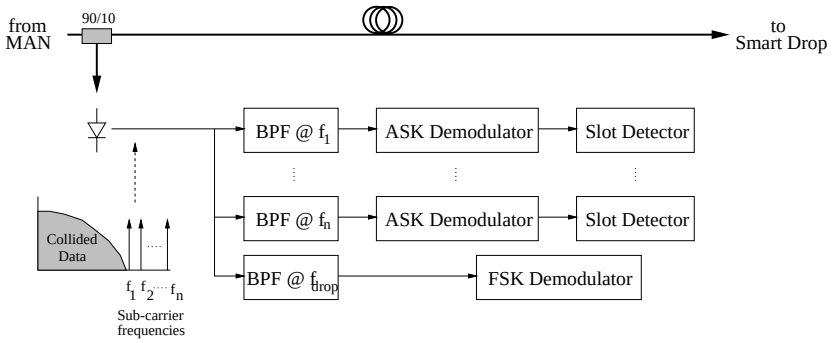


Figure 3.6: Structure of the HORNET slot manager.

3.3 WDM Rings and Access Protocols

In this section, we provide a comprehensive survey of packet-switched ring metro WDM networks. We first discuss a generic WDM ring network architecture from which essentially all studied architectures can be derived with a few modifications. We also introduce a classification of the networks based on the employed MAC protocol. We then survey the networks in the individual categories of our classification.

Most packet-switched ring WDM networks are based on a unidirectional all-optical fiber ring, as shown in Fig. 3.7. At each node an OADM drops a prescribed wavelength from the ring and allows addition of data at any arbitrary wavelength. A node transmits data on the added wavelength while it receives data on the dropped wavelength. Data on the dropped wavelength are removed from the ring and optical-electronically converted. If the number of nodes N is equal to the number of wavelengths W, as depicted in Fig. 3.7 for $N = W = 4$, each node has a dedicated *home channel* for reception. However, in general $N \geq W$ since the number of available wavelengths is limited, e.g., for cost reasons or finite transceiver tuning ranges. With $N \geq W$ the system is referred to as *scalable* since the number of nodes is independent of the number of available wavelengths.

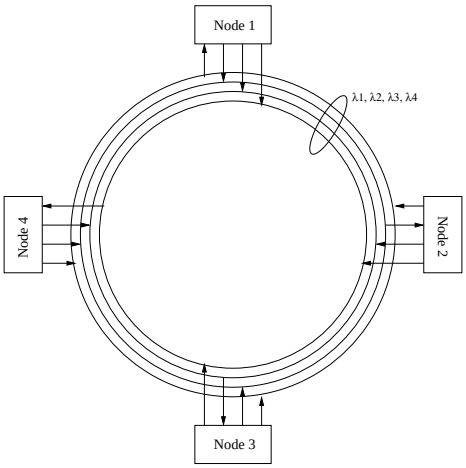


Figure 3.7: Single-fiber network architecture with $N = 4$ nodes and $\Lambda = 4$ wavelength channels.

Each node is equipped with one or more fixed-tuned and/or tunable transmitters and receivers. We adopt the FT^i–TT^j–FR^m–TR^n notation to describe the node architecture, where $i, j, m, n \geq 0$ [58]. That is, each node is equipped with i fixed-tuned transmitters, j tunable transmitters, m fixed-tuned receivers, and n tunable receivers. For example, a TT–FR node structure means that each node has one tunable transmitter and one fixed-tuned receiver.

When a node inserts a packet on a given wavelength while another packet is currently passing the ring on the same wavelength a *channel collision* occurs and both packets are disrupted. With tunable receivers also *receiver collisions*, which are also known as destination conflicts, can occur when a node's receiver is not tuned to the wavelength of an incoming packet. This can happen if the destination node does not know about the transmission or another packet is currently received on a different wavelength. Clearly, both channel and receiver collisions have a detrimental impact on the throughput-delay performance of the network. The degradation of the network performance due to channel or receiver collisions can be mitigated or completely avoided at the architecture and/or protocol level. For example, equipping each node with a receiver fixed tuned to a home channel (either dedicated to a single node or shared by multiple nodes) prevents receiver collisions. Similarly, allocating each node a separate home channel for transmission avoids channel collision at the expense of scalability. In scalable systems, i.e., systems with $N \geq W$, however, each wavelength channel is typically shared by multiple nodes giving rise to channel collisions. Clearly, MAC protocols are needed to govern the access to the wavelength channels and to mitigate or prevent channel (and receiver) collisions.

Packet-switched ring WDM networks can be classified according to a number of different criteria, e.g., unidirectional vs. bidirectional rings or dedicated vs. shared protection [59]. We introduce a classification of the networks according to the MAC protocols that they employ. As illustrated in Fig. 3.8, we introduce the main categories of slotted rings, multitoken rings, and meshed rings.

Slotted ring MAC protocols, in which the time is divided into fixed-length slots, can be further classified into protocols without and with channel inspection, and those making use

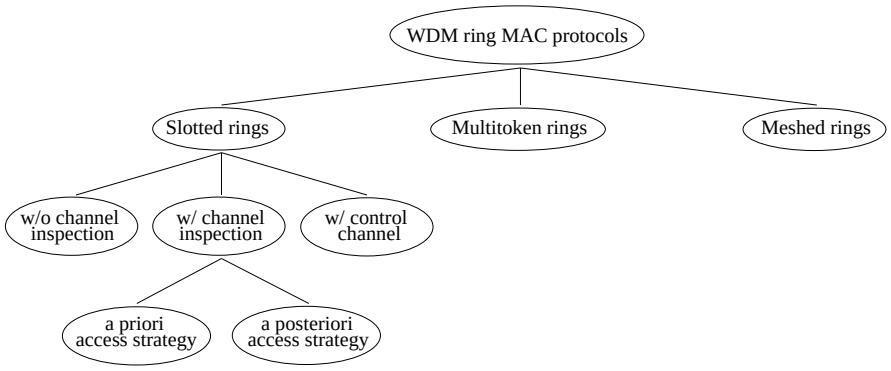


Figure 3.8: Classification of ring WDM network MAC protocols.

of a separate control channel. Protocols with channel inspection determine the status (empty or occupied) of a slot before sending a packet, whereas protocols without channel inspection do not perform such a check of the slot status. A control channel is an additional wavelength channel that is used exclusively for the transmission of control information and does not carry any payload data. MAC protocols with channel inspection use one of two different access strategies: Either an *a priori* or an *a posteriori* access strategy. With *a priori* access the packet to transmit in the upcoming slot is selected before the channel inspection of the slot is completed. This has the advantage that the packet selection can be performed without strict timing constraints. The drawback is that the drop wavelength of the destination of the selected packet may turn out to be occupied in the upcoming slot, in which case the packet can not be transmitted. Also, if some other wavelength is free in this slot, it is not possible to select a different packet for any such free wavelength, resulting in a potential waste of bandwidth. With *a posteriori* access, on the other hand, the packet to transmit in an upcoming slot is selected after the inspection of the slot is completed. This has the advantage that only packets whose destination drop wavelength is empty in the slot are considered. The drawback is that the *a posteriori* packet selection needs to be performed under tight timing constraints since there is only a small fiber delay between the slot inspection and the packet transmission into the slot, as illustrated in Figs. 3.4 and 3.5.

In multitoken rings the time is not slotted. Instead, on each wavelength channel there is a special control packet, the *token*, that travels around the ring. A given node can hold the token for some time duration governed by the MAC protocol and transmit data packet(s) on the corresponding wavelength while it holds the token.

Finally, a meshed ring network is a ring network that is augmented by additional fibers that create short-cuts between prescribed nodes on the ring. Although the meshed ring is strictly speaking not a 'pure' ring network, we include it in our survey for completeness, and because meshed ring networks are closely related to ring networks.

We now comprehensively survey the packet-switched ring metro WDM networks. Our discussion proceeds from left to right in the classification illustrated in Fig. 3.8, i.e., we begin with slotted rings without channel inspection and end with meshed rings.

3.3.1 Slotted Rings

Slotted rings are attractive because the fixed size time slots make it relatively easy to keep all nodes synchronized which can be tricky in high-speed networks. On the downside, the fixed slot size requires additional efforts to transport variable size packet, as we will see shortly.

Slotted Rings without Channel Inspection

A simple way to avoid channel and receiver collisions is the deployment of time division multiple access (TDMA). Time is divided into slots equal to the packet transmission time. Typically, these time slots are of a fixed size with multiple slots circulating at each wavelength on the ring, as illustrated in Fig. 3.9. The slots at different wavelengths are typically aligned. With TDMA, channel and receiver collisions are avoided by statically assigning each slot to a prescribed source-destination pair. Thus, a fixed amount of capacity is allocated to each pair of nodes which is well suited for uniform regular traffic at medium to high loads, but leads to wasted bandwidth and low channel utilization in the case of bursty traffic.

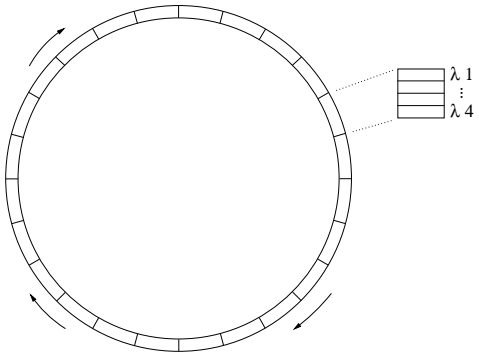


Figure 3.9: Slotted unidirectional WDM ring with $W = 4$ wavelengths.

The only packet-switched network that falls in this category of slotted rings without channel inspection is the Metropolitan Area Wavelength Switched Optical Network (MAWSON) [60, 61, 62]. MAWSON is based on a FT^W–FR or alternatively a TT–FR node architecture. N nodes are connected to the ring via OADMs that use FBGs, as discussed in Sec. 3.2.1, for dropping a different wavelength for reception at each node. In MAWSON, the number of nodes N is equal to the number of wavelengths W and each node has a dedicated home channel, which avoids receiver collisions. In other words, a given wavelength channel interconnects $(N-1)$ source nodes and one destination node. With the FT^W–FR node structure broadcasting and multicasting can be achieved by simultaneously turning on multiple lasers, but only unicasting is considered in the evaluation of the MAC protocol.

Time is divided into fixed-size slots, which are assumed to be aligned across all W wavelengths. Each slot is further subdivided into header and data fields, as shown in Fig. 3.10. The slots on a given wavelength channel are assigned dynamically on demand. To this end, the header of each slot consists of $(N-1)$ *Request/Allocation (R/A) minislots* which are statically preassigned in a TDMA fashion to $(N-1)$ source nodes. Each R/A minislot essentially consists of two fields, one for requests and one for allocations. More precisely, node i ready to send variable-size data packets to node j uses the request field of its assigned R/A minislot

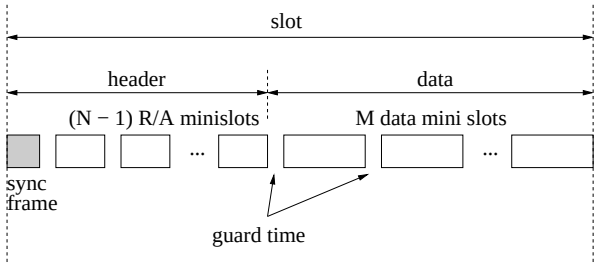


Figure 3.10: Slot structure of Request/Allocation Protocol in MAWSON.

on j's home wavelength channel to make a request. Upon receipt of node i's request, node j allocates one or more data minislots to node i by using the allocation field of its assigned R/A minislot on i's home wavelength. After receiving its allocation node i transmits the data packet using the allocated data minislots.

To save costs the node architecture and protocol of MAWSON are kept simple, e.g., no carrier sensing capabilities are required. Due to in-band signaling no additional control channel and control transceivers are needed. The protocol completely avoids both channel and receiver collisions, achieves good throughput performance, and provides fairness by allocating slots in a round-robin manner. However, the R/A procedure introduces some overhead and additional delay since the request and allocation takes at least one round trip time around the ring.

Slotted Rings with Channel Inspection

In most slotted WDM rings channel collisions are avoided by enabling the nodes to check the status (used/unused) of each slot. Generally, this is done by tapping off some power from the fiber and delaying the slot while the status of each wavelength is inspected in the tapped off signal and electronically processed. A packet can then be inserted in a slot at an unused wavelength. Packets waiting for transmission are stored in *virtual output queues (VOQs)*. Typically, a node maintains separate VOQs either for each destination or for each wavelength. In the latter case packets arriving at a node from the higher layer are put in the VOQ associated with the drop wavelength of the packet's destination. In WDM ring networks it is typically the responsibility of the MAC protocol to select the appropriate VOQ from which to send a packet in a time slot according to a given access strategy. This can be done *a priori*, i.e., without taking the status of the slots into account, or *a posteriori*, i.e., with taking the status of the slots into account. In the *a priori* access strategy each node selects a VOQ prior to inspecting the slot status. Whereas, in the *a posteriori* strategy each node first checks the status of a slot and then selects an appropriate (non-empty) VOQ. We now describe the networks that fall in the category of slotted rings with channel inspection; see Fig. 3.8. We cover both the networks with *a priori* access and the networks with *a posteriori* access in this section, as many networks can be operated with either access strategy.

RINGO We have already presented the network architecture of the RINGO in Section 3.2.2 and now discuss the MAC protocol for RINGO. First, recall from Section 3.2.2 that RINGO uses a FT^W–FR node architecture. Each node has channel inspection capability built with

commercially available components. Nodes execute a multichannel empty-slot MAC protocol which can also be applied to a TT–FR node architecture.

A MAC protocol with *a posteriori* queue selection has been implemented in the RINGO testbed [63]. The number of wavelengths is assumed to be equal to the number of nodes, and each node has one VOQ with first-come-first-served (FCFS) queuing discipline for each wavelength, meaning the oldest packet is sent first. Only the VOQs where the corresponding wavelengths have been found to be empty (unused) are allowed to send data packets in the free time slot. If a TT–FR node architecture is used only one packet can be sent per time slot and the longest among those queues is chosen.

The overhead of the RINGO empty-slot MAC protocol is very small. To identify the status of a given slot, a single bit is sufficient. All wavelengths are used for data transmission and no separate control channel or control transceivers are required. It was demonstrated that all-optical packet-switched ring WDM networks are feasible with currently available technology. However, owing to the fixed slot size, the transmitted packets have to be of fixed size. Note that variable-size packets can be transmitted in slotted rings without segmentation and reassembly by means of buffer insertion techniques exploiting optical delay lines [64]. More precisely, a transmitting node inserts a sufficiently long optical delay line to delay the in-transit ring traffic until the node has completed the transmission of its (variable-size) data packet. In doing so, the collision of the in-transit ring traffic and the locally injected traffic is avoided.

Synchronous Round Robin Synchronous Round Robin (SRR) is another empty-slot MAC protocol for a unidirectional WDM ring network with fixed-size time slots and destination stripping [65, 66, 67]. Each node is equipped with one tunable transmitter and one fixed-tuned receiver (TT–FR), where the transmitter is assumed to be tunable across all W wavelengths on a per-slot basis. If $N = W$ each node has its own home wavelength channel for reception. In the more general case $N > W$ each wavelength is shared by multiple destination nodes [65].

In SRR, each node has $(N - 1)$ separate FIFO VOQs, one for each destination, as shown in Fig. 3.11. SRR uses an *a priori* access strategy. Specifically, each node scans the VOQs in a round-robin manner on a per-slot basis, looking for a packet to transmit. If such a deterministically selected VOQ is nonempty, the first (oldest) packet is transmitted, provided the current slot was sensed empty. If the selected VOQ is empty the first packet from the longest queue among the remaining VOQs is transmitted, again provided the current slot is unused. If the current slot is occupied, i.e., a transmission is not possible as it would result in a channel collision, then no packet is transmitted from the selected VOQ. For the transmission attempt in the next slot, the next VOQ is selected according to the round-robin scanning of SRR. In doing so, under heavy uniform load conditions, when all VOQs are non-empty, the SRR scheduling algorithm converges to round-robin TDMA.

For uniform traffic, SRR asymptotically achieves a bandwidth utilization of 100%. However, the presence of unbalanced traffic leads to wasted bandwidth due to the nonzero probability that the *a priori* access strategy selects a wavelength channel whose slot is occupied while leaving free slots unused. It was shown in [68] that *a posteriori* access strategies avoid this drawback resulting in an improved throughput-delay performance, albeit at the expense of increased complexity.

SRR achieves good performance requiring only local information on the backlog of the VOQ, which also avoid the well-known head-of-line (HOL) blocking problem. Owing to destination

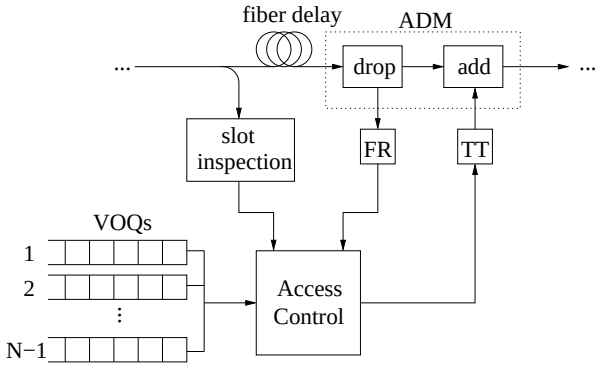


Figure 3.11: SRR node architecture with VOQs and channel inspection capability.

stripping, slots can be spatially reused several times as they propagate along the ring. On the other hand, slot reuse raises fairness control problems, particularly for nonuniform traffic. A node to which a large amount of slots is directed generates a large amount of free slots, and nodes immediately downstream are in a favorable position with respect to other nodes. We will address this fairness problem in Section 3.4.1. Note that in order to provide QoS, SRR requires additional modifications, which are discussed in greater detail in Section 3.4.2.

HORNET The architecture of the HORNET has already been described in Section 3.2.3. Recall that HORNET is a unidirectional destination-stripping ring WDM network with a TT–FR network structure. Similar to SRR, to avoid HOL blocking, each node uses VOQs, one for each wavelength, and both *a priori* and *a posteriori* access strategies can be used. Nodes sense the availability of each slot by monitoring SCM tones. The SCM based carrier-sensing scheme requires fewer hardware components than demultiplexing, separately monitoring, and subsequently multiplexing all wavelengths of the WDM comb, as done in RINGO. More precisely, instead of the demultiplexer, photodiode array, and multiplexer used in RINGO, the HORNET channel inspection scheme requires only a single photodiode.

HORNET's carrier sense multiple access with collision avoidance CSMA/CA MAC protocol initially assumed fixed-size slots which are well suited for the transport of fixed-size packets, e.g., ATM cells [50]. The MAC protocol can be extended to support variable-size IP packets. Two CSMA/CA MAC protocols both supporting variable-size packets are proposed and investigated in [69]. In the first protocol, slots of different sizes circulate along the ring. The slot sizes are chosen according to the predominant IP packet lengths as typically found in traffic measurements. For example, three slots sizes can be chosen such that 40, 552, and 1500 Byte long IP packets are accommodated. A dedicated node controls the size and number of slots such that they match the packet size distribution. A variant of this protocol for a TT–FR^W architecture has been proposed in [70].

The second protocol is unslotted and operates similarly to carrier sense multiple access with collision detection (CSMA/CD), i.e., it features collision detection and backoff. More precisely, when a wavelength is sensed idle, a given node begins to transmit a packet. When another packet arrives on the same wavelength before the transmission is complete, the packet transmission is aborted. In this case, the incomplete packet is marked by adding a jamming signal to the end of the packet. Aborted transmissions are resumed after some backoff time

interval.

A more bandwidth-efficient modification of the second unslotted CSMA/CA protocol was examined in [71]. In the examined *carrier sense multiple access with collision preemption (CSMA/CP) protocol*, variable-size IP packets do not necessarily have to be transmitted in a single attempt. Instead, packets are allowed to be transmitted and received as fragments that are reassembled at the receiver. Thus, successfully transmitted parts of the original *IP* packet are not retransmitted, resulting in a higher channel utilization.

Besides demonstrating the feasibility of the SCM based channel inspection approach, the HORNET project also proved the feasibility of fast tunable transmitters. These allow for replacing arrays of multiple fixed-tuned transmitters with a single tunable transmitter. In HORNET the number of nodes is independent from the number of wavelengths, and it is thus considered scalable. Generally, each wavelength is allowed to be shared by multiple destination nodes with packet forwarding at intermediate nodes, resulting in translucent multihop networks. Note that intermediate nodes not only forward packets towards the destination but also provide signal regeneration in the electrical domain. On the other hand, the CSMA/CA random access protocol does not provide QoS, and the destination stripping gives rise to fairness problems.

Several *a posteriori* buffer selection schemes for the HORNET architecture are studied by Bengi and van As [72, 73]. Recall that in an empty-slot protocol, each unused slot on any wavelength channel can be used for packet transmission by a source node. However, when more than one wavelength channel carries an empty slot in the current slot period, one packet (or equivalently, one VOQ) corresponding to one of the empty channels has to be chosen according to a prescribed selection rule. Due to the short time between channel inspection and packet transmission, the *a posteriori* packet selection process has to be performed at a high speed in the electronic domain, which increases the processing complexity compared to an *a priori* packet selection scheme. Five different *a posteriori* VOQ selection strategies are described and examined in [72]:

- *Random Selection*: The VOQ from which a packet is to be transmitted is selected randomly according to a uniform distribution.
- *Longest Queue Selection*: The longest VOQ is chosen upon buffer contention.
- *Round-Robin Selection*: The VOQ is chosen in a round-robin fashion.
- *Maximum Hop Selection*: The packet (VOQ) associated with the maximum hop distance between source and destination node is selected when buffer contention arises.
- *C–TDMA Selection*: The *channel-oriented TDMA (C–TDMA)* scheme first attempts to select the packet according to a round-robin policy. If that selection would prevent a transmission, either due to an empty VOQ or an occupied slot, then the longest VOQ that allows for a packet transmission is chosen. This scheme is largely equivalent to the SRR scheme with *a posteriori* access; see Section 3.3.1.

It was found that the random and round-robin buffer selection schemes provide a satisfactory compromise between performance and implementational complexity.

FT–TR Rings Jelger and Elmirghani [74] proposed a unidirectional empty-slot WDM ring network that uses source stripping. Each node is equipped with one fixed-tuned transmitter and one tunable receiver (FT–TR). Packets are buffered in a single FIFO transmit queue at each node. In the applied source-stripping scheme, a sender must not reuse the slot it just

marked empty. It was shown that for source-stripping rings this simple mechanism ensures fairness in that a node can not starve the entire network; however the mechanism does not ensure fairness for destination-stripping rings.

The performance of the network was compared for both source and destination stripping in [75]. By means of simulation it was shown that destination stripping clearly outperforms source stripping in terms of throughput, delay, and packet dropping probability.

Clearly, with only one tunable receiver at each node, receiver collisions can occur. Receiver collisions can be avoided in a number of ways. In one approach, arriving packets which find the destination's receiver busy re-circulate on the ring until the receiver of the destination is free, i.e., is tuned to the corresponding wavelength [75]. Alternatively, receiver collisions can be completely avoided at the architecture level by replacing each node's tunable receiver with an array of W fixed-tuned receivers, each operating at a different wavelength (FT–FR^W) [76]. Another proposal to resolve receiver contention is based on optical switched delay lines (SDLs) [77]. A destination node puts all simultaneously arriving packets but one into optical delay lines such that packets can be received sequentially.

Slotted Rings with Control Channel

In slotted ring networks with control channel, the status of the slots is transmitted on a separate control channels (CCs) wavelength. Each node is typically equipped with an additional transmitter and receiver, both fixed tuned to the control wavelength. A separate control channel wavelength enables nodes to exchange control information at high line rates and eases the implementation of enhanced access protocols with fairness control and QoS support, as we will see shortly.

Bidirectional HORNET with SAR–OD An extended version of the original unidirectional TT–FR HORNET ring architecture in which SCM is replaced with a separate control channel wavelength is investigated in [78]. Transmission on the control channel (and data wavelengths) is divided into fixed-size slots. The control channel carries the wavelength availability information such that nodes are able to 'see' one slot into the future. Two counterdirectional fiber rings each carrying W data wavelengths and an additional control channel wavelength operate in parallel. On each ring, every node deploys one fast-tunable transmitter and one fixed-tuned receiver for data, and one transceiver fixed tuned to the control channel wavelength. Thus, the control channel based HORNET network is a CC–FT^2–TT^2–FR^4 system.

A modified MAC protocol able to efficiently support variable-size packets over the bidirectional ring network was examined. This *segmentation and reassembly on demand (SAR-OD)* access protocol aims at reducing the number of segmentation and reassembly operations of variable-size packets. Specifically, the transmission of a packet from a given VOQ starts in an empty slot. If the packet is larger than a single slot, the transmission continues until it is complete or the following slot is occupied, i.e., the packet is segmented only if required to avoid channel collisions. If a packet has to be segmented, it is marked incomplete, and the transmission of the remaining packet segment(s) continues in the next empty slot(s) on the corresponding wavelength. By means of simulation it was shown that SAR-OD reduces the segmentation/reassembly overhead by approximately 15% compared to a less intelligent approach where all packets larger than one slot are segmented irrespective of the state of successive slots.

The control channel based bidirectional HORNET ring network preserves the advantages of the original unidirectional HORNET ring, e.g., scalability and small number of hardware components. Bidirectional dual-fiber rings provide an improved fault tolerance against node/fiber failures and survivability compared to unidirectional single-fiber rings [79, 80]. Furthermore, the control channel can also be used to achieve efficient fairness control, as described in greater detail in Section 3.4.1.

Variable-Size Packets without Segmentation/Reassembly An access protocol for a control channel based slotted ring WDM network that completely avoids segmentation and reassembly of variable-size packets was studied by Bengi in [81, 82]. The access protocol is an extended version of Bengi's original protocol described above in Section 3.3.1. The architecture differs from the control channel based HORNET in that a unidirectional ring is deployed, and each node uses an additional transmitter fixed tuned to the node's drop wavelength, resulting in a CC–FT^2–TT–FR^2 system. The additional transmitter is used to forward dropped packets which are destined to downstream nodes which share the same drop wavelength.

The extended MAC protocol relies on a frame-based slot reservation strategy including reservation of successive slots for data packets longer than the given slot size and immediate access for packets shorter than the slot length. Each node is equipped with two VOQs for each wavelength, one for short packets and one for long packets. The ring is subdivided into multiple *reservation frames* with the frame size equal to the largest possible packet length. In these frames, multiple consecutive slots are reserved to transmit long packets without segmentation. A single reservation control packet containing all reservations circulates on the control channel. Each node maintains a table in which the reservations of all nodes are stored. When the control packet passes, a node updates its table and is allowed to make a reservation. The additional fixed-tuned transmitter is used to forward packets concurrently with transmitting long packets within multiple contiguous slots. Besides the support of long packets via reservation, short packets fitting into one slot are accommodated by means of immediate access of empty and unreserved slots.

The proposed protocol provides immediate medium access for packets shorter than one slot and completely avoids the segmentation and reassembly of longer variable-size packets, resulting in a reduced complexity. The reservation protocol also enables QoS support, as discussed in greater detail in Section 3.4.2. On the other hand, the reservation protocol introduces some delay overhead, and reserved slots on their way back to the source node can not be spatially reused after destination stripping.

Wavelength Stacking The *wavelength stacking* technique studied by Smiljanic *et al.* transmits a packet using all wavelengths of a time slot of a control channel based slotted unidirectional WDM ring [83, 84, 85]. Each node is equipped with one fast-tunable transmitter and one photodiode. Time is divided into slots of duration T_p. The length of a data packet is W time slots. A fast-tunable laser at a given node starts transmission W time slots before its scheduled time slot. As illustrated in Fig. 3.12 for $W = 3$, in each following time slot it transmits data on a different wavelength. The signal passes through the array of fiber gratings separated by delay lines so that the W segments of the data packet transmitted at different wavelengths are aligned in time. The packet is then transmitted to the network on all wavelengths in parallel by setting switch S to the cross state. On the receiver side, the

reverse procedure is performed. A packet is received when switch S is in the cross state and is then unstacked by passing through the same array of fiber gratings and delay lines. Note that a single broadband wavelength-insensitive photodiode without optical filter is sufficient for the packet reception since at most one wavelength needs to be converted from the optical to the electronic domain at any given time. The photodiode converts the optical signal into an electrical signal irrespective of the optical carrier frequency (wavelength).

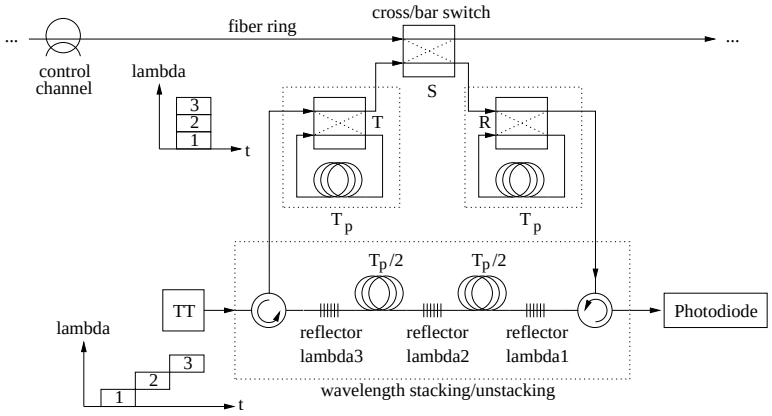


Figure 3.12: Node architecture for wavelength stacking.

Because wavelength stacking takes W time slots, a node needs to decide in advance when to access the medium. A separate wavelength is used as a control channel for the reservations. Time slots are grouped into cycles of length W slots. Each node may transmit and receive at most one packet within each cycle. The switches T and R in Fig. 3.12 synchronize the wavelength stacking and unstacking. The wavelength stacking is completed in the last time slot of a given cycle and the packet is stored in the delay line by setting T in the cross state. A packet is stored as long as switch T is in the bar state. The packet is transmitted to the network by setting switches T and S in the cross state exactly $2W$ time slots after the reservation. Whenever a node recognizes its address on the control channel, it stores the packet in the delay line by setting switches S and R in the cross state $2W$ time slots after the address notification. The node starts unstacking the packet at the beginning of the next cycle by setting switch R in the cross state. Each node removes a packet that it receives as well as its reservation.

The wavelength stacking/unstacking allows a node to simultaneously send/receive data at different wavelengths in the same time slot despite the fact that the node has only one transceiver. The presented node architecture can be used to realize photonic slot routing (PSR) metro WDM networks, where all wavelengths in a given slot (the *photonic slot*) are switched together rather than separately on a per-wavelength basis [86, 87]. However, the quality of the optical signal may suffer from passing the numerous delay lines and switches in a node.

Virtual Circles with DWADMs In the unidirectional slotted ring WDM network presented by Cho and Mukherjee in [88], each node is equipped with a *dynamic wavelength add-drop multiplexer (DWADM)*. As opposed to tunable transmitters and receivers which can operate independently, the input and output wavelengths of a DWADM must be the same, i.e., if the wavelength to receive at a given node s is λ_i, the wavelength to transmit must be the

same wavelength λ_i. Furthermore, if the node has to send to another node d, then node d has to use wavelength λ_i to receive and to send a packet. Thus, virtual circles are created, as depicted in Fig. 3.13, which change dynamically according to varying traffic demands.

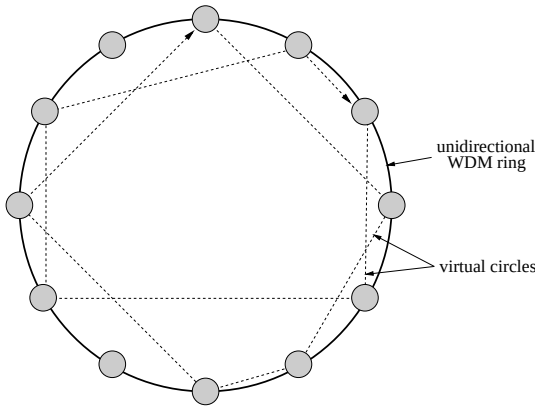


Figure 3.13: Virtual circles comprising nodes whose DWADMs are tuned to the same wavelength.

The ring network uses W data wavelength channels and a separate control wavelength channel. Nodes communicate over the control channel in a TDM fashion to exchange transmission requests and acknowledgements. The $(W + 1)$ wavelengths are divided into three cycles, which are repeated periodically. In the first cycle a control packet sent by a server node collects transmission requests from all nodes. These are processed by the server node, and wavelength assignments/acknowledgements are sent back to the nodes in the second cycle. In the third cycle each node that has been assigned a wavelength tunes the DWADM appropriately and starts the data transmission.

Owing to their relatively simple structure DWADMs are less expensive than tunable transceivers [88]. However, due to their reduced flexibility, the wavelength utilization is typically smaller than in TT–TR systems, where transmitters and receivers can be tuned to any arbitrary wavelength independently.

3.3.2 Multitoken Rings

Slotted WDM ring networks have a number of advantages such as easy synchronization of nodes even at high data rates. Also, they can achieve high channel utilization and low access delay and allow for relatively simple access schemes. However, variable-size packets are difficult to accommodate and, as discussed in Sec. 3.4, explicit fairness control is needed, which can complicate the medium access significantly. In contrast, variable-size packets can be transported in a reasonably fair manner in token rings where the access is controlled by means of a special control packet, the token, which circulates around the ring. The token is passed downstream from node to node. Each node can hold the token up to a prescribed amount of time, during which the node is allowed to send (fixed-size or variable-size) packets. Due to the limited token holding time, fairness is achieved. Furthermore, as opposed to slotted rings, nodes do not have to be synchronized. On the other hand, immediate channel access is not possible and the token rotation time (ring propagation time) may decrease the channel utilization efficiency in high-speed optical networks.

A token based access scheme for a CC–FT^{W+1}–FR^{W+1} unidirectional WDM ring network, the *multitoken interarrival time (MTIT)* access protocol, was examined in [89, 90]. For each data channel, every node has one fixed-tuned transmitter, one fixed-tuned receiver, and one on-off optical switch, as shown in Fig. 3.14. The on-off switches are used to control the flow of optical signals through the ring and prevent re-circulation of the same packet on the ring. Once transmitted by the source node, the packet makes one round trip in the ring and is removed from the network by the same source node, i.e., MTIT employs source stripping. A separate wavelength is used as the control channel for the purpose of access control and ring management. The optical signal on the control channel is separately handled by an additional fixed-tuned transceiver.

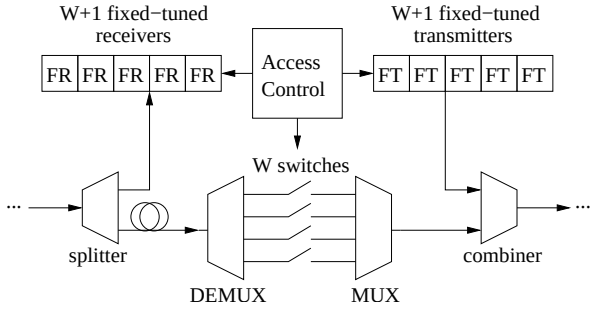


Figure 3.14: MTIT node architecture.

Channel access is regulated by a multitoken approach. Each channel is associated with one token that circulates among the nodes on the control channel and regulates the access to the corresponding data channel. The MTIT protocol controls the token holding time by means of a *target token interarrival time* with value $TTIT$. The $TTIT$ is agreed upon by all nodes connected to the ring at the configuration time of the system. The *token interarrival time* $TIAT$ is defined as the time elapsed between two consecutive token arrivals at the node. Upon a token arrival, the node is allowed to hold the token for a period of time equal to $TTIT - TIAT$. When the token holding time is up, the node must release the token as soon as the currently ongoing packet transmission is completed. A token can also be released earlier if no more packets are left in the node's transmission buffer. Note that concurrent transmissions on distinct channels are possible at the same node when two or more tokens are simultaneously held at the node.

With the FT^W–FR^W node structure, MTIT avoids receiver collisions and allows each node to simultaneously use multiple data wavelength channels. However, the number of transceivers at each node is rather large. MTIT achieves low access delay due to the fact that a node has the opportunity to grab a token more frequently than in conventional token rings where a node has to wait one round-trip time for the next token. A unique feature of MTIT is its capability to self-adjust the relative positions of tokens along the ring circumference and maintain an even distribution of the token position. As a result, the variance of the token inter-arrival time is low, guaranteeing to every node a consistent channel access delay in support of high-priority traffic. On the other hand, the capacity of MTIT is smaller than that of destination-stripping ring networks since source stripping does not allow for spatial wavelength reuse. For uniform traffic it was shown that MTIT achieves high bandwidth efficiency and low access delay for varying packet sizes even in relatively large (thousands of kilometers) networks.

Both bandwidth efficiency and access delay improve with the number of wavelengths used in the ring.

3.3.3 Meshed Rings

In unidirectional ring WDM networks with source stripping, packets are removed by the source node and each transmission requires a full circulation of the packet on the ring. The network capacity is limited by the aggregate capacity of all wavelengths. The network capacity of unidirectional ring networks can be increased with destination stripping where a transmission is propagating only on the ring segment between the corresponding pair of source and destination nodes. Due to spatial reuse, multiple simultaneous transmissions can take place on each wavelength. For uniform traffic, the mean distance between source and destination is half the ring circumference. As a consequence, two simultaneous transmissions can take place at each wavelength on average, resulting in a network capacity that is 200% as large as that of unidirectional rings with source stripping. In bidirectional rings, the network capacity can be further increased by means of *shortest path routing*, where a given packet is sent on that ring which provides the shortest distance to the corresponding destination. For uniform traffic the mean distance between source and destination is only a quarter of the ring circumference. Therefore, the aggregate capacity of bidirectional destination-stripping ring networks is increased by 400% compared to unidirectional source-stripping ring networks. The capacity of bidirectional ring WDM networks can be further increased by meshing the ring, which is discussed next.

The *Scalable Multi-channel Adaptable Ring Terabit Network (SMARTNet)* [91, 92, 93, 94] achieves a significant increase in the network capacity over the bidirectional destination-stripping ring by adding fiber short-cuts that connect certain nodes. More precisely, SMARTNet is based on a bidirectional slotted ring network with shortest path routing and destination stripping. Each node is connected to both rings and has a FT^W–FR^W structure, which allows a node to simultaneously transmit and receive data on different wavelengths. All wavelengths are divided into fixed-size slots whose length is equal to the transmission time of a fixed-size packet plus a header for indicating the slot status. Medium access is governed by means of an empty-slot protocol.

In addition to the N nodes, K equally spaced wavelength routers, each with four pairs of input/output ports, are deployed in the bidirectional ring. Wavelength routers are used to provide short-cuts in that data packets do not have to pass through the ring nodes that are between two interconnected routers. Specifically, two input/output ports of each wavelength router are used to insert the router into the bidirectional ring, the other two pairs of ports are used for creating bidirectional links (chords) to the two Mth neighboring routers. Routers $r_{[(k+M) \bmod K]}$ and $r_{[(k-M) \bmod K]}$ are said to be the Mth neighboring routers of router r_k on the ring, where $k = 0, 1, \ldots, K-1$. Fig. 3.15 depicts a meshed ring with $K = 6$ wavelength routers, each connected to its two $M = 2$nd neighboring routers.

Each wavelength router is characterized by a wavelength routing matrix that determines to which output port each wavelength from a given input port is routed. The wavelength routing matrix is chosen such that the average distance between each source-destination pair is minimized with a minimum number of required wavelengths. For example, an optimal set of wavelength paths for $K = 4$, $M = 2$, and $W = 3$ is shown in Fig. 3.16.

SMARTNet is able to significantly increase the capacity of a bidirectional ring network with shortest path routing and destination stripping. For uniform traffic it was shown that

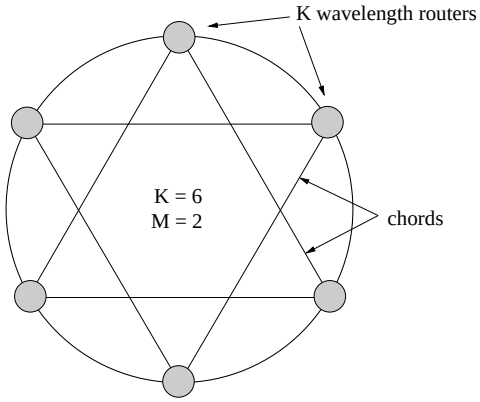


Figure 3.15: SMARTNet: Meshed ring with $K = 6$ wavelength routers, each connected to its $M = 2$nd neighboring routers.

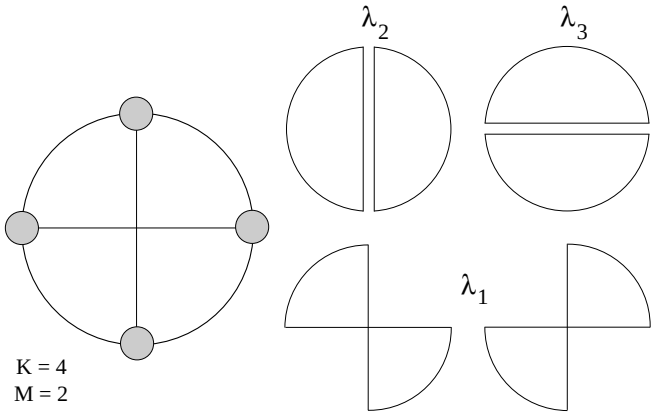


Figure 3.16: Wavelength paths in a meshed ring with $K = 4$ and $M = 2$, using $W = 3$ wavelengths.

a meshed ring with $K = 6$ wavelength routers and $M = 2$ increases the network capacity by 720% compared to unidirectional source stripping rings. Thus, the capacity of meshed rings is 80% larger than that achieved by non-meshed bidirectional ring networks with destination stripping at the expense of additional wavelength routers and chords which add to the network costs.

3.4 Fairness Control and QoS Support

Several of the aforementioned access protocols were extended in order to achieve fairness QoS support. In this section, we discuss these protocol extensions in greater detail. Note that fairness control and QoS support in RPR is discussed in Chapter 9.

3.4.1 Fairness Control

In general, the bandwidth of a network is shared by all nodes. Each node ready to send data should have the same opportunity to transmit data. As we have seen in the preceding section, most of the packet-switched ring WDM networks are based on a unidirectional ring; see Fig. 3.7. In this architecture, each wavelength can be considered a unidirectional bus terminating at a prescribed destination, as illustrated in Fig. 3.17. In an empty-slot access protocol, upstream nodes have a better-than-average chance to receive an empty slot for transmission, while downstream nodes have a worse-than-average chance. At heavy traffic this can lead to *starvation* of downstream nodes since they 'see' slots which are mostly used by upstream nodes. To avoid starvation, the transmission rate of nodes has to be controlled in order to achieve fairness among all nodes. However, restricting nodes in their transmission decreases the channel utilization. In general, there is a tradeoff between fairness and channel utilization.

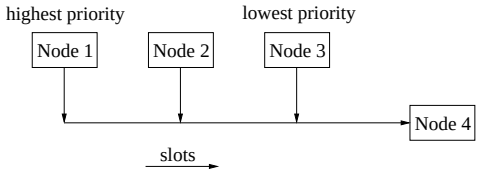


Figure 3.17: Medium access priorities in ring networks.

We now comprehensively survey the fairness mechanisms that have been developed for slotted ring networks.

MMR

Since SRR (see Section 3.3.1) is not able to enforce fairness, a fairness control algorithm is typically superimposed on SRR. The *Multi-MetaRing (MMR)* fairness algorithm is used on top of SRR in [95]. The MMR algorithm adapts a mechanism originally proposed for the MetaRing high-speed electronic metropolitan area network [39, 40, 96]. Fairness in the MetaRing is achieved by circulating a control message, named SAT (short for SATisfied). Each node is assigned a maximum number of packets to be transmitted between two SAT visits, this maximum number of packets is the node's *quota* or *credit*. Each node normally forwards the SAT message on the ring with no delay, unless it is not SATisfied in the sense that it has not transmitted the permitted number of packets since the last time it forwarded the SAT. The SAT is delayed at unSATisfied nodes until SATisfaction is obtained, i.e., either the node packet buffer is empty or the permitted number of packets has been transmitted.

In the *MMR Single SAT (MMR-SS)*, a single SAT message regulates the transmissions of all nodes on all wavelength channels. Each node can transmit up to K packets to each destination since the last SAT visit. Each SATisfied node forwards the SAT to the upstream node. Thus, the SAT logically rotates in the opposite direction with respect to data (although the physical propagation is co-directional). With this scheme the SAT propagation delays are very large since the SAT message has to traverse almost the entire network to reach the upstream node. Alternatively, the *MMR Multiple SAT (MMR-MS)* uses one SAT message for each wavelength. It was shown in [97] that this MMR-MS scheme is generally the preferable extension of the MetaRing fairness control scheme to a WDM ring.

M–ATMR

The access protocol of Section 3.3.1 suffers from fairness problems due to destination stripping. In [98] Bengi and van As adopted an extension of the well-established Asynchronous Transfer Mode Ring (ATMR) fairness protocol to the multiple channel WDM ring case; this extension is the *multi-channel ATMR (M-ATMR)*. In M-ATMR each node gets a prescribed number of transmission credits for each destination. When a node has used all its credits or has nothing to send, it transitions into the inactive state. In order to properly apply the credit reset mechanism, every node has to know which node was the last active node. To achieve this, each active node overwrites a so-called *busy address field* in the header of every incoming slot with its own address (the busy address field may be included into the SCM header of each WDM wavelength channel). Thus, a node receiving a slot with its own busy address knows that all the other nodes are inactive. If the last active node detects inactivity of all the other nodes, it generates a reset immediately after its own transmission. The reset mechanism causes the nodes to reset their credits to the predefined values. This way, it is guaranteed that every node uses a maximum number of slots between two subsequent reset cycles. It was shown in [98] that the M-ATMR fairness protocol applied for best-effort traffic provides throughput and delay fairness for both uniform and client/server traffic scenarios.

DQBR

The *Distributed Queue Bidirectional Ring (DQBR)* fairness protocol [78] for the control channel based HORNET of Section 3.3.1 is an adaptation of the DQDB protocol. The DQBR fairness protocol works as follows. In each control channel frame, a bit stream of length W bits, called the *request bit stream*, follows the wavelength-availability information. When a node on the network receives a packet in VOQ w, the node notifies the upstream nodes about the packet by setting bit w in the request bit stream in the control channel that travels upstream with respect to the direction the packet will travel. All upstream nodes take note of the requests by incrementing a counter called *request counter* (RC). Each node maintains a separate RC for each wavelength. Thus, if bit w in the request bit stream is set, RC w is incremented. Each time a packet arrives to VOQ w, the node stamps the value in RC w onto the packet and then clears the RC. The value of this stamp is called *wait counter* (WC). After the packet reaches the head of the VOQ, if the WC equals n it must allow n empty frames to pass by for downstream packets that were generated earlier. When an empty frame passes by the node on wavelength w, the WC for the packet at the head of VOQ w is decremented (if the WC equals zero, the RC w is decremented). Not until the WC equals zero can the packet be transmitted. The counting system ensures that the packets are sent in the order in which they arrived to the network.

The performance of the DQBR fairness control scheme was investigated for a 25-node HORNET network by means of simulation for two traffic scenarios in [78]. In the first traffic scenario, variable-size packet traffic was uniformly randomly generated by the nodes. The traffic generated for node 18 was 1.5 times the capacity of wavelength 18. It was demonstrated that with DQBR the throughput is equal for all nodes, whereas without DQBR the nodes close to node 18 have difficulties sending packets to node 18. In the second traffic scenario, unbalanced traffic was considered. Specifically, node 10 had 9.33 Gbit/s of traffic arriving to its queue destined for node 18, node 11 had 4.67 Gbit/s destined for node 18, and all other nodes had little traffic. The wavelength could support only 10 Gbit/s, so it was heavily overloaded. It

was found, that without DQBR the nodes close to node 18 are unable to transmit packets on wavelength 18, whereas with DQBR all nodes have an equal ratio of throughput to load for wavelength 18.

3.4.2 QoS Support

Many applications, e.g., multimedia traffic, require QoS with respect to throughput, delay, and jitter. To meet these requirements, networks typically provide different service classes, e.g., constant bit rate (CBR) and variable bit rate (VBR). In general, in WDM networks, traffic with stringent throughput, delay, and jitter requirements is supported by means of circuit switching via reservation of network resources, resulting in *guaranteed* QoS. On the other hand, to provide QoS to bursty traffic more efficiently, nodes process and forward packets with different priorities while benefiting from statistical multiplexing, leading to *statistical* QoS. In the following, we review different approaches for providing QoS in metro WDM ring networks.

SR^3

SRR with Reservations (SR^3) is derived from the SRR (see Section 3.3.1) and MMR (see Section 3.4.1) protocols and allows nodes to reserve slots, thereby achieving a stronger control on access delays [99]. The SR^3 protocol can be used in conjunction with SRR and MMR, requiring a marginal algorithmic complexity increase with no additional signaling messages.

In SR^3, time is subdivided into successive periods called *reservation frames*. Each reservation frame consists of P SRR frames. Each node can reserve up to P slots for a given destination per reservation frame, i.e., at most one slot per destination per SRR frame. Reservations are effective when all network nodes have become aware of the other nodes' reservations. SAT messages are used to broadcast the reservation information. Each SAT distributes information regarding current reservations on the channel it regulates. Each SAT contains a *Reservation Field* (SAT–RF) which is subdivided into $(N-1)$ subfields; each subfield is assigned to a particular node for reservations. If node i needs to reserve h, $1 \leq h \leq P$, slots per reservation frame on channel j, it waits until it receives the j–SAT; it then forwards the reservation request after properly setting the ith SAT–RF subfield to the value h. The j–SAT visits all nodes during the next tour of the multi-ring. By the time node i receives again the j–SAT, all nodes in the network are aware of the request of node i. Node i can thus update its reservation request on channel j every time it releases the j–SAT.

It was shown in [99] that SR^3 guarantees a throughput-fair access to each node. Moreover, the bandwidth left unused by guaranteed services can be shared by best-effort traffic very effectively. Even for the basic best-effort service, that requires no service guarantee, the reservation scheme can be very beneficial; the average and variability of the access delays are greatly reduced when slots are reserved, leading to improved performance and fairness. The reservation scheme can also be extended to *multiple* service classes. It was shown in [100], that in an unbalanced multiclass traffic scenario, a very good separation of the different traffic classes is obtained; the performance of higher-priority traffic is largely unaffected by lower-priority traffic, even when the latter one is grown to overload conditions.

Reservation Scheme for QoS Support

For QoS support in the WDM ring network of Section 3.3.1, a connection-oriented protocol based on connection setup and termination was proposed in [72]. In order to enable connection-oriented packet transmission for real-time services, the ring is subdivided into so-called *connection frames*. The real-time connections are established by reserving equally spaced slots within successive connection frames such that each destination node can be reached by a prescribed slot on the corresponding wavelength. Best-effort data traffic is transmitted in slots that are unreserved and empty. It was shown that this QoS approach is able to meet the delay requirements almost deterministically. Note that this scheme allows for reserving only one fixed-size slot per frame, i.e., only fixed-size packets are supported, similar to the SR^3 scheme.

A similar reservation scheme for providing QoS was presented in [72, 98]. In addition to the W normal VOQs each node has W real-time VOQs. Packets in the real-time VOQs are transmitted via connections in equally spaced, reserved slots. At each wavelength the ring is subdivided into frames each consisting of N/W slots, one slot per destination node receiving on that wavelength. A single reservation slot carries a connection setup and a connection termination field, each consisting of N bits on the subcarrier. When a node sets a bit in the setup field, the slot to the corresponding destination is reserved in each frame. After one circulation of the reservation slot, all nodes are aware of the reservation and the setup flag is cleared. All nodes keep track of the reservations by maintaining a table that is updated when the reservation slot passes. To free the reserved slots the same set/circulation/reset procedure is performed with the corresponding bit in the termination field.

MTIT - QoS With Lightpaths

The MTIT protocol of Section 3.3.2 can be extended to support not only packet switching but also circuit switching with guaranteed QoS [101]. The proposed solution allows for the all-optical transmission of packets with source stripping and circuits via a *tell-and-go* establishment of lightpaths (wavelength routes) with destination stripping. The lightpath establishment technique sets up a point-to-point connection between the source and the destination as follows. The on-off switches (see Fig. 3.14) at both the source and the destination corresponding to the lightpath wavelength are set in the off state. As a consequence, the data transmission is restricted to the ring segment between the source and destination nodes. This allows downstream nodes following the destination node to spatially reuse the wavelength channel.

Each node maintains a *Local Lightpath Table* (LLT) for all active lightpaths that is updated each time a token passes. A *Token Lightpath Table* (TLT) is transmitted with each token to broadcast the changes of lightpath deployment on the ring on the wavelength associated with the token. Each token consists of two lists, the so-called add-list for circuit setup and the so-called delete-list for circuit teardown. Specifically, a node holding a token can set up a lightpath to a destination node at the token's wavelength by making an entry in the add-list of the token. The path to the destination must not be occupied by another lightpath. A lightpath is torn down by the source by making an entry in the delete-list of the token. Assuming uniform traffic with Poisson arrivals and exponentially distributed message lengths it was analytically shown that an acceptable throughput/access delay performance can be achieved and that the achievable system throughput grows and access delay decreases as the

number of wavelengths increases.

3.5 Conclusions

In this chapter, we have provided an up-to-date overview of previous work addressing the metro gap with a focus on of packet-switched ring WDM networks. The current goal of the research on ring WDM networks is to develop designs that overcome the emerging metro gap between high-speed local clients (and networks) and the very-high-speed backbone networks. To overcome this metro gap, the ring networks need to efficiently use the wavelength resources, to be easily upgradeable (and scalable), and to flexibly support varying traffic loads and packet formats in a fair and cost-effective manner. For the networks we surveyed we attempt to give a qualitative assessment of how the developed networks address the metro gap issues and to outline open areas for future research efforts. Toward this end, in Table 3.5 we contrast the surveyed networks in terms of node structure, scalability, packet removal as well as support for variable-size packet fairness and QoS. We also consider the focus and perspective of the research efforts and the method of performance evaluation. For the HORNET and Bengi networks we consider the versions without and with control channel separately.

We see from the table that among the networks not having a control channel, the TT–FR node structure is most common. Indeed, this node structure is relatively simple and effective for unicast packet transmissions. For multicast traffic, which requires multiple transmissions on the different drop wavelengths of the fixed–tuned receivers, the FT^W structure has the advantage that these transmissions can be conducted simultaneously. With the TT structure, on the other hand, multiple sequential transmissions are required. As noted in the table, for a control channel based network, an FT–FR transceiver that is used exclusively for control is added. It may be worthwhile to investigate the cost-effectiveness tradeoffs between operating these dedicated control components and control wavelength channel, on the one hand, and conducting the control over the data transceiver and data channels, on the other hand. In conjunction with this question it may be of interest to explore whether the control channel and control components could be efficiently used to also carry some data traffic, e.g., multi- and broadcast data traffic that has to reach a large number of receivers, similar to control traffic. As we observe from the table the entire single fiber node structure is duplicated for the dual-fiber HORNET. An important direction for future research is to investigate effective protection strategies for such multi-fiber rings, as well as the scaling to additional rings for very-high capacity networks, e.g., similar to [102].

We see from the table that all protocol and concept oriented research efforts (as well as the HORNET testbed) allow for easy scalability in the number of nodes. The proof-of-concept testbeds MAWSON and RINGO, on the other hand, are at present limited to as many nodes as there are wavelength channels. There appears to be a need for more testbed activity on scalable networks.

All networks, except for the token ring network, allow for destination removal (stripping) and can thus exploit spatial wavelength reuse. Spatial wavelength reuse is not possible in the source stripping token ring. However, the source stripping in conjunction with token passing does have several advantages, such as easy support for fairness and QoS. Clearly the challenge for ring networks is to achieve the efficiency of spatial wavelength reuse while at the same time providing QoS and fairness for variable-size packets. As surveyed in this chapter and indicated in Table 3.5, a number of techniques have recently been developed to support some

combination of support for variable-size packets, fairness, and QoS in the different destination stripping networks and this area appears to continue to be a very active research area.

We note from Table 3.5 that the developed networks have been evaluated either by analysis, simulation, or experimentation, or a combination of analysis/simulation or simulation/experiment. There appears to be a need to complement the experiment (and experiment/simulation) evaluations with formal analysis, which may lead to fundamental insights that can enhance the considered testbed implementations. Similarly, it may be worthwhile to test the concept and protocol developments that have so far been evaluated by analysis and simulation in future testbeds.

	MAWSON	RINGO	SRR	HORNET
Research Focus	Testbed + Protocol	Testbed	Protocol	Testbed + Protocol
Special Feat.	Technically Simple	–	–	–
Node Structure	FT^W–FR	FT^W–FR, TT–FR	TT–FR	TT–FR
Scalability	$N = W$	$N = W$	$N \geq W$	$N \geq W$
Packet Removal	Dest.[1]	Dest.	Dest.	Dest.
Var. Packet Size	Reservation	–	–	Var. Size Slots
Fairness Control	Not Required	–	MMR	–
QoS Support	–	–	CBR + VBR	–
Perf. Evaluation	Sim.	–	Analy. + Sim.	Sim.
References	[60][61]	[48][49][63]	[65][66][97] [95][100][99]	[50][51][69]

	Bengi *et al.*	Jelger *et al.*	CC HORNET	CC Bengi
Research Focus	Protocol	Protocol	Testbed + Protocol	Protocol
Special Feat.	–	–	Bidirectional	–
Node Structure	HORNET	FT–FR^W, FT–TR	CC–TT^2/FT^2–FR^4	CC–TT/FT^2–FR^2
Scalability	$N \geq W$	$N \geq W$	$N \geq W$	$N \geq W$
Packet Removal	Dest.	Dest.	Dest.	Dest.[1]
Var. Packet Size	–	–	Reduced Fragment.	Reservation
Fairness Control	M-ATMR	–	DQBR	–
QoS Support	–	–	–	CBR
Perf. Evaluation	Analy. + Sim.	Analy. + Sim.	Sim.	Sim.
References	[72][73][98]	[76][74][75]	[78]	[81][82]

	Smiljanić *et al.*	Cho *et al.*	MTIT	SmartNET
Research Focus	Architec. + Prot.	Concept	Protocol	Concept
Special Feat.	Wavel. Stacking	Virtual Circles	Token Ring	Meshed Ring
Node Structure	CC–TT/FT–FR^2	CC–DWADM	CC–FT^{W+1}–FR^{W+1}	FT^W–FR^W
Scalability	$N \geq W$	$N \geq W$	$N \geq W$	$N \geq W = 5$
Packet Removal	Dest.	Dest.[1]	Source	Dest.
Var. Packet Size	–	–	Yes	–
Fairness Control	–	–	Not Required	–
QoS Support	CBR	–	CBR	–
Perf. Evaluation	Analy.	Sim.	Analy. + Sim.	Analy.
References	[84][85]	[88]	[89][90][101]	[91][92]

Table 3.1: Overview of surveyed packet-switched ring WDM networks. ([1]not explicitly addressed, but possible)

Chapter 4

Ring vs. Star Topology

CLEARY, key to overcoming the bandwidth bottleneck in metropolotan areas is the deployment if WDM technology. Single-channel optical systems cannot provide the huge capacity required to satisfy the increasingly higher bandwidth demands. As we have seen in Chapter 3, the majority of metro WDM systems is based in a ring topology. However, there is also a number of WDM systems with a star topology. Very little is known about the relative performance differences of WDM ring and star networks (the only performance comparison we are aware of the delay comparison between ring and bus networks [103]). To find out more about the specific strengths and shortcomings of each of the two approaches we conduct a comprehensive comparison of a state-of-the-art WDM ring network with a state-of-the-art WDM star network. In particular, we compare time-slotted WDM ring networks (both single-fiber and dual-fiber) with tunable-transmitter and fixed-receiver (TT–FR) nodes and an AWG (see Section 6.3.1) based single-hop star network with tunable-transmitter and tunable-receiver (TT–TR) nodes. We evaluate mean aggregate throughput, relative packet loss, and mean delay by means of simulation for Bernoulli and self-similar traffic models for unicast traffic with uniform and hot-spot traffic matrices as well as for multicast traffic. Our results quantify the fundamental performance characteristics of ring networks vs. star networks and vice versa, as well as their respective performance limiting bottlenecks. (Note that these insights are provide guidance for formulating the research question in Chapter 5.)

This chapter is structured as follows. In the following section we describe the architectures and MAC protocols of the considered single-fiber and dual-fiber ring networks. In Section 4.2, we describe the architecture and MAC protocol of the considered AWG based star network. In Section 4.3, we present our comparative simulations of ring and star WDM networks. We consider uniform and non-uniform traffic matrices for Bernoulli and self-similar traffic. We study the impact of fairness control on the performance of ring networks. We also compare the performance of ring and star networks for multicasting traffic. We summarize our findings in Section 4.4.

4.1 Slotted Ring WDM Network

When considering the various ring networks reviewed in Chapter 3, the most common WDM ring architecture seems to be a slotted ring based on the all-optical TT–FR node structure. Therefore, we chose this architecture as our representative ring WDM network. We consider both unidirectional and bidirectional rings and in the latter case the access protocol takes

advantage of shortest path routing.

4.1.1 Network Architecture

The ring connects N nodes and Λ wavelength channels are deployed. We consider initially the single-fiber network which connects all nodes with a single unidirectional fiber [72] and then the dual-fiber network which connects all nodes with two counter-directional fibers [78]. In the single-fiber network, the fiber bandwidth is divided into Λ wavelength channels. Each channel is divided into fixed-length time slots whose boundaries are synchronized across all wavelengths. The slot duration equals the transmission time of a fixed-size packet. (We note that variable-size packets can be accommodated on ring networks using for instance the mechanisms studied in [78].) Each node is equipped with one tunable transmitter and one fixed-tuned receiver (TT–FR). A node can send packets on any wavelength, while it is able to receive packets only on a preassigned *drop wavelength*. For $N = \Lambda$ each node has its own separate *home channel* for reception, as shown in Fig. 3.7 for $N = \Lambda = 4$. For $N > \Lambda$ each wavelength is shared by several nodes for the reception of packets. Specifically, the destination nodes $j = i + n \cdot \Lambda$ with $n \in \{0, 1, \ldots, \left\lceil \frac{N}{\Lambda} \right\rceil - 1\}$ share the same drop wavelength i, $i \in \{1, 2, \ldots, \Lambda\}$. Consequently, nodes sharing the same drop wavelength have to forward packets toward the destination node, resulting in *multihopping*. The destination node takes the packet from the ring (destination stripping). With this destination stripping, wavelengths can be *spatially reused* by downstream nodes, leading to an increased network capacity. To avoid HOL blocking each node deploys $(N - 1)$ VOQs, one for each destination node. Each VOQ holds up to B packets.

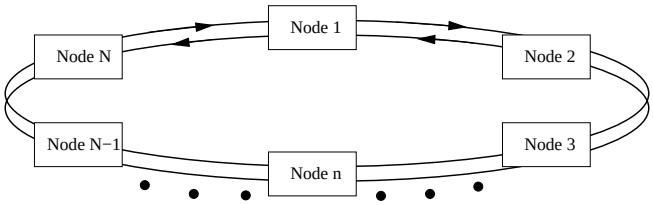


Figure 4.1: Dual-fiber ring network architecture.

In the dual-fiber ring network, the N nodes are interconnected with two counter-directional fiber rings, as illustrated in Fig. 4.1. To investigate the effect of the counter-directionality in the fibers we assign $\Lambda/2$ wavelength channels to each fiber, for a total of Λ channels, as in the single-fiber ring network. In a dual-fiber network the node structure of the single-fiber network is typically duplicated, i.e., there are one TT and one FR for each fiber [78]. The TT^2–FR^2 node structure allows a node to send and to receive two packets simultaneously. Each node has a home channel on each of the fibers.

4.1.2 MAC Protocol

In this section we outline the MAC protocol employed in the considered ring networks. To control the access of the nodes to the slots on the wavelength channels, every slot on each wavelength is accompanied by control information. This control information indicates whether the slot is empty or occupied by a data packet (wavelength availability information). If

the slot is occupied, the control information also gives the destination address of the packet occupying the slot. The control information may be transmitted on a separate control channel (as for instance in [78, 104, 105]) or in a subcarrier multiplexed header (as for instance in [106, 51]). We describe the principles of the control information transmission by discussing the control channel approach. For details on the subcarrier multiplexing approach we refer the interested reader to the corresponding references. With the control channel approach, the control information in a given slot on the control channel corresponds to the status of the data wavelength channels in the next slot. This is illustrated for the wavelength availability information (bits) in Fig. 4.2. (The destination node information is not shown to avoid overcrowding this illustrative figure.) If a bit is set to one the corresponding wavelength channel is occupied in the next slot, and otherwise it is empty.

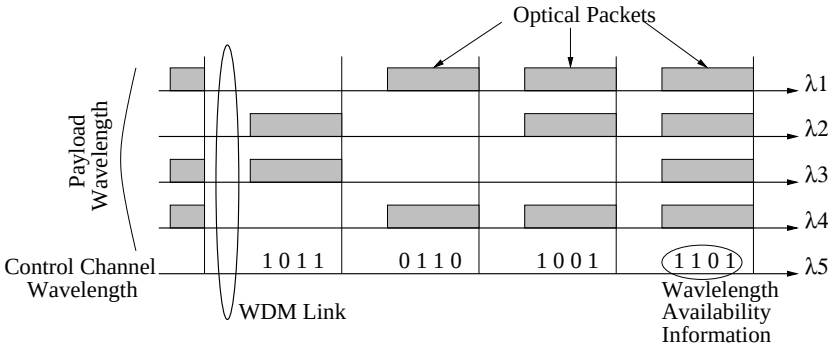


Figure 4.2: Control information transport on control channel: control information in a slot corresponds to data (payload) wavelength occupancy in next slot.

We note that with the control channel approach there is an additional wavelength channel for control on each fiber, for a combined total of $\Lambda + 1$ wavelength channels in the single-fiber ring network and $\Lambda + 2$ wavelength channels in the dual-fiber ring network. In addition, there is a fixed-tuned transceiver (FT–FR) at each node for each fiber to transmit and monitor the control information on the control channel on each fiber.

For the packet transmission, we consider the so-called *a posteriori* access strategy, which we outline at first for the single-fiber to highlight the main points. With the *a posteriori* access strategy, a node first checks the availability status of each slot on all wavelengths by inspecting the control information and then selects the appropriate VOQ. (This *a posteriori* strategy gives generally better performance at the expense of higher complexity compared with the *a priori* access strategy[72].) The node has to wait until an empty slot arrives on one (or more) wavelength channel(s). When an arriving slot is empty on one (or more) wavelength channel(s) the node can use this slot to transmit one packet from one of the corresponding VOQs. In the considered single-fiber ring architecture with per-destination VOQs, buffer selection is necessary if (*i*) $N = \Lambda$ and multiple channels have an empty slot, or if (*ii*) $N \geq \Lambda$ and at least one channel has an empty slots since a node can only transmit one packet at any given time with its single transmitter. Among various buffer selection schemes we choose the *longest queue (LQ)* selection scheme. With the LQ scheme, the longest VOQ (i.e., VOQ with the largest occupancy) is chosen. When there is a tie, the queue with the lowest index $j \in \{1, 2, \ldots, (N-1)\}$ is chosen. The motivation behind this LQ scheme is load balancing among the queues in the system, which increases the node and network throughput

at acceptable system complexity [72].

In the dual-fiber ring network, the packets are transmitted in similar fashion with the *a posteriori* access strategy. The two main adaptations of the access strategy outlined for the single-fiber network to the dual-fiber network are (*i*) that a node can transmit up to two packets simultaneously, and (*ii*) that a packet can be transmitted in either direction along the ring. Different strategies for choosing the direction are studied in Section 4.3.3.

We remark that the packet transmissions according to the wavelength availabilities require fast-tunable transmitters with a tuning time that is a small fraction of the slot duration, which is on the order of a few microseconds for typical scenarios, see 4.3.1. This requirement will be fulfilled by the recently reported fast-tunable transmitters with tuning times on the order of a few nanoseconds, see for instance [107, 108].

4.2 AWG Star WDM Network

Most star WDM networks for the metro area are based on the broadcast-and-select PSC (see Section 6.3.1) [109, 110]. Star WDM networks based on the wavelength-routing AWG have recently attracted attention both for metropolitan area networks [111, 112, 113, 114] and national-scale networks [115, 116]. It was shown in [117] that AWG based single-hop networks clearly outperform their PSC based counterparts in terms of throughput, delay, and packet loss due to spatial wavelength reuse. Therefore, in our comparison we consider a single-hop star WDM network that is based on a wavelength-routing AWG. For extensive surveys on physical star networks, the interested reader is referred to [58, 118, 119, 120].

4.2.1 Network Architecture

As shown in Fig. 4.3, the star network is based on a $D \times D$ AWG used as a wavelength-routing device. A wavelength-insensitive $S \times 1$ combiner is attached to each AWG input port and a wavelength-insensitive $1 \times S$ splitter is attached to each AWG output port. Thus, the network connects $N = D \cdot S$ nodes. Each node is equipped with a laser diode (LD) and a photodiode (PD) for data transmission and reception, respectively. Both data transmitter and receiver are tunable over Λ wavelengths which are not preassigned to nodes (TT–TR). Similar to the ring network, each node has $(N-1)$ VOQs, one for each destination. Again, each VOQ holds up to B packets. The number of available wavelengths Λ span R adjacent FSRs of the underlying AWG, each FSR consists of D contiguous wavelength channels, i.e., $\Lambda = R \cdot D$, as illustrated in Fig. 4.4. Note that the AWG allows for spatial wavelength reuse. As a result, the Λ wavelengths can be simultaneously applied at each of D AWG input ports, for a total of $D \cdot \Lambda$ wavelength channels connecting the D AWG input ports with the D AWG output ports. Also, note that there are R wavelength channels connecting each AWG input-output port pair.

The MAC protocol makes use of a control channel to broadcast control information. This control channel can be implemented (*i*) as an inband control channel by exploiting the spectral slicing of a broadband light source in conjunction with spectrum spreading of the control signals [111], or as an out-of-band control channel, e.g., by running a PSC in parallel to the AWG [121]. In our explanation of the basic principles of the MAC protocol in the AWG star we focus on the out-of-band control channel; we refer the interested reader to [111] for details on the inband control channel. We only note here in brief that the capacity of the inband control channel is limited to a few Mbit/s due to the physical limitations of the employed broadband light source and spectrum spreading. For the out-of-band control channel, each node can be

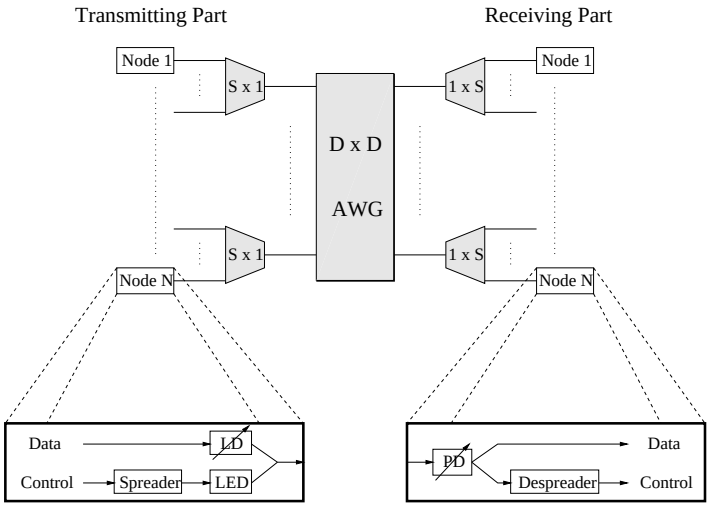


Figure 4.3: Architecture of AWG based star WDM network.

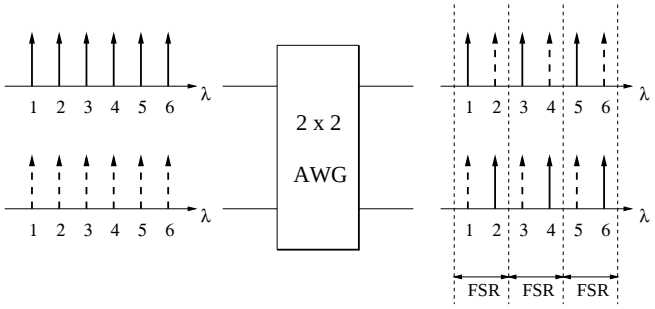


Figure 4.4: Illustration of wavelength routing in AWG with $D = 2$ input and output ports when $R = 2$ FSRs are used. Each FSR provides one wavelength channel between an input-output port pair. A total of $D \cdot D \cdot R = D \cdot \Lambda = 2 \cdot 4$ wavelength channels connect the input ports to the output ports.

connected via an additional fixed-tuned transceiver and fiber pair to a PSC which is operated in parallel with the AWG. This requires more hardware than the inband signaling approach but the control channel capacity is much higher. Additionally, the PSC part of the network can be used for protection of the single point of failure of star networks, as studied in [121].

4.2.2 MAC Protocol

In the considered MAC protocol, wavelengths are assigned dynamically *on demand* such that any pair of nodes is able to communicate in one *single hop*. The applied MAC protocol is an *attempt-and-defer reservation* protocol, i.e., a data packet is sent after a successful reservation, which is conducted with a control packet. The reservation protocol avoids both channel and receiver collisions of data packets. We consider a MAC protocol with data packet aggregation. That is, a single control packet makes a reservation for all (fixed-size) data packets that are destined to a given destination node (i.e., are buffered in the corresponding VOQ), thus forming *variable-size* data packet aggregates. The data packet aggregate is kept in the VOQ until it has been transmitted.

With out-of-band signaling, the basic time unit on the control channel is the control slot. The length of the control slot is equal to the time required to send a control packet over the out-of-band control channel. For our comparisons in this chapter we consider the following packet transmission strategy with data packet aggregation. Each node employs the LQ buffer selection scheme to determine for which VOQ to send control packets. More specifically, in each control slot, each node selects the VOQ with the largest number of unscheduled packets, forms a data packet aggregate from the packets, and prepares a corresponding control packet. The control slots for control packet transmission are not fixed assigned. Instead, the control packets are sent on a contention basis using slotted ALOHA. Specifically, a given node sends its prepared control packet with probability $1/N$ in a given control slot.

After the one-way end-to-end propagation delay (i.e., half the end-to-end round-trip time) a transmitted control packet is received and collected by every node (including the sending node). This allows each node to maintain *global knowledge* of all the other nodes' activities and a node also learns whether its own control packet collided in the control packet contention or not. All nodes periodically process the successfully received control packets by executing the same first-come-first-served and first-fit scheduling algorithm, which we adopt since scheduling in very-high-speed optical networks must have low complexity. The scheduling algorithm tries to schedule the variable-size data packet aggregates within the scheduling window of pre-specified length. Note that all the nodes need to execute the scheduling algorithm on the collected control packets at the same time to ensure that all nodes compute the same transmission schedule and preserve global knowledge about the ongoing data packet aggregate transmissions.

If the control packet collided in the control packet contention, or the scheduling of the data packet within the scheduling window fails, the source node retransmits the control packet in the next control slot, provided the corresponding VOQ is still the longest VOQ. Also, note that VOQs for which a control packet is currently on its way (so it is not yet known whether it will be successful or not) are not considered in the LQ selection.

4.3 Performance Comparison

In this section, we conduct a detailed quantitative comparison between the state-of-the-art ring and star metro WDM networks described in the preceding two sections. In our performance comparison we focus on the packet level performance metrics, i.e., throughput, packet delay, and packet loss, which we define in Subsection 4.3.1, in which we also describe our simulation set-up.

Prior to proceeding to our detailed investigations of the packet level performance we briefly note that the ring and star networks have specific advantages and pose specific challenges at the photonic level and for implementation. We briefly review the photonic level issues arising from transmission impairments and insertion losses, which are important considerations for the choice of network topology. In particular, in ring networks, the insertion losses of the wavelength multiplexers and demultiplexers (which are typically based on AWGs) used in all-optical node architectures may limit the power budget and thereby the number of nodes that can be traversed without signal regeneration [122, 123]. Furthermore, ring networks are affected by ASE noise, which may accumulate over long all-optical paths with multiple amplifiers. This accumulated ASE noise along with other impairments, such as fiber nonlinearities and crosstalk, may significantly degrade the signal quality, see for instance [124]. Techniques

to mitigate these effects are under development; the ASE noise accumulation, for instance, can be reduced by employing variable optical attenuators [125]. Also, the signal regeneration at a ring node forwarding traffic to other nodes on its drop wavelength can overcome these limitations. In the single-hop star network, on the other hand, the AWG is passed only once between each pair of source and destination nodes. Thus, the insertion loss of the AWG and the transmission impairments do typically not severely restrict the scale of the network, and allow in fact for a national scale single-hop network [126, 127].

Another critical issue for the operation of packet-switched WDM networks is synchronization at the slot level. Ring networks allow for relatively simple synchronization even at extremely high data rates, see for instance[100]. In star networks, on the other hand, the slot synchronization is more challenging due to the distributed nature of the network nodes and the possibly different distances of the nodes to the hub of the star network. Techniques to achieve synchronization in PSC based star networks have been studied extensively, see for instance [128, 129, 130][131, Sec. 7.2.1], and can be extended to the AWG based star network in a straightforward manner.

4.3.1 Simulation Set-up and Performance Metrics

By default we consider typical metro networks interconnecting $N = 64$ nodes that are equidistant to each other on the circumference of a ring with a diameter of 91.67 km. The parameters used in both networks are summarized in Table 4.1, those specific to the star are listed in Table 4.2. The nodes are interconnected by a ring WDM network or a star WDM network with $\Lambda = 8$ wavelength channels. (In the dual-ring network there are four wavelength channels on each fiber plus two control wavelength channels, one on each fiber, resulting in a total of 10 wavelength channels.) Each wavelength channel operates at a line rate of 2.5 Gbit/s (OC-48) and the propagation speed on the optical fiber link is set to $2 \cdot 10^5$ km/s. The packet size is fixed to 1500 byte, which is one of the dominant packet sizes in the Internet as well as the maximum packet size (maximum transfer unit (MTU)) of Ethernet. The corresponding slot duration is 4.8 μs (1500 byte/2.5 Gbit/s). For the star network we set the size of a control packet to 2 byte (which is sufficient to accommodate source and destination address and length of data aggregate) and the speed of the out-of-band control channel to 333 Mbit/s which is easily feasible.

Bernoulli and self-similar traffic are considered. In both cases, the average packet generation rate at each given node is σ, $0 \leq \sigma \leq 1$. More precisely, at a given node in each slot a new packet is independently generated for each of the other $(N-1)$ nodes with probability $\sigma/(N-1)$. A newly generated packet is put in the corresponding VOQ of the destination node (or dropped if the VOQ is full). Similarly, for each of the destination VOQs, self-similar packet traffic with Hurst parameter 0.75 is generated from ON/OFF processes with Pareto distributed on-duration and geometrically distributed off-duration [132]. For both types of traffic the N nodes in the network generate on average $N \cdot \sigma$ packets per slot. In addition to the uniform traffic, where a packet generated by a given node is destined to any one of the other $(N-1)$ nodes with equal probability $1/(N-1)$, we consider non-uniform hot-spot traffic in Section 4.3.4.

In our performance evaluation we consider the mean aggregate throughput, the relative packet loss, as well as the mean packet delay.

- The *mean aggregate throughput* is defined as the mean number of source node transmitters sending in the network in steady state. (Note that in the dual-fiber network

Description	Symbol	Default Value
Network Diameter	Δ	91.67 km
Number of Nodes	N	64
Number of Wavelengths	Λ	8
Data Rate	C	2.5 Gbit/s
Propagation Speed	c	$2 \cdot 10^5$ km/s
Packet (Slot) Size	L	1500 byte
Slot Duration	–	4.8 μs
VOQ Size (per Dest.)	B	64 Packets

Table 4.1: Network parameters: Default values for both ring and star network.

Description	Symbol	Default Value
AWG Degree	D	8
Splitter/Combiner Degree	S	8
Number of Used FSRs	R	1
Control Packet Size	–	2 byte
Control Channel Speed (Out-of-Band)	–	333 Mbit/s
Control Slot Duration	–	48 ns
Scheduling Window Size	–	200 Slots

Table 4.2: Parameters specific to star: Default values.

a source node can transmit up to two packet simultaneously as opposed to the star network with at most one active transmitter per source node.) We also study the mean throughput for individual source-destination node pairs, which equals the probability that the considered source node is transmitting a packet to the considered destination node in steady state. (Note that packets forwarded by intermediate nodes along the ring do not count towards the measured throughput; only the transmission of the original source node contributes to the throughput.)

- The *relative packet loss* in the network is defined as the ratio of the total number of dropped packets and the total number of generated packets in the network. For some scenarios we also study the relative packet loss of individual source-destination node pairs, which is the ratio of the total number of dropped packets of a given source-destination node pair and the total number of packets generated for the source-destination node pair.
- We define the *mean packet delay* in the network as the time period elapsed from the generation of a packet to the complete reception of the packet in ms in steady state.

We estimate the defined performance metrics from discrete event simulations. Each simulation was run for 10^6 slots (including a warm-up phase of 10^5 slots). For Bernoulli traffic we obtained 95% confidence intervals on the performance metrics using the method of batch means. The 95% confidence intervals are too small to be seen in the figures.

4.3.2 Fairness Control in Ring Network

Due to the ring symmetry and the applied destination release, each node has a better-than-average access to channels leading to certain destination nodes and a worse-than-average

access to channels leading to other destinations [65]. Spatial reuse may cause *starvation*, which occurs when a node is constantly being covered by up-stream ring traffic and thus is not able to access the ring for very long periods of time [133]. This fairness problem has received considerable attention in the literature [78, 65, 72, 134, 135].

In this comparative study, for both single-fiber and dual-fiber ring networks we use the fairness control described in [72] which is a modified form of ATMR [135] (see Section 3.4.1). (We note that a modified fairness control of DQDB called DQBR in [78] for dual-fiber ring networks. However, the DQBR scheme can not be directly employed in the considered network which uses destination stripping with spatial wavelength reuse and $N > \Lambda$.) The used fairness control represents a credit allocation scheme and provides fair channel access by means of a distributed credit mechanism and a cyclic reset scheme based on a monitoring approach. The fairness control algorithm works as follows. Initially, each node is allocated a predefined credit, referred to as *window size* W, and is set to the active state. The node status (active or inactive) for a channel is included in a so-called *busy address field* in the control information sent on the control channel. Each node decreases the window size whenever it uses a free slot to send a packet. If the node is still in the active state, i.e., if the window size is larger than zero, the node sets the busy address field to the node's address. When the window size reaches zero, the node changes its state to the inactive state, i.e., the node is not allowed to send any data using the wavelength and leaves the busy address field unchanged. Thus, all nodes in the network can see which nodes are in the active state. If a node receives a slot with busy address field set to the node itself, the node knows that all other nodes are in the inactive state. The node then immediately sends a reset message to all other nodes by setting the so-called *reset-request field* in the control information on the control channel and resets its window size to the predefined window size W. The node sends the reset message only once and waits for the reset message to circulate around the entire ring network. When the reset message is received by the node which sent the reset message, the message is stripped from the ring. When a node receives a reset-request, the node sets its status to the active state, sets the window size for the channel to the predefined window size W and forwards the reset-request. This algorithm is invoked on all Λ channels at each node.

The window size W specifies the credit/quota of usable slots and determines the duration of the activity cycles. Thus, W represents the main parameter of the fairness control. Fig. 4.5 depicts the throughput performance of the single-fiber ring network for uniform self-similar traffic without fairness control as well as with fairness control with different window size $W = \{50,\ 300,\ 500,\ 700,\ 1000\}$.

We observe and re-confirm the well-known trade-off between (throughput) fairness and aggregate network performance, i.e., fairness control degrades the aggregate throughput performance of the network. We observe that a medium window size $W \in \{300,\ 500\}$ achieves the largest mean aggregate throughput. Whereas choosing a small or large W leads to a reduced throughput performance. This reduced performance is mainly due to interruptions in the transmissions caused by very frequent or infrequent consumption of the complete quota, as detailed in [136].

Unless otherwise noted, we employ for all the following simulations the ATMR fairness control with a window size of $W = 500$ in the single-fiber ring network, and a window size of $W = 1000$ in the dual-fiber ring network (which was found to give good performance for this larger window size, see [136] for details).

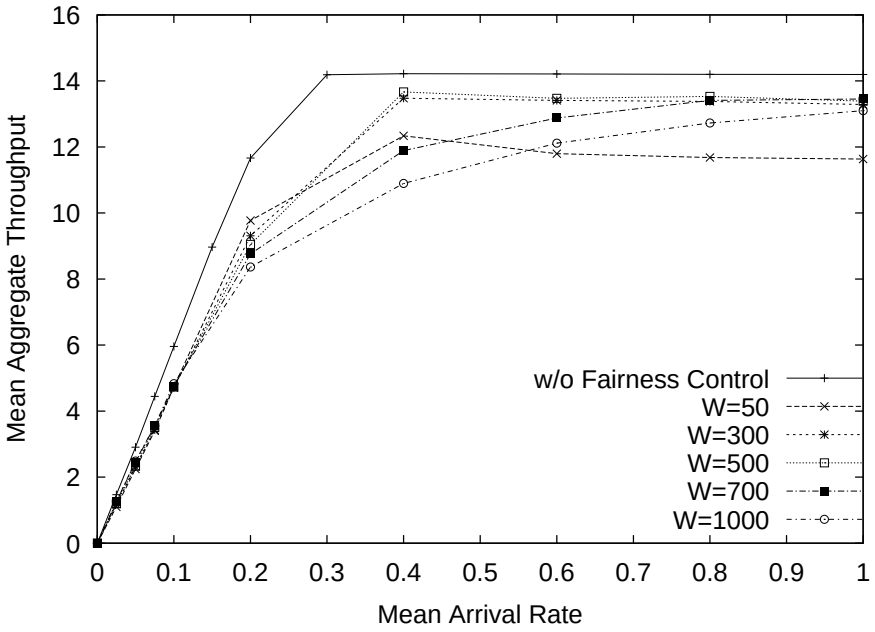


Figure 4.5: Mean aggregate throughput of single-fiber ring network for uniform self-similar traffic with $W = \{50,\ 300,\ 500,\ 700,\ 1000\}$.

4.3.3 Uniform (Balanced) Traffic Scenario

We first compare the performance of the ring and star networks for uniform (balanced) traffic, where a given source node sends a generated packet to any of the remaining $(N-1)$ nodes with equal probability $1/(N-1)$. We consider both Bernoulli and self-similar traffic.

In the dual-fiber ring network each packet can be sent on either of the two counter-directional rings. Two algorithms for choosing the direction are compared. With the first algorithm (Alg. 1) a packet is sent in the first empty slot that appears on either of the two rings. In the second algorithm (Alg. 2) the packet is sent on that ring which provides the smaller hop distance to the corresponding destination. We observe from Fig. 4.6 that with Alg. 1 the throughput improves only very slightly compared to a single-fiber ring, while Alg. 2 roughly doubles the number of concurrent transmissions at medium to high traffic loads. This is because with Alg. 2 the mean hop distance is reduced by 50% and the spatial wavelength reuse is increased by a factor of two. We have also found that Alg. 2 gives smaller delays than Alg. 1, see [136] for details. Based on these findings, we use Alg. 2 in the remainder of the chapter.

Clearly, in the single-hop star network the mean hop distance is minimum (unity). The degree of spatial wavelength reuse is controlled by the AWG degree D since all Λ wavelengths can be used at each port simultaneously, as illustrated in Fig. 4.4. Therefore, the star network with $D = 8$, $S = 8$, and $R = 1$ is able to achieve about twice the throughput of the setup with $D = 4$, $S = 16$, and $R = 2$ which is also reflected by the results of the simulation in Fig. 4.8. Note that for full connectivity the maximum AWG degree is limited by $D \leq \Lambda$, resulting in an upper bound on the spatial wavelength reuse. Moreover, to fully exploit the capacity of the AWG at least $S \geq \Lambda$ nodes have to be attached to each port (cf. Figs. 4.3

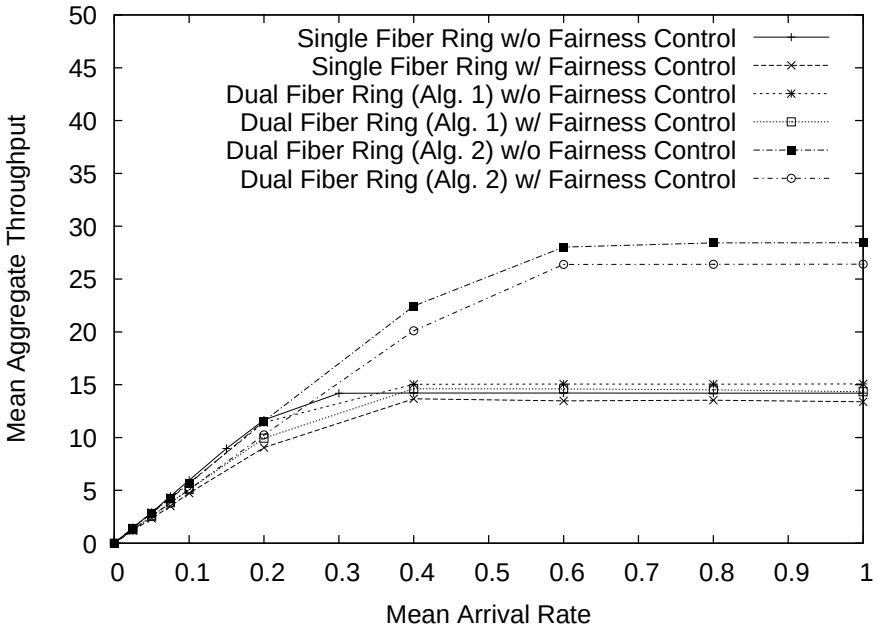


Figure 4.6: Mean aggregate throughput of ring networks for uniform self-similar traffic.

and 4.4). When comparing the out-of-band signaling with the inband signaling approach (for which we consider a physically feasible speed of 3.33 Mbit/s, see [136] for details) it turns out that the latter one severely limits the performance of the star network. The capacity of the inband channel does not suffice to (re)transmit all control packets leading to a much smaller throughput and to a significantly increased delay as shown in Figs. 4.8 and 4.12 (the delay with inband signaling quickly shoots up and levels out around 380 ms, which extends beyond the delay range shown in Fig. 4.12). In contrast, the external control channel provides sufficient capacity, leading to only few collisions and to aggregates consisting mostly of only a single packet. In the following sections only the star with external control channel based on an AWG of degree $D = 8$ is considered.

When comparing the results of the star with those of the ring in Figs. 4.7–4.12 the latter one is clearly outperformed in terms of all measured performance metrics. The difference in the throughput of the networks, shown in Figs. 4.7 and 4.8, reflect the theoretical capacity limits of $2 \cdot \Lambda = 16$ in the single-fiber ring, $4 \cdot \Lambda = 32$ in the bidirectional ring, and $D \cdot \Lambda = 64$ in the star. In the star, up to about $\sigma = 0.8$ nearly all packets are scheduled by the first control packet leading to a small delay, depicted in Figs. 4.11 and 4.12, which is close to the theoretical minimum of two times the propagation delay of the ring diameter and nearly no packet is lost. The packet loss shown in Figs. 4.9 and 4.10 also illustrates the difference between Bernoulli and self-similar traffic. The latter one is bursty and leads to VOQ overflows even if the network is in principle able to handle the offered amount of traffic and the VOQs are relatively short. As the networks saturate the difference between the two traffic models vanishes. In the following we only consider the more realistic self-similar traffic model.

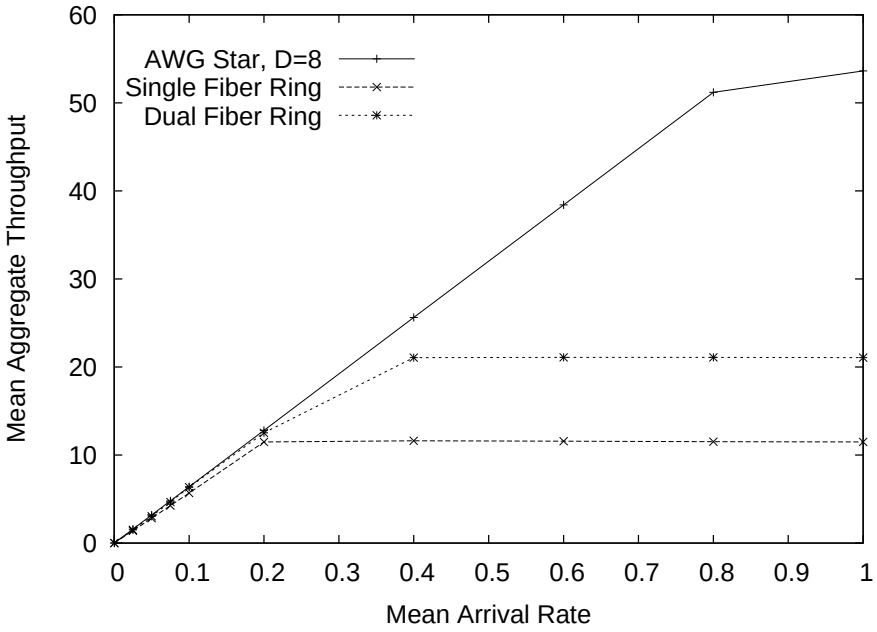


Figure 4.7: Mean aggregate throughput of star and ring networks for uniform Bernoulli traffic.

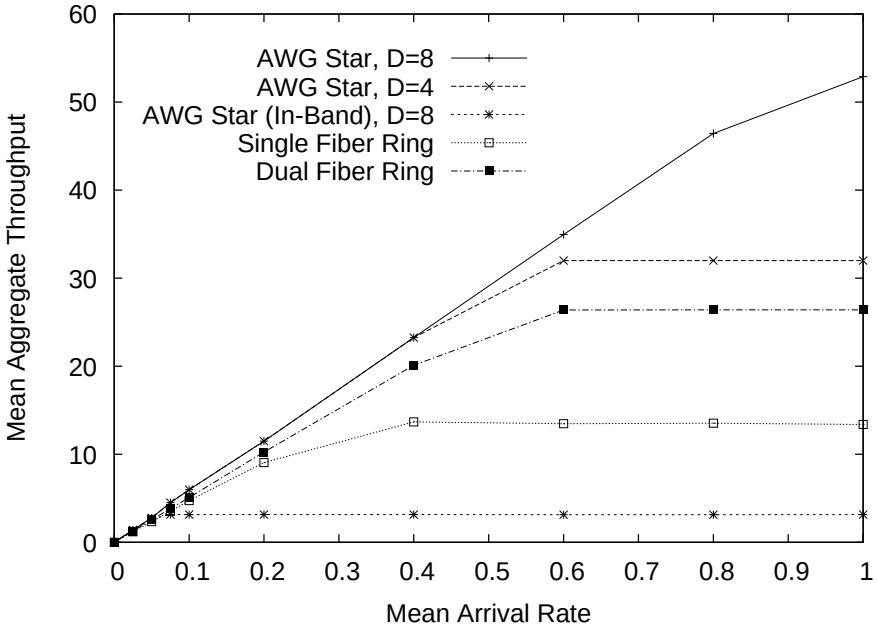


Figure 4.8: Mean aggregate throughput of star and ring networks for uniform self-similar traffic.

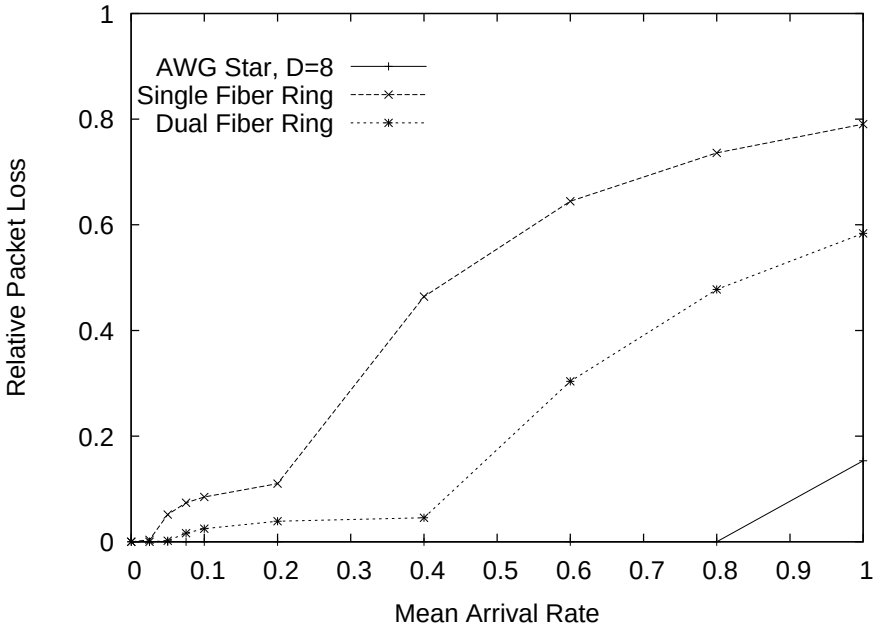


Figure 4.9: Relative packet loss of star and ring networks for uniform Bernoulli traffic.

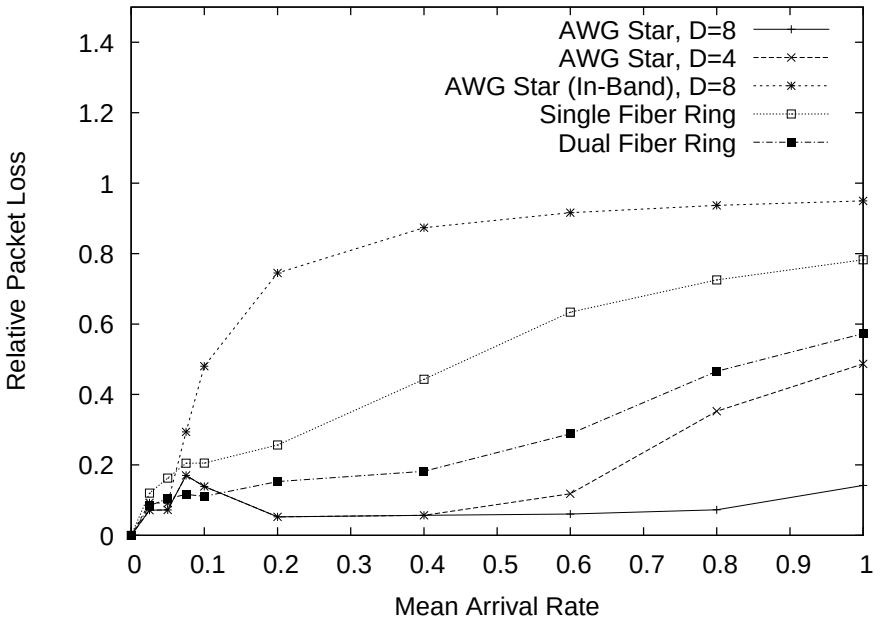


Figure 4.10: Relative packet loss of star and ring networks for uniform self-similar traffic.

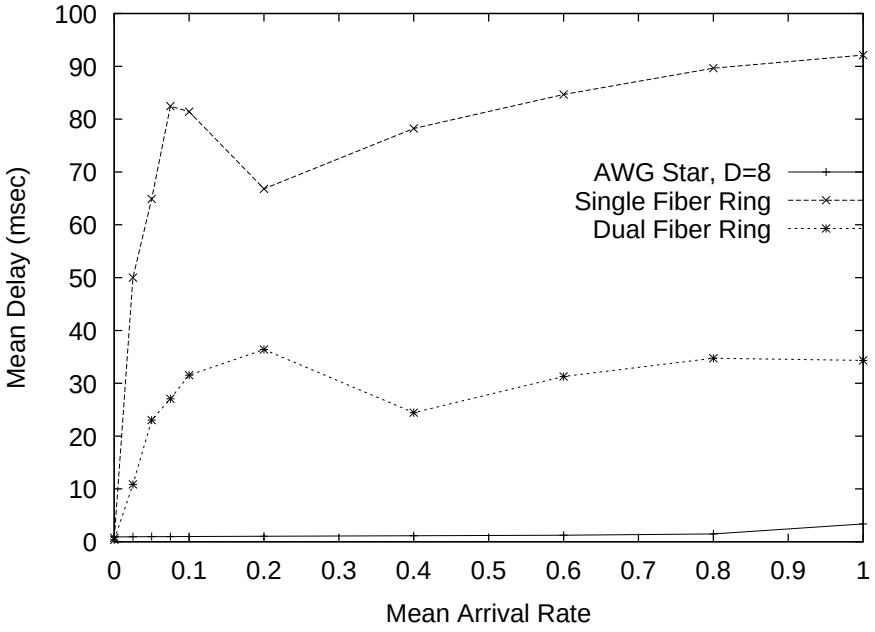


Figure 4.11: Mean delay of star and ring networks for uniform Bernoulli traffic.

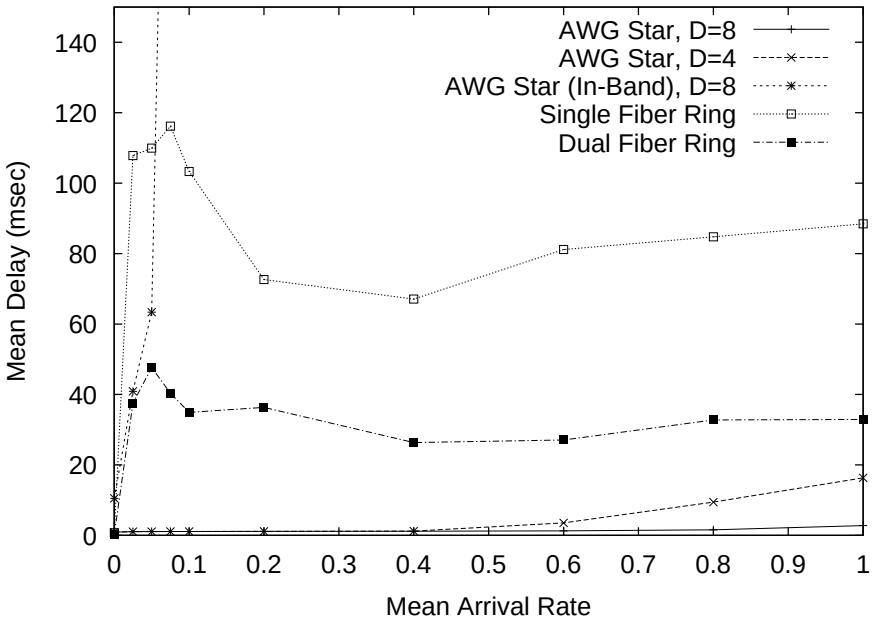


Figure 4.12: Mean delay of star and ring networks for uniform self-similar traffic.

4.3.4 Non-uniform (Unbalanced) Traffic Scenario

We now focus on self-similar traffic and compare the ring and star network performance for a non-uniform (unbalanced) traffic mix consisting of a uniform traffic component and a client-server traffic component. Specifically, we assume to have one hot-spot (either server or POP), while the remaining $(N-1)$ nodes act as identical clients. A client sends a fraction h of the traffic to the hot-spot, while the remaining fraction $(1-h)$ of the traffic is equally distributed among the other $(N-2)$ clients. Note that $h = 1/(N-1)$ corresponds to uniform traffic only as discussed above. We assume that the server generates as much traffic as all $(N-1)$ clients together and set $\sigma = 0.4$. Similar to the uniform traffic scenario, the total load offered to the network is $N \cdot \sigma$. In the ring networks, the ATMR fairness control is employed as discussed above.

The performance results are shown in Figs. 4.13, 4.14, and 4.15. As h increases the throughput steadily decreases. For $h = 1$, the throughput of the star network is roughly equal to two, which corresponds to the one transmitter plus one receiver in the hot-spot. The throughput in the ring networks for $h = 1$ is only roughly half the combined number of transmitters and receivers in the hot-spot (two in single-fiber ring network and four in dual-fiber ring network). This is due to a degeneration of the ATMR fairness control for $h = 1$ which can be overcome by a modified fairness control, see [136] for details. Interestingly, we observe that for moderate h values in the range from approximately 0.5 to 0.8, the throughput in the single-fiber ring network is relatively close to the throughput in the dual-fiber network. We also observe that with an increasing fraction h of hotspot traffic, the relative packet loss increases steadily. This is mainly due to the limited capacity of the hot-spot's transceiver(s), which results in most of the packets destined to or originating from the server to be lost.

We observe from Fig. 4.15 that there are marked differences in the delay. In the star, the data aggregates corresponding to hot-spot traffic are mostly of the maximum size and experience a large delay. However, the client-to-client traffic is still transported efficiently via numerous data aggregates, mostly of the minimum size. Therefore, the delay does not increase significantly until there is rarely any more client-to-client communication. In the ring, client-to-client traffic, except that on the hot-spot's home channel, does not experience an increased delay. In contrast, any packet transmitted on the congested home channel of the hot-spot and packets originating from the hot-spot (transmitter bottleneck) are additionally delayed. Therefore, as the fraction of hot-spot traffic increases the delay also increases. Note that for $h = 1.0$, the delay in the ring networks is smaller than in the case of uniform traffic (Fig. 4.12), which is due to the degeneration of the ATMR fairness control for $h = 1$, see [136] for details.

Next, we fix $h = 0.3$ and study fairness among the individual source-destination node pairs, with node 1 functioning as server. As shown in Fig. 4.16, the star network provides throughput fairness among all clients due to the random control packet contention and the first-come-first-served scheduling. The server achieves a larger mean throughput which is desirable since it has much more data to send than the clients. In contrast, Fig. 4.17 re-confirms the fairness problems in ring networks. The hot-spot, node 1, leaves most slots at its drop-wavelength empty. The succeeding nodes use these slots and achieve a high throughput to the hot-spot at the expense of the nodes further downstream, for which no capacity is left. Only nodes downstream to the nodes which share the same drop wavelength with the hot spot, get a chance to send to the server using free slots. The source-destination pairwise metrics in the single-fiber ring network with fairness control are shown in Figs. 4.18 and 4.19.

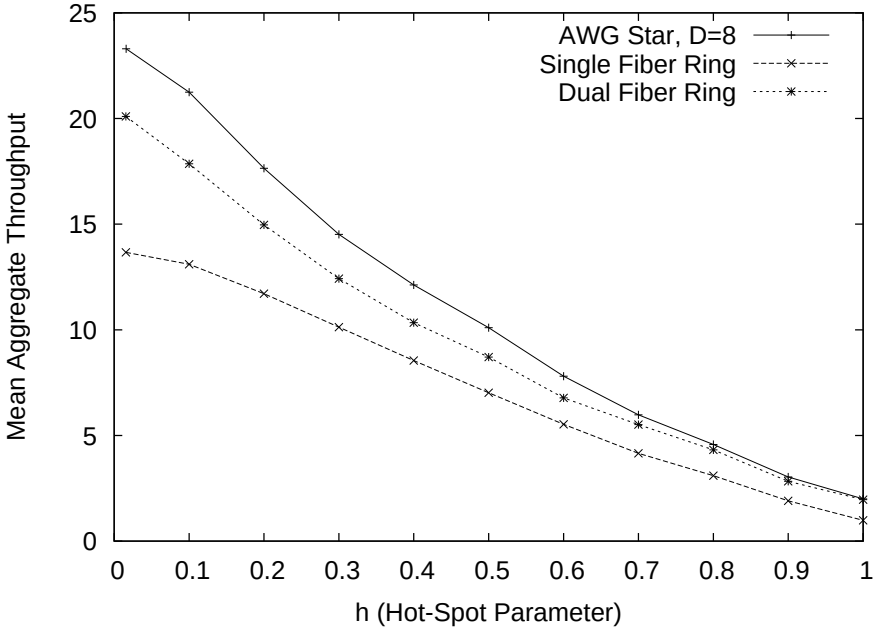


Figure 4.13: Mean aggregate throughput as a function of the fraction of hot-spot traffic h with $\sigma = 0.4$, fixed

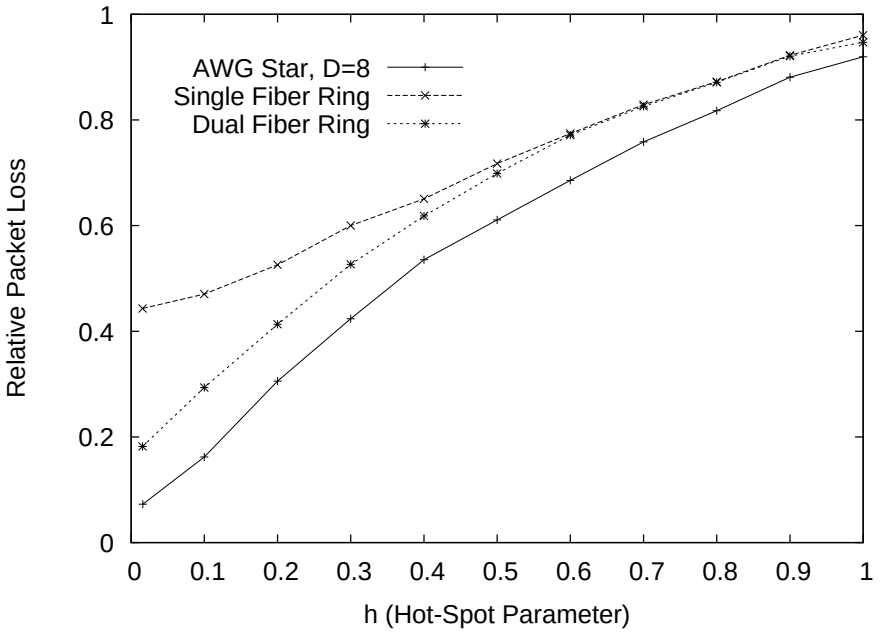


Figure 4.14: Relative packet loss as a function of the fraction of hot-spot traffic h with $\sigma = 0.4$, fixed

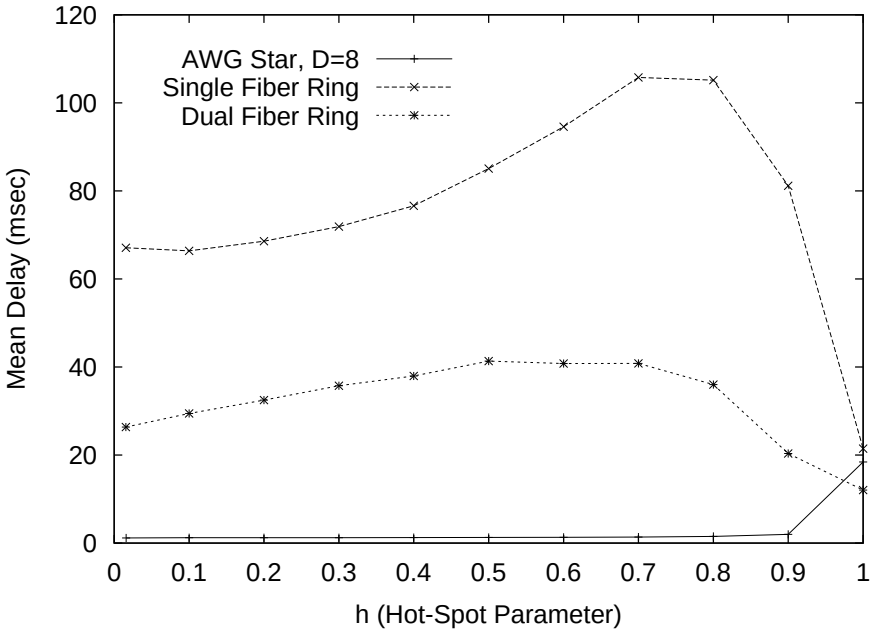


Figure 4.15: Mean delay as a function of the fraction of hot-spot traffic h with $\sigma = 0.4$, fixed

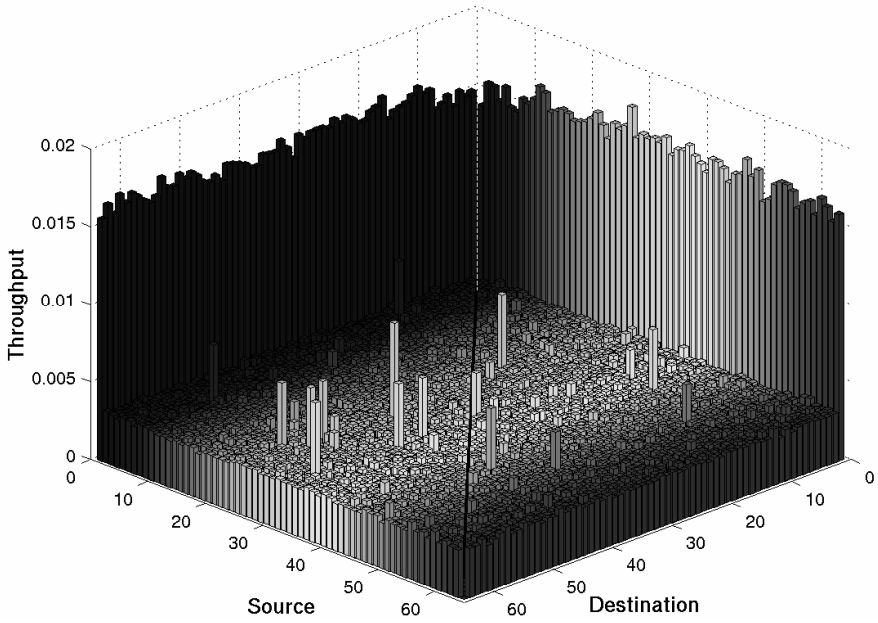


Figure 4.16: Pairwise mean aggregate throughput of AWG star network for self-similar hot-spot traffic with $\sigma = 0.4$ and $h = 0.3$

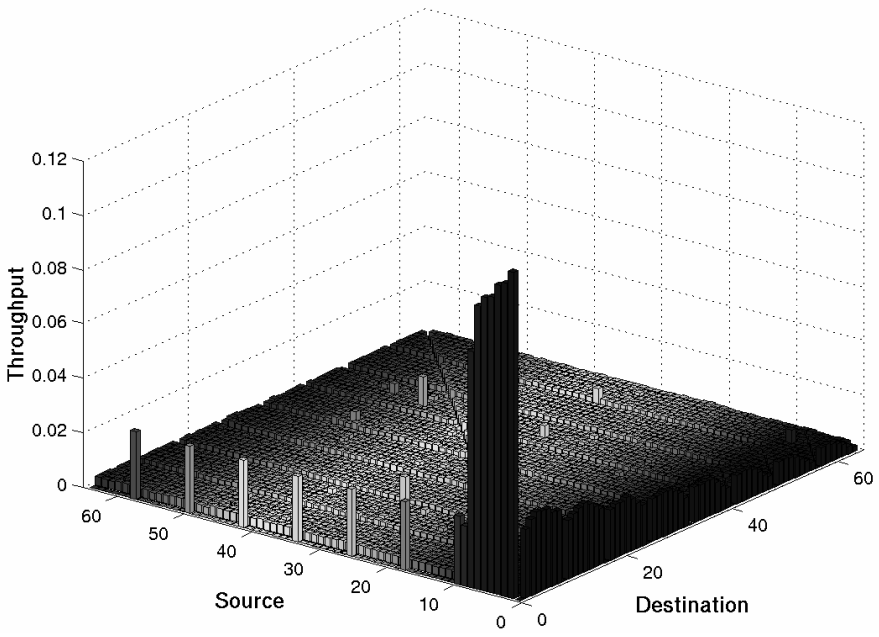


Figure 4.17: Pairwise mean aggregate throughput of single-fiber ring network without fairness control for self-similar hot-spot traffic with $\sigma = 0.4$ and $h = 0.3$

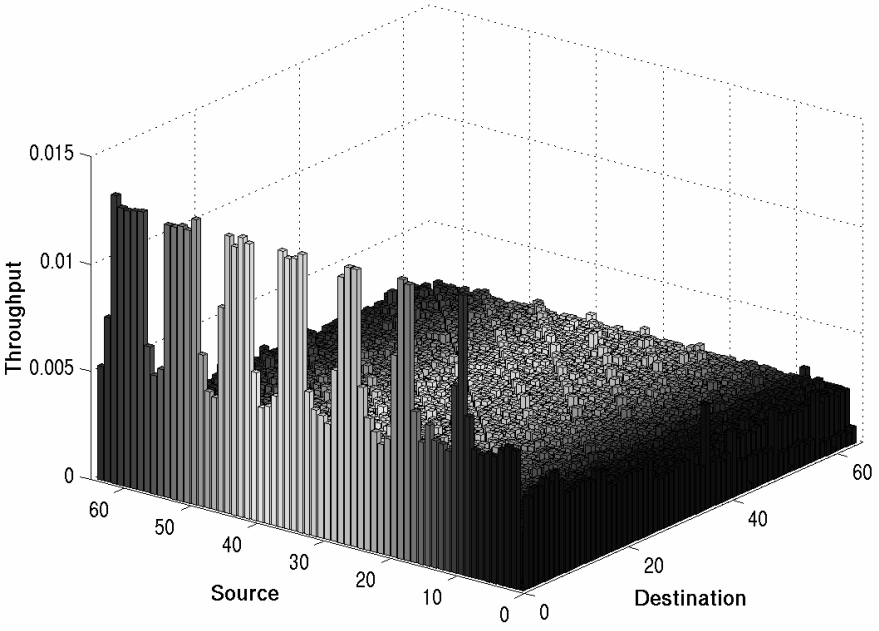


Figure 4.18: Pairwise mean aggregate throughput of single-fiber ring network with fairness control for self-similar hot-spot traffic with $\sigma = 0.4$ and $h = 0.3$

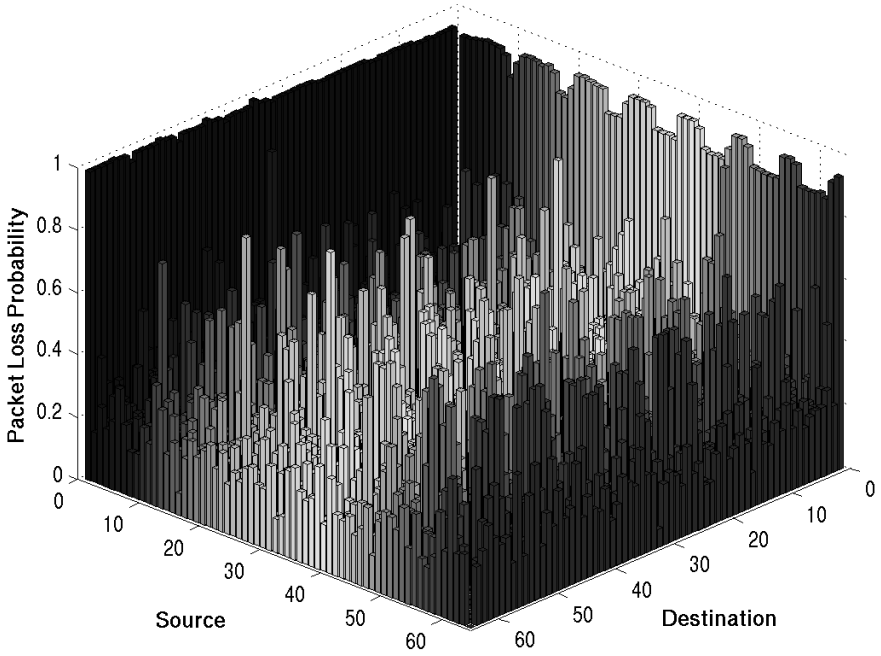


Figure 4.19: Pairwise packet loss probability of single-fiber ring network with fairness control for self-similar hot-spot traffic with $\sigma = 0.4$ and $h = 0.3$

Compared to Fig. 4.17, the applied fairness control scheme balances the network.

4.3.5 Multicast Traffic

In this section we compare the performance of the ring and star WDM networks for multicast (multi-destination) traffic. Multicast traffic is expected to account for a significant portion of the traffic in future metro WDM networks due to applications such as video conferencing, tele-medicine, distributed games, and content distribution. Multicasting in ring network has received relatively little attention [48, 137], similarly, multicasting in the AWG based metro WDM network has received only limited attention so far [138]. In this section we first outline the modifications to the node architectures and MAC protocols in the ring and star network to accommodate a mix of multicast and unicast traffic. We then define the considered performance metrics for this traffic mix and present the results of our comparisons.

In both the single-fiber ring and the star network, each node is equipped with Λ buffers (each of capacity B packets) for multicast traffic (in the dual-fiber network each node has $\Lambda/2$ multicast buffers), in addition to the $N-1$ buffers (each of capacity B packets) for unicast traffic. The multicast buffers are operated as follows. First recall that in the single-fiber ring network each wavelength is the home channel for N/Λ nodes (in the dual-fiber ring network each wavelength on a given fiber is the home channel of $2 \cdot N/\Lambda$ nodes). Thus a packet transmission on a given wavelength can reach up to N/Λ destination nodes of a multicast in the single-fiber ring network (or up to $2 \cdot N/\Lambda$ nodes in the dual-fiber network). Now, one of the Λ multicast buffers is assigned to each of the Λ wavelengths in the single-fiber ring network.

(In the dual-fiber ring network one multicast buffer is assigned to each of the $\Lambda/2$ home channels.) The destination nodes of a given multicast are partitioned into up to Λ groups in the single-fiber ring (up to $\Lambda/2$ groups in the dual-fiber ring) according to the different home channels of the destination nodes. A copy of the multicast packet is generated for each group of destination nodes and placed in the corresponding multicast buffer. If all nodes of a given multicast share the same home channel, then only one packet copy is generated and placed in the corresponding multicast buffer. If a multicast has destination nodes on each of the home channels, then Λ packet copies are generated, and one each is placed in the Λ multicast buffers in the single-fiber network (in the dual-fiber network $\Lambda/2$ copies are generated and placed). In the star network one of the Λ multicast buffers in a given node is assigned to each of the Λ wavelengths or equivalently the $\Lambda = D$ destination splitters (for the considered scenario with $R = 1$, generally, if $R \geq 1$ then R multicast buffers would be assigned to each destination splitter). The destination nodes of a multicast are partitioned according to the splitters that the destination nodes are attached to. If all destination nodes are attached to the same splitter, then one packet copy is placed in the corresponding multicast buffer. If a multicast has destinations at all splitters, then Λ packet copies are generated and one each is placed in the Λ multicast buffers.

The MAC protocol for multicast packet (copies) works as follows. The addresses of the intended destination nodes of a given multicast packet copy on a given home channel are included in the control information corresponding to the packet. Each node monitors its home channel as described in Section 4.1.2. When a node receives a data packet, it checks whether there are additional destinations downstream. If so, the node forwards the packet to the downstream nodes. If the node is the last destination, then it takes the packet off the ring. To keep the delays small for multicasts, which inherently use the bandwidth more efficiently than unicasts, we send the multicast packet copies using Alg. 1 (see Section 4.3.3) in the direction that has the first vacant slot. We employ the LQ buffer selection in both ring and star network. We count one for a unicast packet. For a given multicast packet copy we count the number of intended destination nodes that it will reach, i.e., up to N/Λ on the single-fiber ring and star networks, and up to $2 \cdot N/\Lambda$ on the dual-fiber ring network. This counting scheme tends to give a multicast packet copy higher priority according to the number of destination that it reaches.

In the star network, multicast packet copies are not combined into aggregates and the scheduler schedules a multicast packet copy transmission only if all intended destination nodes at the respective splitter are free.

In the performance evaluation for mixed unicast and multicast traffic, we consider the following performance metrics.

- The *mean aggregate receiver throughput* is defined as the number of receivers that are receiving a packet destined to them in steady state.
- The *mean aggregate transmitter throughput* is defined as for unicast traffic in Section 4.3.1.
- The *mean aggregate multicast throughput* is defined as the mean number of multicast completions per slot. The multicast throughput is equal to the ratio of the mean transmitter throughput to the mean number of packet copy transmissions required to reach all intended destination nodes of a given multicast packet. The multicast throughput thus measures the multicast efficiency of each packet copy transmission.
- The *mean packet delay* is defined similar to the unicast traffic scenario in Section 4.3.1.

For a multicast packet, however, we consider the individual delays until the complete reception of the packet by the individual receivers. The individual delays experienced until a given multicast packet is received by its destination nodes are all individually counted when evaluating the mean packet delay.

- The *relative packet loss* is defined as for unicast traffic, with the differences that (*i*) a generated multicast packet with m destination nodes counts as m generated packets, and (*ii*) a multicast packet copy destined to n destination nodes that finds its multicast buffer full and is dropped counts as n dropped packets.

In our simulations for mixed unicast and multicast traffic we consider uniform self-similar traffic with a fraction p_m of multicast traffic. More specifically, each node generates new packets as in the case of unicast traffic. With probability p_m a given newly generated packet becomes a multicast packet. For a given multicast packet the number of destination nodes is drawn independently randomly from a uniform distribution over $[1, N-1]$, and the destination nodes are distributed uniformly randomly over the other $N-1$ nodes. For the simulations reported here the fraction of multicast traffic is set to $p_m = 0.3$. The window size for fairness control is set as for unicast traffic to $W = 500$ in the single-fiber network and $W = 1000$ for the dual-fiber network. Figs. 4.20, 4.21, and 4.22 give the mean aggregate receiver, transmitter, and multicast throughput as a function of the mean arrival rate σ. A number of observation are in order. First, we observe that the dual-fiber ring network achieves close to twice the receiver throughput of the single-fiber ring. This is because a packet copy transmitted on a given wavelength in the dual-fiber ring network can reach up to twice the number of nodes compared to a packet copy transmitted on the single-fiber ring. Hence the difference in receiver throughput between single-fiber and dual-fiber ring despite both having roughly the same transmitter throughput and multicast throughput. We also observe that in both ring networks the transmitter throughput and multicast throughput have a slight peak around $\sigma = 0.2$ and then level off as the traffic load is increased further. At the same time the receiver throughput continues to increase. This is because the LQ buffer selection policy tends to give priority to multicast packets, especially at increasing network loads when the buffers tend to get filled (and packet loss becomes large, see Fig. 4.23). To see this, recall that we count the number of destination nodes of the multicast packet copies in the LQ buffer selection, as is natural and reasonable. In the single-fiber ring network each multicast packet copy is destined on average to $N/(2 \cdot \Lambda) = 4$ destination nodes ($2 \cdot N/(2 \cdot \Lambda) = 8$ in the dual-fiber network). Thus a unicast packet buffer completely filled with B packets has about the same priority in the buffer selection as a multicast buffer filled to a quarter of its capacity in the single-fiber network (one eighth in the dual-fiber network). As the multicast buffers are filled up to higher levels they are given priority over unicast packet buffers. More specifically, priority is given to the multicast buffer holding the packet copies with the largest number of destinations.

For the star network, on the other hand, we observe that transmitter and receiver throughput as well as multicast throughput continue to increase as the traffic load increases. The receiver throughput, however, stays well below the levels reached by the dual-fiber ring. This is due to the combined dynamics of buffer selection and data packet scheduling in the star network. Similar to the ring network, the multicast packet copies are given priority in the LQ selection of the VOQs for which control packets are sent. The multicast data packet copies, however, are more difficult to schedule than unicast packets, as the copies require all intended receivers at a given splitter to be free in the same slot. Therefore, the scheduling of multicast

packets becomes difficult, especially as the traffic load increases. As a consequence, at high traffic loads the star network tends to transmit unicast packets (and a few multicast packet copies with a small number of destination nodes). (In ongoing work we are addressing this bias against multicast packets, one strategy is to partition the intended receivers at a splitter into subgroups.) The unicast packets are transmitted with moderate delay as they tend to experience some delay until they are selected in the buffer selection, but are then quickly scheduled. The few multicast packet copies that do eventually succeed in the scheduling, however, experience a very large delay. As a result the average delays are significantly larger for the mix of unicast and multicast traffic in the star network (see Fig. 4.24) compared to the delays for unicast traffic (see Fig. 4.12).

In contrast, we observe that the delays for the mix of unicast and multicast traffic in the ring networks (see Fig. 4.24) are smaller than the corresponding delays for unicast traffic (see Fig. 4.12). This is because multicast packet copies (especially those with many destinations) are given priority in the buffer selection as explained above and thus tend to experience relatively small queue build-up and queuing delays. At the same time, multicast packets with many destinations have a large impact on the average packet delay in the network, resulting in the small delays observed in Fig. 4.24.

Overall, we observed from Fig. 4.24 that the dual-fiber ring network gives the smallest delays. The delays in the star network are roughly twice as large for a wide range of traffic loads. In the single-fiber ring network there is a hump in the delay for moderate traffic loads (with roughly $0.2 \leq \sigma \leq 0.6$), which is due to the transmission of unicast packets that have experienced relatively large delays, see [136] for details.

Generally, we may conclude that for multicast traffic the ring networks have the advantage that there are no receiver conflicts. A multicast packet copy that is transmitted in an empty slot is delivered to all its intended destinations around the ring without requiring any coordination of the receivers. In contrast in the TT–TR star network, destination conflicts tend to make the scheduling of multicast packet copies difficult. As we have observed in this section, the combination of these effects results in a significantly improved performance of the dual-fiber ring network over the star network. Throughout this section we focused on the aggregate network performance for mixed unicast and multicast traffic. As we noted in the interpretations of our results, unicast and multicast traffic experience different dynamics in the considered networks with the ring networks having a bias in favor of multicast traffic and the star network having a bias in favor of unicast traffic. In ongoing work we study the fair treatment of these traffic types.

4.4 Conclusions

We have compared the performance of state-of-the-art WDM star and WDM ring metro networks. We considered an AWG based star network with a TT–TR node architecture, as well as the all-optical single-fiber ring with TT–FR nodes and the counter-directional dual-fiber ring with a TT^2–FT^2 node structure. In addition, fixed-tuned transceivers (FT–FR) are used in the ring networks for the transmission of control information over the control channel and for the out-of-band signaling in the AWG star network. We considered WDM ring networks with a slotted time structure and with the ATMR fairness control.

Our main finding is that the AWG star network with out-of-band signaling clearly outperforms the ring networks in terms of throughput, packet loss, and delay for unicast traffic. In

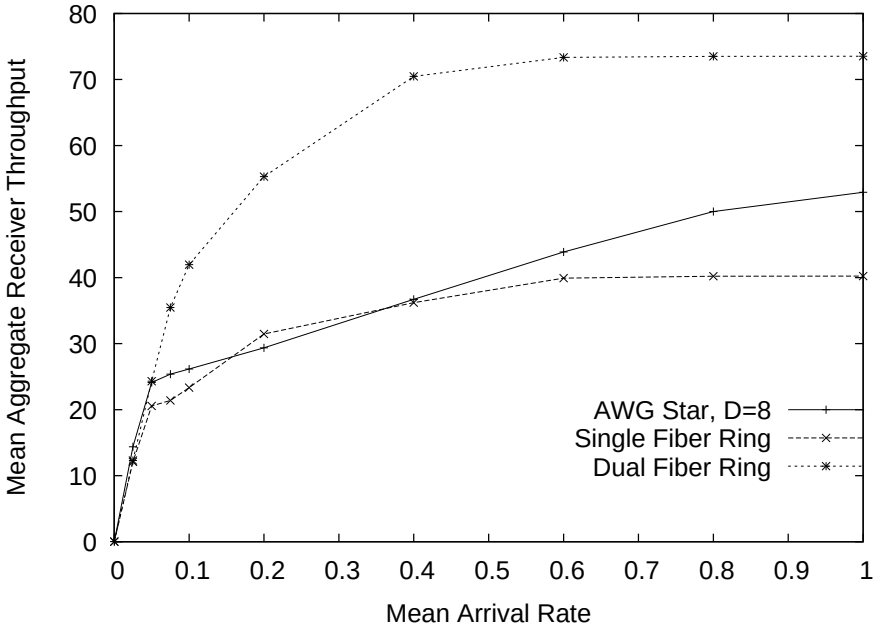


Figure 4.20: Aggregate receiver throughput for uniform self-similar traffic with $p_m = 30\%$ multicast traffic

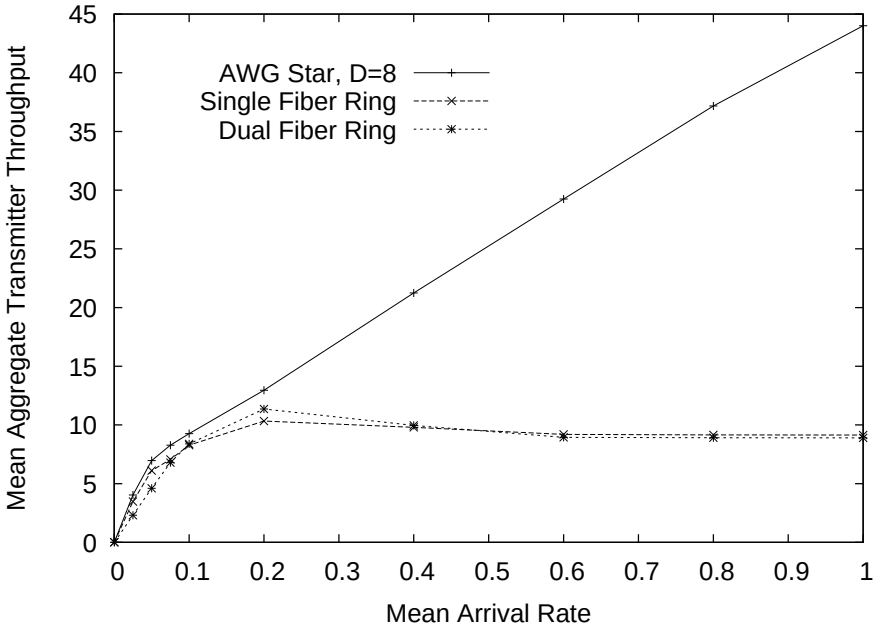


Figure 4.21: Aggregate transmitter throughput for uniform self-similar traffic with $p_m = 30\%$ multicast traffic

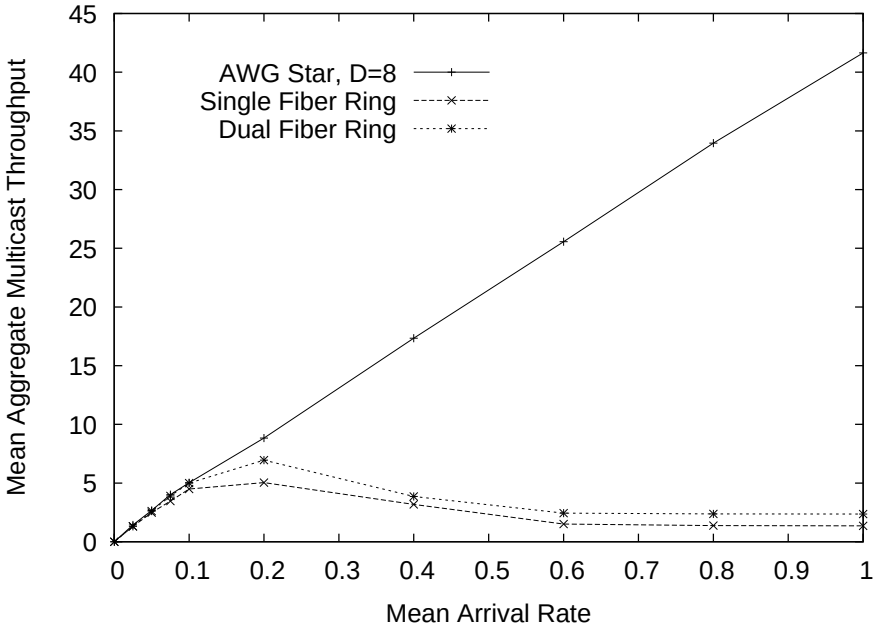


Figure 4.22: Aggregate multicast throughput for uniform self-similar traffic with $p_m = 30\%$ multicast traffic

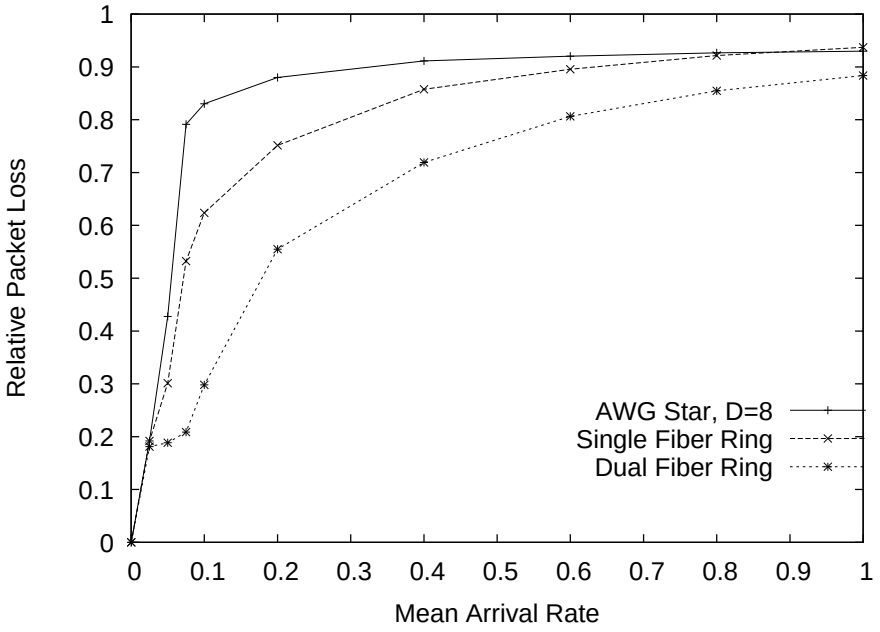


Figure 4.23: Relative packet loss for uniform self-similar traffic with $p_m = 30\%$ multicast traffic

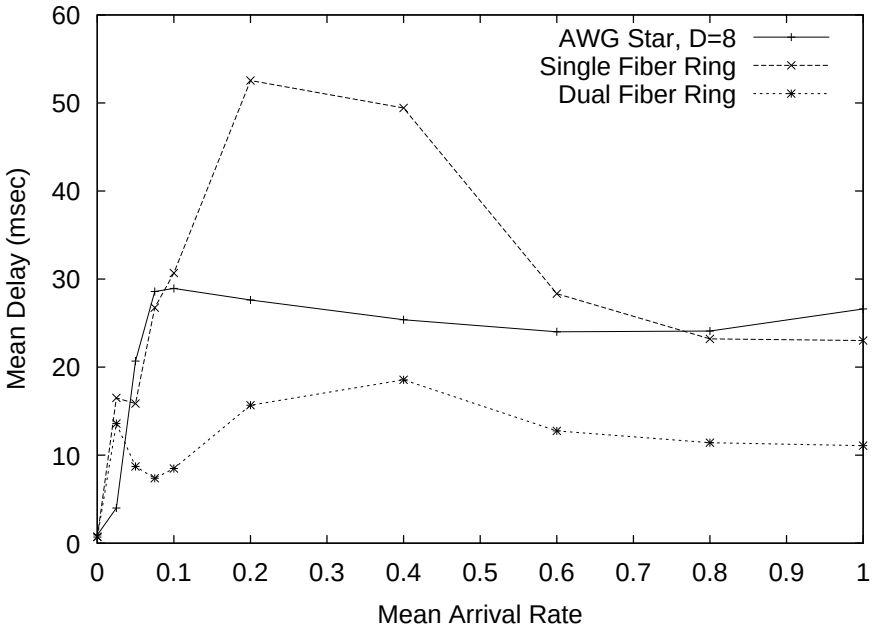


Figure 4.24: Mean aggregate delay for uniform self-similar traffic with $p_m = 30\%$ multicast traffic

addition, the star's reservation protocol naturally provides fairness and support for variable-size packets. However, the dual-fiber ring network outperforms the AWG based star network for multicast traffic. As an architecture being relatively new to the metro area, future research should focus on solving the technological challenges of the AWG based star network to get closer to the market. For example, fast-tunable receivers are not yet commercially available, a cost efficient external control channel is required, and cost-effective protection strategies need to be developed.

All-optical WDM ring networks, on the other hand, have proven to be technically feasible in various testbeds. The dual fiber infrastructure from existing SONET/SDH based solutions can probably be reused for a cost-effective upgrade to the all-optical dual fiber ring, also featuring protection. Furthermore, techniques for accommodating variable-size packets in WDM ring networks are being studied, see for instance [78], and comparing them with the transport of variable-size packets in the star networks is an interesting avenue for future work. To be competitive with the star in terms of performance, future research should focus on developing new architectures allowing for a higher degree of spatial wavelength reuse.

In hot-spot traffic scenarios both the ring and the star's performance is mostly limited by the capacity of the hot-spot's transceivers. This requires special attention in the design of the access protocol and the node structure.

Chapter 5

Motivation of Our Approach

IN this chapter we combine the insights gained in the previous chapters and formulate the research question tackled in the remainder of this work. At the beginning of Chapter 3 we identified three main developments addressing the metro gap, namely (*i*) Data over SONET/SDH (DoS), (*ii*) IEEE 802.17 Resilient Packet Ring (RPR), and (*iii*) all-optical WDM rings. In Chapter 4 we additionally introduced (*iv*) single-hop WDM star networks. We evaluate the potential to overcome the problems in the metro area of each of these approaches by discussing the performance with respect to the metro requirements defined in Chapter 1. As we summarize and compare our findings for each approach, it turns out that packet-switched ring networks, especially RPR, fulfill the requirements relatively well but suffer from performance problems, as evaluated in detail in Chapter 4. This leads to the idea of developing *a performance enhancing technique for optical packet-switched ring networks.*

WDM star networks, on the other hand, provide huge capacity but suffer from a number of practical limitations. Overall, when considering the metro requirements, many strengths of RPR seem to be weaknesses of WDM star networks and vice versa. Therefore, we argue that the combination of a packet switched ring network with a WDM single-hop star network, i.e., a hybrid ring-star architecture where the star is used to performance upgrade the ring, is a promising approach for future MANs. This idea becomes our research question. The goal is to combine both topologies in a way that the specific strengths of ring and star are maintained while eliminating their individual shortcomings.

The chapter is structured as follows. We first discuss the four main metro approaches with respect to the metro requirements. We then summarize the results and derive our research question. Finally, we review related work on ring performance enhancing techniques and hybrid-ring star architectures.

5.1 Main Approaches

In the following, we discuss DoS, RPR, WDM rings, and WDM star networks with respect to the metro requirements summarized in Table 2.3. For the last two approaches, this discussion is relies on a generic architecture whenever possible and on the representative architectures described in Chapter 4.1 and 4.2, if not. As this work focuses on metro networks on the level of network/node architectures and MAC protocols, we do not address the aforementioned requirements 'manageability' as well as 'reliability and modularity'. Whenever the term 'connection' is used below it generically refers to a bandwidth demand between two nodes,

which could be both a circuit (e.g., TDM voice) or a flow of packets (e.g., GbE).

5.1.1 Data over SONET/SDH

SONET/SDH has originally been designed only for TDM voice traffic. Data services and *multi-protocol support* are enabled using the GFP. Remember from Section 3.1.2 that the GFP-F describes how data packets are mapped into SONET/SDH TDM streams which enables the transport packet based protocols such as IP. The GFP-T enables transparency for block coded data used in storage area networking protocols such as Fibre Channel and ESCON. The price for multi-protocol support in SONET/SDH is the additional electronic processing required for mapping the data into and out of the high speed and relatively complex TDM structure.

Optical transparency is possible in SONET/SDH in the sense that an intermediate node can optically bypass a wavelength channel if no lower bandwidth circuits are locally added to or dropped from that wavelength. Optical bypassing is advantageous since it reduces the node complexity. Related to the issue of optical bypassing in SONET/SDH is the so-called 'traffic grooming' problem [30]. It is an optimization problem where lower bandwidth circuits are routed through the network in a way that the number of bypasses is maximized to reduce the overall complexity of the network. However, this generally does not result in end-to-end transparency and the wavelengths still carry SONET/SDH TDM streams and are therefore not transparent to other modulation formats. Provisioning full wavelength channels to customers end-to-end is still not so common for several reasons. When adding a wavelength, parts of the network might need to be brought down, the power must be rebalanced in the whole network, and optical transparency forbids performance monitoring. Bad signal quality due to lack of 3R regeneration is also a problem that needs to be considered.

In terms of *differentiated SLAs and QoS levels*, SONET/SDH only supports one option that is constant delay with low jitter and no data loss. This results from the fact that for each data connection a dedicated circuit with a bandwidth larger or equal to the possible peak rate of the data connection is reserved in all TDM streams along the path from source to destination. Intermediate nodes only perform add-drop multiplexing operations which do not introduce any delay jitter or data loss. Unfortunately, the circuit switching concept does not allow for lower QoS classes for best-effort traffic.

Since setting up or changing circuits is such a complex task, *fast provisioning* of new connections is impossible and usually takes several weeks to months.

Sub-rate provisioning is enabled in SONET/SDH by VC. Traditionally, SONET/SDH has a strict data rate hierarchy. For instance, a GbE connection would have to be transported via an OC-48 circuit at a data rate of 2.5 Gbit/s what results in a channel utilization of not more than 40%. The next smaller SONET/SDH data rate would be OC-12 at 622 Mbit/s, which does not suffice to carry a GbE connection. Using VC, the bandwidth assignment is much more flexible and in the example above the remaining 1.5 Gbit/s of the OC-48 circuit could be allocated to other connections, such as several voice circuits and another GbE connection. Another out of many other options for provisioning a GbE connection is to virtually concatenate two OC-12 channels resulting in a 1.25 Gbit/s circuit. On the downside, VC requires lots of electronic processing for multiplexing and demultiplexing the individual connections into the TDM frames.

One of SONET/SDH's major shortcomings is achieving a *high bandwidth utilization*. Altough VC helps to improve the network utilization by providing each connection exactly the bandwidth it needs, SONET/SDH is still inefficient for the ever increasing amount of bursty

data traffic since it does not feature statistical multiplexing. Each data connection has to be provisioned for it's peak rate resulting in a poor network utilization due to the strong variations in the data rate. For instance, to offer GbE service over SONET/SDH, a circuit of 1 Gbit/s needs to be statically provisioned, altough the mean data rate of a typical GbE connection is much smaller. However, if the data rate of a connection changes relatively slowly, i.e., within minutes or hours, the provisioned bandwidth can be adapted using the LCAS.

The *scalability* requirement also raises some problems in SONET/SDH. Capacity upgrades usually require expensive 'forklift operations' where large fractions of the equipment need to be replaced which involves high costs and interruption of normal operation. SONET/SDH based metro networks are usually based on a ring topology limiting the geographical scalability to the area covered by the ring. To add network nodes, additional circuits must be rolled-out through the network which is a time consuming and expensive task. Each connection between each pair of nodes requires a separate circuit and if the traffic pattern requires that each node communicates with many other nodes, too many circuits might be required in total. Large parts of the network need to be reconfigured and as the number of circuits grows central offices might run out of rack space.

The same properties that limit SONET/SDH's scalability also degrade the *efficiency for different traffic patterns*. The network can be statically configured to support a specific traffic matrix, e.g., hot-spot or uniform traffic, by rolling-out circuits between the nodes according to the demands. But, again, if the number of nodes is high and a connection is required between most pairs of nodes an inadequately high number of circuits might be required to satisfy all demands. Additionally, the static circuit configuration only supports one specific traffic pattern. As the traffic pattern changes, costly reconfiguration is required which involves expensive addition, removal, or relocation operations of equipment and also results in network downtimes. For coping with smaller, temporary changes in the traffic pattern, for instance resulting from time-of-day variations, the LCAS is a promising approach. The LCAS enables the network operator to change the bandwidth of individual circuits, depending on the current demand. Note that the LCAS only provides a mechanism to adapt the bandwidth of existing circuits not for the setup or tear-down of circuits. Also, to adapt a connections bandwidth automatically, intelligent control mechanisms are required which are difficult to design.

One of SONET/SDH's key features and strengths is *survivability*. The technology is on the market for a long time and SONET/SDH equipment is generally considered very robust and mature. In fact, the 50 ms benchmark originates from SONET/SDH's APS mechanism. Usually 50% of the available bandwidth is reserved as spare capacity (so-called 1+1 and 1:1 protection). If a failure occurs, the system switches to the spare capacity and recovers the full bandwidth within 50 ms. The disadvantage of this concept is that not more than 50% of the available bandwidth can be utilized under normal operation. However, altough still not so common today, the spare bandwidth can be used for best-effort traffic with no survivability guarantees. More differentiated protection strategies can be realized using so-called 1:N protection where multiple connections share a common backup connection. In this case, the full bandwidth is recovered as long as only one of the N connections fails. Unfortunately, this kind of protection is not suited for SONET/SDH systems with ring topology dominantly deployed in the metro area. Also, 1:N protection does not allow for load balancing if more than one connection fails.

In terms of *cost-efficiency* the system can be considered as a low-first-cost solution because the existing SONET/SDH network still forms the core of the network and is evolutionary upgraded by the DoS extensions. However, as detailed above, the system has significant short-

comings and cannot be considered as a future proof migration path towards next-generation metro networks. The system complexity increases relatively fast as the network and the number of circuits grows, reconfiguring the system is hard, the utilization of the available resources is low, and the DoS extensions do not provide low-cost service for increasingly important best-effort traffic. Capacity upgrades usually require expensive forklift operations which does not support the 'pay-as-you-grow' philosophy. Furthermore, the ring topology does not enable a smooth migration towards future all-optical mesh architectures and the complex multiplexing hierarchy requires lots of electronic processing capacity.

To sum up, the GFP, VC, and the LCAS are significant improvements to traditional SONET/SDH, but the system is still dominated by the circuit-switching paradigm suffers from the resulting inefficiencies for data traffic.

5.1.2 Resilient Packet Ring

In contrast to SONET/SDH, RPR implements the packet-switching paradigm. Note that packets in RPR are also called frames. RPR's frame format is protocol independent and enables *multi-protocol support* by encapsulating data originating from arbitrary protocols into a so-called payload field that is part of each frame. If the data to be transmitted corresponds to a stream, for instance TDM voice traffic, the stream is consecutively splitted into fragments and handled the same way as packet based traffic. Furthermore, frames corresponding to a data stream are marked as so-called Class A high priority traffic which features low delay, low jitter, and zero data loss. Therefore, all frames arrive at the destination node regularly and the payload data can be reassembled to the original TDM stream.

Nodes in RPR are connected via bidirectional fibers and OEO conversion is performed at each node. Due to the OEO conversion RPR does not allow for *optical transparency*, i.e., optically bypassing intermediate nodes or end-to-end provisioning of transparent wavelength channels. On the other hand, OEO conversion also means 3R signal regeneration and results in relaxed requirements on the transmission layer. As the distances between the ring nodes are relatively short, there is usually no need for expensive optical amplifiers or dispersion compensation and larger geographic areas can be covered compared to a network where several intermediate nodes are optically bypassed. Another advantage of OEO conversion is that it enables monitoring of the signal quality and collection of traffic statistics.

RPR features three different traffic priority classes to provide *differentiated SLAs and QoS levels*. While class A traffic is considered as mission critical traffic and does not experience significant delay jitter nor any data loss, traffic of the the lowest class, class C, is transported in a best-effort manner without any performance guarantees. (RPR's traffic classes are discussed in in Section 9.1.1.)

Fast provisioning of connections, for instance a GbE connection between two customer sites, is possible almost instantly as no circuits need to be rolled out like in SONET/SDH. As long as there is sufficient bandwidth available on the ring, the customer can directly be attached to GbE ports in the RPR nodes closest to the customer sites. Since no connection setup or tear-down is required in a packet-switched network, services like 'dialing-for-bandwidth' can be easily implemented in RPR.

Sub-rate provisioning is natural in a packet-switched network, due to statistical multiplexing each connection claims exactly the bandwidth it needs.

Statistical multiplexing also results in a *high bandwidth utilization* for bursty data traffic as opposed to a circuit-switched network. In contrast to SONET/SDH, best-effort connections

do not need to be provisioned according to their peak rate. For instance, the mean date rate of a GbE connection is usually only a fraction of the peak rate of 1 Gbit/s. Therefore, several GbE connections can usually share a link with a capacity much smaller than their aggregate peak rate with only few data loss. Generally, as the number of data connections on a link increases, the required bandwidth converges to the aggregate mean data rate of all connections as opposed to a circuit-switched network where all connections are provisioned according to their peak rate. Therefore, RPR achieves a significantly higher bandwidth utilization than SONET/SDH. Additionally, RPR's destination removal and shortest path routing features further increase the network capacity. Unfortunately, like in any ring network, the capacity per node decreases asymptotically with the number of ring nodes. As the number of nodes grows, the fraction of transit traffic that needs to be forwarded increases, and a lower fraction of each node's transmission capacity can be used to send the node's own local traffic. For instance, if the line rate is 2.5 Gbit/s on each of the two counterdirectional rings, the network capacity for uniform traffic is approximately 20 Gbit/s (see Section 5.3.1). This capacity is almost completely independent of the number of nodes, i.e., if the ring consists of ten nodes each node sends own traffic at a rate of 2 Gbit/s, but if the ring consists of 20 nodes this rate reduces to 1 Gbit/s.

Due to this problem, the *scalability* of the number of nodes is limited. However, from an operational point of view, increasing the number of nodes is a relatively easy task. Since RPR's topology discovery protocol automatically updates the topology information in each node, the original network nodes do not need to be reconfigured as new nodes are inserted. The network capacity can be scaled by increasing the line rate or by deploying multiple rings in parallel using WDM. However, the former case requires to replace a large fraction of the existing node equipment while in the latter case multiple nodes must be deployed per central office which requires lots of expensive central office space and increases power consumption. Both scenarios involve high cost, especially is the ring consists of a large number of nodes. Therefore, the capacity is only scalable cost efficiently in small rings. As in any ring network, the geographical scalability is limited to the area covered by the fiber infrastructure.

Packet-switching results in a good *efficiency for different traffic patterns*. RPR copes with any traffic pattern, ranging from uniform to hot-spot traffic, as long as the capacity of no link is exceeded, and does not need to be reconfigured if the traffic matrix changes. This is a big advantage over SONET/SDH, where the static circuit configuration only matches a specific traffic pattern and a high number of circuits is required for distributed traffic patterns where each node communicates with many others. However, as shown in Section 7.2.2, RPR's capacity effectively reduces to 50% for asymmetric hot-spot traffic compared to uniform traffic.

In terms of *survivability*, RPR's wrapping and steering mechanism provides 50 ms recovery from single link or node failures, similar to SONET/SDH. Multiple failures cannot be fully recovered. For instance, if the fiber in both directions is cut in two different places, the ring is divided into two disjoint segments and does no longer provide full connectivity between all nodes. Also, as there is no spare capacity reserved to protect the ring, only part of the full ring capacity can be recovered in case of a failure. If bandwidth gets scarce, traffic of lower priority classes is dropped first. Therefore, RPR supports multiple survivability options. To guarantee survivability for mission-critical traffic the total amount of this type of traffic should not exceed the capacity that remains after recovery from a failure.

Considering *cost-efficiency*, RPR can be regarded as a 'low-first-cost' solution as it can be implemented within an existing SONET/SDH environment with relatively less effort by reserving a certain amount of SONET/SDH TDM bandwidth around the ring and deploying

RPR interface cards in each SONET/SDH ring node. While extending the number of nodes is relatively simple, capacity upgrades are costly and the geographical scalability is limited. This makes it hard to consider RPR as a 'pay-as-you-grow' solution. Furthermore, the fixed ring topology does not provide a smooth migration path towards future all-optical mesh networks.

Overall, RPR overcomes many of SONET/SDH's and DoS's shortcomings for data traffic. Most importantly RPR features better network utilization and simplified OA&M due to dissociating from the circuit-switching paradigm.

5.1.3 All-Optical Packet-Switched WDM Ring

In many aspects the all-optical packet-switched WDM ring is similar to RPR. The main differences result from the fact that most all-optical ring architectures rely on a slotted time structure and that wavelength channels are optically bypassed at intermediate nodes (see Section 3.3). The higher the number of deployed wavelengths, the more nodes are optically bypassed.

Multi-protocol support is realized by segmenting data packets or TDM streams into fragments of a fixed size equal to the optical slot size, and reassembling them at the destination. Alternatively, if the slot size is very large, each slot is filled up with multiple data packets. For transporting mission-critical traffic such as voice circuits, optical slots are periodically reserved resulting in a data channel with low delay jitter and no information loss, similar to a circuit.

As most intermediate nodes are optically bypassed, the all-optical WDM ring provides *optical transparency* to some degree. If the number of wavelengths is larger or equal to the number of ring nodes, each node is assigned a dedicated wavelength for receiving information. Using this so-called home channel, each pair of nodes communicates via an optically transparent wavelength path, enabling transparency to different modulation formats. Transparent wavelengths channels between customer premises cannot be provisioned because the slotted time structure requires all nodes to follow the MAC rules. If the number of nodes is larger than the number of wavelengths, several groups of nodes share the same home channel for receiving information. In this case, a node receiving data not destined to itself has to forward the slot towards the destination which involves OEO conversation. In this case the network is translucent, i.e., only partially tansparent.

For enabling *differentiated SLAs and QoS levels* in a slotted WDM ring two mechanisms have been proposed (see Section 3.4.2). Both support the transport of mission-critical traffic with low delay jitter and no packet loss by periodically reserving time slots and can be extended to support multiple QoS classes.

Resulting from the fact that the slotted WDM ring implements the packet-switching paradigm, the same holds for *fast provisioning* of connections and *sub-rate provisioning* as discussed above for RPR. Connections can be provisioned instantly and a wide range of data rates is handled efficiently.

Also in terms of achieving a *high bandwidth utilization* RPR and the slotted WDM ring perform very similar and the previous discussion above is also valid here. However, there is one significant difference: Because many nodes are optically bypassed and no or only few OEO conversion is performed in the WDM ring, the networking equipment is utilized much more efficiently. Each node forwards no or only few packets electronically and uses almost all of its transmission capacity to send its own ingress traffic. For instance, a resilient packet ring that operates at 10 Gbit/s per ring has the same capacity as a WDM ring with four 2.5 Gbit/s

channels per ring, but in the former case 10 Gbit/s need to be electronically processed per ring direction per node versus 2.5 Gbit/s in the optically transparent WDM ring. Fewer electronic processing reduces equipment cost. As in RPR, the electronic forwarding burden increases asymptotically with the number of nodes but it is generally lower due to optical bypassing and can be further reduced by increasing the number of wavelengths. When looking at the access protocol's impact on the utilization, the fact that all data must be fragmented to fit into the fixed size time slots leads to some overhead as each fragment is preceeded by some header information.

While optical bypassing reduces equipment cost, the special wavelength interconnection pattern between the nodes of an optically transparent WDM ring limits its *scalability*. The architecture is most efficient if the number of nodes is a multiple of the number of wavelengths. That makes it difficult to add a specific number of nodes or wavelengths to adapt to changes in geographic or bandwidth demands. Furthermore, and different from RPR, changing the number of nodes requires a network wide reconfiguration of all nodes on a hardware level. Due to the fact that several nodes are optically bypassed the network needs to be more carefully engineered to reduce signal degradation, costly optical amplifiers and dispersion compensation might be required, and the circumference of the fiber ring is limited. This further reduces the limited geographical scalability of the ring topology.

The *efficiency for different traffic patterns* is the same as in RPR with the difference that in the optically transparent WDM ring architecture the efficiency for hot-spot traffic can be increased by equipping the hot-spot node with multiple transceivers. Similar to RPR, *survivability* is achieved by steering packets away from the failed link or node using the opposite transmission direction. It has been shown that the worst case downtime for a connection in a 100 km ring is about 1 ms [139]. Multiple failures generally cannot be recovered and the capacity degradation in case of a failure is the same as in RPR.

Concerning *cost-efficiency*, the system can be considered as a 'low-first-cost' solution since the existing fiber structure can be reused and the nodes require relatively few equipment. Due to the scalability problems in terms of capacity, the number of nodes and geographic dispersion the system cannot be considered as a future-proof 'pay-as-you-grow' solution.

Overall, when comparing the WDM to RPR, ring fixed size time slots and optical bypassing enable very high speed metro rings at the price of limited scalability.

5.1.4 Single-Hop WDM Star Network

Similar to the all-optical WDM ring, the single-hop WDM star is based on a slotted time structure. Depending on the slot size, multiple packets with the same destination are either aggregated into a single slot or packets are fragmented into multiple consecutive slots. This enables the transport of variable size packets and, in conjunction with the QoS mechanism for mission-critical traffic discussed below, enables *multi-protocol support*.

One of the star's strengths is *optical transparency*. In fact, each pair of nodes communicates via optically transparent lightpaths. Each pair of nodes can use a different modulation format and optically transparent wavelength channels can be provisioned to customers. The only constraint is that collisions of wavelengths at the central AWG and/or PSC must be avoided.

Differentiated SLAs and QoS levels are enabled by periodically reserving time slots, similar to the WDM ring. Furthermore, when configuring the network's software parameters, there is a tradeoff between high capacity and low delay. This relationship can be exploited to optimize the network for the current traffic requirements depending on the time of the day [140]. For

instance, measurements show that web traffic, which benefits from low delays, is dominant in the evening while at late night the exchange of large files between servers raises a demand for high network capacity.

Similar to the other discussed packet-switched networks, the WDM star features *fast provisioning* and *sub-rate provisioning* of bandwidth.

However, the main advantage of the star over other topologies is that it achieves a very *high bandwidth utilization*. As all pairs of nodes communicate directly in a single hop, no capacity is wasted for forwarding traffic. Depending on the configuration of the network, all nodes can communicate simultaneously resulting in the highest possible network capacity which, for uniform traffic, is equal to the aggregate capacity of all nodes' transmitters. Compared to RPR, where the capacity for uniform traffic is fixed to approximately eight times the equipment data rate, this is a big improvement, especially for a large number of nodes. However, to fully exploit the star's capacity, collisions of packets from different source nodes must be avoided. This is achieved by performing pretransmission coordination in conjunction with a reservation mechanism. The nodes must provide sufficient electronic processing capacity to be able to run this mechanism at high data rates.

In terms of *scalability* is must be considered that the network achieves the highest capacity per transceiver if both the number of nodes attached to each AWG port and the number of wavelengths are equal to the number of AWG ports. In this case all nodes can communicate simultaneously using the minimum number of wavelengths. To make future upgrades of the number of nodes more efficient, it is a good strategy to use a larger than necessary AWG with a corresponding increased number of wavelengths in the initial network deployment. Then, as nodes are added, the network approaches the optimum configuration. If nodes are added beyond that point the network is still remains more efficient as if a smaller AWG and number of wavelengths has been deployed initially. If at any point a larger AWG needs to be deployed to provide sufficient capacity for an increased number of nodes, all nodes must be upgraded to support a higher number of wavelengths which involves high cost. Similarly, scaling the network's capacity by upgrading the data rate also affects all nodes and prohibits cost-efficient upgrades. Note that the network capacity 'automatically' increases as nodes are added. The geographic scalability is improved compared to a ring since the star topology allows to connect to nodes geographically spread over a wide area. However, the total network diameter is limited by the fact that no electronic signal regeneration is performed on the all-optical path between source and destination. To cover larger areas costly optical amplifiers as well as dispersion compensation are required.

Similar to the previously discussed packet-switched networks, the WDM star provides *efficiency for different traffic patterns* and supports any traffic pattern as long as the capacity of no transmitter or receiver is exceeded. For handling static, asymmetric demands more efficiently, for instance hot-spot traffic, multiple nodes or nodes with multiple transceivers can be deployed in central offices with higher than average traffic demands.

Speaking of *survivability*, the central AWG and/or PSC represents a so-called single point of failure, i.e., a failure in this component makes the whole network inoperable. As a solution the so-called AWG||PSC network has been proposed, where AWG and PSC are deployed in parallel and protect each other against failures. The PSC additionally provides an efficient control channel required by the MAC protocol during failure free operation [121][141]. Generally, in case of a fiber cut, all nodes attached to that fiber are disconnected, but the remaining part of the network remains functional and maintains its full capacity. For instance, if the fiber between a node and the corresponding combiner/splitter pair fails, only this node is

disconnected. If the fiber between a combiner/splitter pair and the AWG fails, all nodes attached to this AWG port are disconnected.

Concerning *cost-efficiency*, the WDM star cannot be considered as a 'low-first-cost' solution compared to the formerly discussed ring networks for which at least parts of the existing SONET/SDH infrastructure can be reused. On the other hand, nodes can be added in a 'pay-as-you-grow' manner while the topological constraints mentioned above should be respected to ensure high efficiency. The nodes require sufficient processing capacity to run the reservation mechanism and to (re)assemble the packets. Geographically, the star topology can be extended more easily to cover new areas and a migration towards a mesh network is more feasible compared to the ring topology, but still limited.

In summary, the WDM star provides lots of capacity due to the fact that all nodes communicate in a single hop and interesting features like optical transparency or geographical scalability. A disadvantage from the architectural perspective is the high number of optical fibers required to deploy the star.

5.2 Research Question

When reflecting over the previous discussion of the individual approaches to address the 'metro gap', which is summarized in Table 5.1, it gets clear that the DoS does not overcome SONET/SDH's limitations in the metro area. While they are certainly an improvement for circuit-switched long-haul networks, the fundamental shortcomings for metro networks mentioned in Chapter 1 persist: Capacity upgrades still require expensive 'forklift operations', bursty data traffic results in poor bandwidth utilization, provisioning of new circuits takes lots of time, and the high system complexity results in high equipment and OA&M cost. Currently, networks based on the packet-switching paradigm seem to be more suitable for efficiently handling the ever increasing amount of bursty data traffic and to provide the flexibility necessary to cope with largely unpredictable and varying traffic patterns.

Table 5.1 shows that the two packet-switched ring networks, RPR and the all-optical WDM ring, have similar properties. Some differences result from the fact that optical bypassing in the WDM ring enables higher capacities compared to RPR, while the latter features a more mature access protocol and is already commercially available. Since both packet-switched rings are relatively similar, answering the question which of the previously discussed approaches is most suitable for future metro networks comes down to a comparison of these rings to single-hop WDM star networks. In terms of performance, the investigation in Chapter 4 has shown that the star network clearly outperforms the ring in terms of throughput, delay, and packet loss. On the other hand, when looking at other factors besides these performance metrics, the discussion in this chapter shows that both ring networks, especially RPR, meet the metro requirements relatively well and, overall, better than the star. For instance, RPR features sophisticated QoS support, efficient failure recovery mechanisms, and integration within existing SONET/SDH environments. This leads to the idea of providing a means to improve the performance of optical packet-switched ring networks while at the same time maintaining the specific strengths, in other words, *a performance upgrade for optical packet-switched ring networks*. Note that, alternatively, our goal could also be to overcome the star's limitations with respect to the metro requirements while maintaining its high performance. In fact, as Table 5.1 shows, ring and star are complementary in many aspects, for instance in terms of forwarding overhead, surviability, or cost for the initial deployment. This leads to a refine-

ment of our idea of providing a performance upgrade for packet rings, namely to combine both topologies to a hybrid ring-star architecture in a way that the strengths of either topology are maintained thereby eliminating the shortcomings of the other. In different words, *we use a single-hop WDM star network as a performance upgrade for optical packet-switched ring networks.*

Remember that the star's capacity generally increases with the number of nodes. Therefore, in our approach, the number of ring nodes that are connected to the star is not fixed but a parameter used to scale the overall capacity of the network. Fig. 5.1 shows such a hybrid architecture, where a subset of the ring nodes is connected to a single-hop star network. In Chapter 6 we will define the details of this architecture as well as a corresponding MAC protocol and call the resultant ring-and-star based network *'RINGOSTAR'*.

Note that for the ring part of the network we rely on RPR. First, RPR better complements the star compared to the all-optical WDM ring as the WDM ring features a less mature MAC protocol, which is also a weakness of the star. Additionally, the single channel resilient packet ring benefits more from a capacity upgrade than the WDM ring. Furthermore, at the time of writing, RPR is commercially available and starts being deployed in metropolitan areas making it a more interesting candidate for a performance upgrade. However, our performance upgrade can in principle be applied to any packet-switched ring network.

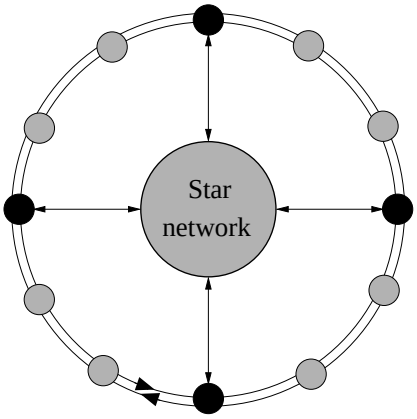


Figure 5.1: Hybrid ring-star architecture consisting of bidirectional packet-switched ring and single-hop WDM star network.

5.3 Related Work

As our goal is to develop a performance upgrade for optical packet-switched ring networks, we now provide a historical perspective on previous work on ring performance enhancing techniques. Furthermore, we discuss previous work on hybrid ring-star architectures.

5.3.1 Optical Single-Channel Ring Networks

In optical single-channel ring networks each fiber link provides one single communication channel. Optical single-channel ring networks belong to the first generation of *opaque* optical networks where OEO conversion takes place at each node [142]. In particular, the so-called

Requirement	DoS	RPR	WDM Ring	WDM Star
Multi-Protocol-Support	Yes	Yes	Yes	Yes
Optical Transparency				
Bypassing Nodes	Partly	No	Partly	Yes[1]
Mod. Format Transp.	No	No	Partly	Yes
Wavelength Provisioning	Difficult	No	No	Yes
Differentiated SLAs & QoS				
Circuit Emulation	Yes[2]	Yes	Yes	Yes
Best-Effort	No	Yes	Yes	Yes
Additional Classes	No	Yes	Partly	No[3]
Fast Provisioning	No	Yes	Yes	Yes
Sub-rate Provisioning	Yes	Yes	Yes	Yes
High Bandwidth Utilization				
Statistical Multiplexing	No	Yes	Yes	Yes
Forwarding Overhead	Moderate	High	Low	None
Scalability				
Number of Nodes	Limited	Good	Limited	Moderate
Capacity	Limited	Limited	Moderate	Moderate
Geographically	Limited	Limited	Limited	Good
Different Traffic Patterns				
Uniform	Good	Good	Good	Good
Hot-spot	Good	Moderate	Good	Good
Dynamically Changing	Limited[4]	Moderate	Good	Good
Survivability				
Unused Spare Bandwidth	Yes	No	No	No recovery,
Full Bandwidth Recovery	Yes	No	No	failure
Sub-50 ms recovery	Yes	Yes	Yes	only affects
Multiple Failure Recovery	Limited	Limited	Limited	network
Multiple Surv. Classes	Yes	Yes	No	locally
Cost-efficiency				
Low-First-Cost	Yes	Yes	Partly[5]	No
Pay-As-You-Grow	No	Partly[6]	No	Yes
Node Complexity	High	Moderate	Low	Moderate
Migration Towards Mesh	No	No	No	Limited

Table 5.1: Comparison of individual metro approaches. ([1]single hop, [2]real circuits, [3]potentially possible, [4]using LCAS, [5]fiber reuse, [6]no. nodes)

Cambridge ring and *buffer insertion* rings have attracted considerable attention and have influenced the development of next-generation optical ring networks, as we will see shortly.

The Cambridge ring is a *unidirectional* ring network [143]. Channel access is based on the *empty slot* principle. The channel is divided into time slots. At the beginning of each slot a bit indicates whether the slot is used (occupied) or not (empty). To transmit, a node must wait until an empty slot arrives. Having filled an empty slot with a packet, the node waits until the slot returns and marks it empty. In other words, the Cambridge ring deploys *source stripping* where the source node takes the transmitted packet from the ring.

The buffer insertion ring is also a *unidirectional* ring network [38]. Each node is equipped with three electrical FIFO buffers: A reception, a transmission, and an insertion buffer. The reception and transmission buffers store packets that are destined to or originate from the corresponding node. The insertion buffer temporarily stores the incoming ring traffic in the electrical domain in order to allow the local node to transmit a packet onto the ring. To prevent packet loss on the ring, ring traffic is given priority as soon as the insertion buffer fills up, i.e., the buffer occupancy rises above a certain threshold. In both unidirectional ring networks the maximum hop distance equals $h_{max} = N - 1$, where N denotes the number of nodes in the network. As opposed to the Cambridge ring, however, packets are removed from the buffer insertion ring by the receiving node (rather than the transmitting node). This so-called *destination stripping* enables downstream nodes to *spatially reuse* bandwidth and decreases the mean hop distance of the ring network. For uniform traffic, i.e., each node generates the same amount of traffic and a given packet is destined to any of the $(N-1)$ nodes with equal probability $1/(N-1)$, the mean hop distance of destination stripping rings is given by

$$\overline{h} = \frac{1}{N-1} \sum_{j=1}^{h_{max}} j = \frac{1}{N-1} \sum_{j=1}^{N-1} j = N/2. \tag{5.1}$$

Due to the decreased mean hop distance and spatial reuse the average throughput of the network is improved by a factor of up to two.

The most important legacy standard for optical single-channel ring networks is FDDI, which is based on IEEE 802.5 Token Ring [144]. Unlike the aforementioned ring networks, FDDI deploys a dual-fiber architecture forming a *bidirectional* ring network. The counter-rotating fiber rings can be used for concurrent transmission and provide protection against a single link or node failure by means of ring wrapping. In FDDI nodes deploy source stripping and at most one node (the node that holds the token) can transmit packets. Consequently, FDDI is not able to provide spatial reuse of bandwidth (no matter whether with or without early token release).

The so-called MetaRing overcomes these limitations. MetaRing is a bidirectional full-duplex ring operating either in the buffer insertion mode for variable-size packets or slotted mode for fixed-size packets/cells where the slot size equals the transmission time of a fixed-size packet/cell [40]. Nodes deploy destination stripping. Furthermore, packets are transmitted via the *shortest path* by choosing the appropriate ring. With destination stripping and shortest path routing the maximum hop distance is equal to $h_{max} = \lceil (N-1)/2 \rceil$. For uniform traffic the mean hop distance of bidirectional destination stripping rings with shortest path routing is given by

$$\overline{h} = \frac{2}{N-1} \sum_{j=1}^{\lfloor \frac{N-1}{2} \rfloor} j + \frac{(N-1) \bmod 2}{N-1} \left\lceil \frac{N-1}{2} \right\rceil \tag{5.2}$$

$$= \begin{cases} \frac{N+1}{4} & \text{if } N \text{ odd} \\ \frac{N^2}{4(N-1)} & \text{if } N \text{ even,} \end{cases} \tag{5.3}$$

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote the ceiling function and floor function, respectively. Due to the decreased mean hop distance the spatial reuse factor equals four, i.e., for uniform traffic the average throughput is improved by a factor of up to four.

In MetaRing a backlogged node may start a packet transmission at any given time, provided the insertion buffer (in buffer insertion operation mode) or the current slot (in empty slot operation mode) is empty, i.e., MetaRing gives priority to in-transit ring traffic. Note that in either mode nodes can suffer from *starvation*, which happens if a node is constantly being covered by upstream ring traffic and thus is not able to access the ring, giving rise to fairness problems. The fairness control in MetaRing makes use of a control signal, termed SAT, which circulates on each ring and regulates the transmission quota/credit of every node. A node forwards the SAT signal upstream with no delay, unless it is not SATisfied or starved. A node is considered starved if it could not send the permitted number of packets since the last time it has forwarded the SAT signal. The node is SATisfied if between two SAT signals the node has sent at least l packets or if all packets present in its transmission buffer when the previous SAT was sent upstream, were transmitted. When the node receives a SAT and it is SATisfied, it will forward the SAT upstream. If the node is not SATisfied, it will hold the SAT until it is SATisfied and then forwards the SAT upstream. After a node forwards the SAT, it can send k more packets, where $k \geq l$.

Note that various aspects of the above mentioned ring networks can also be found in the IEEE 802.17 Resilient Packet Ring standard for high-performance packet switched optical single-channel MANs [145][37].

5.3.2 Ring WDM Upgrades

Clearly, by using WDM each fiber link provides multiple communication channels, each operating at a different wavelength. There is a large number of architectures and access protocols proposed for next-generation ring WDM networks which can be used for upgrading optical single-channel rings. We have comprehensively reviewed these networks in Section 3.3. In the following, we focus on the merits and shortcomings of the major previously reported improvements of optical ring networks. Specifically, we concentrate on how the capacity of ring WDM networks can be increased. (We use the term 'capacity' to refer to the 'maximum achievable aggregate throughput' of the network, if not denoted otherwise, for uniform traffic.)

All-Optical Node Structures

Instead of OEO converting all signals at each node, *all-optical* node structures have been proposed which leave the data packets in the optical domain while processing the packet header information in the electrical domain to decide whether to drop or forward the data packet. In doing so, only packets destined for the local node have to be optical-electronically converted while in-transit traffic remains in the optical domain. Note, however, that these so-called OOO nodes do not necessarily provide logical optical bypasses, as discussed in the subsequent section. For instance, in order to prevent channel collisions, each node needs to monitor the status of all wavelength channels prior to and/or while injecting packets into the ring [51][48]. In other words, each node needs to inspect and process the status (busy or idle) of all wavelength channels.

Optical Bypassing and Traffic Grooming

With *optical bypassing* each node has to inspect/process only a subset of the wavelengths while the remaining wavelengths pass through the node untouched. This helps alleviate the computational burden and reduce the number of electronic port cards at bypassed nodes. When a wavelength is not dropped at a node an electronic card (including receiver) is not required for that wavelength. The required number of electronic port cards can be further reduced by *grooming* the lower rate traffic such that a smaller number of wavelengths need to be dropped and the electronic processing is decreased at each node [146][147][30]. More importantly, optical bypassing enables the design of *logical* topologies which are embedded on the physical ring network, see for example [28][88]. By optically bypassing more nodes the mean hop distance of logical topologies is decreased. A small mean hop distance is advantageous in that the forwarding burden is alleviated and the number of required interfaces is decreased at nodes.

Note that in logical topologies, the *logical* maximum and mean hop distance between nodes is decreased, but the *physical* path remains unchanged. Hence, traffic consumes the same amount of bandwidth resources no matter whether optical bypassing is provided or not. As a consequence, *in WDM ring networks, either with or without optical bypassing, the spatial reuse factor is no larger than in their single-channel counterparts.* For example, in bidirectional single-channel and WDM rings, both deploying shortest path routing and destination stripping, the spatial reuse factor is no larger than four for uniform traffic.

Meshed Rings

The spatial reuse factor in bidirectional WDM rings can be increased by providing *alternate* physical paths in addition to the fiber rings, resulting in so-called *meshed rings* [91][92]. Meshed rings are based on two counter-rotating WDM fiber rings, each carrying W wavelength channels. All nodes are equipped with multiple fixed-tuned transceivers for simultaneous transmission/reception on multiple wavelength channels on both rings. Nodes deploy shortest path routing and destination stripping. In addition to the ring nodes, K wavelength routers are equally distributed among the nodes on the bidirectional ring. Counter-directional pairs of fiber, so-called chords, are used to interconnect the wavelength routers. More precisely, each chord interconnects two different wavelength routers. The wavelength routers provide cross-connect routing of specified wavelength channels across the chordal links. In doing so, *physical short-cuts* are created which allow to send data packets skipping all intermediate ring nodes between two connected wavelength routers. Thus, instead of traveling along the ring, data packets take the short-cut, thereby consuming fewer ring bandwidth resources. Consequently, compared to non-meshed rings packet transmissions are bounded to smaller ring segments and more transmissions can simultaneously take place on the ring, resulting in an increased spatial reuse.

It was shown in [91][92] that meshed rings achieve a significantly larger capacity than non-meshed bidirectional destination stripping ring networks. For uniform traffic a meshed ring using $K = 6$ wavelength routers and $W = 5$ wavelengths increases the network capacity by 720% compared to unidirectional source stripping rings, which translates into a spatial reuse factor of 7.2. Thus, the capacity of the meshed ring is 80% larger than that achieved by non-meshed bidirectional destination stripping ring networks (whose spatial reuse factor equals 4), at the expense of additional multiple wavelength routers and chordal fiber links. We do

not discuss the analysis of the mean hop distance of meshed rings here (the interested reader is referred to [91][92] for more details). Instead, we provide an intuitive understanding why chords are limited in further decreasing the mean hop distance of meshed rings. Recall that each chord provides a short-cut between two wavelength routers. Each chord interconnects a different pair of wavelength routers. In general, a data packet has to traverse multiple chords in order to reach its destination. This is due to the fact, that each chord acts as a stand-alone short-cut of limited range in that each chord provides a short-cut to a single wavelength router which in general is not the one closest to the final destination node. As a consequence, data packets generally travel along *multiple hops* on their way from a given source node to a given destination node. As we will see shortly, our approach avoids multihopping by interconnecting all chords through a central hub in a *single-hop* subnetwork such that a given wavelength router is able to use its locally attached chord to get access to the central hub and thereby to all other chords attached to the hub. The subnetwork provides single-hop short-cuts among all wavelength routers attached to the hub. Thus, a given source node is able to send packets to the wavelength router that is closest to the corresponding destination node, resulting in a decreased mean hop distance and an increased capacity. (Note that multihopping in meshed rings could also be prevented by interconnecting all wavelength routers in a full mesh. As opposed to our approach, however, this method requires a prohibitively large number of fibers (chords) for interconnecting multiple wavelength routers.)

The architecture presented in the next chapter, RINGOSTAR, is an entirely different approach to WDM upgrade RPR and optical single-channel rings. In our approach, we use *one single* wavelength router functioning as a central hub as opposed to meshed rings which use multiple wavelength routers placed on the bidirectional ring. Furthermore, we do not apply WDM on the ring as done in the aforementioned WDM upgrades of optical single-channel ring networks. Instead, WDM is used only on the central wavelength router based single-hop star network while leaving the peripheral fiber rings unchanged.

5.3.3 Hybrid Ring-Star Architectures

The combination of star and ring configurations to form hybrid network topologies has already been addressed to some extent previously.

A modification of FDDI to a hybrid ring-star architecture is presented in [148]. The proposal targets at increasing the capacity of FDDI for broadband CBR traffic, which for instance results from high-quality video transmissions. A subset of the ring nodes, the so-called high-speed nodes, are allowed to send CBR traffic at rates significantly higher than regular ring nodes. The ring is divided into segments and each segment consists of several regular ring nodes but only few high-speed nodes. Both ends of each segment are connected to a central high-speed switching unit. I.e., if there are N segments, the original fiber ring is physically transformed to N loops which are interconnected by the central switching unit. Traffic from regular ring nodes arriving from one segment is simply forwarded to the next segment by the switching unit. Therefore, logically all regular ring nodes are still connected with a ring topology. Traffic from high-speed stations, however, is switched directly to the segment containing the high-speed destination station. This reduces the mean hop distance between high-speed stations and therefore reduces the forwarding burden of the ring nodes resulting in an increased network capacity for both regular and high-speed traffic. The work focuses mainly on discussing the MAC protocol and implementational aspects, a performance evaluation is not presented. Note that the proposed network features QoS support by providing

different services classes for best effort data traffic as well as for delay and loss sensitive CBR traffic.

Bellcore's *Star-Track* switch is formed from two internal networks, an optical PSC based broadcast-and-select single-hop star WDM network and an electronic unidirectional token based control ring [149][150]. To access the star network, each node has one FT and one tunable receiver TR. The control token ring is used for making reservations. After one ring round-trip propagation delay, data packets are sent across the star. Star-Track does not allow for immediate ring access due to the token based protocol. Moreover, the PSC as a broadcast device does not support spatial wavelength reuse, as opposed to the wavelength-routing AWG.

A hybrid star-ring network based on *multiple* central wavelength routers in parallel was proposed in [151]. *All* ring nodes are connected to the central wavelength routers by either one or two pairs of fiber (so-called spokes). In addition, ring nodes are interconnected by a small number of fibers around the circumference carrying protection-switched traffic to standby spokes as well as residual working wavelength channels. The use of additional fibers in a ring around the periphery of the multiple-star network is one of the key features that allows total fiber quantities to be minimized. It was shown that for a single path failure and uniform traffic, fiber requirements are less than for a WDM ADM ring, while providing greater resilience to multiple path failures. The work focused primarily on path and wavelength router protection strategies and did not specify any MAC protocol. Furthermore, the architecture does not deploy splitters for enabling optical multicasting.

A multilevel star-ring architecture consisting of a star network on the upper level and *multiple concatenated* ring subnets on the lower level was studied in [152]. The upper level star network ensures high network capacity and its weakness in reliability is overcome by the concatenated ring subnets with self-healing capabilities. The work concentrates on the physical transmission limitations rather than protocols. Again, a MAC protocol for such a modified star-ring architecture was not provided and investigated.

RINGOSTAR differs from the above mentioned ring-star architectures in a number of ways. As we will see in the next chapter, RINGOSTAR deploys *one single* wavelength router (AWG) in parallel to a broadcast device (PSC) with attached *splitter/combiner* pairs. Only a *subset* of the ring nodes are directly connected to the star network. The MAC protocol allows for *immediate* medium access on the *bidirectional* ring. The integrated ring-star network forms a *single-level* architecture.

5.4 Conclusions

We have discussed the four main approaches for future metro networks, namely DoS, RPR, WDM rings, and WDM star networks with respect to the metro requirements. It turned out that DoS does not overcome SONET/SDH's limitations in the metro area, while RPR meets the metro requirements relatively well. However, the single-channel RPR network suffers from relatively low performance, especially when compared to the WDM star that provides huge capacity but suffers from a number of practical limitations. This leads to the idea of combining a single channel RPR like packet-switched ring with a single-hop WDM star network which serves as a performance upgrade for the ring. The goal is to combine the strengths of either topology while overcoming the weaknesses of the other. After declaring this idea as our research question, we reviewed related work in ring performance enhancements and hybrid ring-star architectures.

Chapter 6

RINGOSTAR

IN this chapter, we propose and examine a novel ring-and-star based architecture which we call 'RINGOSTAR'. This hybrid network not only aims at combining the aforementioned strengths of both ring and star configurations while avoiding their drawbacks but also follows an entirely new direction to WDM upgrade optical single-channel rings. Instead of deploying WDM on the ring, RINGOSTAR uses WDM on the central AWG/PSC based single-hop star network, thereby exploiting the large spatial wavelength reuse capability of the wavelength-routing AWG. Generally, in RINGOSTAR only a *subset* of ring nodes is directly connected to the star network, resulting in less fiber requirements and node interfaces compared to a 'pure' star network. Furthermore, by using 'dark fibers', which are abundantly available in most metro areas, no additional fiber structure needs to be installed, thus avoiding costly construction work and manpower. For the ring part of the hybrid architecture we rely on RPR. However, note that RINGOSTAR is in principal suitable for any packet-switched ring network. No major modifications of RPR's basic protocol and mechanisms are required. Only the subset of nodes attached to the star needs to be equipped with additional hardware and software while all remaining nodes may operate without any modification. The proposed upgrade therefore allows for an evolutionary upgrade of existing RPR networks in that subset of nodes can be upgraded in a pay-as-you-grow manner according to traffic demands and cost constraints. We introduce the novel concept of *proxy stripping* which is used to route ring traffic on single-hop short-cuts across the star subnetwork rather than the peripheral ring, which dramatically decreases the mean hop distance and therefore increases the capacity. By means of mathematical analysis of the system we show that by WDM upgrading and interconnecting only 64 nodes of a 256-node RINGOSTAR network the mean hop distance is less than 5% of that of bidirectional rings with destination stripping and shortest path routing.

The chapter is organized as follows. We first provide a description of the underlying RPR network in Section 6.1. In Sections 6.3 and 6.4 we describe our novel hybrid architecture and the corresponding MAC protocol (or 'access protocol') in detail. Finally, we show that the proposed network features a significantly reduced mean hop distance compared to unidirectional, bidirectional, and meshed WDM rings in Section 6.5. The reduced mean hop distance translates in a significantly larger network capacity which is analyzed in detail in Chapter 7.

6.1 Resilient Packet Ring

RPR has been previously discussed from a more general perspective in Section 3.1.3. Here, we focus on details of the architecture and on the access protocol. Various aspects of the single-channel ring networks discussed in Section 5.3 can also be found in IEEE 802.17 Resilient Packet Ring. As illustrated in Fig. 3.1, RPR is a bidirectional dual-fiber ring network with OEO conversion at each of the N nodes. Every node is equipped with two FTs and two FRs, one for each fiber ring. Destination stripping in conjunction with shortest path routing improves the spatial reuse of bandwidth significantly. Note, however, that the path selection in specified in the Institute of Electrical and Electronics Engineers (IEEE) standards document is not necessarily shortest path routing. Higher layers (such as IP) may explicitly specify the 'best' direction/ring to each destination, including shortest path [153]. Broadcasting is achieved by means of source stripping. Each node has separate transit and station queues for either ring, as depicted in Fig. 3.1. Specifically, for each ring a node has one or two transit queues termed primary transit queue (PTQ) and secondary transit queue (STQ), one transmission queue termed stage queue, one reception queue, and one addMAC queue which stores control packets generated by the local node. All queues implement FIFO queues. The queue structure is illustrated in Fig. 6.1. The next packet that is sent on the ring is chosen from one of these queues according to the following arbitration mechanism. This mechanism represents the basis for RPR's QoS support which is discussed along with RPR's three different traffic classes A, B, and C in Section 9.1.1.

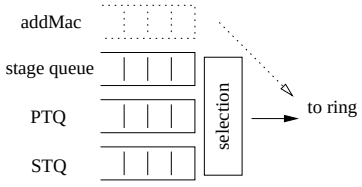


Figure 6.1: Queue structure of an RPR node for one ring direction.

RPR nodes operate in one of two modes: (*i*) *single-queue* mode or (*ii*) *dual-queue* mode. In single-queue mode, the transit path only consists of the PTQ. If the PTQ is not full, highest priority is given to addMAC traffic. At the absence of local control traffic, priority is given to in-transit ring traffic over station traffic. In dual-queue mode, the transit path comprises both PTQ and STQ. The PTQ is used for exclusively for class A traffic while the STQ stores packets belonging to class B and C traffic. In dual-queue mode, if both PTQ and STQ are not full, highest priority is given to addMAC traffic (similar to single-queue mode). If there is no local control traffic, PTQ traffic is served always first. If the PTQ is empty, the local transmission queue (stage queue) is served until STQ reaches a certain queue threshold. When the STQ reaches that threshold, STQ in-transit ring traffic is given priority over station traffic such that in-transit packets are not lost due to buffer overflow. Thus, the transit path is lossless and a packet put on the ring is not dropped at downstream nodes. In summary, the scheduling algorithms to arbitrate service among transit and station queues for both single-queue and dual-queue modes at RPR nodes are described by pseudo-code in Fig. 6.2.

Furthermore, RPR defines fairness control algorithms that specify how a congested downstream node can throttle the transmission rate of upstream nodes by sending fairness control packets upstream (fairness control is addressed in Chapter 9.2.)

```
// single-queue mode (PTQ only)
while (true) { // endless loop
    if (PTQ_not_full) {
        if (addMac_not_empty) {
            send_packet_from_addMac();
            continue; // reenter loop
        }
    }
    if (PTQ_not_empty) {
        send_packet_from_PTQ();
        continue; // reenter loop
    }
    if (stage_not_empty)
        send_packet_from_stage();
}
```

```
// dual-queue mode (PTQ and STQ)
while (true) { // endless loop
    if (PTQ_not_full) {
        if (STQ_not_full) {
            if (addMac_not_empty) {
                send_packets_from_addMac();
                continue; // reenter loop
            }
        }
    }
    if (PTQ_not_empty) {
        send_packet_from_PTQ();
        continue; // reenter loop
    }
    if (STQ_nearly_full) {
        send_packet_from_STQ();
        continue; // reenter loop
    }
    if (stage_not_empty) {
        send_packet_from_stage();
        continue; // reenter loop
    }
    if (STQ_not_empty)
        send_packet_from_STQ();
}
```

Figure 6.2: Pseudo-code for queue arbitration in RPR.

RPR provides a number of advantageous performance features. Among others, the counter-rotating rings provide protection against any single link or node failure and the dual-queue operation mode enables service differentiation, e.g., guaranteed QoS. Moreover, due to OEO conversion at each node, 3R signal regeneration can be provided in the electrical domain which enables optically unamplified transmission between network nodes such that no expensive optical amplifiers are required.

6.2 Proxy Stripping

To improve spatial reuse and bandwidth efficiency of RPR we propose to augment the bidirectional ring by a *single-hop* star subnetwork, as illustrated in Fig. 6.3 (a) for $N = 12$ ring nodes. A subset of $P \leq N$ ring nodes are connected to the single-hop star subnetwork, preferably by bidirectional pairs of *dark fiber*. Note that recently most conventional carriers, a growing number of public utility companies, and new network operators make use of their right of ways especially in metropolitan areas to build and offer so-called dark-fiber networks. These dark-fiber providers have installed a fiber infrastructure that exceeds their current needs. The unlit fibers provide a cost-effective way to build very high capacity networks or upgrade the capacity of existing (ring) networks. Buying one's own dark fibers is a promising solution to reduce network costs as opposed to leasing bandwidth which is an ongoing expense. Nodes can be attached to the single-hop star subnetwork one at a time in a pay-as-you-grow manner according to given traffic demands. The hub of the single-hop star network may be a PSC, an AWG, or a combination of both. For more details on various AWG and PSC based single-

hop star network and node architectures along with MAC protocols the interested reader is referred to [154]. Nodes attached to the star subnetwork perform *proxy stripping*, a novel packet stripping technique developed in RINGOSTAR.

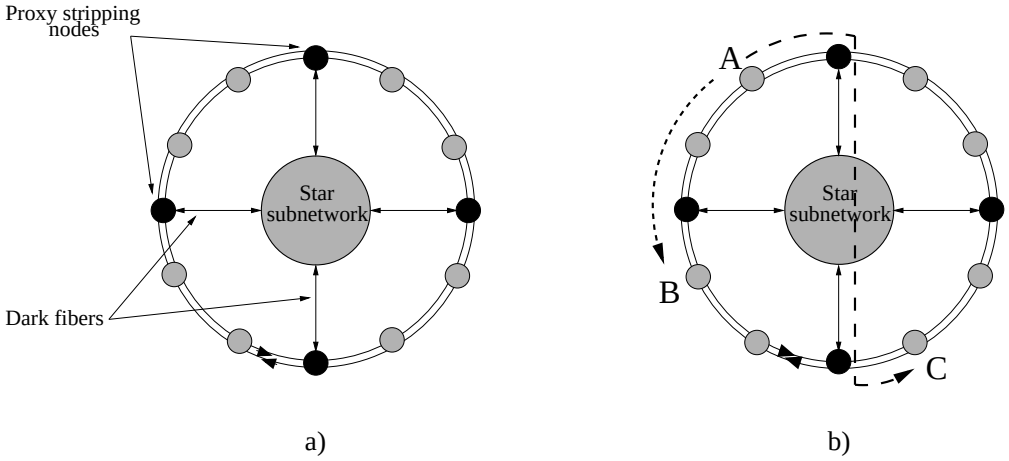


Figure 6.3: Proxy stripping technique: (a) RPR with $N = 12$ nodes, where $P = 4$ of them are interconnected by a dark-fiber single-hop star subnetwork, (b) proxy stripping in conjunction with destination stripping and shortest path routing for source node A and destination nodes B and C.

Proxy stripping is illustrated in Fig. 6.3 (b). Recall from Section 6.1 that in RPR spatial reuse is achieved by means of shortest path routing and destination stripping, as shown in Fig. 6.3 (b) for source node A and destination node B. Note that only source node A (shortest path routing) and destination node B (destination stripping) are involved, but the intermediate node attached to the star subnetwork performs simple forwarding on the ring. In this case, the node attached to the star subnetwork does not pull packets destined for node B from the ring and does not send them across the star subnetwork since the path on the counterclockwise ring is the shortest path between nodes A and B in terms of hops. If, however, the short-cuts of the star subnetwork provide a shorter path than either peripheral fiber ring intermediate nodes attached to the star subnetwork perform proxy stripping instead of simple forwarding. Proxy stripping makes use of RPR's shortest path routing and destination stripping features. As shown in Fig. 6.3 (b) for source node A and destination node C, node A sends its packets destined for node C to its closest proxy-stripping node (shortest path routing). Now, instead of simply forwarding the packets on the clockwise peripheral ring the proxy-stripping node pulls the packets from the ring and sends them across the single-hop star subnetwork to the proxy-stripping node closest to destination node C by using the MAC protocol of the given star subnetwork. The receiving proxy-stripping node forwards the packets on the shortest path along the counterclockwise ring towards node C which finally takes the packet from the ring (destination stripping). Practically, proxy stripping can be done by monitoring an arriving packet's source and destination MAC addresses and making a look-up in each proxy-stripping node's topology database in order to decide whether a given packet has to be proxy stripped or not. The topology database is built and continously updated by using RPR's built-in topology discovery protocol (see Section 9.1.3) [155].

By means of proxy stripping the single-hop short-cuts of the star subnetwork are exploited to decrease the mean hop distance and diameter of the network. Thus, packet transmissions

require fewer bandwidth resources on the ring, resulting in an increased spatial reuse and an improved throughput-delay performance.

6.3 Architecture

Clearly, there are several ways to add fiber links to the bidirectional ring. Among others, individual pairs of ring nodes can be interconnected by means of fiber short-cuts [94] or all ring nodes can be interconnected via a central hub node that consists of multiple working and stand-by wavelength routers [151]. In this work only a *subset* of ring nodes are interconnected by means of a single-hop star WDM network. The star network's hub consists of an AWG in parallel with a single broadcast PSC. It was shown in [141] that using a star coupler in parallel with a wavelength router not only protects the wavelength router and thus avoids the single point of failure of the star subnetwork but also combines the respective strengths of wavelength routing devices (wavelength router) and wavelength insensitive devices (star coupler) in an efficient manner.

6.3.1 Building Blocks

Let us first briefly describe the functionality of the underlying building blocks used in the proposed network architecture, which are depicted in Fig. 6.4 (a)–(f):

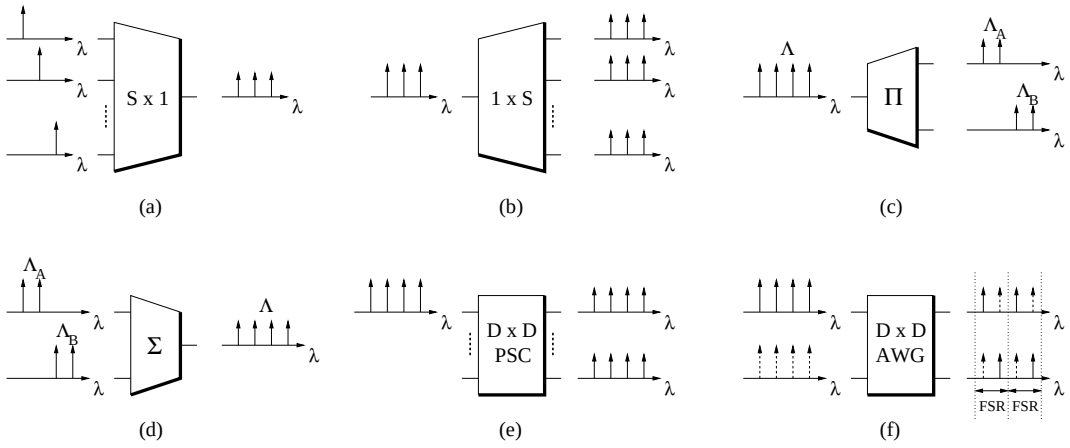


Figure 6.4: Architectural building blocks: (a) $S \times 1$ combiner, (b) $1 \times S$ splitter, (c) waveband partitioner, (d) waveband departitioner, (e) $D \times D$ passive star coupler (PSC), and (f) $D \times D$ arrayed-waveguide grating (AWG) with $D = 2$.

- (a) *Combiner*: An $S \times 1$ combiner has S input ports and 1 output port, where $S \geq 1$. It collects wavelength channels from all S input ports and combines them onto the common output port. To avoid channel collisions at the output port of the combiner, the collected wavelength channels must be different. Thus, a given wavelength channel can be used only at one of the S input ports at any time.
- (b) *Splitter*: A $1 \times S$ splitter has 1 input port and S output ports, where $S \geq 1$. It equally distributes all incoming wavelength channels to all S output ports. Hence, a given wavelength channel can be received at all S output ports.

- (c) *Waveband partitioner*: A waveband partitioner Π has 1 input port and 2 output ports. It partitions an incoming set of Λ contiguous wavelength channels into two wavebands (subsets of wavelength channels) of Λ_A and Λ_B contiguous wavelength channels, where $1 \leq \Lambda_A, \Lambda_B \leq \Lambda$ and $\Lambda = \Lambda_A + \Lambda_B$. Each waveband is routed to a different output port.
- (d) *Waveband departitioner*: A waveband departitioner Σ has 2 input ports and 1 output port. It collects two different wavebands consisting of Λ_A and Λ_B contiguous wavelength channels from the upper and lower input port, respectively. The combined set of Λ wavelength channels is launched onto the common output port, where $1 \leq \Lambda_A, \Lambda_B \leq \Lambda$ and $\Lambda = \Lambda_A + \Lambda_B$.
- (e) *Passive star coupler (PSC)*: A $D \times D$ PSC has D input ports and D output ports, where $D \geq 1$. It works similar to a $D \times 1$ combiner and $1 \times D$ splitter interconnected in series. Accordingly, it collects wavelength channels from all D input ports and equally distributes them to all D output ports. Similar to the splitter, a given wavelength channel can be received at all D output ports and, similar to the combiner, to avoid channel collisions at the output ports a given wavelength channel can be used only at one of the D input ports at any time.
- (f) *Arrayed-waveguide grating (AWG)*: A $D \times D$ AWG has D input ports and D output ports, where $D \geq 1$. Without loss of generality, we consider a 2×2 AWG to explain the properties of an AWG. Fig. 6.4 (f) illustrates a scenario where four wavelengths are fed into both AWG input ports. Let us first consider only the upper input port. The AWG routes every second wavelength to the same output port. This period of the wavelength response is called free spectral range (FSR). In our example, there are two FSRs, each containing two wavelengths. Generally, the FSR of a $D \times D$ AWG consists of D contiguous wavelengths, i.e., the physical degree of an AWG is identical to the number of wavelengths per FSR. As depicted in Fig. 6.4 (f), this holds also for the lower AWG input port. Note that the AWG routes wavelengths such that no collisions occur at the AWG output ports, i.e., each wavelength can be applied at all AWG input ports simultaneously. In other words, with a $D \times D$ AWG each wavelength channel can be spatially reused D times, as opposed to the PSC. Also, note that each FSR provides one wavelength channel for communication between a given pair of AWG input and output ports. Hence, using R FSRs allows for R simultaneous transmissions between each AWG input-output port pair and the total number of wavelength channels available at each AWG port is given by $R \cdot D$, where $R \geq 1$.

6.3.2 Network Architecture

As shown in Fig. 6.5, the proposed network consists of the RPR bidirectional *ring subnetwork* and a *star subnetwork*:

Ring Subnetwork

The RPR ring subnetwork interconnects $N \geq 1$ nodes which are subdivided into two subgroups of $N_{rs} = D \cdot S$ ring-and-star homed nodes, and $N_r = N - N_{rs}$ ring homed nodes, with $D \geq 1$ and $S \geq 1$. The N_{rs} ring-and-star homed nodes are equally spaced among the N_r ring homed nodes on the ring, as illustrated in Fig. 6.5 for $N = 16$ and $N_{rs} = D \cdot S = 2 \cdot 2 = 4$ (and $N_r = N - N_{rs} = 12$). Unlike the ring homed nodes, the ring-and-star homed nodes are also

attached to the star subnetwork. The ration N/N_{rs} denotes the number of ring homed nodes between two adjacent nodes ring-and-star homed nodes, including one of the two ring-and-star homed nodes, and is assumed to be an integer.

Star Subnetwork

The star subnetwork is based on a central hub which consists of a $D \times D$ AWG in parallel with a $D \times D$ PSC, where $D \geq 1$. Each ring-and-star homed node i, $i = 1, \ldots, N_{rs}$, has a home channel λ_i on the PSC, i.e., a unique wavelength channel λ_i on which node i receives data transmitted over the PSC. In addition, there is a control wavelength channel λ_c on the PSC. Consequently, there are $\Lambda_{PSC} = N_{rs} + 1 = D \cdot S + 1$ wavelength channels on the PSC, which make up the PSC waveband. The AWG waveband consists of $\Lambda_{AWG} = D \cdot R$ contiguous data wavelength channels, where $R \geq 1$ denotes the number of used FSRs of the underlying $D \times D$ AWG. A total of $\Lambda = \Lambda_{AWG} + \Lambda_{PSC}$ contiguous wavelength channels are operated in the star subnetwork (as further detailed in Section 6.4).

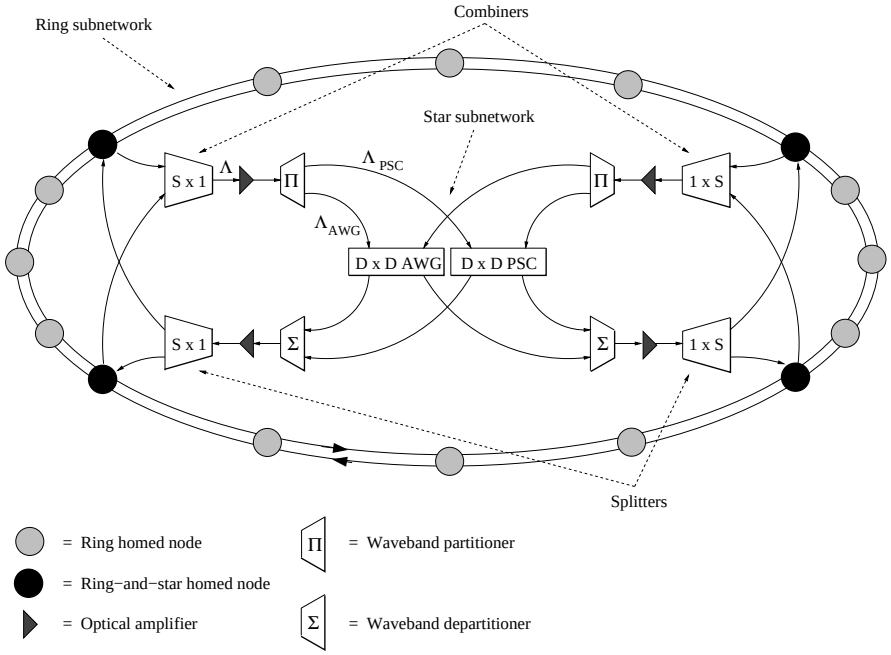


Figure 6.5: Network architecture with $N = 16$ nodes, where $N_{rs} = D \cdot S = 2 \cdot 2 = 4$ are ring-and-star homed nodes and $N_r = N - N_{rs} = 12$ are ring homed nodes. There are $\Lambda_{PSC} = D \cdot S + 1 = 2 \cdot 2 + 1 = 5$ wavelengths on the PSC, $\Lambda_{AWG} = D \cdot R = 2 \cdot R$ wavelengths on the AWG, for a total of $\Lambda = \Lambda_{PSC} + \Lambda_{AWG} = 5 + 2 \cdot R$ wavelengths in the star subnetwork.

The signals from S ring-and-star homed nodes on the Λ wavelength channels are transmitted on S distinct fibers to a $S \times 1$ combiner, which combines the signals onto the Λ wavelength channels of one fiber leading to a waveband partitioner. The waveband partitioner partitions the set of Λ wavelengths into the AWG and PSC wavebands, which are fed into an AWG and PSC input port, respectively. The signals from the opposite AWG and PSC output ports are collected by a waveband departitioner and then equally distributed to the S ring-and-star homed nodes by a $1 \times S$ splitter. If necessary, optical amplifiers are used between combiner

and partitioner as well as splitter and departitioner to compensate for attenuation and insertion losses of the star subnetwork. A total of D of these arrangements, each consisting of combiner, amplifier, waveband partitioner, waveband departitioner, amplifier, and splitter, are used to connect all $N_{rs} = D \cdot S$ ring-and-star homed nodes to the central hub.

6.3.3 Node Architecture

Next, let us take a closer look at the structure of both ring homed and ring-and-star homed nodes.

Ring Homed Node

The architecture of ring homed nodes is identical to that of RPR nodes described in Section 6.1. As shown in Fig. 6.6, every ring homed node is equipped with two FTs and two FRs, one for each ring. Both FT and FR operate at the single wavelength channel of the corresponding ring. Each ring homed node has separate transit and station queues for either ring.

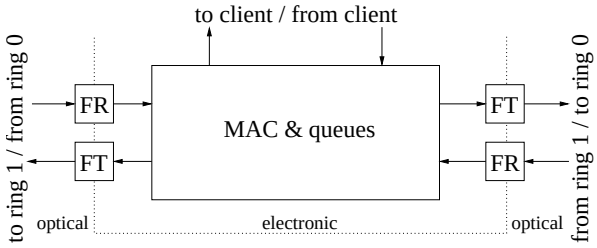


Figure 6.6: Ring homed node: Architecture (same as RPR node).

Fig. 6.7 depicts path and queue selection that has already been discussed in Section 6.1 in more detail. Remember that for each direction a ring homed node has one or two transit queues (here two), one transmit queue, and one receive queue (the additional addMAC queue for control is not shown). For a packet arriving from the client first the appropriate ring direction is determined, usually according to the shortest path, then the packet is stored in the stage queue corresponding to that direction. A packet arriving from the ring is either stripped from the ring and put in the receive queue if the packet is destined to the node itself or put in one of the two transit queues to be forward on the ring. The service among transmit and transit queues, i.e., choosing the next packet to send on the ring, is arbitrated according to the scheduling algorithms reviewed in Section 6.1.

Ring-and-Star Homed Node

Fig. 6.8 depicts the architecture of a ring-and-star homed node with PSC data channel λ_i, where $\lambda_i \in \{1, 2, \ldots, D \cdot S\}$. Each ring-and-star homed node has the same number and type of transceivers and queues as a ring homed node for transmission and reception on both rings. In addition, each ring-and-star homed node has several transceivers which are attached to the star subnetwork by means of a pair of outgoing and incoming fibers. The outgoing fiber is connected to a combiner and the incoming fiber is connected to the splitter which is attached to the opposite AWG-PSC input ports.

As shown in Fig. 6.8, for control transmission on the star subnetwork each ring-and-star homed node is equipped with a FT tuned to the control wavelength channel λ_c of the PSC

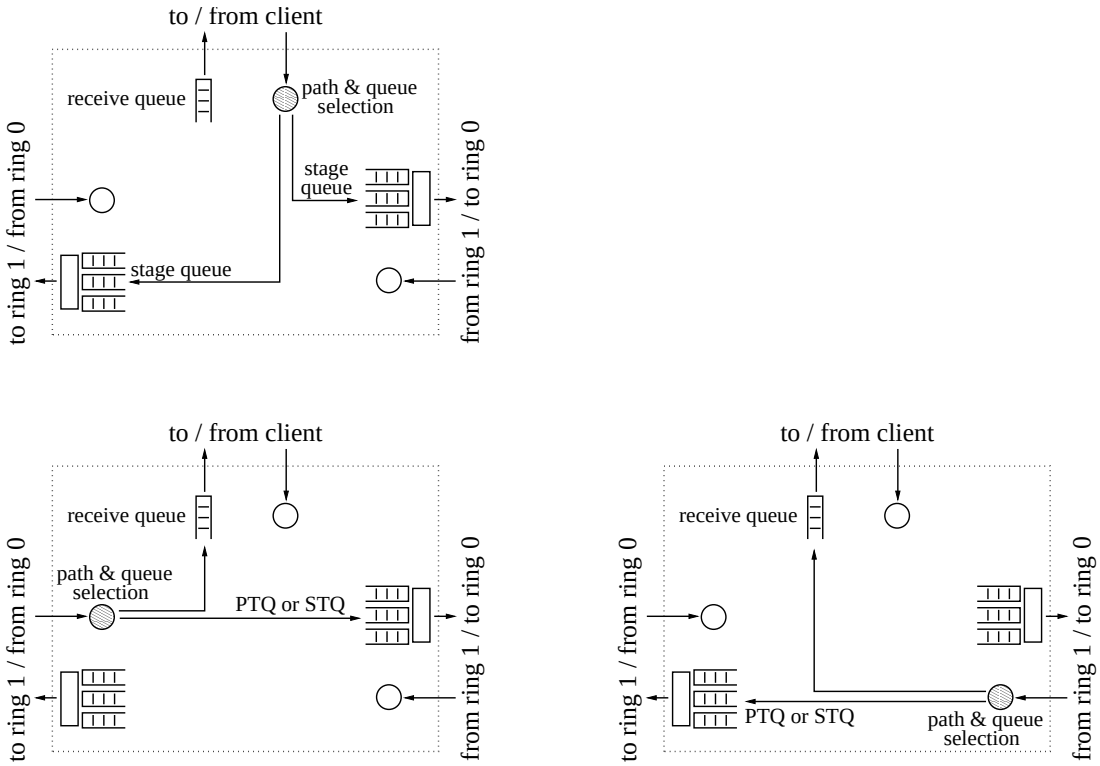


Figure 6.7: Ring homed node: Queue structure and path and queue selection for a packet arriving from the client or from either of the two rings (same as in RPR).

waveband, which consists of $\Lambda_{PSC} = 1 + D \cdot S$ wavelength channels. The remaining $D \cdot S$ wavelength channels of the PSC waveband and all $\Lambda_{AWG} = D \cdot R$ wavelength channels of the AWG waveband are accessed for data transmission by a TT whose tuning range equals $D \cdot S + \Lambda_{AWG} = D(S + R)$. Similarly, for control reception on the star subnetwork each ring-and-star homed node is equipped with a FR tuned to the control wavelength channel λ_c of the PSC waveband. For data reception on the PSC each ring-and-star homed node has a separate FR operating at its own dedicated *home channel* $\lambda_i \in \{1, 2, \ldots, D \cdot S\}$. Each data wavelength channel of the PSC waveband is dedicated to a different ring-and-star homed node for reception. Data packets transmitted on PSC data wavelength channels do not suffer from receiver collisions (a receiver collision occurs when the receiver of the intended destination node is not tuned to the wavelength channel on which the data packet was sent by the corresponding source node). Moreover, on the wavelength channels of the AWG waveband, data packets are received by a TR whose tuning range equals $\Lambda_{AWG} = D \cdot R$. All transceivers of the star subnetwork are connected to the station queues. Note that the required tuning range of the tunable receiver (Λ_{AWG}) is smaller than that of the tunable transmitter ($D \cdot S + \Lambda_{AWG}$). These requirements take into account the current state-of-the-art of tunable transceivers. While fast tunable transmitters with a relatively large tuning range have been shown to be feasible [107, 53], tunable receivers are less mature in terms of tuning time and/or tuning range.

Fig. 6.9 depicts the queue structure of a ring-and-star homed node and illustrates the path and queue selection for packet arriving from the client, the star, or one of the two rings. Equal

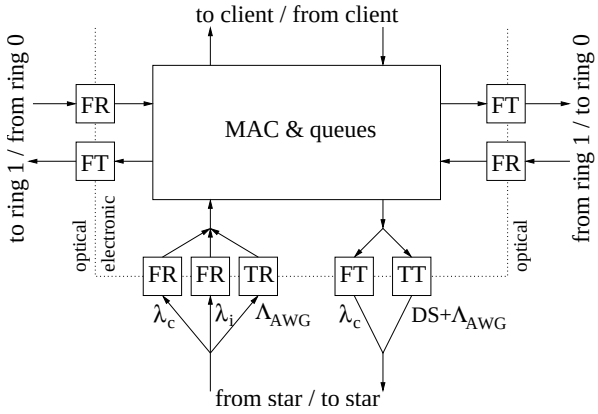


Figure 6.8: Ring-and-star homed node: Architecture with home channel $\lambda_i \in \{1, 2, \ldots, D \cdot S\}$.

to a regular RPR node, each ring-and-star homed node has for each ring direction the following queues: one or two ring transit queues (depending on the operation mode), one ring transmit queue, and one receive queue (the additional queue for control is not shown). An additional queue structure consisting of stage queue, PTQ, and STQ is deployed for the star. To send locally generated traffic to the star subnetwork the ring-and-star homed node puts the packet in the star transmit queue. Furthermore, packets that are pulled from the ring (coming in from both directions of the ring) and forwarded onto the star subnetwork are placed in the star transit queue (single-queue mode) or one of two star transit queues according to their priority (dual-queue mode). Traffic that is received from the star subnetwork and needs to be forwarded on either ring is placed in the ring transit queue of the appropriate direction. Alternatively, if the packet received from the star is destined to the nodes itself the packet is put into the receive queue (The additional queue for sending control packet on the star subnetwork is not shown.) Each of the three queue structures stage queue, PTQ, and STQ are arbitrated according to the policy described in Section 6.1 to select the next packet to send on the ring or star, respectively.

6.4 Access Protocol

Below we discuss RINGOSTAR's MAC protocol, or 'access protocol', which controls usage of the wavelength channels.

6.4.1 Wavelength Assignment in Star Subnetwork

Fig. 6.10 illustrates how the Λ contiguous wavelength channels of the star subnetwork are used for control and data transmission. Time is divided into frames which are repeated periodically. Each frame consists of $F \geq D \cdot S$ slots, where one slot is equal to the transmission time of a control packet (function and format of a control packet are defined in the following subsection). As shown in Fig. 6.10, all $D \cdot S$ home channels of the PSC waveband and all wavelength channels of the AWG waveband are used for data transmission. All these data wavelength channels are not statically assigned to nodes. Instead, access to these wavelength channels is arbitrated by broadcasting control packets on the control wavelength channel λ_c of the PSC prior to transmitting data packets, as explained in greater detail in the following

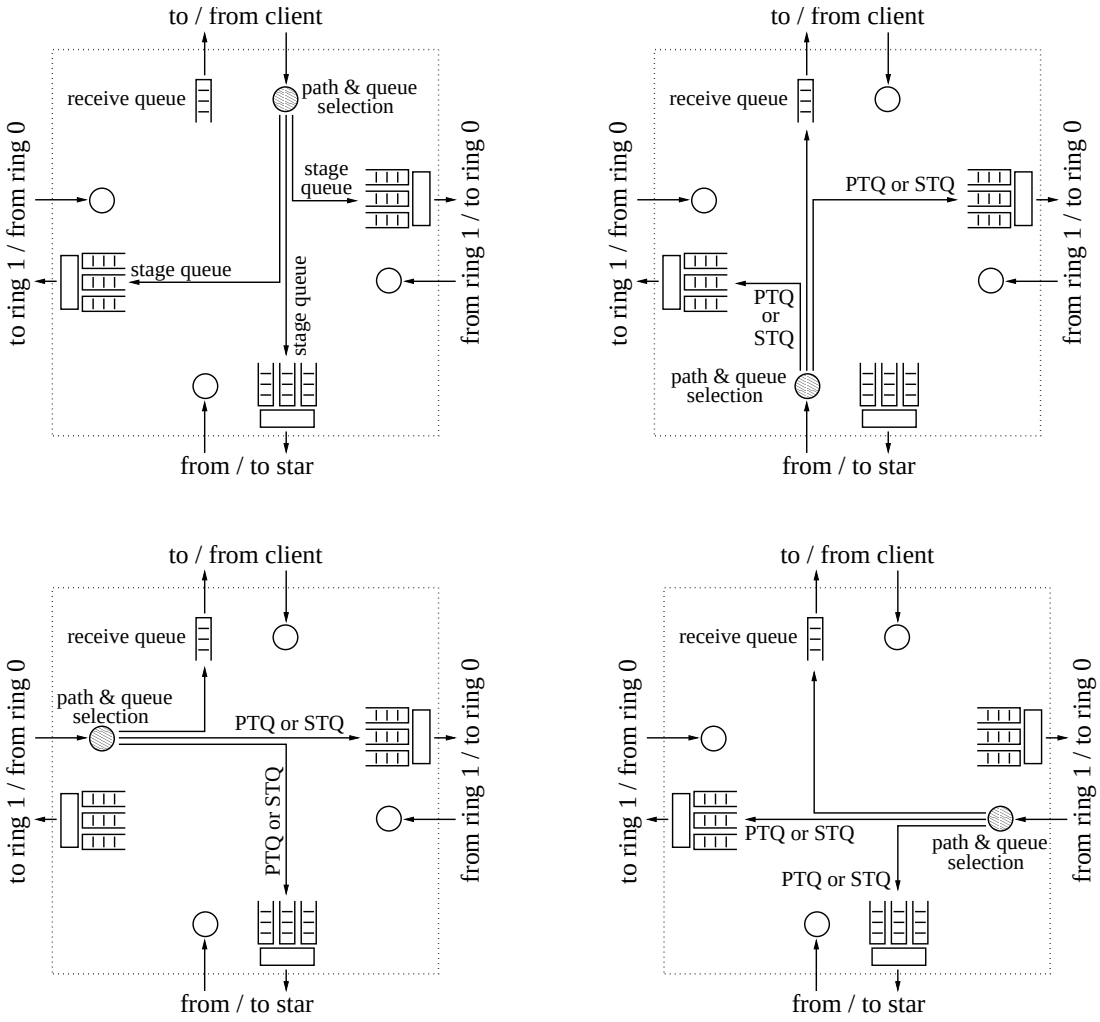


Figure 6.9: Ring-and-star homed node: Queue structure and path and queue selection for a packet arriving from the client, from the star, or from either of the two rings.

subsection. Control packets are allowed to be sent on λ_c during the first $D \cdot S$ slots of each frame. More precisely, each of these $D \cdot S$ slots is dedicated to a different ring-and-star homed node such that channel collisions of control packets are avoided. The remaining $(F - D \cdot S)$ slots of each frame can be used for data transmission on λ_c. Note that data packets sent during these slots on λ_c are received by all ring-and-star homed nodes by using their receiver fixed tuned to λ_c. Thus, these slots allow for broadcasting in the star subnetwork.

6.4.2 Wavelength Access

All N ring and ring-and-star homed nodes use the single-queue or dual-queue scheduling algorithm to arbitrate service among transit and station queues of the ring subnetwork, as outlined in Section 6.3.3. In the following, we consider unicast (point-to-point) traffic. Each of the N nodes sends data packets on the shortest path to the corresponding destination node. Next, we specify the shortest path routing for both ring and ring-and-star homed nodes. Let one hop denote the distance between two adjacent nodes. Adjacent nodes can either be two

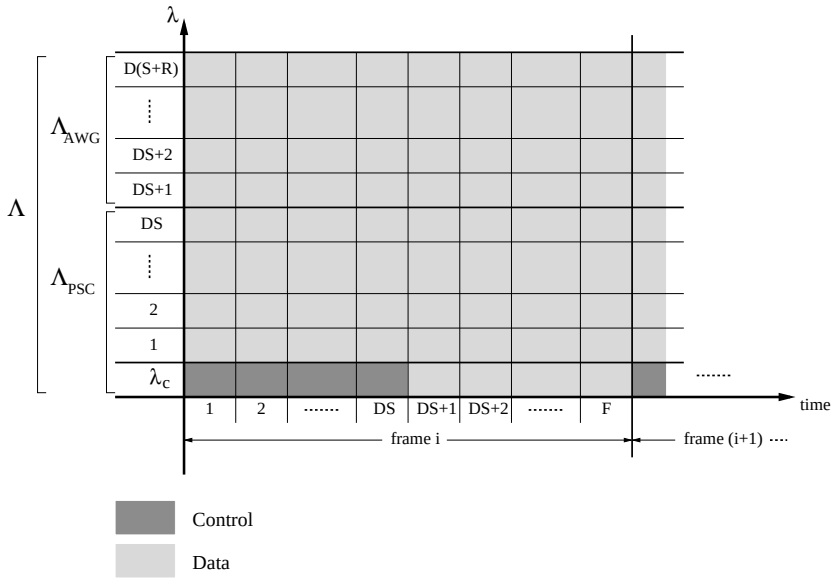


Figure 6.10: Wavelength assignment in star subnetwork.

neighboring nodes on the ring or two nodes interconnected via the single-hop star subnetwork (this holds only for ring-and-star homed nodes). We define the following variables for a given pair of source node s and destination node d, where $s, d \in \{0, 1, \ldots, N-1\}$:

- h_{s_rs}: Hop distance between source node s and the closest ring-and-star homed node.
- h_{d_rs}: Hop distance between destination node d and the closest ring-and-star homed node.
- $h^{ring}_{min_{s\_d}}$: Minimum hop distance between source node s and destination node d on the ring *without* using the short-cuts of the star subnetwork.
- $h^{star}_{min_{s\_d}}$: Minimum hop distance between source node s and destination node d on the ring *with* using the short-cuts of the star subnetwork. Note that $h^{star}_{min_{s\_d}} = h_{s\_rs} + 1 + h_{d\_rs}$.

Ring Homed Nodes

Generally speaking, if the hop distance between a given source ring homed node s and destination node d is 'small enough', the ring homed node sends the data packet(s) on the ring without using the short-cuts of the star subnetwork. More precisely, if $h^{ring}_{min_{s\_d}} \leq h^{star}_{min_{s\_d}}$, then source node s sends the data packet(s) to destination node d along the ring on the shortest path by choosing the appropriate fiber ring. Destination node d takes the transmitted data packet(s) from the ring. Note that in this case intermediate ring-and-star homed nodes forward the data packet(s) on the ring rather than sending them across the star subnetwork. However, if $h^{ring}_{min_{s\_d}} > h^{star}_{min_{s\_d}}$, i.e., if the short-cuts of the star subnetwork form a shorter path between nodes s and d than either peripheral fiber ring, the source node s sends the data packet(s) to its closest ring-and-star homed node. Note that the chosen direction does not necessarily have to be the same as that used in shortest path routing on the ring. The corresponding ring-and-star homed node pulls the data packet(s) from the ring, as described in greater detail in the following.

Ring-and-Star Homed Nodes

To pull data packet(s) from the ring, ring-and-star homed nodes perform the *proxy stripping* technique. With proxy stripping, a given ring-and-star homed node pulls only data packets from the ring whose source and destination addresses satisfy the condition $h^{ring}_{min_{s\_d}} > h^{star}_{min_{s\_d}}$. A ring-and-star homed node puts data packet(s) pulled from the ring in one of the two corresponding star transit queues belonging to its TT that is attached to the star subnetwork. The star transit queue of the TT is chosen according to the priority of the pulled data packet(s). The service among these two star transit queues that store in-transit traffic coming from the ring and the star transmit queue that stores locally generated traffic is arbitrated by applying the same scheduling algorithms as used on the ring (see Section 6.3.3). That is, ring-to-star in-transit traffic is given priority over star traffic locally generated by the proxy-stripping node. Similar to the transit queues on the ring, the star transit queues thus provide a lossless path for in-transit traffic. To guarantee losslessness the star subnetwork needs to be dimensioned properly, as analyzed in greater detail in Section 8.3.2. Depending on the traffic pattern as well as the number and location of the proxy-stripping nodes the amount of proxy-stripped traffic may become large. To provide lossless delivery of proxy-stripped packets the star subnetwork in general needs to operate at a higher line rate than the ring subnetwork (see Section 8.3.2). Alternatively, each proxy-stripping node may be equipped with more than one star data transceiver, each operating at the same line rate as the ring transceivers. For more details on star WDM networks with multiple transceivers at each node the interested reader is referred to [154]. Note that the star transit queues (as well as the star station queues) of each ring-and-star homed node need to be added to the RPR MAC layer.

Prior to transmitting a data packet, the corresponding ring-and-star homed node broadcasts a control packet on λ_c to all N_{rs} ring-and-star homed nodes in its assigned slot of the upcoming frame by using its FT. The control packet consists of three fields: (*i*) destination address of the ring-and-star homed node that is closest to destination node d, (*ii*) length of the corresponding data packet, and (*iii*) priority of the corresponding data packet. After announcing the data packet in its assigned control slot, the ring-and-star homed node transmits the corresponding data packet on the home channel λ_i of the addressed ring-and-star homed node in the subsequent L slots by using its TT, where $\lambda_i \in \{1, 2, \ldots, D \cdot S\}$ and L denotes the length of the data packet in number of slots. Data packets are sent within the same frame as the corresponding control packet and have a maximum length of $(F - D \cdot S)$ slots, i.e., $1 \leq L \leq F - D \cdot S$. We note that due to this assumption a small fraction of each home channel λ_i is not used at the beginning of each frame. However, this could be easily avoided by letting nodes send data packets across the boundary of adjacent frames. After an end-to-end propagation delay of the PSC of the star subnetwork all ring-and-star homed nodes receive the broadcast control packet by using their FRs fixed tuned to λ_c. The corresponding data packet is successfully received at the addressed ring-and-star homed node by using its FR fixed tuned to λ_i, unless one or more other ring-and-star homed nodes have transmitted data packets on λ_i in at least one of the aforementioned L slots. In the latter case, all involved data packets are assumed to be corrupted due to (channel) collision and have to be retransmitted. Collided data packets are kept in the queues until the transmission is successful. Note that due to the fact that control packets are sent collisionfree all ring-and-star homed nodes are aware of the original order of the corresponding data packets. As a consequence, even though collided data packets need to be retransmitted, the receiving ring-and-star homed nodes are able to restore the original sequence of data packets and thus maintain in-order packet delivery.

The retransmission of collided data packets works as follows. Due to the dedicated access control of the control channel λ_c, collisions of control packets are prevented. Therefore, for collided data packets no control packets have to be retransmitted. Instead, each ring-and-star homed node is able to find out which transmitted data packets have experienced channel collision by processing the previously (successfully) transmitted control packets. More precisely, the index j, $1 \leq j \leq D \cdot S$, of the used control slot and the destination and length fields of the control packet enable each ring-and-star homed node to determine whether the corresponding data packet has collided or not. Collided data packets are not retransmitted on the home channels of the PSC but across the AWG by using one of the Λ_{AWG} wavelength channels. Given the index j of the control slot, which uniquely identifies not only the given source ring-and-star homed node but more importantly also the input port of the AWG to which it is attached, together with the destination field of the corresponding control packet all ring-and-star homed nodes are able to determine the wavelength in each FSR of the AWG which provides a single-hop connection between the corresponding pair of source and destination ring-and-star homed nodes. The actual retransmissions on the chosen wavelength channels are scheduled in a distributed fashion by all ring-and-star homed nodes. The scheduling starts at the beginning of frame $(i+1)$ upon receiving the control packets in frame i after one end-to-end propagation delay of the PSC of the star subnetwork. At the end of every frame each ring-and-star homed node collects all control packets belonging to collided data packets. By using each control packet's priority field, each ring-and-star homed node first processes all high-priority control packets and then all low-priority control packets. Control packets of the same priority class are randomly chosen for scheduling. All ring-and-star homed nodes deploy the same random algorithm and same seed and therefore build the same schedule. Note that randomizing the scheduling counteracts the static control slot assignment, resulting in an improved fairness among the ring-and-star homed nodes. Otherwise, source nodes with a smaller index j would be more successful in the scheduling than nodes with a larger index. The corresponding data packets of the selected control packets are scheduled on a first-fit basis starting from the lowest possible wavelength channel at the earliest possible time. The data packet is retransmitted on the corresponding AWG wavelength channel at the scheduled time. After the successful retransmission of a given data packet across the AWG, the corresponding ring-and-star homed receiving node puts the data packet in the receive queue if the data packet is destined for itself. Otherwise, the ring-and-star homed node forwards the data packet on the ring towards the destination node d on the shortest path by using the appropriate fiber ring and placing the data packet in the corresponding ring transit queue. Destination node d finally takes the data packet from the ring. We note that the aggregated length of the collided packets can be larger than $F - D \cdot S$. This fact does not pose any problems since the retransmission takes place over the AWG where transmissions are permitted to be scheduled across frame boundaries.

Besides pulling data packets from the ring and forwarding them on the ring, ring-and-star homed nodes also generate traffic. Note that in this case we have $h_{s\_rs} = 0$. Again, if $h^{ring}_{min_{s\_d}} \leq h^{star}_{min_{s\_d}}$, then the ring-and-star homed source node s transmits the data packet on that fiber ring which provides the shortest path to destination node d. Otherwise, if $h^{ring}_{min_{s\_d}} > h^{star}_{min_{s\_d}}$, then the ring-and-star homed source node s sends the data packet across the star subnetwork to the corresponding ring-and-star homed node which is either the destination itself or forwards the data packet onwards to node d via the shortest path ring. (Re)transmission and reception of the data packet on the star subnetwork are done in the same way as explained above.

6.5 Discussion

In this section we show that RINGOSTAR significantly reduces the mean hop distance between the nodes compared to both bidirectional rings with shortest path routing and meshed rings. The reduced mean hop distance translates into higher spatial reuse on the ring and therefore to an increased capacity. More intuitively speaking, the star provides shortcuts for transmissions which normally would require to traverse a large number of ring links. The skipped ring links can be used for other transmissions simultaneously.

In the computation of the mean hop distance of RINGOSTAR we assume *uniform* traffic, i.e., each node generates the same amount of traffic and a given data packet is destined to any of the $(N-1)$ nodes with equal probability $1/(N-1)$. (The assumption of uniform traffic is realistic in metro core networks with any-to-any traffic demands between central offices [156]. By assuming uniform traffic we are also able to compare the mean hop distance of RINGOSTAR to that of unidirectional and bidirectional rings of Eqs. (5.1) and (5.3), respectively.)

The mean hop distance $\overline{h}$ of RINGOSTAR is given by

$$\begin{aligned} \overline{h} &= E[hops] && (6.1)\\ &= \frac{1}{N(N-1)} \sum_{i=0}^{N-1} \sum_{j=0, j\neq i}^{N-1} \min\left\{h^{ring}_{min_{i\_j}}, h^{star}_{min_{i\_j}}\right\} && (6.2)\\ &= \frac{1}{N(N-1)} \sum_{i=0}^{N-1} \sum_{j=0, j\neq i}^{N-1} \min\left\{h^{ring}_{min_{i\_j}}, h_{i\_rs} + 1 + h_{j\_rs}\right\}. && (6.3) \end{aligned}$$

By exploiting the architectural symmetry of RINGOSTAR Eq. (6.3) becomes

$$\overline{h} = \frac{D \cdot S}{N(N-1)} \sum_{i=0}^{\frac{N}{D\cdot S}-1} \sum_{j=0, j\neq i}^{N-1} \min\left\{h^{ring}_{min_{i\_j}}, h_{i\_rs} + 1 + h_{j\_rs}\right\}, \quad (6.4)$$

where

$$h^{ring}_{min_{i\_j}} = \min\{|i-j|, N-|i-j|\} \quad (6.5)$$

and

$$h_{l\_rs} = \min\left\{l \bmod \frac{N}{D\cdot S}, \frac{N}{D\cdot S} - \left(l \bmod \frac{N}{D\cdot S}\right)\right\}, \quad (6.6)$$

with $l \in \{i, j\}$.

Fig. 6.11 depicts the mean hop distance $\overline{h}$ vs. the number of nodes N for RINGOSTAR with $D \cdot S \in \{4, 8, 16, 32, 64, 128, 256\}$, the unidirectional ring with destination stripping (see Eq. (5.1), and the bidirectional ring with destination stripping and shortest path routing (see Eq. (5.3)). Clearly, for all types of network $\overline{h}$ increases with increasing N. However, note that the slope of the curves differs for the two rings and the various configurations of RINGOSTAR. The unidirectional ring features the largest mean hop distance and slope. Due to its dual-fiber structure and shortest path routing the bidirectional ring provides a mean hop distance and slope that are approximately 50% smaller than those of the unidirectional ring. We observe from Fig. 6.11 that in RINGOSTAR $D \cdot S = 4$ ring-and-star homed nodes are sufficient to decrease the mean hop distance and slope significantly compared to unidirectional and bidirectional rings. A small mean hop distance improves the network capacity by alleviating

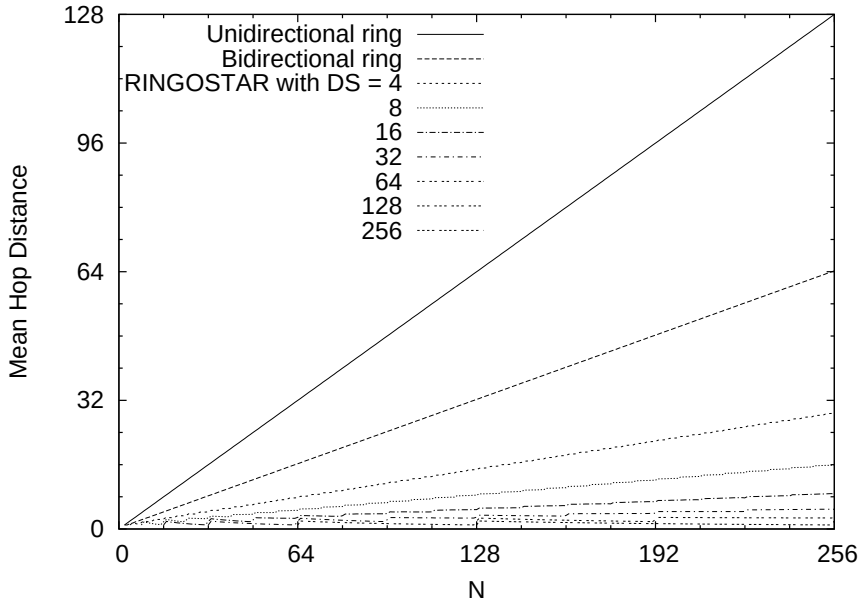


Figure 6.11: Mean hop distance $\overline{h}$ of unidirectional ring with destination stripping, bidirectional ring with destination stripping and shortest path routing, and RINGOSTAR with different $D \cdot S \in \{4, 8, 16, 32, 64, 128, 256\}$ vs. number of nodes N.

the forwarding burden of each node. With a smaller slope new nodes can be added to the network without deteriorating the mean hop distance significantly. As shown in Fig. 6.11, increasing the number of ring-and-star homed nodes $D \cdot S$ up to 64 further decreases $\overline{h}$. However, attaching more than 64 ring nodes to the central star does not further decrease $\overline{h}$ significantly. This fact is also illustrated in Fig. 6.12 which depicts the mean hop distance $\overline{h}$ of RINGOSTAR vs. the number of ring-and-star homed nodes $D \cdot S$ for a fixed number of nodes N (for comparison we also show $\overline{h}$ of both unidirectional and bidirectional rings which are independent from $D \cdot S$, of course). To demonstrate the potential of the proxy stripping technique we choose a rather large value of $N = 256$ which is the maximum number of nodes supported by RPR. As mentioned above, connecting only a few nodes to the star network results in RINGOSTAR outperforming its ring counterparts clearly.

Network type	Mean Hop Distance
Unidirectional Ring	128.0
Bidirectional Ring	64.25
RINGOSTAR w/	
DS = 4	28.7941
DS = 8	15.9
DS = 16	8.7
DS = 32	4.91176
DS = 64	2.97059
DS = 128	1.98824
DS = 256	1.0

Table 6.1: Mean hop distance in RINGOSTAR: Numerical values for $N = 256$.

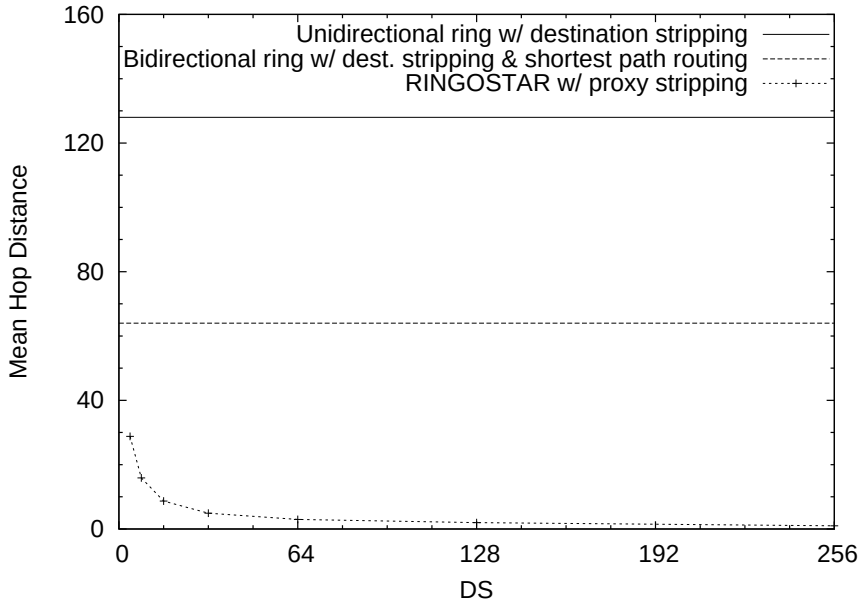


Figure 6.12: Mean hop distance $\overline{h}$ of unidirectional ring with destination stripping, bidirectional ring with destination stripping and shortest path routing, and RINGOSTAR with proxy stripping vs. $D \cdot S$ for $N = 256$.

To quantify the benefit of proxy stripping, Table 6.1 lists the mean hop distance of both ring networks and RINGOSTAR with different $D \cdot S$ values for fixed $N = 256$. Due to destination stripping the mean hop distance of the unidirectional ring is equal to 128, i.e., half the maximum hop distance. By using an additional counterdirectional fiber ring in conjunction with shortest path routing the mean hop distance $\overline{h}$ is further reduced by a factor of approximately 2 in the bidirectional ring network, resulting in $\overline{h} = 64.25$. RINGOSTAR with $D \cdot S = 4$ achieves a mean hop distance of $\overline{h} = 28.7941$ which translates into a reduction of the mean hop distance by a factor of more than 2 compared to the bidirectional ring. Similarly, for $D \cdot S = 64$ the mean hop distance is equal to $\overline{h} = 2.97059$ which corresponds to an improvement by a factor of more than 21. Thus, by WDM upgrading only $\frac{64}{256} = 25\%$ of the ring nodes and attaching them to the star subnetwork the mean hop distance is less than 5% of that of the bidirectional ring. Note that in RINGOSTAR the minimum achievable mean hop distance $\overline{h} = 1.0$ is obtained if all 256 nodes are attached to the star subnetwork. In this case, each pair of source and destination nodes can communicate in one single hop at the expense of WDM upgrading and interconnecting all nodes via the star subnetwork.

6.6 Conclusions

We have proposed the RINGOSTAR, a multichannel extension of RPR in particular and optical single-channel ring networks in general by using WDM. Most previously reported multichannel extensions deploy WDM on the ring. All these WDM extensions have in common that all nodes have to be WDM upgraded, be it by arrays of fixed-tuned transceivers, tunable transceivers, wavelength multiplexers and demultiplexers. Furthermore, applying WDM on the ring achieves only a limited spatial reuse of wavelengths and thus a limited increase of

capacity.

Our proposed multichannel extension follows an entirely different direction to WDM upgrade RPR and optical single-channel ring networks. In our approach, only a subset of ring nodes need to be upgraded with a single tunable transceiver. The subset of ring nodes are interconnected through a passive AWG and PSC based wavelength-routing single-hop star network by using dark fibers which are abundantly available in metropolitan areas. Unlike previous multichannel extensions, we deploy WDM on the star subnetwork rather than on the ring. The resultant hybrid ring-star architecture, termed RINGOSTAR, provides an evolutionary and cost-effective dark-fiber WDM upgrade in that it builds on the single-channel network and node architecture. In doing so, RINGOSTAR benefits from the performance enhancing techniques of RPR, e.g., destination stripping, shortest path routing, service differentiation, QoS support, electronic packet processing and signal regeneration. Owing to its hybrid architecture, RINGOSTAR is able to combine the merits or ring topology (fault tolerance) and single-hop star topology (high bandwidth utilization, inherent transparency). By using the novel concept of proxy stripping, data packets are sent on single-hop short-cuts across the star subnetwork. As a result, in RINGOSTAR the overall mean hop distance is dramatically decreased and the capacity is significantly increased due to improved spatial wavelength reuse on both star and ring subnetworks.

Of course, the gained capacity is not for free. The higher the targeted capacity, the more nodes must be connected to the star subnetwork. Each of these nodes must be upgraded with tunable transceivers and the star's access protocol requires additional processing capacity. An AWG and/or PSC, splitters, and combiners must be deployed, and all nodes and components must be interconnected by fibers. However, by means of analysis we have found that RINGOSTAR clearly outperforms unidirectional, bidirectional, and meshed ring networks, in which *all* nodes must be upgraded, in terms of mean hop distance. As we will see in the following chapters the reduced mean hop distance translates into a significantly increased capacity. Therefore RINGOSTAR can be regarded as being cost-effective, which also supported by the fact that dark fiber is used to build the star. Another tradeoff between costs and performance concerns the additional delay introduced by the pretransmission coordination required for each transmission over the star subnetwork. In our approach, the delay is reduces by deploying a PSC in parallel to the AWG.

Chapter 7

Proxy Stripping

IN the previous chapter we introduced *proxy stripping*, RINGOSTAR's underlying the performance enhancing mechanism. The calculation of the mean hop distance already indicated the huge potential proxy stripping provides for increasing the capacity of packet-switched optical networks like RPR. In this chapter we will evaluate and compare the performance of RPR-like networks with and without proxy stripping in detail. As we want to isolate the impact of the proxy stripping mechanism on the ring's performace, the star network is idealized to provide infinite capacity and constant delay (resulting from the finite propagation speed of the optical signal). Therefore, the performance results presented in this chapter must be regarded as an upper bound demonstrating the potential of the proxy stripping as a performance enhancing technique. Note, however, that the performance results obtained for RINGOSTAR in Chapter 8 come very close to this idealized setting.

The performance results presented in this chapter are obtained by means of probabilistic analysis, supported by verifying computer simulations. The analysis considers arbitrary propagation delays, arbitrary packet length distribution, and arbitrary traffic matrices. In particular we consider uniform, hot-spot, symmetric, and asymmetric traffic demands, which are the most common traffic patters in the metro area, as discussed in the context of 'efficiency for different traffic patterns' in Section 2.3.2.

7.1 Analysis

In this section, we analyze the throughput-delay performance of RPR-like networks, i.e., single-channel bidirectional buffer insertion rings with destination stripping and shortest path routing, both with and without proxy stripping.

7.1.1 Notation

Fig. 7.1 depicts the bidirectional ring topology. The symbols (+) and (–) denote the clockwise and counterclockwise directions of the ring, respectively. The number of ring nodes equals N, with P of them acting as proxy stripping nodes, where $2 \leq P \leq N$. The proxy stripping nodes are equally spaced among the ring nodes at the position $i = 0, n, 2n, \ldots, N-n$, where $n = N/P$.

Next, for a given node i we define i' and i'' as follows:

$$i' = i \bmod N \tag{7.1}$$

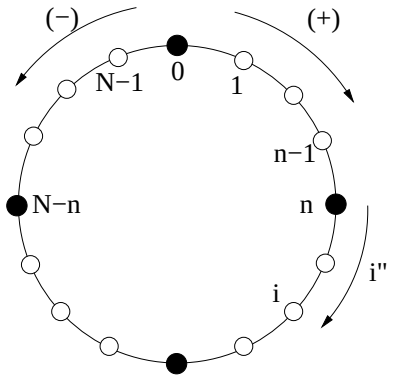


Figure 7.1: Notation for ring direction and position of ring nodes.

and

$$i'' = i \bmod n, \tag{7.2}$$

where i'' denotes the distance between a given node i and the closest proxy stripping node in (–) direction, as shown in Fig. 7.1. The distance between a given node i and the closest proxy stripping node in (+) direction is given by $n - i''$. The position of the proxy stripping nodes in (–) and (+) directions next to node i equals $i - i''$ and $i - i'' + n$, respectively. The distance and position of both proxy stripping nodes next to node i are summarized in Table 7.1.

Distance to Next Proxy Node in (–) Direction	i''
Distance to Next Proxy Node in (+) Direction	$n - i''$
Index of Next Proxy Node in (–) Direction	$i - i''$
Index of Next Proxy Node in (+) Direction	$i - i'' + n$

Table 7.1: Distance and index of proxy stripping nodes next to node i.

Let $f(i)$ denote the value of a given performance metric at node i, e.g., the waiting time an arriving data packet experiences in the transit queue of node i. To sum up the individual values of contiguous ring nodes we introduce the following definition:

$$\sum_{i=a}^{b*} f(i) := \begin{cases} \sum_{i=a'}^{b'} f(i) & \text{if } (a' \le b') \wedge (b' - a' \le \frac{N}{2}) \\ \sum_{i=a'}^{b'+N} f(i') & \text{if } (a' > b') \wedge (b' + N - a' \le \frac{N}{2}) \\ 0 & \text{else.} \end{cases} \tag{7.3}$$

Note that the above starred sum eases the notation by including the discontinuity at the transition from $i = N - 1$ to $i = 0$ in a convenient way. Otherwise, this transition would always have to be treated as a special case below. The starred sum equals 0 if the lower summation index is larger than the upper one, which is the case if the sum covers more than $N/2$ contiguous ring nodes. The double-starred sum does not have this restriction and is defined in this section as

$$\sum_{i=a}^{b**} f(i) := \begin{cases} \sum_{i=a'}^{b'} f(i) & \text{if } a' \le b' \\ \sum_{i=a'}^{b'+N} f(i') & \text{if } a' > b'. \end{cases} \tag{7.4}$$

7.1.2 Assumptions

In our analysis we make the following assumptions:

- *Single-queue mode*: We examine the single-queue mode of RPR, i.e., each node is equipped with one PTQ but no STQ. In addition, each node has a single transmit queue.
- *Infinite buffer size*: The size of both the PTQ and the transmit queue at each node is infinite, i.e., there is no packet loss due to buffer overflow.
- *Proxy stripping*: Packets that are proxy stripped from the ring are put into the star transmit queue of the corresponding proxy stripping node. Packets that arrive from the star and need to be forwarded on the shortest path towards their destination are put in the corresponding transit queue of the receiving proxy stripping node.
- *Propagation delay*: The nodes are equally spaced on the ring. The propagation delay between two adjacent ring nodes is given by τ. Thus, the round-trip time (RTT) of the RPR ring equals $N \cdot \tau$.
- *Unicast traffic*: We consider unicast traffic, i.e., all data transmissions are point-to-point.
- *Poisson packet arrival process*: The packet arrival process at the transmit queue of node i is Poisson with a mean arrival rate of $\lambda(i)$ packets per time unit, where $0 \leq i \leq N-1$. Note that the Poisson arrival rates of different nodes do not necessarily have to be the same.
- *Arbitrary packet length distribution*: We consider variable-size packets with an arbitrary packet length distribution, where $E[T_p]$ denotes the mean packet transmission time in time units.
- *Arbitrary traffic matrix*: A packet arriving at source node i is destined for node j with probability $p(i,j)$, where $0 \leq p(i,j) \leq 1$ and $0 \leq i,j \leq N-1$. Thus, packets destined for node j arrive at the transmit queue of node i with a mean arrival rate of $\lambda(i,j) = \lambda(i) \cdot p(i,j)$. For each source-destination node pair (i,j) the amount of offered traffic is specified by the traffic matrix, whose elements are given by $\rho(i,j) = \lambda(i,j) \cdot E[T_p]$.

7.1.3 Performance Metrics

The performance of the networks is evaluated in terms of mean delay and mean aggregate throughput which are defined as follows:

- *Mean Delay*: The mean delay denotes the average time period between packet arrival at the source node and packet reception at the destination node in steady state. The mean delay is given in time units.
- *Mean Aggregate Throughput*: The mean aggregate throughput denotes the mean number of transmitting nodes in steady state.

7.1.4 RPR with Proxy Stripping

The mean delay is equal to the weighted sum of the mean delay $d(i,j)$ of each source-destination node pair (i,j). The weights are the elements of the traffic matrix and represent

the amount of traffic from source node i to destination node j. The mean delay d is given by

$$d = \frac{1}{\rho_{tot}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \rho(i,j) d(i,j), \tag{7.5}$$

with

$$\rho_{tot} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \rho(i,j). \tag{7.6}$$

Depending on whether proxy stripping occurs or not, the mean delay of source-destination node pair (i,j) is obtained as

$$d(i,j) = \begin{cases} d_{ring}(i,j) & \text{if } h_{ring}(i,j) \leq h_{rs}(i) + 1 + h_{rs}(j) \\ d_{rs}(i) + d_{star} + d_{sr}(j) & \text{else,} \end{cases} \tag{7.7}$$

where

$$h_{ring}(i,j) = \min\{|i-j|, N - |i-j|\} \tag{7.8}$$

and

$$h_{rs}(l) = \min\{l'', n - l''\}, \quad l \in \{i,j\}. \tag{7.9}$$

To see this, recall from Section 6.2 that proxy stripping does not take place if the path on the peripheral ring is shorter than or equal to that on the short-cuts of the star subnetwork in terms of hops, i.e., $h_{ring}(i,j) \leq h_{rs}(i)+1+h_{rs}(j)$. Otherwise, packets undergo proxy stripping. The hop distance between a given node i and the two neighbor proxy stripping nodes and the hop distance between source node i and destination node j on the ring are illustrated in Fig. 7.2. These distances are used to determine $h_{ring}(i,j)$ and $h_{rs}(l)$ in Eqs. (7.8) and (7.9), respectively. As illustrated in Fig. 7.3, with proxy stripping $d(i,j)$ equals $d_{ring}(i,j)$, which

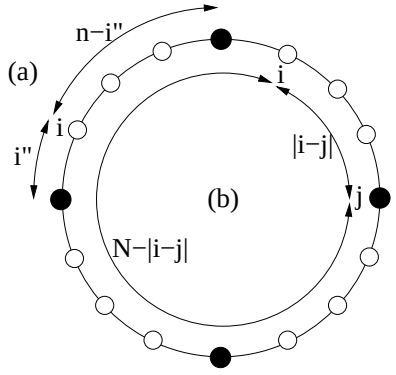


Figure 7.2: Hop distances: (a) Between node i and neighbor proxy stripping nodes and (b) between source node i and destination node j (in both directions).

denotes the mean delay encountered on the shortest ring path between source node i and destination node j. Without proxy stripping, $d(i,j)$ equals $d_{rs}(i)+d_{star}+d_{sr}(j)$, where $d_{rs}(i)$ denotes the mean delay encountered between source node i and its closest proxy stripping node, d_{star} denotes the time period required for transmitting the corresponding proxy-stripped packet across the star subnetwork, and $d_{sr}(j)$ denotes the mean delay encountered between

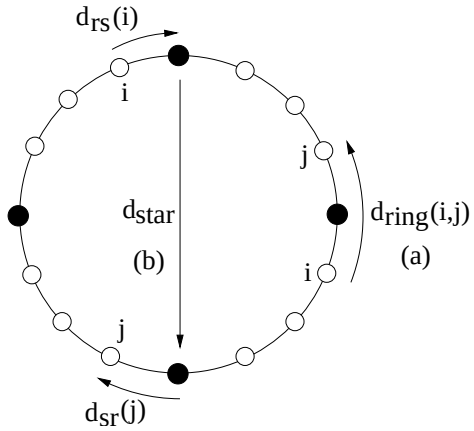


Figure 7.3: Mean delay of source-destination node pair (i, j): (a) Without proxy stripping and (b) with proxy stripping.

destination node j and its closest proxy stripping node. Next, we need to calculate $d_{ring}(i, j)$, $d_{rs}(i)$, $d_{sr}(j)$, and d_{star}.

The mean delay $d_{ring}(i, j)$ for a ring-only transmission without proxy stripping is composed of the mean waiting time $w_t(i)$ encountered at the transmit queue of source node i, the mean packet transmission time $E[T_p]$, the link propagation delay τ, and the mean waiting time $w_r(k)$ encountered at the transit queues of nodes k between source node i and destination node j. The mean delay $d_{ring}(i, j)$ is given by

$$d_{ring}(i,j) = \begin{cases} d^{+}_{ring}(i,j) & \text{if } [(i<j) \wedge (j-i<\frac{N}{2})] \vee \\ & \quad [(i>j) \wedge (i-j>\frac{N}{2})] \\ d^{-}_{ring}(i,j) & \text{if } [(i<j) \wedge (j-i>\frac{N}{2})] \vee \\ & \quad [(i>j) \wedge (i-j<\frac{N}{2})] \\ \frac{1}{2}d^{+}_{ring}(i,j) + \frac{1}{2}d^{-}_{ring}(i,j) & \text{if } |i-j| = N/2 \\ 0 & \text{if } (i=j), \end{cases} \tag{7.10}$$

with

$$d^{+}_{ring}(i,j) = w^{+}_{t}(i) + E[T_p] + \tau + \sum_{k=i+1}^{j-1}{}^{*} (w^{+}_{r}(k) + \tau) \tag{7.11}$$

and

$$d^{-}_{ring}(i,j) = w^{-}_{t}(i) + E[T_p] + \tau + \sum_{k=j+1}^{i-1}{}^{*} (w^{-}_{r}(k) + \tau). \tag{7.12}$$

As depicted in Fig. 7.4, two different cases have to be considered for either direction, which is indicated by the upper index (+) and (–), respectively. In Eq. (7.10), the first and second line of the first 'if' correspond to (a) and (b) in the figure and the first and second line of the second 'if' correspond to (c) and (d) in the figure (the third and fourth 'if' are not illustrated in the figure).

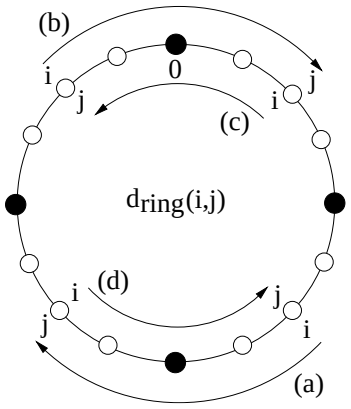


Figure 7.4: Mean delay $d_{ring}(i,j)$ of a ring-only transmission without proxy stripping between source node i and destination node j.

Similarly, the mean delay $d_{rs}(i)$ is given by

$$d_{rs}(i) = \begin{cases} d_{rs}^{-}(i) & \text{if } 0 < i'' < \frac{n}{2} \\ d_{rs}^{+}(i) & \text{if } i'' > \frac{n}{2} \\ \frac{1}{2}d_{rs}^{-}(i) + \frac{1}{2}d_{rs}^{+}(i) & \text{if } i'' = \frac{n}{2} \\ 0 & \text{if } i'' = 0, \end{cases} \tag{7.13}$$

with

$$d_{rs}^{-}(i) = w_t^{-}(i) + E[T_p] + \tau + \sum_{k=i-i''+1}^{i-1} (w_r^{-}(k) + \tau) \tag{7.14}$$

and

$$d_{rs}^{+}(i) = w_t^{+}(i) + E[T_p] + \tau + \sum_{k=i+1}^{i-i''+n-1} (w_r^{+}(k) + \tau). \tag{7.15}$$

Note that if n is even nodes in the middle of two neighbor proxy stripping nodes have the same hop distance to both proxy stripping nodes. In this case, nodes in the middle split their traffic equally and transmit the same amount of traffic in both directions.

Packets arriving from the star are put in the transit queue of the receiving proxy stripping node and are forwarded towards their destination node j. The forwarded packets traverse all transit queues of the intermediate nodes between the corresponding proxy stripping node and destination node j. Accordingly, the mean delay $d_{sr}(j)$ is given by

$$d_{sr}(j) = \begin{cases} d_{sr}^{+}(j) & \text{if } 0 < j'' < \frac{n}{2} \\ d_{sr}^{-}(j) & \text{if } j'' > \frac{n}{2} \\ \frac{1}{2}d_{sr}^{+}(j) + \frac{1}{2}d_{sr}^{-}(j) & \text{if } j'' = \frac{n}{2} \\ 0 & \text{if } j'' = 0, \end{cases} \tag{7.16}$$

with

$$d_{sr}^{+}(j) = \sum_{k=j-j''}^{j-1} (w_r^{+}(k) + \tau) \tag{7.17}$$

and

$$d_{sr}^{-}(j) = \sum_{k=j+1}^{j-j''+n} (w_r^{-}(k) + \tau). \tag{7.18}$$

The mean delay d_{star} depends on the access control used in the star subnetwork. For random and preallocation access control d_{star} is given by

$$d_{star} = E[T_p^{star}] + \frac{N \cdot \tau}{\pi}, \tag{7.19}$$

where $E[T_p^{star}]$ denotes the mean packet transmission time on the star subnetwork and $N\tau/\pi$ denotes the propagation delay of the star subnetwork. For reservation access control with pretransmission coordination via the ring d_{star} is given by

$$d_{star} = N \cdot \tau + E[T_p^{star}] + \frac{N \cdot \tau}{\pi}, \tag{7.20}$$

where $N \cdot \tau$ represents the RTT of the ring. Note that in Eqs. (7.19) and (7.20) we assume that the star subnetwork provides sufficient capacity such that the waiting time at the star transmit queues of the proxy stripping nodes is negligible. This assumption is motivated by the fact that in this work we aim at demonstrating the potential of the proxy stripping technique rather than addressing the design of a specific star subnetwork.

Next, for the above expressions of $d_{ring}(i,j)$, $d_{rs}(i)$, and $d_{sr}(j)$ we need to calculate the mean waiting time in the transmit queue $w_t^{\pm}(i)$ and the mean waiting time in the transit queue $w_r^{\pm}(i)$ at node i. Under the assumption that the packet arrival process at the transit queue is Poisson, the mean waiting times in both the transmit queue and transit queue were analyzed in [157] for the case of unidirectional rings. By extending these results to our bidirectional ring we obtain $w_t^{\pm}(i)$ as

$$w_t^{\pm}(i) = \frac{(\rho_r^{\pm}(i) + \rho_t^{\pm}(i))E[T_p^2]}{2(1 - \rho_r^{\pm}(i) - \rho_t^{\pm}(i))(1 - \rho_r^{\pm}(i))E[T_p]} \tag{7.21}$$

and $w_r^{\pm}(i)$ as

$$w_r^{\pm}(i) = \frac{\rho_t^{\pm}(i)E[T_p^2]}{2(1 - \rho_r^{\pm}(i))E[T_p]}, \tag{7.22}$$

where $\rho_t^{\pm}(i)$ and $\rho_r^{\pm}(i)$ denote the amount of traffic arriving at the ring transmit queues and the ring transit queues of both directions (+) and (–) at node i, respectively (to be defined shortly). In Section 7.2 we show by means of extensive verifying simulations that our analysis provides very accurate results despite the simplifying assumption of Poisson arrivals at transit queues. Next, we calculate the amount of traffic arriving at the ring transmit queues $\rho_t^{\pm}(i)$ and the ring transit queues $\rho_r^{\pm}(i)$ at node i for both directions (+) and (–).

Ring Transmit Queues

The amount of traffic $\rho_t^{\pm}(i)$ which arrives at the ring transmit queue of node i and corresponds to the direction towards the closest proxy stripping node is composed of the ring-only traffic $\rho_t^{r\pm}(i)$ for that direction and all traffic $\rho_t^{out}(i)$ that is sent via the star subnetwork. For the other direction, $\rho_t^{\pm}(i)$ comprises the ring-only traffic for that direction. If n is even and the node i is located between two adjacent proxy stripping nodes, i.e., $i'' = n/2$, the star traffic

is equally split and sent in both directions. The total amount of traffic originating from node i equals $\rho_t^{\pm}(i) = \rho_t^{+}(i) + \rho_t^{-}(i)$, where $\rho_t^{+}(i)$ and $\rho_t^{-}(i)$ are given by

$$\rho_t^{+}(i) = \begin{cases} \rho_t^{r+}(i) & \text{if } 0 \le i'' < n - i'' \\ \rho_t^{r+}(i) + \frac{1}{2}\rho_t^{out}(i) & \text{if } i'' = n - i'' \\ \rho_t^{r+}(i) + \rho_t^{out}(i) & \text{if } i'' > n - i'' \end{cases} \tag{7.23}$$

and

$$\rho_t^{-}(i) = \begin{cases} \rho_t^{r-}(i) & \text{if } (i'' = 0) \vee (i'' > n - i'') \\ \rho_t^{r-}(i) + \frac{1}{2}\rho_t^{out}(i) & \text{if } i'' = n - i'' \\ \rho_t^{r-}(i) + \rho_t^{out}(i) & \text{if } 0 < i'' < n - i''. \end{cases} \tag{7.24}$$

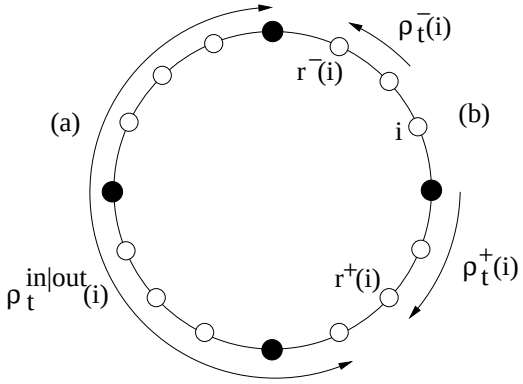


Figure 7.5: Destination nodes reached by source node i (a) with proxy stripping and (b) without proxy stripping.

As depicted in Fig. 7.5, source node i sends packets directly on the ring without proxy stripping up to destination node $r^{-}(i)$ in (–) direction and up to destination node $r^{+}(i)$ in (+) direction. The remaining destination nodes are reached by means of proxy stripping. The nodes $r^{\pm}(i)$ are given by

$$r^{+}(i) = \begin{cases} i + \lceil \frac{n}{2} \rceil & \text{if } i'' < \lfloor \frac{n}{2} \rfloor \\ i + n & \text{if } i'' = \lfloor \frac{n}{2} \rfloor \\ i - i'' + n + \lceil \frac{n}{2} \rceil & \text{if } i'' > \lfloor \frac{n}{2} \rfloor \end{cases} \tag{7.25}$$

and

$$r^{-}(i) = \begin{cases} i - i'' - \lceil \frac{n}{2} \rceil & \text{if } i'' < \lceil \frac{n}{2} \rceil \\ i - n & \text{if } i'' = \lceil \frac{n}{2} \rceil \\ i - \lceil \frac{n}{2} \rceil & \text{if } i'' > \lceil \frac{n}{2} \rceil. \end{cases} \tag{7.26}$$

Fig. 7.6 shows the three different cases in the calculation of $r^{\pm}(i)$. In Eqs. (7.25) and (7.26) the first 'if' corresponds to (a), the second to (b), and the third to (c) in the figure. Given $r^{\pm}(i)$, we obtain $\rho_t^{r+}(i)$, $\rho_t^{r-}(i)$, and $\rho_t^{out}(i)$ as

$$\rho_t^{r+}(i) = \sum_{j=i+1}^{r^{+}(i)} {}^{*}\, \rho(i,j) - \begin{cases} \frac{1}{2}\rho(i, r^{+}(i)) & \text{if } |i - r^{+}(i)| = N/2 \\ 0 & \text{else} \end{cases} \tag{7.27}$$

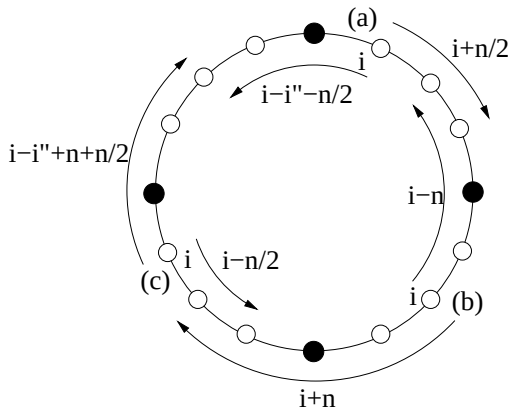


Figure 7.6: Ring segments which are reached by source node i without proxy stripping.

$$\rho_t^{r-}(i) = \sum_{j=r^-(i)}^{i-1*} \rho(i,j) - \begin{cases} \frac{1}{2}\rho(i,r^-(i)) & \text{if } |i-r^-(i)| = N/2 \\ 0 & \text{else,} \end{cases} \tag{7.28}$$

and

$$\rho_t^{out}(i) = \begin{cases} 0 & \text{if } (n=\frac{N}{2}) \wedge [(i''=\lfloor \frac{n}{2} \rfloor) \vee (i''=\lceil \frac{n}{2} \rceil)] \\ \sum_{j=r^+(i)+1}^{**r^-(i)-1} \rho(i,j) & \text{else.} \end{cases} \tag{7.29}$$

Ring Transit Queues

The amount of traffic $\rho_r^{\pm}(i)$ which arrives at the ring transit ring queue of node i is composed of the forwarded ring-only traffic $\rho_r^{r\pm}(i)$ and the traffic $\rho_r^{s\pm}(i)$ forwarded either from or to the star, depending on the position of node i. Hence, $\rho_r^{\pm}(i)$ is given by

$$\rho_r^{\pm}(i) = \rho_r^{r\pm}(i) + \rho_r^{s\pm}(i). \tag{7.30}$$

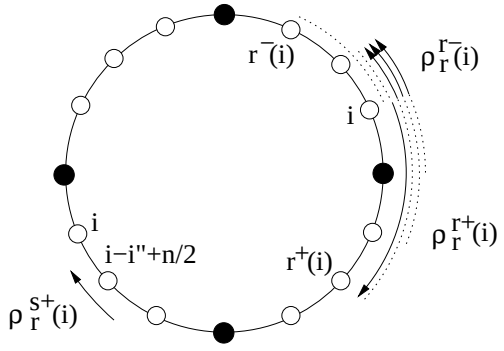


Figure 7.7: Illustration of forwarded ring-only traffic.

As illustrated on the right-hand side of Fig. 7.7, the forwarded ring-only traffic $\rho_r^{r\pm}(i)$ is composed of the traffic that originates from the nodes 'before' node i and is destined for

nodes 'behind' node i. In the figure, each arrow corresponds to a different node 'before' node i and covers the nodes 'behind' node i. Thus, for each direction $\rho_r^{r\pm}(i)$ is given by

$$\rho_r^{r+}(i) = \sum_{k=r^-(i)}^{i-1*} \left(\sum_{l=i+1}^{r^+(k)*} \rho(k,l) - \begin{cases} \frac{1}{2}\rho(k,r^+(k)) & \text{if } |k-r^+(k)| = N/2 \\ 0 & \text{else} \end{cases} \right) \tag{7.31}$$

and

$$\rho_r^{r-}(i) = \sum_{k=i+1}^{r^+(i)*} \left(\sum_{l=r^-(k)}^{i-1*} \rho(k,l) - \begin{cases} \frac{1}{2}\rho(k,r^-(k)) & \text{if } |k-r^-(k)| = N/2 \\ 0 & \text{else} \end{cases} \right). \tag{7.32}$$

Similarly, the traffic $\rho_r^{s\pm}(i)$ forwarded from (or to) the star is composed of the aggregate traffic $\rho_t^{in}(k)$ (or $\rho_t^{out}(k)$ of Eq. (7.29)) of all nodes between the closest proxy stripping node 'before' and all nodes 'behind' node i, where $\rho_t^{in}(k)$ is given by

$$\rho_t^{in}(k) = \begin{cases} 0 & \text{if } (n = \frac{N}{2}) \wedge [(k'' = \lfloor \frac{n}{2} \rfloor) \vee (k'' = \lceil \frac{n}{2} \rceil)] \\ \sum_{j=r^+(k)+1}^{**r^-(k)-1} \rho(j,k) & \text{else.} \end{cases} \tag{7.33}$$

For each direction $\rho_r^{s\pm}(i)$ is given by

$$\rho_r^{s+}(i) = \begin{cases} \sum_{k=i+1}^{i-i''+\lfloor n/2 \rfloor} \rho_t^{in}(k) & - \begin{cases} \frac{1}{2}\rho_t^{in}(i-i''+\frac{n}{2}) & \text{if } n \text{ even} \\ 0 & \text{if } n \text{ odd} \end{cases} \quad \text{if } 0 \le i'' < \lfloor \frac{n}{2} \rfloor \\ \sum_{k=i-i''+\lceil n/2 \rceil}^{i-1} \rho_t^{out}(k) & - \begin{cases} \frac{1}{2}\rho_t^{out}(i-i''+\frac{n}{2}) & \text{if } n \text{ even} \\ 0 & \text{if } n \text{ odd} \end{cases} \quad \text{if } i'' > \lceil \frac{n}{2} \rceil \\ 0 & \text{if } (i'' = \lfloor \frac{n}{2} \rfloor) \vee (i'' = \lceil \frac{n}{2} \rceil) \end{cases} \tag{7.34}$$

and

$$\rho_r^{s-}(i) = \begin{cases} \sum_{k=i+1}^{i-i''+\lfloor n/2 \rfloor} \rho_t^{out}(k) & - \begin{cases} \frac{1}{2}\rho_t^{out}(i-i''+\frac{n}{2}) & \text{if } n \text{ even} \\ 0 & \text{if } n \text{ odd} \end{cases} \quad \text{if } 0 < i'' < \lfloor \frac{n}{2} \rfloor \\ \sum_{k=i-i''+\lceil n/2 \rceil}^{i-1} \rho_t^{in}(k) & - \begin{cases} \frac{1}{2}\rho_t^{in}(i-i''+\frac{n}{2}) & \text{if } n \text{ even} \\ 0 & \text{if } n \text{ odd} \end{cases} \quad \text{if } (i'' = 0) \vee (i'' > \lceil \frac{n}{2} \rceil) \\ 0 & \text{if } (i'' = \lfloor \frac{n}{2} \rfloor) \vee (i'' = \lceil \frac{n}{2} \rceil). \end{cases} \tag{7.35}$$

Note that proxy stripping nodes and nodes located in the middle of adjacent proxy stripping nodes do not forward any star traffic.

Star Transmit Queues

To evaluate the forwarding burden caused by proxy stripping, we also calculate the amount of traffic arriving at proxy stripping nodes. The amount of traffic $\rho_s(i)$ which arrives at the star transmit queue of node i, $i = 0, n, 2n, \ldots, (N-n)$, consists of the traffic to be stripped from both rings and the traffic generated and sent by the proxy stripping node itself via the star subnetwork. Thus, $\rho_s(i)$ is given by

$$\rho_s(i) = \rho_r^{s+}((i-1)') + \rho_t^{out}((i-1)') + \rho_t^{out}(i) + \rho_t^{out}(i+1) + \rho_r^{s-}(i+1), \tag{7.36}$$

where ρ_t^{out}, ρ_r^{s+}, and ρ_r^{s-} are given in Eqs. (7.29), (7.34), and (7.35), respectively.

7.1.5 RPR without Proxy Stripping

In this section, we analyze the throughput-delay performance of RPR without proxy stripping. As opposed to the above analysis, in RPR without proxy stripping there is no star subnetwork and the above equations are modified as follows. Eq. (7.7) reduces to $d(i,j) = d_{ring}(i,j)$ and Eq. (7.5) becomes

$$d = \frac{1}{\rho_{tot}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \rho(i,j) d_{ring}(i,j). \tag{7.37}$$

The expressions for the waiting times $w_t^{\pm}(i)$ and $w_r^{\pm}(i)$ of Eqs. (7.21) and (7.22) also hold for RPR without proxy stripping. However, the calculation of $\rho_t^{\pm}(i)$ and $\rho_r^{\pm}(i)$ is different since in RPR without proxy stripping there is no star subnetwork. For the ring transmit queues, Eqs. (7.23) and (7.24) reduce to $\rho_t^{+}(i) = \rho_t^{r+}(i)$ and $\rho_t^{-}(i) = \rho_t^{r-}(i)$, respectively. Furthermore, Eqs. (7.27) and (7.28) have to be slightly modified as follows

$$\rho_t^{r+}(i) = \sum_{j=i+1}^{r^+(i)*} \rho(i,j) - \begin{cases} \frac{1}{2}\rho(i,r^+(i)) & \text{if } N \text{ even} \\ 0 & \text{if } N \text{ odd} \end{cases} \tag{7.38}$$

and

$$\rho_t^{r-}(i) = \sum_{j=r^-(i)}^{i-1*} \rho(i,j) - \begin{cases} \frac{1}{2}\rho(i,r^-(i)) & \text{if } N \text{ even} \\ 0 & \text{if } N \text{ odd,} \end{cases} \tag{7.39}$$

where $r^+(i)$ and $r^-(i)$ are given by

$$r^+(i) = i + \lfloor N/2 \rfloor \tag{7.40}$$

and

$$r^-(i) = i - \lfloor N/2 \rfloor . \tag{7.41}$$

Note that without proxy stripping all nodes are reached via the ring. Consequently, the borders of ring-only transmissions $r^{\pm}(i)$ cover the whole ring now. Each direction covers one half of the ring corresponding to shortest path routing and traffic from source node i destined for the opposite ring node is equally split and sent in both directions. Similarly, for the ring transit queues Eq. (7.30) reduces to $\rho_r^{\pm}(i) = \rho_r^{r\pm}(i)$. In addition, Eqs. (7.31) and (7.32) are modified as follows

$$\rho_r^{r+}(i) = \sum_{k=r^-(i)+1}^{i-1*} \left(\sum_{l=i+1}^{r^+(k)*} \rho(k,l) - \begin{cases} \frac{1}{2}\rho(k,r^+(k)) & \text{if } N \text{ even} \\ 0 & \text{if } N \text{ odd} \end{cases} \right) \tag{7.42}$$

and

$$\rho_r^{r-}(i) = \sum_{k=i+1}^{r^+(i)-1*} \left(\sum_{l=r^-(k)}^{i-1*} \rho(k,l) - \begin{cases} \frac{1}{2}\rho(k,r^-(k)) & \text{if } N \text{ even} \\ 0 & \text{if } N \text{ odd} \end{cases} \right). \tag{7.43}$$

7.2 Results

In this section, we conduct numerical investigations of the throughput-delay performance of RPR both with and without proxy stripping for different traffic matrices. The default network parameters are set as follows: Line rate of each ring equals 2.5 Gbit/s, signal propagation

delay equals $2/3 \cdot c_o$, where $c_o = 3 \cdot 10^8$ m/s, and circumference of the bidirectional ring equals 100 km, i.e., the RTT of the ring is constant and equals $N \cdot \tau = 3l/2c_o$. For the packet size we use the approximately trimodal distribution that is typically found in IP networks, as shown in Table 7.2. (We note that the emergence of new applications, e.g., content distribution network (CDN) and media streaming, may result in different packet length distributions on specific links. However, on a large number of links the typical trimodal packet length distribution is still valid [158].) Without loss of generality we set $E[T_p^{star}] = 0$. To verify the accuracy of our analytical model we have also conducted extensive simulations. In each simulation we have generated 10^6 packets including a warm-up phase of 10^5 packets. Using the method of batch means we calculated the 95% confidence intervals for the performance metrics.

40 byte	50% of Packets
552 byte	30% of Packets
1500 byte	20% of Packets

Table 7.2: Trimodal packet length distribution.

In the following, we examine the throughput-delay performance of RPR under uniform, hot-spot, and asymmetric traffic and pay particular attention to the impact of proxy stripping on the performance of RPR.

7.2.1 Uniform Traffic

Under uniform traffic a given node sends a generated packet to any other node with equal probability $1/(N-1)$. Recall from Section 2.3.2 that uniform traffic is typically found in metro core rings. Figs. 7.8 and 7.9 depict the mean delay (given in integer multiples of the RTT) vs. mean aggregate throughput (number of simultaneously transmitting nodes) both without and with proxy stripping for different number of nodes $N \in \{8, 16, 256\}$. As shown in Fig. 7.8, without proxy stripping the mean delay is equal to one fourth of the RTT at light loads since for uniform traffic packets traverse one fourth of the ring on the average without experiencing any significant queuing delay. For an increasing offered load the channel utilization increases until all bandwidth resources are fully utilized. Under high channel utilization nodes have to wait for a longer time period to find the channel idle, resulting in an increased delay. The maximum mean aggregate throughput of RPR is given by the ratio of number of links divided by the mean hop distance $2N/\overline{h}$, where $\overline{h}$ is given in Eq. (6.3) of Section 6.5. We observe from Fig. 7.8 that RPR achieves a maximum mean aggregate throughput of seven to eight, depending on N. Note that for small N the analytical and simulative results match perfectly while for an increasing N the simulation provides a slightly larger throughput. This is due to the fact that we assumed Poisson packet arrivals at the transit queue of each node. With increasing N the error caused by this simplifying assumption is accumulated, resulting in a more pronounced discrepancy between analysis and simulation, where the analysis slightly underestimates the more realistic simulation results.

Fig. 7.9 shows the impact of proxy stripping on the throughput-delay performance of RPR using $P \in \{2, 4\}$ proxy stripping nodes. Interestingly, using $P = 2$ proxy stripping nodes increases the maximum mean aggregate throughput of RPR only for $N = 8$. In contrast, for $N = 16$ and in particular $N = 256$ using $P = 2$ proxy stripping nodes slightly deteriorates the throughput-delay performance of RPR. This is because with proxy stripping, source ring nodes send some of their packets to the closest proxy stripping nodes rather than directly

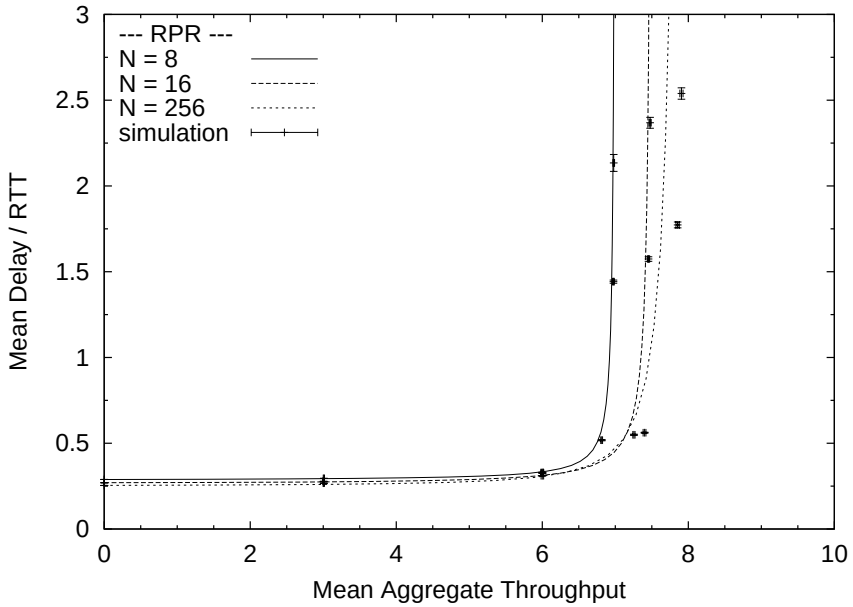


Figure 7.8: Mean delay vs. mean aggregate throughput of RPR without proxy stripping for uniform traffic with different $N \in \{8, 16, 256\}$.

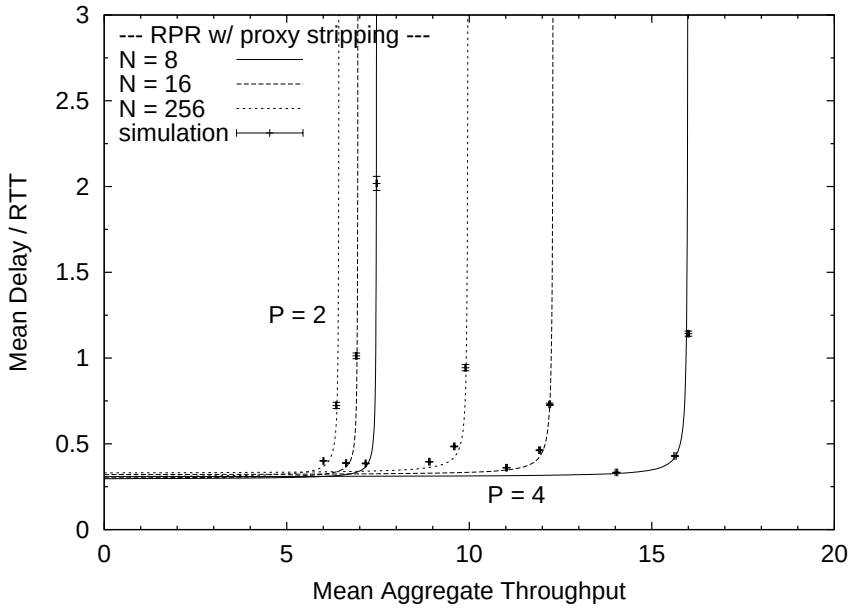


Figure 7.9: Mean delay vs. mean aggregate throughput of RPR with $P \in \{2, 4\}$ proxy stripping nodes for uniform traffic with different $N \in \{8, 16, 256\}$.

to the corresponding destination nodes. As a consequence, the proxy stripping nodes form a hot spot whose attached ring fibers get more congested with increasing traffic load. These congested fiber links prevent ring nodes from sending more data packets, resulting in a decreased throughput and a slightly increased delay. Clearly, with increasing N and $P = 2$ fixed the congestion becomes more severe. The congestion on the fiber links close to the proxy stripping nodes can be mitigated by increasing the number of proxy stripping nodes, as depicted in Fig. 7.9 for $P = 4$. We observe that the throughput of RPR using $P = 4$ proxy stripping nodes is better than that of RPR without proxy stripping for all $N \in \{8, 16, 256\}$. Note that in Fig. 7.9 analysis and simulation results match very well. This is due to the fact that with proxy stripping data packets are sent via the short-cuts of the star subnetwork and thus traverse fewer ring transit queues on the average. Consequently, the error due to the assumed Poisson arrival at transit queues in the analysis is less pronounced.

Fig. 7.10 shows the throughput-delay performance of RPR for different number of proxy stripping nodes $P \in \{4, 8, 16, 32, 64\}$ and $N = 256$ fixed. Obviously, the throughput of RPR is dramatically improved by increasing P. For instance, by interconnecting $32/256 = 12.5\%$ of the nodes via a star subnetwork, i.e., $P = 32$, a maximum mean aggregate throughput of approximately 75 is achieved. Compared to Fig. 7.8, this translates into a throughput improvement by a factor of almost ten. As shown in Fig. 7.10, the throughput performance of RPR can be further improved by increasing P at the expense of more star transceivers and dark fibers. Note that at light loads the mean delay is slightly larger than one fourth of RTT. This is due to the queuing delays encountered at the ring transit queues of the hot-spot proxy stripping nodes.

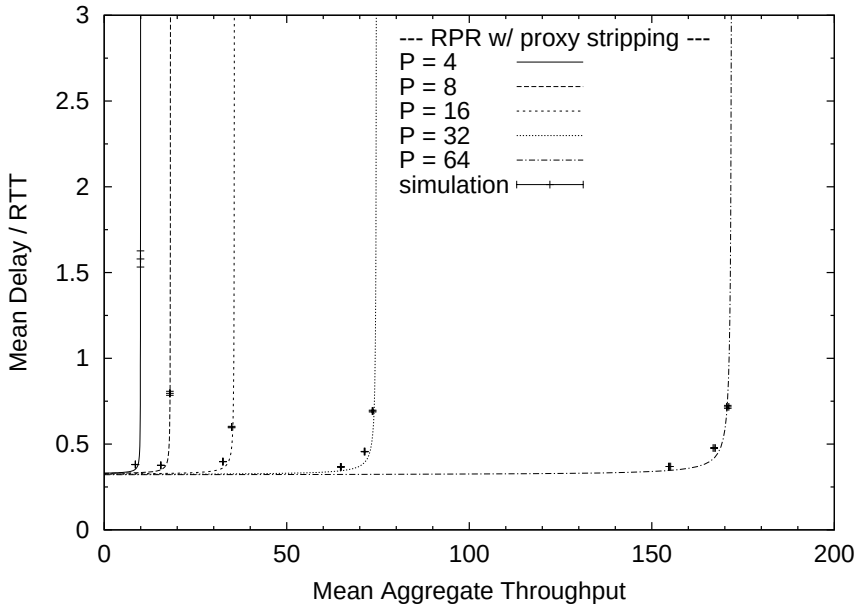


Figure 7.10: Mean delay vs. mean aggregate throughput of RPR with $P \in \{4, 8, 16, 32, 64\}$ proxy stripping nodes for uniform traffic with $N = 256$.

So far, we have considered star subnetworks without pretransmission coordination overhead, e.g., preallocation and random access protocols. That is, the transmission of proxy-

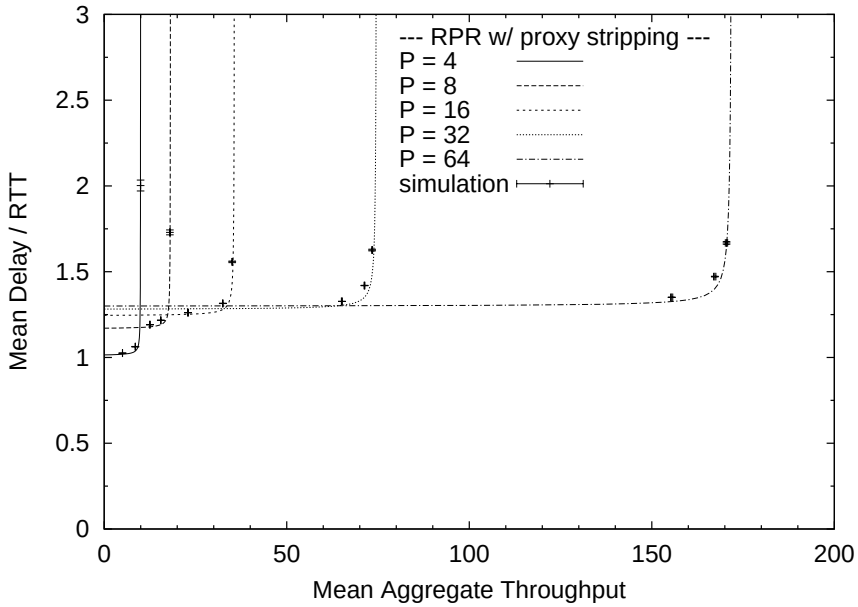


Figure 7.11: Mean delay vs. mean aggregate throughput of RPR with $P \in \{4, 8, 16, 32, 64\}$ proxy stripping nodes and pretransmission coordination for uniform traffic with $N = 256$.

stripped packets across the short-cuts of the star subnetwork did not imply any reservation overhead. Fig. 7.11 depicts the throughput-delay performance of RPR if channel access on the star subnetwork is arbitrated by using a reservation protocol with pretransmission coordination for $P \in \{4, 8, 16, 32, 64\}$ and $N = 256$. Recall from Section 7.1.4 that with pretransmission coordination control packets are broadcast along either ring prior to data transmission, resulting in an overhead of $N \cdot \tau$ time units (RTT). Consequently, the throughput-delay curves are shifted towards higher delay values, as depicted in Fig. 7.11. We observe from the figure that with increasing P the mean delay increases. This is because with larger P more packets are proxy stripped, leading to an increased amount of control traffic and thus larger pretransmission coordination overhead. In the rest of this chapter, we consider star subnetworks without pretransmission coordination.

7.2.2 Hot-Spot Traffic

Next, we investigate RPR and the impact of proxy stripping on its throughput-delay performance under hot-spot traffic. Recall from Section 2.3.2 that hot-spot (hubbed) traffic is typically found in metro edge rings. We define our hot-spot traffic matrix as follows. Let node $i = 0$ be the hub node (hot spot). Each node i, $1 \leq i \leq N-1$, generates the same amount of traffic ρ, where $\rho \geq 0$. A given node i, $1 \leq i \leq N-1$, sends a generated packet to the hot spot with probability h, $0 \leq h \leq 1$, and to any other of the remaining $(N-2)$ nodes with equal probability $(1-h)/(N-2)$. To examine both symmetric and asymmetric hot-spot traffic we introduce the parameter α which controls the traffic generated by the hot spot and the remaining $(N-1)$ nodes. Specifically, the amount of traffic generated by hot spot $i = 0$ and destined for any of the remaining $(N-1)$ nodes is equal to $\alpha \cdot h \cdot \rho$, where $0 \leq \alpha \leq 1$. The amount of traffic generated by node i, $1 \leq i \leq N-1$, is multiplied by $(1-\alpha)$. Thus, we

have

$$\rho(0,j) = \alpha \cdot h \cdot \rho, \quad \text{if } 1 \leq j \leq N-1, \tag{7.44}$$

$$\rho(i,0) = (1-\alpha) \cdot h \cdot \rho, \quad \text{if } 1 \leq i \leq N-1, \tag{7.45}$$

and

$$\rho(i,j) = (1-\alpha) \cdot (1-h) \cdot \frac{1}{N-2} \cdot \rho, \quad \text{if } 1 \leq i,j \leq N-1, \tag{7.46}$$

where $\alpha, h \in [0,1]$. Table 7.3 shows different types of traffic used in the subsequent numerical investigations and the corresponding values of α and h.

Traffic type	α	h
Symmetric Uniform	0.5	$1/(N-1)$
Symmetric Hot-Spot	0.5	1.0
Asymmetric Hot-Spot (Data Distribution)	1.0	1.0
Asymmetric Hot-Spot (Data Collection)	0	1.0

Table 7.3: Generic traffic model.

In this section, we concentrate on symmetric traffic with $\alpha = 0.5$. That is, a given node and the hub node generate the same amount of traffic destined for each other. In Figs. 7.12 and 7.13 we examine the throughput-delay performance of RPR without and with proxy stripping under hot-spot traffic and compare it to that obtained under uniform traffic for $N = 256$. Fig. 7.12 illustrates the throughput-delay performance of RPR without proxy stripping for $h \in \{1/(N-1), 0.5, 1.0\}$. For uniform traffic, i.e., $h = 1/(N-1) = 1/255$, the maximum mean aggregate throughput is upper bounded by eight, as discussed above. However, for non-uniform traffic the performance of RPR decreases dramatically. For $h = 1.0$, i.e., all nodes send packets only to the hub, the maximum aggregate throughput equals four, which is half of that obtained under uniform traffic. Also, we observe that for a mixed traffic scenario with $h = 0.5$, i.e., 50% of the generated packets are destined to the hub while the other 50% are equally distributed among the remaining $(N-2)$ destination nodes, the throughput performance of RPR is still decreased significantly. The throughput deterioration of RPR under non-uniform traffic is due to the fact that packets traverse more intermediate nodes and thus consume more bandwidth resources compared to uniform traffic. As a result, fewer nodes can transmit simultaneously which translates into a decreased mean aggregate throughput.

Fig. 7.13 shows the throughput-delay performance of RPR using $P = 32$ proxy stripping nodes for both uniform and non-uniform traffic. Under non-uniform traffic we observe the opposite behavior in RPR with proxy stripping compared to RPR without proxy stripping. We observe that under non-uniform traffic sending proxy-stripped traffic across the short-cuts of the star subnetwork increases the maximum mean aggregate throughput dramatically. Note that for $h = 1.0$ proxy stripping increases the maximum mean aggregate throughput of RPR by a factor of more than 30.

7.2.3 Asymmetric Traffic

Next, we examine asymmetric hot-spot traffic. In Figs. 7.14 and 7.15 we investigate the throughput-delay performance of RPR without and with proxy stripping under hot-spot traffic

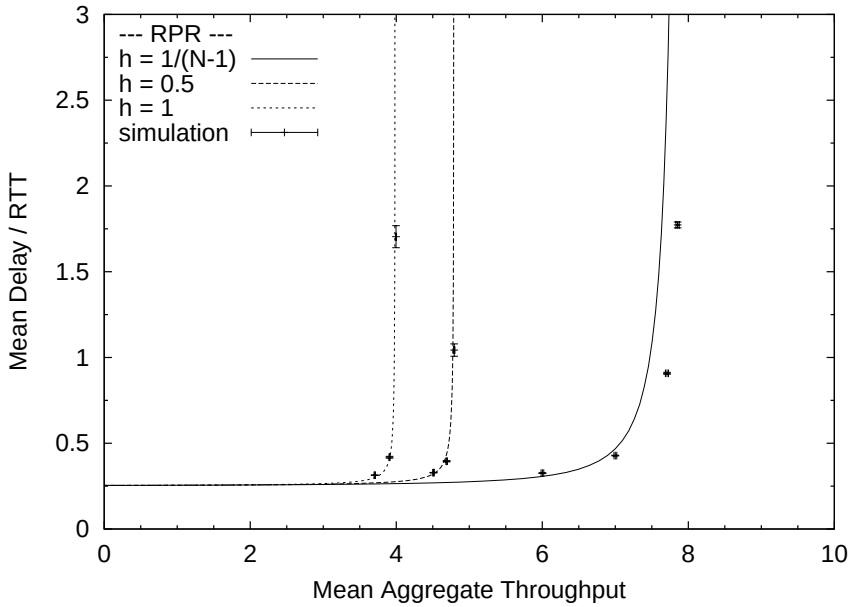


Figure 7.12: Mean delay vs. mean aggregate throughput of RPR without proxy stripping for symmetric hot-spot traffic with $h \in \{1/(N-1), 0.5, 1.0\}$, $\alpha = 0.5$, and $N = 256$.

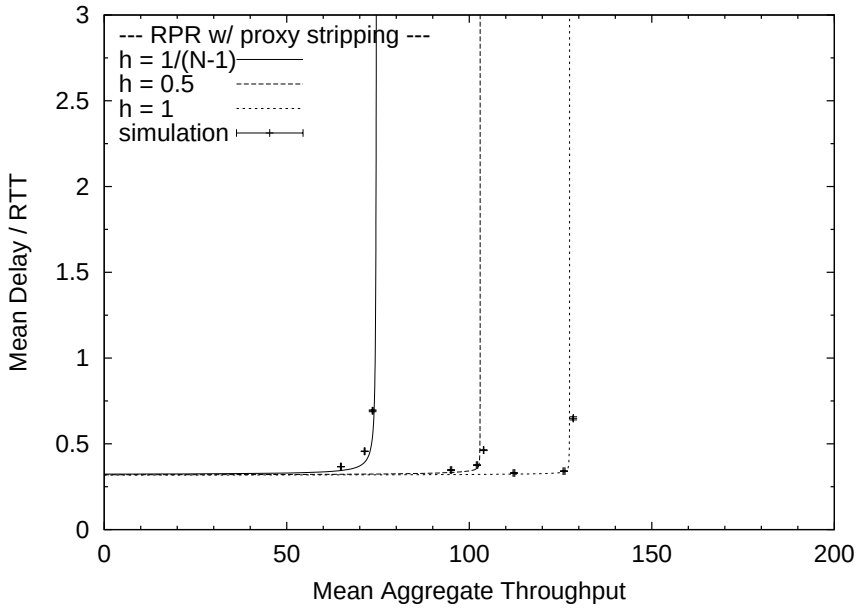


Figure 7.13: Mean delay vs. mean aggregate throughput of RPR with $P = 32$ proxy stripping nodes for symmetric hot-spot traffic with $h \in \{1/(N-1), 0.5, 1.0\}$, $\alpha = 0.5$, and $N = 256$.

with $h = 1.0$ for $N = 256$. Again, in RPR with proxy stripping we set $P = 32$. In both figures we consider $\alpha \in \{0, 0.5, 1.0\}$. Recall that $\alpha = 0.5$ corresponds to symmetric traffic. The other two cases $\alpha = 0$ and $\alpha = 1.0$ represent asymmetric traffic between hub node and regular ring nodes. More precisely, with $\alpha = 0$ the hub generates no traffic while the remaining $(N - 1)$ nodes generate only traffic destined for the hub. This traffic scenario corresponds to data collection. Conversely, with $\alpha = 1.0$ only the hub generates traffic for the remaining $(N - 1)$ nodes while the latter ones are completely idle. This traffic scenario corresponds to data distribution. We observe that due to the symmetry of the architecture both data collection and data distribution achieve the same maximum mean aggregate throughput which is half of that obtained under symmetric traffic. As shown in Fig. 7.14, for $\alpha \in \{0, 1.0\}$ the mean aggregate throughput of RPR without proxy stripping is not more than two since the the hub deploys two transceivers, one for each fiber ring. In contrast, in RPR with proxy stripping the mean aggregate throughput is more than sixty and thus dramatically larger than two for both data collection and distribution. This is because apart from using the ring the hub node also sends/receives data via the star subnetwork, leading to a dramatically increased mean aggregate throughput.

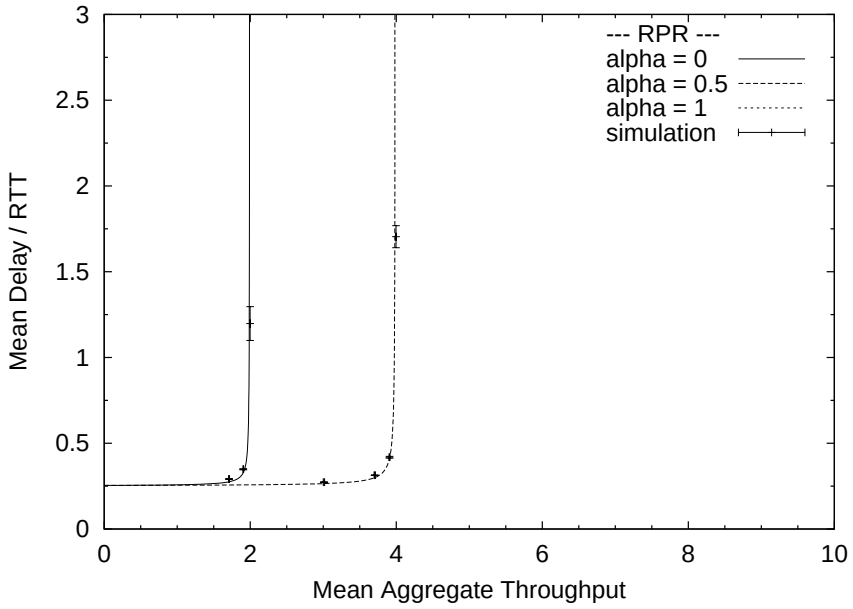


Figure 7.14: Mean delay vs. mean aggregate throughput of RPR without proxy stripping for asymmetric hot-spot traffic with $\alpha \in \{0, 0.5, 1.0\}$, $h = 1.0$, and $N = 256$.

7.2.4 Dimensioning of Star Subnetwork

In this section, we investigate the forwarding burden caused by proxy stripping and the resultant capacity requirements of the star subnetwork in greater detail. Recall from Section 7.1.4 that the star subnetwork was assumed to provide sufficient capacity to carry proxy-stripped traffic. In the following, we quantify the capacity requirements of the star subnetwork which must be met in order to avoid a bandwidth bottleneck. To this end, we consider proxy stripping node $i = 0$ under both symmetric uniform and symmetric hot-spot traffic, i.e., $\alpha = 0.5$

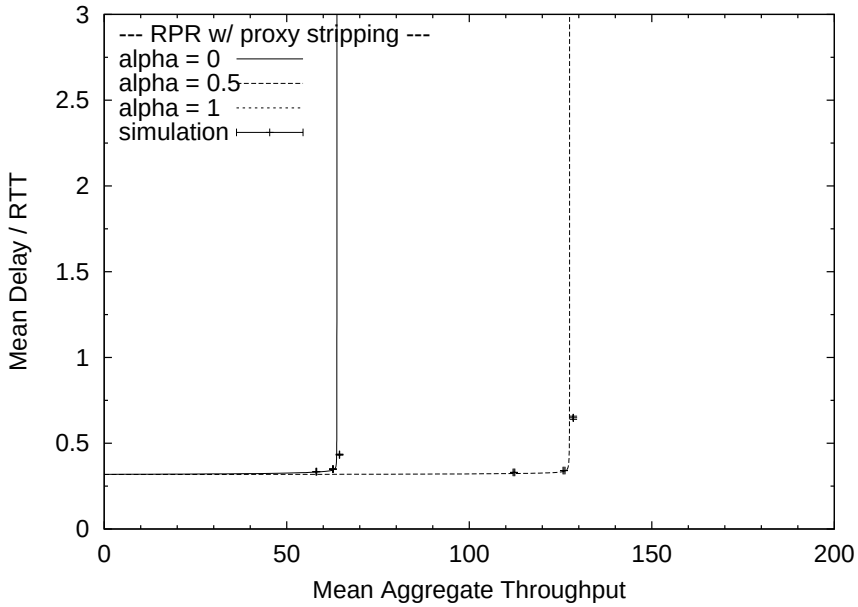


Figure 7.15: Mean delay vs. mean aggregate throughput of RPR with $P = 32$ proxy stripping nodes for asymmetric hot-spot traffic with $\alpha \in \{0, 0.5, 1.0\}$, $h = 1.0$, and $N = 256$.

and $h = 1/(N-1)$ or $h = 1.0$, respectively. To measure the capacity requirement of node $i = 0$ we use the ratio of star transceiver load and ring transceiver load. Note that this performance measure indicates the required star transmission rate normalized by the arrival rate of one ring. The star transceiver load at node $i = 0$ is identical to the amount of traffic $\rho_s(0)$ which arrives at the star transmit queue of node $i = 0$ (given by Eq. (7.36) of Section 7.1.4). The ring transceiver load at node $i = 0$ is identical to the amount of traffic which arrives at one of both ring transit queues of node $i = 0$. We choose the ring transit queue that belongs to the counterclockwise fiber ring. Thus, the ring transceiver load at node $i = 0$ is composed of all traffic coming from the transmit and transit queues of neighbor node $i = 1$. The ring transceiver load at node $i = 0$ is thus equal to the sum $\rho_t^{r-}(1) + \rho_t^{out}(1) + \rho_r^{r-}(1) + \rho_r^{s-}(1)$ (where the individual terms are given by Eqs. (7.28), (7.29), (7.32), and (7.35) of Section 7.1.4, respectively).

Figs. 7.16 and 7.17 depict the ratio of star transceiver load and ring transceiver load at node $i = 0$ vs. number of nodes N for symmetric uniform and hot-spot traffic with $P \in \{4, 8, 16, 32, 64\}$. Note that under uniform traffic node $i = 0$ is one of P proxy stripping nodes and thus represents the traffic present at the remaining $(P-1)$ proxy stripping nodes as well. From Fig. 7.16 we observe that the ratio increases for larger P under uniform traffic. This is because with an increasing P more nodes are attached to the star subnetwork. Therefore, more nodes communicate with each other via the star subnetwork, resulting in an increased star traffic volume. Also, we observe that for a given P the ratio decreases with increasing N. This is due to the fact that with P fixed and increasing N more nodes communicate with each other via the ring rather than the star subnetwork. As a consequence, the ring traffic increases and the star traffic decreases, leading to a smaller ratio. In summary, for uniform traffic it appears to be reasonable to use a moderate number of proxy stripping nodes P

compared to the number of nodes N. In doing so, the traffic load is well balanced between the ring and the star subnetwork. Moreover, choosing a moderate number of proxy stripping nodes P requires fewer dark fibers and star transceivers, each operating at a line rate that is slightly larger than that of the ring transceivers.

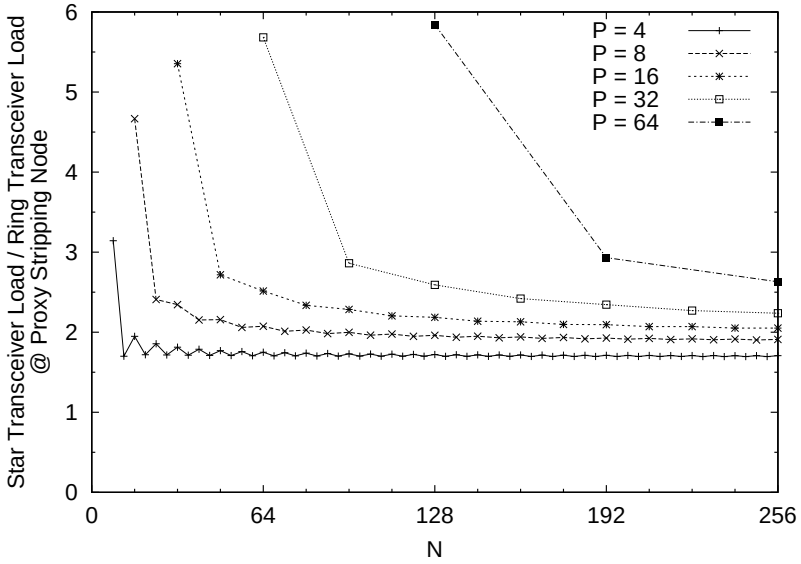


Figure 7.16: Ratio of star transceiver and ring transceiver loads vs. number of nodes N at proxy stripping node $i = 0$ for symmetric uniform traffic ($\alpha = 0.5$, $h = 1/(N-1)$) with $P \in \{4, 8, 16, 32, 64\}$.

Next, let us consider the ratio under hot-spot traffic, as shown in Fig. 7.17. Again, we observe that with increasing P the ratio becomes larger. Note, however, that under hot-spot traffic the ratio is significantly larger than under uniform traffic. This is because now all nodes have traffic destined only for hot-spot node $i = 0$, which in terms of hops is best reached via the short-cuts of the star subnetwork. To use these short-cuts, regular ring nodes send their hot-spot traffic towards their closest proxy stripping node, which then transmits the traffic directly to node $i = 0$ across the star subnetwork. Due to the lack of traffic between regular nodes the utilization of the ring is rather small compared to the star subnetwork. As a result, the ratio is much larger under hot-spot than uniform traffic. Furthermore, we observe from Fig. 7.17 that for a given P the ratio does not decrease for increasing N. Instead, for a given P there are certain values of N which provide a smaller or larger ratio, where the difference between the small and large ratios gets more pronounced with increasing P. Note that for each value of P the oscillations between small and large ratios get gradually smoother with increasing N. The reason for this is as follows. The number of regular ring nodes next to hub node $i = 0$ is equal to $(N/P - 1)$ in each direction. Among these nodes, $\lceil (N/P - 1)/2 \rceil$ nodes send their packets to node $i = 0$ via the ring while the remaining $\lfloor (N/P - 1)/2 \rfloor$ make use of the star subnetwork. Now, by gradually increasing $(N/P - 1)$ every second node sends its hot-spot traffic to node $i = 0$ either directly on the ring or via the star subnetwork. As a result, only one of the transceiver loads at node $i = 0$ is increased, i.e., either star or ring transceiver load, while the other one remains unchanged, resulting in the oscillations of the

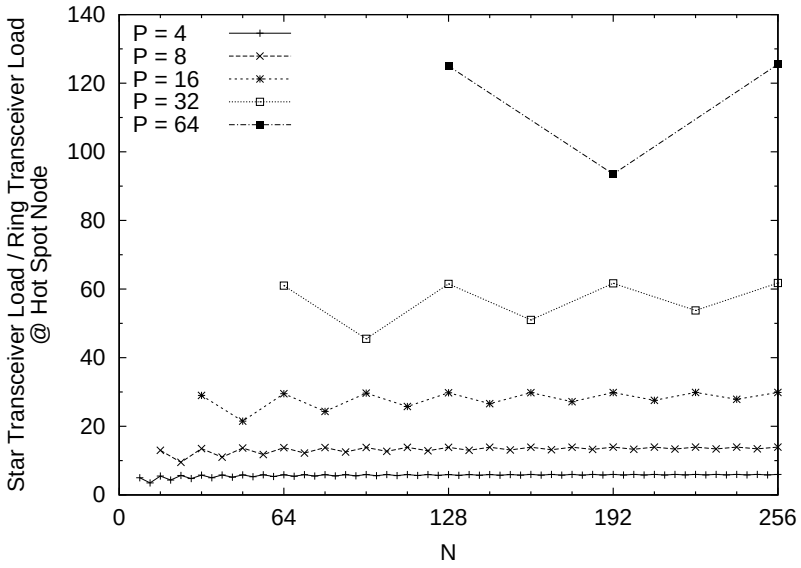


Figure 7.17: Ratio of star transceiver and ring transceiver loads vs. number of nodes N at hot-spot node $i = 0$ for symmetric hot-spot traffic ($\alpha = 0.5$, $h = 1.0$) with $P \in \{4, 8, 16, 32, 64\}$.

ratio. The oscillations get smoother since the relative traffic contribution of each newly added node gets smaller for increasing $(N/P - 1)$.

Given the number of ring nodes N, number of proxy stripping nodes P, and traffic type (uniform, non-uniform), the star subnetwork can be designed such that the above mentioned ratio of star transceiver and ring transceiver loads is satisfied. For a small ratio and/or small P, the star subnetwork may consist of a PSC with one star transceiver at each proxy stripping node. Whereas for a large ratio and/or large P each proxy stripping node may be equipped with an array of transceivers attached to a wavelength-routing AWG based star subnetwork which provides a large number of communication channels due to extensive spatial wavelength reuse, as discussed in Section 6.3.1.

7.3 Conclusions

As discussed in Section 2.3.2, metro networks often consist of interconnected metro core and metro edge rings whose traffic demands are completely different. While traffic in core rings is approximately uniform edge rings carry strongly hubbed 'hot-spot' traffic.

We have shown by means of probabilistic analysis and verifying simulations that proxy stripping increases the maximum mean aggregate throughput of RPR-like networks both under uniform and hot-spot traffic significantly. E.g., for uniform traffic, interconnecting 32 of 256 ring nodes via a star subnetwork increases the maximum mean aggregate throughput of RPR by a factor of almost ten. Another interesting observation is that the throughput-delay performance of RPR decreases dramatically under non-uniform symmetric and asymmetric traffic. In metro edge rings with hot-spot traffic demands the maximum aggregate throughput of RPR reduces to half of that obtained under uniform traffic.

However, while proxy stripping can result in a significantly improved performance, we

have also seen that the number of proxy stripping nodes P must be chosen properly. If P is chosen too small the fiber links close to the proxy stripping nodes get congested and form a bottleneck, resulting in an underutilized star subnetwork and a deteriorated overall throughput-delay performance. One approach to alleviate this congestion might be the use of transparent proxy stripping where ring nodes are not aware of proxy stripping nodes and thus do not create this type of hot-spot traffic on the ring. On the other hand, a large P requires too many star transceivers and dark fibers for interconnecting the proxy stripping nodes. A moderate number of proxy stripping nodes appear to provide a reasonable trade-off between throughput-delay performance improvement of RPR and costs.

This chapter also provides us guidelines how to dimension the star network. Looking at Fig. 7.16, a good rule of thumb seems to be that for uniform traffic the star transceivers should provide twice as much transmission capacity as the ring transceivers. This can also be motivated intuitively: The major part of a proxy nodes star traffic is forwarded from or to the two ring interfaces and therefore cannot exceed their aggregate capacity. Under hot-spot traffic it is advantageous to equip the hot-spot with multiple star interfaces. In contrast to RPR, in this case the performance even improves.

Chapter 8

Protectoration

AFTER evaluating the performance of the proxy stripping mechanism in an idealized setting in the previous chapter, we now proceed to incorporate RINGOSTAR's specific star architecture and MAC protocol into the analysis. As we target to make the performance evaluation more realistic we also consider link and node failures. Recall from Section 2.3.2 that optical networks must remain operable in case of failures, preferably with no or only few performance deterioration.

We propose a novel hybrid fault recovery technique termed '*protectoration*' that aims at combining the benefits of protection and restoration. More precisely, protectoration exploits the fast recovery time of protection mechanisms and the bandwidth efficiency of restoration mechanisms. Our proposed resilience technique enables RPR to recover from *multiple* link and node failures. Moreover, protectoration does not require any major modifications of RPR's MAC protocol as it makes use of RPR's own survivability mechanisms wrapping and steering. Thus, protectoration follows our strategy to provide an evolutionary upgrade of existing RPR networks.

This chapter is organized as follows. In the following section we review related work on failure recovery techniques for optical networks. After discussing RPR's failure recovery mechanism in Section 8.2.1 we describe the protectoration multiple-failure recovery technique in Sections 8.2.2 and 8.2.3. In Section 8.3 we extend the probabilistic analysis presented in the previous chapter to incorporate the star subnetwork's access protocol and the protectoration technique. Numerical results for various network configurations and failure scenarios are presented in Section 8.4. We provide verifying results for the analysis as well as additional results for self-similar traffic and finite buffers obtained by supplementary computer simulations. Section 8.5 concludes this chapter.

8.1 Related Work on Failure Recovery

The underlying principles and fundamental techniques used for achieving survivability in optical fiber single-channel and WDM networks are discussed in [159] and [160], respectively. An overview of fault management in WDM mesh networks is provided in [161]. The reported fault recovery techniques are categorized into protection or restoration techniques. For more detailed studies of various path and link protection and restoration techniques for WDM networks the interested reader is referred to [162, 163, 164, 165, 166, 167, 168, 169, 170] and the references therein. Restoration schemes for IP-over-WDM networks with GMPLS based

control signaling were investigated in [171] while protection schemes for these networks were studied in [172]. The provisioning of different levels of fault recovery has recently been studied, see for instance [173, 174, 175, 176] and references therein. Aside from fiber cuts, different equipment failures, e.g., transponder failure, can occur in optical networks. Both fiber and equipment protection schemes and their implementation aspects were examined in [177] and [178], respectively. For more details on interworking aspects between layers 2 and 3 of joint protection/restoration in IP-centric optical WDM networks the interested reader is referred to [179, 180]. The graph-theoretical aspects of augmented ring networks that deploy additional short-cut links to the ring have attracted considerable attention [181].

Most previously proposed fault recovery techniques for optical networks are either protection or restoration techniques. In this chapter, we describe and examine a hybrid fault recovery technique for optical ring networks that aims at combining the recovery time of protection (wrapping) and the bandwidth efficiency of restoration by using additional fiber short-cuts. We note that the recently reported *pre-configured* cycles (p-cycles) [182, 183] and the generalized *pre-cross-connected trail (PXT)* [184] have the same goal. However, both p-cycles and PXTs are designed for wide area networks (WANs) with a mesh rather than a ring topology.

8.2 Protectoration Protocol

In this section, we describe and discuss the protectoration fault recovery technique as an extension to the architecture and access protocol described in Section 6.3 and Section 6.4, respectively. Note that as long as there are no failures, the network operates the same as before.

Aside from link and node failures, other network elements may fail. In the star subnetwork, splitters, amplifiers, combiners, waveband partitioners/departitioners, PSC, or AWG may go down. Note that the various failures affect the network differently. For instance, while a fiber cut between a given ring-and-star homed node and the attached combiner disconnects only a single node from the star subnetwork, the entire star subnetwork goes down if the central hub fails, i.e., if both AWG and PSC fail. In the following, we assume that each node is able to detect any type of failure in both ring and star subnetworks. For a more detailed discussion on available techniques for fault detection in the ring and star subnetwork we refer the interested reader to [185] and [141], respectively.

8.2.1 Fault Recovery in RPR

Let us first briefly review the wrapping and steering techniques of RPR. Fig. 8.1 depicts an RPR bidirectional ring with $N = 16$ nodes, including a pair of source and destination nodes. The source node sends its data packets in clockwise direction since this direction provides the shortest path in terms of hops. For illustration, we assume that a single fiber cut has occurred right before the destination node. Upon detection of the link failure, the node on the left-hand side of the fiber cut wraps the traffic away from the link failure in the counterclockwise direction. In addition, the node on the left-hand side of the fiber cut broadcasts a control packet in the counterclockwise direction in order to inform all other nodes about the link failure. The wrapped traffic travels all the way back to the source node. Upon learning that a link failure has occurred, the source node steers the traffic away from the fiber cut and sends all traffic in the counterclockwise direction. Note that the two protection techniques wrapping

and steering lead to a rather inefficient use of bandwidth resources due to (*i*) the round-trip between source node and wrapping node without getting closer to the destination node, and (*ii*) the secondary path (counterclockwise direction in our example) which is longer than the primary path in terms of hops. Furthermore, in case of an additional link or node failure on the secondary path the source node would be unable to send traffic to the destination node since the ring network would be divided into two disjoint subrings, one containing the source node and the other one the destination node. Thus, the protection techniques of the bidirectional RPR ring network are able to recover only from a single link failure. Likewise, the RPR ring is able to recover only from a single node failure.

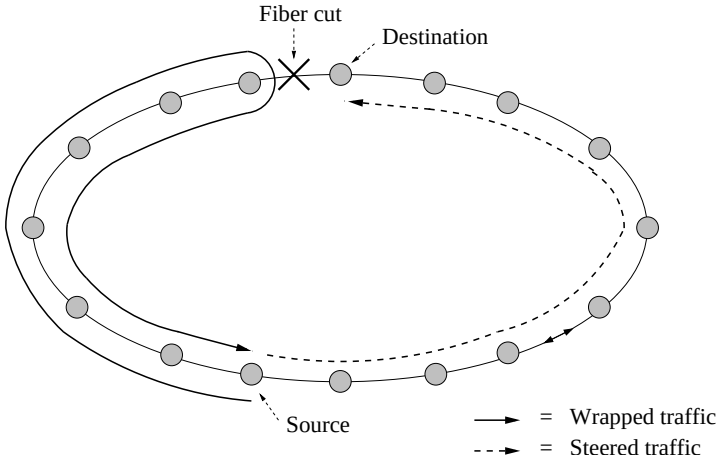


Figure 8.1: RPR bidirectional ring with $N = 16$ nodes using wrapping and steering in the event of a fiber cut.

Next, we explain the protectoration technique in greater detail. The protectoration technique builds on the wrapping and steering techniques of RPR and thus provides an evolutionary upgrade of RPR. Moreover, protectoration makes RPR resilient against multiple link and node failures in an efficient manner, as we shall see shortly. In the following subsection 8.2.2, we first consider link and node failures only in the ring subnetwork while the star subnetwork is assumed to work properly. In subsection 8.2.3, we also take failures in the star subnetwork into account.

8.2.2 Failures Only in Ring Subnetwork

Fig. 8.2 depicts an RPR bidirectional ring with $N = 16$ nodes, where $N_{rs} = 4$ ring-and-star homed nodes are interconnected via the star subnetwork of Section 6.3.2 and $N_r = N - N_{rs} = 12$ are conventional ring homed nodes. Recall from Section 6.2 that the ring-and-star homed nodes perform proxy stripping. Again, for illustration we consider a pair of source and destination nodes and a single fiber cut, as shown in Fig. 8.2. The source node sends all data packets intended for the destination node to its closest proxy stripping node, which in turn forwards the data packets across the single-hop star subnetwork to the proxy stripping node that is closest to the destination node. Upon detection of the fiber cut, the data packets are wrapped and return to the proxy stripping node closest to the destination node. Now, instead of forwarding the wrapped traffic to the source node on the counterclockwise ring (as done in conventional RPR), the corresponding proxy stripping node sends the wrapped data packets

across the single-hop star subnetwork to that proxy stripping node which is on the other side of the fiber cut. The latter proxy stripping node receives the data packets from the star subnetwork and forwards them to the destination node on the counterclockwise ring. Note that the former proxy stripping node (on the left-hand side of the link failure) deploys not only proxy stripping but also steering of wrapped traffic. Thus, wrapped traffic does not have to go back all the way to the source node but is sent along a single-hop path to another proxy stripping node. Compared to Fig. 8.1, in the event of a single fiber cut, wrapped traffic does neither make a round-trip between source and destination nodes nor take any long secondary path. As a result, wrapped traffic consumes significantly fewer bandwidth resources on the ring network, resulting in a more efficient use of bandwidth.

After learning about the fiber cut, the source node, and its closest reachable ring-and-star homed node, in case proxy stripping results in an intact source-to-destination route with a smaller hop count, transmit the data packets along a different path. In our example, after learning that the fiber cut has occurred the source node sends the data packets intended for the destination node to its closest proxy stripping node, which in turn sends the proxy stripped packets on the star subnetwork directly to the proxy stripping node on the right-hand side of the destination node. The latter proxy stripping node finally forwards the data packets to the destination node on the counterclockwise ring. Hence, protectoration requires significantly fewer bandwidth resources on the ring than conventional steering which would use the peripheral counterclockwise ring that is significantly longer than the short-cuts of the star subnetwork in terms of hops.

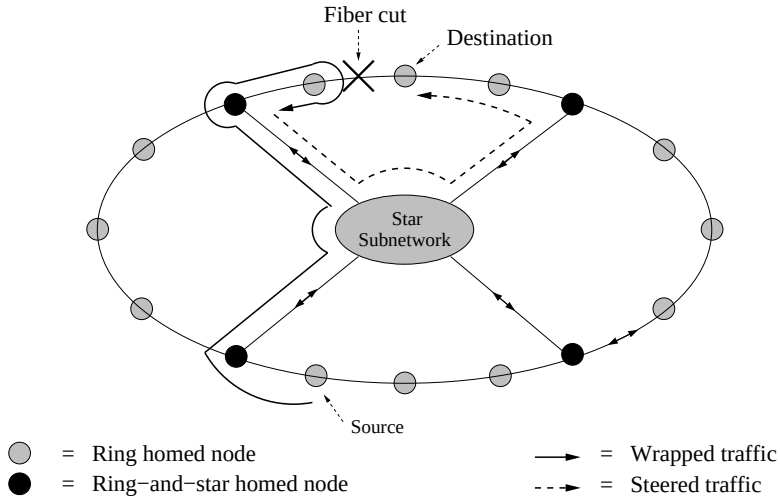


Figure 8.2: RPR bidirectional ring with $N_r = 12$ ring homed nodes and $N_{rs} = 4$ ring-and-star homed nodes using protectoration in the event of a fiber cut.

Note that a ring-and-star homed node is able to determine which packets have to be proxy stripped by using the source and destination addresses available in RPR's MAC address fields of each packet. These MAC addresses together with the direction a given packet comes from, enable each ring-and-star homed node to determine whether a given data packet has undergone wrapping or not. If wrapping has taken place, the corresponding ring-and-star homed node recomputes the shortest path taking the link failure into account and sends a given wrapped packet along the updated shortest path. Depending on the updated shortest path, the corresponding ring-and-star homed node forwards a given wrapped data packet

either on the ring or star subnetwork. Note that each node computes the shortest path by using its topology database. Each node maintains and updates its topology database by means of RPR's built-in topology discovery protocol (see Section 9.1.3).

Clearly, single-failure scenarios also include node failures apart from link failures. While a single failed ring homed node triggers the same procedure as above, special attention has to be paid to a failed ring-and-star homed node. If a ring-and-star homed node goes down it is not further available for proxy stripping traffic from the ring subnetwork and forwarding traffic coming from the star subnetwork. In this case, the two ring homed nodes adjacent to the failed proxy stripping node detect the failure and inform the remaining nodes by sending control packets. After learning about the failed proxy stripping node the remaining nodes do not send traffic to the failed ring-and-star homed node. Instead, the neighboring proxy stripping nodes of the failed proxy stripping node take over its role of proxy stripping regular traffic and steering wrapped traffic.

Next, let us consider multiple failures in the ring subnetwork. If there are multiple link and/or node failures on the ring subnetwork, nodes can use the star subnetwork to bypass the failures. Thus, with an intact star subnetwork multiple link and/or node failures on the ring subnetwork may occur simultaneously without losing full connectivity. Note, however, that full connectivity in the event of multiple failures is only guaranteed if no more than one link or node failure occurs between each pair of ring-and-star homed nodes. Otherwise, one or more nodes between a given pair of ring-and-star homed nodes are disconnected from the network.

8.2.3 Failures in Both Ring and Star Subnetworks

Failures in the star subnetwork include fiber cuts and nonfunctional network devices such as failed combiners/splitters, waveband (de)partitioners, AWG, PSC, and amplifiers. Depending on the failure, only one, a subset, or all ring-and-star homed nodes are disconnected from the star subnetwork. More precisely, a fiber cut between a given ring-and-star homed node and the combiner/splitter port to which it is attached disconnects only the ring-and-star homed node from the star subnetwork. If a given combiner/splitter, amplifier, waveband (de)partitioner, or any fiber between these devices goes down, all S corresponding ring-and-star homed nodes are disconnected from the star subnetwork. If the central hub (AWG and PSC) goes down, the connectivity of the star subnetwork is entirely lost, reducing the network to a conventional bidirectional RPR ring network. If a given ring-and-star homed node detects that it is disconnected from the star subnetwork it is unable to send and receive traffic to and from the star subnetwork. After detecting disconnection, the affected ring-and-star homed node informs all remaining nodes by broadcasting a control packet on either ring and acts subsequently as a conventional ring homed node. Failures in the ring subnetwork are handled as described above.

8.2.4 Discussion

The bidirectional RPR ring network with its two protection techniques wrapping and steering is able to guarantee full connectivity only if no more than one link or node failure occurs. Full connectivity also in the event of multiple link and/or node failures can be achieved by interconnecting several ring nodes via a star subnetwork. In doing so, the ring is divided into several segments, each comprising the nodes between two adjacent ring-and-star homed

nodes. Each segment is able to recover from a single link or node failure without losing full connectivity of the network. Thus, the number of fully recoverable link and/or node failures is identical to the number of ring-and-star homed nodes, provided that there is no more than one failure in each segment.

Similar to RPR ring networks, both ring and ring-and-star homed nodes perform wrapping and steering. In addition, ring-and-star homed nodes also perform proxy stripping. By means of proxy stripping, wrapped traffic is sent across single-hop short-cuts to the neighboring ring-and-star homed node, thereby bypassing the link or node failure(s) of the corresponding ring segment(s). As opposed to the RPR bidirectional ring, wrapped data packets do not have to travel back to the corresponding source node. Instead, steering is also done by the ring-and-star homed node that receives wrapped data packets by sending the wrapped data packets across the single-hop short-cuts of star subnetwork rather than along the ring subnetwork. In doing so, the corresponding ring-and-star homed node restores the network connectivity in a more efficient manner. After learning that a failure has occurred, the source node steers the traffic along the updated shortest path by capitalizing on proxy stripping. Consequently, steered traffic does not have to travel on the longer secondary path along the peripheral ring, requiring fewer bandwidth resources and resulting in an improved bandwidth efficiency.

The proposed multiple-failure recovery technique combines the fast recovery time of *protection* (wrapping) and the bandwidth efficiency of *restoration* (steering together with proxy stripping). Accordingly, we call this hybrid approach *protectoration*.

8.3 Analysis

In this section, we develop an analytical model for investigating the protectoration technique in terms of stability, utilization, and throughput-delay performance. We also address the dimensioning and identify the bottlenecks of the network. We note that in our analysis we do not take fairness control into account. The obtained results are intended to give the maximum achievable throughput-delay performance of the protectoration technique and to provide an upper bound that enables the performance comparison of fairness control mechanisms, e.g., [186, 187, 188].

8.3.1 Assumptions

In our analysis, we make the following assumptions:

- *Single-queue mode*: We examine the single-queue mode of RPR, i.e., each node is equipped with one PTQ but no STQ on either ring. For sending proxy-stripped traffic across the star subnetwork, each ring-and-star homed node is equipped with an additional star transit queue.
- *Infinite FIFO queues*: All queues are assumed to be FIFO queues of infinite capacity, i.e., there is no packet loss due to buffer overflow. In particular, assuming an infinite PTQ is well suited to model the lossless transit path of RPR. (In our verifying simulations we use finite-size FIFO queues which provide very good matches between analysis and simulation results.)
- *Propagation delay*: The N nodes are equally spaced on the ring. The propagation delay between two adjacent ring nodes is given by τ. Thus, the RTT of the RPR ring equals $N \cdot \tau$. The propagation delay of the PSC and AWG star subnetwork is equal to

τ_{PSC} and τ_{AWG}, respectively. Both τ_{PSC} and τ_{AWG} are assumed to be the same for all ring-and-star homed nodes. All propagation delays are given in slots.

- *Unicast traffic*: We consider unicast traffic, i.e., all data transmissions are point-to-point.
- *Packet generation process*: At node i the average number of locally generated packets that are destined for node j per frame is equal to $\sigma(i,j) \geq 0$. For the stability analysis in Section 8.3.2 the packet generation process is assumed to be stationary and ergodic. For the delay analysis in Section 8.3.4 the packet generation process is assumed to be Poissonian.
- *Packet length distribution*: We consider variable-size data packets with a length of L slots, $1 \leq L \leq F - D \cdot S$. The packet length distribution is independent from source node i and destination node j. Let T be a random variable denoting the packet transmission time (in slots), and let $E[T]$ denote its mean.

8.3.2 Stability and Dimensioning

Let us introduce the following definitions. For locally generated traffic at node i we define:

- $\sigma_t^+(i)$ denotes the mean number of locally generated packets at node i per frame which are sent in the clockwise direction of the ring subnetwork.
- $\sigma_t^-(i)$ denotes the mean number of locally generated packets at node i per frame which are sent in the counterclockwise direction of the ring subnetwork.
- $\sigma_t^*(i)$ denotes the mean number of locally generated packets at node i per frame which are sent directly across the star subnetwork (holds only for ring-and-star homed nodes).

Similarly, for in-transit traffic at node i we define:

- $\sigma_r^+(i)$ denotes the mean number of packets arriving at node i per frame which are forwarded in the clockwise direction of the ring subnetwork.
- $\sigma_r^-(i)$ denotes the mean number of packets arriving at node i per frame which are forwarded in the counterclockwise direction of the ring subnetwork.
- $\sigma_r^*(i)$ denotes the mean number of packets arriving at node i per frame which are forwarded on the star subnetwork (holds only for ring-and-star homed nodes).

We note that if node i is a ring-and-star homed node the quantities $\sigma_r^+(i)$, $\sigma_r^-(i)$, and $\sigma_r^*(i)$ account for in-transit traffic that comes from and goes to both the ring subnetwork and the star subnetwork, e.g., $\sigma_r^+(i)$ for a ring-and-star homed node i accounts for both the traffic that arrives from the ring subnetwork and is to be forwarded in the clockwise direction over the ring as well as the traffic that arrives from the star subnetwork and is to be forwarded in the clockwise direction over the ring.

Next, let $p_{ij}(e)$ denote the probability that a data packet, that is generated at node i and is destined to node j, traverses a given (directed) fiber link e of the ring subnetwork between two adjacent nodes. Similarly, for ring-and-star homed nodes k and l let $p_{ij}(k,l)$ denote the probability that a data packet, that is generated at node i and is destined to node j, traverses the star subnetwork from k to l. The calculation of the probabilities $p_{ij}(e)$ and $p_{ij}(k,l)$ depends on the status of the network and the applied routing. As explained in Section 8.2.2, all nodes deploy shortest path routing, noting that the presence of link and/or node failures may affect the shortest path since failed links and nodes can no longer be traversed. The fault scenarios under consideration are single and multiple link and/or node failures. If for a given pair of source and destination nodes there exist multiple shortest paths, the traffic load

is balanced among the multiple shortest paths equally. (Alternatively, the tie could be broken by arbitrating the routing based on the indices of both source and destination nodes. We note that this assumption of load balancing is for analytical simplicity. To guarantee in-order packet delivery a given source node has to send all packets to the corresponding destination node along the same path by choosing one of the multiple shortest paths.) To determine $p_{ij}(e)$ and $p_{ij}(k,l)$ for a given scenario, each link that is on the shortest path(s) is weighed by the probability with which it is used by source node i and destination node j. Hence, if there is a single shortest path between nodes i and j all links belonging to the shortest path have a weight of one whereas the remaining links which are not part of the shortest path have a weight of zero. Otherwise, if there are multiple shortest path between a given pair of source and destination nodes each link belonging to a shortest path is weighed by a factor of one over the number of shortest paths. (We do not provide explicit expressions for $p_{ij}(e)$ and $p_{ij}(k,l)$ here, but note that for a given scenario either with or without failures the probabilities can be easily calculated by means of a computer program.)

Now, let i^+ denote the link on the ring subnetwork between node i and its neighboring node in clockwise direction and i^- denote the link on the ring subnetwork between node i and its neighboring node in counterclockwise direction. We then obtain:

$$\sigma_t^+(i) = \sum_j p_{ij}(i^+) \cdot \sigma(i,j) \tag{8.1}$$

$$\sigma_t^-(i) = \sum_j p_{ij}(i^-) \cdot \sigma(i,j) \tag{8.2}$$

$$\sigma_t^*(i) = \sum_{j,l} p_{ij}(i,l) \cdot \sigma(i,j) \tag{8.3}$$

$$\sigma_r^+(i) = \sum_{\substack{k,j \\ k \neq i}} p_{kj}(i^+) \cdot \sigma(k,j) \tag{8.4}$$

$$\sigma_r^-(i) = \sum_{\substack{k,j \\ k \neq i}} p_{kj}(i^-) \cdot \sigma(k,j) \tag{8.5}$$

$$\sigma_r^*(i) = \sum_{\substack{k,j,l \\ k \neq i}} p_{kj}(i,l) \cdot \sigma(k,j). \tag{8.6}$$

By using Eqs. (8.1)–(8.2) and (8.4)–(8.5) the traffic loads on the ring subnetwork are calculated as follows:

$$\rho_a^b(i) = \frac{E[T]}{F} \cdot f \cdot \sigma_a^b(i), \tag{8.7}$$

where $a \in \{r,t\}$, $b \in \{+,-\}$, and f denotes the ratio of the line rate of the star subnetwork and the line rate of the ring subnetwork. Note that in general the star subnetwork needs to operate at a higher line rate than the ring subnetwork in order to cope with the proxy-stripped traffic of both fiber rings, i.e., $f \geq 1$.

Given this, we are now able to formulate the stability conditions of the ring subnetwork, the PSC, and the AWG. The stability condition of the ring subnetwork is given by

$$\rho_t^b(i) + \rho_r^b(i) < 1, \tag{8.8}$$

which needs to hold for every node i, $0 \leq i \leq N-1$ and direction $b \in \{+,-\}$. Noting that a given ring-and-star homed node k can send at most one packet per frame over the PSC, the

stability condition of the PSC is given by

$$\sigma_t^*(k) + \sigma_r^*(k) < 1, \tag{8.9}$$

which needs to hold for each ring-and-star homed node k, $k = 1, 2, \ldots, D \cdot S$.

Under the assumption that the stability conditions for the ring subnetwork and PSC hold, we proceed to determine the stability condition of the AWG. Let α_{kl}, $k, l = 1, 2, \ldots, D \cdot S$ and $k \neq l$, denote the mean number of data packets to be transmitted from (ring-and-star homed) node k to (ring-and-star homed) node l on the PSC per frame, which is given by

$$\alpha_{kl} = \sum_{i,j} p_{ij}(k, l) \cdot \sigma(i, j). \tag{8.10}$$

Note that $\sum_l \alpha_{kl} = \sigma_t^*(k) + \sigma_r^*(k)$, which is less than one for the assumed stable network by the stability condition of the PSC, see Eq. (8.9). Moreover note that a given ring-and-star homed node k can send at most one packet per frame to another ring-and-star homed node l, hence α_{kl} is equivalent to the probability that node k has a packet to send to l in a given frame. For simplicity, we assume that all packets sent by two or more nodes to the same receiver across the PSC within the same frame collide and need to be retransmitted (this simplifying assumption still provides quite accurate results, as we will see in Section 8.4). Assuming independence among the ring-and-star homed nodes, the probability P that in a given frame a collision occurs at receiver l is given by

$$P = 1 - \prod_{k=1}^{D \cdot S} (1 - \alpha_{kl}) - \sum_{k=1}^{D \cdot S} \alpha_{kl} \cdot \prod_{\substack{j=1 \\ j \neq k}}^{D \cdot S} (1 - \alpha_{jl}), \tag{8.11}$$

where $\alpha_{kk} = 0$ and $l = 1, 2, \ldots, D \cdot S$. Note that $\prod_{k=1}^{D \cdot S}(1 - \alpha_{kl})$ is the probability that the (ring-and-star homed) node l does not receive a data packet. Also, note that

$$\sum_{k=1}^{D \cdot S} \alpha_{kl} \cdot \prod_{\substack{j=1 \\ j \neq k}}^{D \cdot S} (1 - \alpha_{jl}) =: r_l \tag{8.12}$$

is the probability that (ring-and-star homed) node l receives exactly one data packet, which we denote by r_l. To see this note that with probability α_{kl}, node k has a packet for node l in a given frame, and with probability $\prod_{j=1,\ j \neq k}^{D \cdot S}(1 - \alpha_{jl})$ none of the other ring-and-star home nodes j, $j = 1, \ldots, D \cdot S$, $j \neq k$, has a packet for l in the frame, i.e., the transmission from k to l proceeds without collision. Moreover, note that with the approximating assumption that a node can receive at most one packet per frame without collision over the PSC, r_l is equivalent to the mean number of packets that are transmitted in a given frame without a collision to ring-and-star homed node l, $l = 1, 2, \ldots, D \cdot S$.

Out of the offered load $\sum_{k=1}^{D \cdot S} \alpha_{kl}$ (in mean number of packets per frame) to ring-and-star homed node l per frame, the load r_l is sent across the PSC and the remaining load $\sum_{k=1}^{D \cdot S} \alpha_{kl} - r_l$ is sent across the AWG. As a consequence, we obtain two stability conditions of the AWG. The first stability condition is given by

$$\frac{E[T]}{F} \cdot \left(\sum_{k=1}^{D \cdot S} \alpha_{kl} - r_l \right) < 1, \tag{8.13}$$

which needs to hold for all $l,\ l = 1, 2, \ldots, D \cdot S$ and accounts for receiver collisions but does not consider the limited number of available wavelength channels of the AWG. The second stability condition of the AWG takes the limited number of wavelength channels into account and is given by

$$\frac{E[T]}{F} \cdot \sum_{k \in \mathcal{K}_\iota} \sum_{l \in \mathcal{L}_\omega} \left\{ \alpha_{kl} - \alpha_{kl} \cdot \prod_{\substack{j=1 \\ j \neq k}}^{D \cdot S} (1 - \alpha_{jl}) \right\} < R, \tag{8.14}$$

where $\mathcal{K}_\iota$ and $\mathcal{L}_\omega$ denote the two subsets of ring-and-star homed nodes which are attached to AWG input port ι and AWG output port ω, respectively, with $\iota, \omega \in \{1, 2, \ldots, D\}$. This condition needs to hold for all AWG input-output port pairs $\iota,\ \omega \in \{1, 2, \ldots, D\}$. To understand this second stability condition, note that $\sum_{k \in \mathcal{K}_\iota} \sum_{l \in \mathcal{L}_\omega} \alpha_{kl}$ is the mean number of packets to be sent by nodes attached to AWG input port ι to the nodes attached to AWG output port ω per frame, and $\sum_{k \in \mathcal{K}_\iota} \sum_{l \in \mathcal{L}_\omega} \alpha_{kl} \cdot \prod_{\substack{j=1 \\ j \neq k}}^{D \cdot S} (1 - \alpha_{jl})$ is the mean number of packet that are sent per frame without a collision over the PSC between these considered nodes and hence do not require transmission over the AWG.

The network is stable if and only if all four stability conditions (8.8), (8.9), (8.13), and (8.14) are satisfied. If one or more stability conditions can not be satisfied then the network becomes unstable. Thus, for a given traffic load the network has to be dimensioned such that all four stability conditions are satisfied.

8.3.3 Utilization and Bottleneck

In this section, we briefly describe how in a stable network the channel utilization of the ring and star subnetworks can be found by using Eqs. (8.8), (8.9), (8.12), and (8.14), respectively. The utilization of the (data) channel on the ring subnetwork at node i, $0 \leq i \leq N - 1$, in clockwise and counterclockwise direction is equal to $\rho_t^+(i) + \rho_r^+(i)$ and $\rho_t^-(i) + \rho_r^-(i)$, respectively. In the star subnetwork, the utilization of the control channel λ_c at node k equals $\sigma_t^*(k) + \sigma_r^*(k)$, where $k = 1, 2, \ldots, D \cdot S$. The utilization of the PSC home data channel λ_l of ring-and-star homed node l is approximately given by $E[T] \cdot r_l / F$, where $l = 1, 2, \ldots, D \cdot S$. Moreover, the utilization of the R data channels available between AWG input port ι and AWG output port ω is approximately given by

$$\frac{E[T] \cdot \sum_{k \in \mathcal{K}_\iota} \sum_{l \in \mathcal{L}_\omega} \alpha_{kl} \cdot \left(1 - \prod_{\substack{j=1 \\ j \neq k}}^{D \cdot S} (1 - \alpha_{jl}) \right)}{F \cdot R}, \tag{8.15}$$

where $\mathcal{K}_\iota$ and $\mathcal{L}_\omega$ denote the two subsets of ring-and-star homed nodes which are attached to AWG input port ι and AWG output port ω, respectively, with $\iota, \omega \in \{1, 2, \ldots, D\}$.

Note that the utilization of the various network elements enables the identification of the bottleneck. Clearly, the bottleneck of the network is identical to the network element with the largest utilization.

8.3.4 Delay Analysis

In this section, we analyze the mean delay of the network for Poisson traffic. The mean waiting times in both the transmit queue and transit queue were analyzed in [157] for the case of unidirectional rings. By extending these results to our bidirectional ring subnetwork

we obtain for node i, $0 \leq i \leq N-1$, the waiting time (in slots) in the ring transmit queues of both directions approximately as

$$d_t^{\pm}(i) = \frac{(\rho_r^{\pm}(i) + \rho_t^{\pm}(i)) \cdot E[T^2]}{2 \cdot (1 - \rho_r^{\pm}(i) - \rho_t^{\pm}(i)) \cdot (1 - \rho_r^{\pm}(i)) \cdot E[T]} \tag{8.16}$$

and the waiting time (in slots) in the ring transit queues of both directions approximately as

$$d_r^{\pm}(i) = \frac{\rho_t^{\pm}(i) \cdot E[T^2]}{2 \cdot (1 - \rho_r^{\pm}(i)) \cdot E[T]}. \tag{8.17}$$

At each ring-and-star homed node the buffering and sending of data packets (and control packets) across the PSC of the star subnetwork is modelled as an M/D/1 queue, where the service time is equal to one frame. Hence, for ring-and-star homed node k, $k = 1, 2, \ldots, D \cdot S$, the waiting time in the star transmit queue that stores locally generated traffic is approximately given by

$$d_t^*(k) = \frac{\sigma_r^*(k) + \sigma_t^*(k)}{2 \cdot (1 - \sigma_r^*(k) - \sigma_t^*(k)) \cdot (1 - \sigma_r^*(k))} \tag{8.18}$$

and the waiting time in the star transmit queue that stores proxy-stripped traffic is approximately given by

$$d_r^*(k) = \frac{\sigma_r^*(k) + \sigma_t^*(k)}{2 \cdot (1 - \sigma_r^*(k))}. \tag{8.19}$$

Note that both $d_t^*(k)$ and $d_r^*(k)$ are given in frames.

Data packets that are sent across the PSC of the star subnetwork experience a propagation delay of τ_{PSC}. If a given data packet suffers from a channel collision on the PSC it will be scheduled for retransmission across the AWG by all ring-and-star homed nodes in a distributed manner. The fraction of traffic β that is sent across the AWG is given by

$$\beta = \frac{\sum_l \left(\sum_k \alpha_{kl} - r_l\right)}{\sum_{i,j} \sigma(i,j)}. \tag{8.20}$$

A given data packet that is sent across the AWG experiences a certain scheduling delay prior to the propagation delay of the AWG τ_{AWG}. We assume that the scheduling delay is significantly smaller than the combined propagation delay $\tau_{PSC} + \tau_{AWG}$ and neglect it in the following. Note that this assumption appears to be reasonable since the very high-speed star subnetwork operates at a line rate that is by a factor of f larger than that of the ring subnetwork.

By weighing the different waiting times and propagation delays with the probabilities $p_{ij}(e)$ and $p_{ij}(k,l)$ we obtain the mean delay on the ring subnetwork and the PSC of the star subnetwork for any pair of source node i and destination node j as follows. If node i is a ring homed node, the mean delay D_{ij} between source node i and destination node j in slots is given by

$$\begin{aligned} D_{ij} \;=\; & p_{ij}(i^+) \cdot d_t^+(i) + p_{ij}(i^-) \cdot d_t^-(i) + E[T] + \tau + \\ & + \sum_{k \neq i,j} \left[p_{ij}(k^+) \cdot (d_r^+(k) + \tau) + p_{ij}(k^-) \cdot (d_r^-(k) + \tau)\right] + \\ & + \sum_{k,l} p_{ij}(k,l) \cdot (d_r^*(k) \cdot F + \tau_{PSC}) \,. \end{aligned} \tag{8.21}$$

If node i is a ring-and-star homed node, the mean delay D_{ij} between source node i and destination node j in slots is given by

$$\begin{aligned} D_{ij} &= p_{ij}(i^+) \cdot d_t^+(i) + p_{ij}(i^-) \cdot d_t^-(i) + E[T] + \tau + \\ &\quad + \sum_{k \neq i,j} \left[p_{ij}(k^+) \cdot (d_r^+(k) + \tau) + p_{ij}(k^-) \cdot (d_r^-(k) + \tau) \right] + \\ &\quad + \sum_{\substack{k,l \\ k \neq i}} p_{ij}(k,l) \cdot (d_r^*(k) \cdot F + \tau_{PSC}) + \\ &\quad + \sum_{l} p_{ij}(i,l) \cdot (d_t^*(i) \cdot F + \tau_{PSC}) \,. \end{aligned} \tag{8.22}$$

By taking into account the additional delay encountered by traffic sent on the AWG of the star subnetwork the mean delay of the network D for a given scenario is given by

$$D = \frac{\sum_{i,j} \sigma(i,j) \cdot D_{ij}}{\sum_{i,j} \sigma(i,j)} + \beta \cdot \tau_{AWG}, \tag{8.23}$$

where β is given in Eq. (8.20). Note that the mean delay D in Eq. (8.23) is for a given scenario with certain probabilities $p_{ij}(e)$ and $p_{ij}(k,l)$. The mean delay D is given in slots.

8.4 Results

In this section, we numerically investigate the performance of the protectoration technique for Poisson and self-similar traffic. Throughout our investigations we consider *uniform* unicast traffic which is typically found in metro core networks [156]. More precisely, at node i the average number of locally generated packets that are destined for node j per frame equals $0 \leq \sigma \leq 1$, if $i \neq j$, and $\sigma = 0$, if $i = j$, where $i, j = 1, 2, \ldots, N$. The parameters are set to the following default values: $D = 8$, $S = 1$, $R = 1$, and $F = 400$. The length L of (data) packets is uniformly distributed over the interval of $[1, F - D \cdot S] = [1, 392]$ slots, where one slot is four bytes (octets) long (we consider 4 byte sufficient for accommodating destination address, length, and priority fields in a control packet). The ring operates at a line rate of 2.5 Gbit/s and has a circumference of 100 km. Considering cut-thru forwarding on the ring subnetwork the RTT of the ring subnetwork is given by RTT = 100 km $/(2 \cdot 10^5$ km/s). For the star subnetwork we set $\tau_{AWG} = \tau_{PSC} = \frac{100 \text{ km}/\pi}{2 \cdot 10^5 \text{ km/s}}$. Thus, the RTT of the ring subnetwork is assumed to be π times as large as the one-way end-to-end propagation delay of the star subnetwork.

8.4.1 Poisson Traffic

To verify the accuracy of our analysis we provide additional simulation results. As opposed to the analysis, in our simulations we also account for the access delay on the PSC control channel and the scheduling delay on the AWG data channel of the star subnetwork. In each simulation we have generated 10^6 packets including a warm-up phase of 10^5 packets. Using the method of batch means we calculated the 95% confidence intervals for both mean delay and mean aggregate throughput. The mean delay is given in multiples of the RTT of the ring subnetwork and the mean aggregate throughput is equal to the mean number of transmitting nodes in steady state.

Operation without Failures

Let us first consider the network operating without failures. Figs. 8.3–8.5 depict the mean delay vs. mean aggregate throughput for a fixed number of $N_{rs} = D \cdot S = 8$ ring-and-star homed nodes and different numbers of nodes $N \in \{8, 16, 32, 64, 128, 256\}$. We consider different speed-up factors f from $f = 1$, i.e., both star and ring subnetworks operate at the same line rate, in Fig. 8.3 to $f = 16$, i.e., the line rate of the star subnetwork is sixteen times as large as the line rate of the ring subnetwork and thus equals 40 Gbit/s, in Fig. 8.5. The individual curves are obtained by increasing the packet generation probability σ from values close to zero to values that result in very large delays. Focusing for now on Fig. 8.3 we observe that for $N = N_{rs} = 8$ we obtain the lowest delay and largest throughput. For $N = N_{rs}$ all nodes are attached to the star subnetwork and can communicate with each other in a single hop. In particular, each node sends traffic to its two adjacent nodes via the ring subnetwork and to the other nodes via the star subnetwork. At small traffic loads no significant queuing occurs and the mean delay is mainly dictated by the propagation delay encountered on both the ring and star subnetworks. Due to the fact that the propagation delay between two adjacent ring nodes (RTT/8) is smaller than the propagation delay of the star subnetwork (RTT/π), the mean delay is smaller than RTT/π at small traffic loads. For $N > 8$ only a subset of nodes is attached to the star subnetwork and packets need to increasingly traverse multiple nodes on their way from source node to destination node. At small traffic loads the mean delay for $N > 8$ is bounded by the propagation delay from the source node to the closest ring-and-star homed node, which does not exceed RTT/16, plus the propagation delay of the star subnetwork (RTT/π), plus the propagation delay from the destination node to the closest ring-and-star homed node, which does not exceed RTT/16. As shown in Fig. 8.3, for all values of N the mean delay and mean aggregate throughput increase with increasing traffic loads (packet generation probabilities σ). For $N \in \{8, 16, 256\}$ we provide verifying simulation results. Analytical and simulation results match quite well. At small traffic loads the simulation provides a slightly larger mean delay than the analysis. This is because the simulation accounts for the additional access delay on the PSC control channel and the scheduling delay on the AWG data channel of the star subnetwork, as opposed to the analysis. We also observe from Fig. 8.3 that the maximum mean aggregate throughput slightly decreases with increasing number of nodes N. This is because with an increasing number of nodes N and a fixed number N_{rs} of ring-and-star homed nodes, each ring-and-star homed node collects short-cut traffic from an increasing number of ring-homed nodes. This results in increased loads on the ring segments connecting the ring-homed nodes to the ring-and-star homed nodes and the star subnetwork, which in turn makes these ring segments and the star subnetworks the bottlenecks in the network.

Comparing Figs. 8.3–8.5 we observe that increasing the speed-up factor of the star subnetwork to $f = 4$ significantly increases the maximum mean aggregate throughput for the entire range of number of nodes N. On the other hand, further increasing the speed-up factor to $f = 16$, increases the throughput levels significantly for a small number of nodes N = 8 and 16 while for a larger number of nodes $N \geq 32$ there is only a minor increase in the throughput.

The explanation for these dynamics is as follows. For $f = 1$ the star subnetwork is the primary bottleneck in the network, which is relieved by increasing the speed-up factor to $f = 4$. As the speed-up factor is further increased to $f = 16$, the ring segments connecting the ring homed nodes to the ring-and-star homed nodes become the primarily bottleneck,

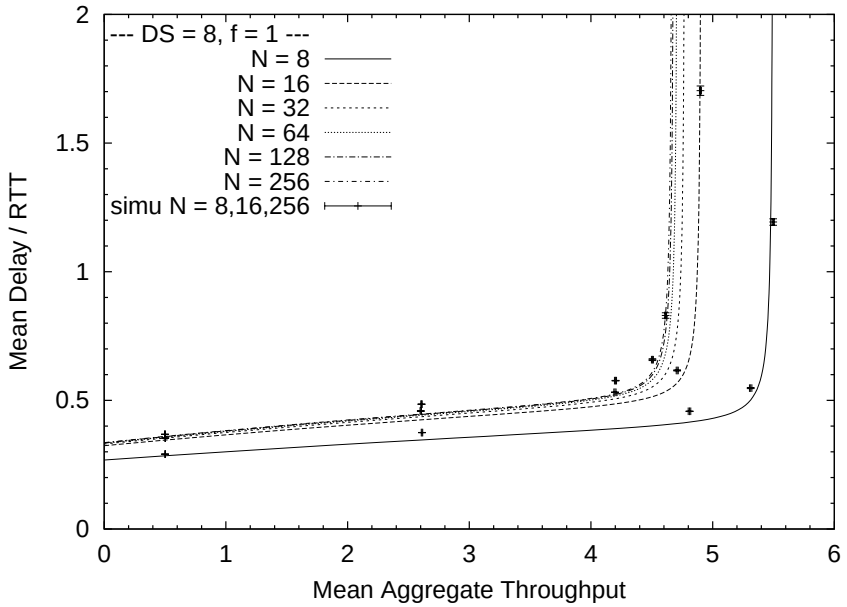


Figure 8.3: Mean delay vs. mean aggregate throughput with $N_{rs} = D \cdot S = 8$ ($D = 8$, $S = 1$) and $f = 1$ for different $N \in \{8, 16, 32, 64, 128, 256\}$.

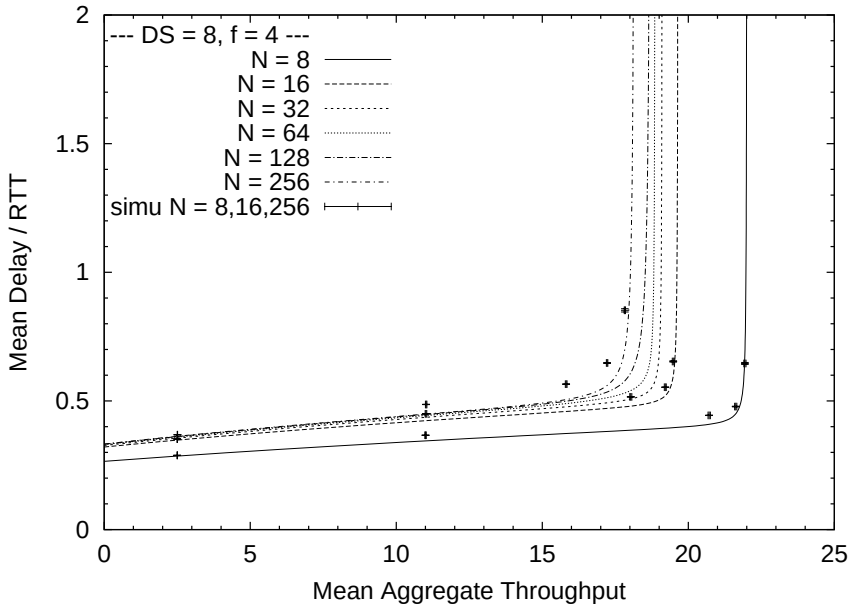


Figure 8.4: Mean delay vs. mean aggregate throughput with $N_{rs} = D \cdot S = 8$ ($D = 8$, $S = 1$) and $f = 4$ for different $N \in \{8, 16, 32, 64, 128, 256\}$.

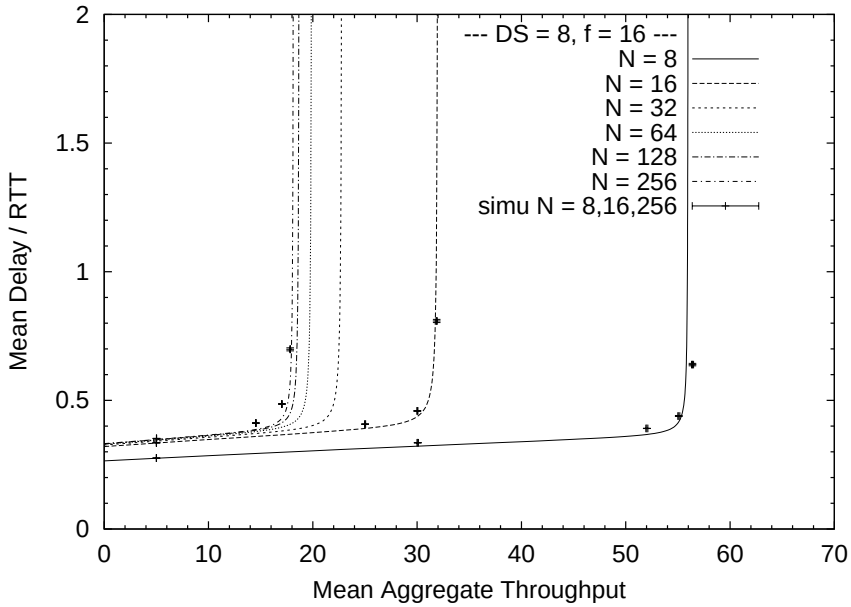


Figure 8.5: Mean delay vs. mean aggregate throughput with $N_{rs} = D \cdot S = 8$ ($D = 8$, $S = 1$) and $f = 16$ for different $N \in \{8, 16, 32, 64, 128, 256\}$.

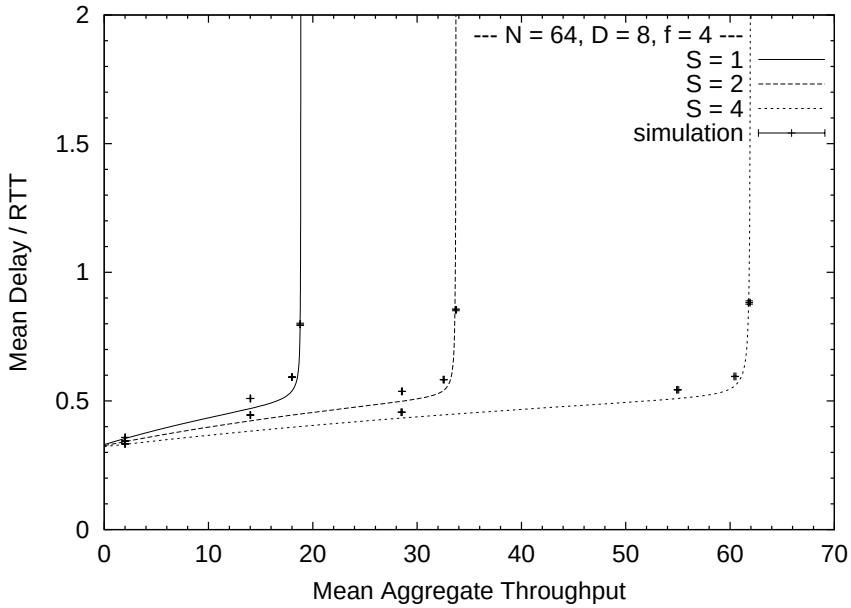


Figure 8.6: Mean delay vs. mean aggregate throughput with $N = 64$, $D = 8$, and $f = 4$ for different $S \in \{1, 2, 4\}$.

especially for an increasing number of ring-homed nodes between ring-and-star homed nodes. This bottleneck prevents the short-cut traffic from reaching the star subnetwork.

To capitalize on the capacity of the star subnetwork the bottleneck on the ring subnetwork has to be mitigated. This may be done by reducing the amount of collected short-cut traffic at each ring-and-star homed node. Clearly, this can be achieved by increasing the number of ring-and-star homed nodes $N_{rs} = D \cdot S$ for a given number of nodes N. Fig. 8.6 depicts the mean delay vs. mean aggregate throughput with $N = 64$, $D = 8$, and $f = 4$ for different $S \in \{1, 2, 4\}$. By connecting more nodes to the star subnetwork the congestion on the ring subnetwork at each ring-and-star homed node is alleviated and the star subnetwork can be utilized to a larger extent, resulting in a dramatically improved throughput-delay performance at the expense of connecting a larger number of nodes to the star subnetwork.

Note that there exist additional approaches to mitigate the bottlenecks on the star and ring subnetworks and to improve the throughput-delay performance of the network. For instance, the capacity of the PSC control channel could be increased by assigning additional control slots to ring-and-star homed nodes during the last $(F - D \cdot S)$ slots of each frame on the control channel. The capacity of the ring subnetwork could be increased by operating more than one wavelength in either direction by means of WDM. These modifications are left for future work.

Operation with Ring Failures

After gaining some insight into the failure-free operation of the network we now investigate the protectoration technique in the presence of various link and node failures. Besides a single failure, we consider also multiple failures in both ring and star subnetworks. To assess their impact on the network operation, we compare throughout our investigations the network performance of the various failure scenarios with that of the failure-free scenario. The locations of the failures are chosen as follows. Starting with a single (link or node) failure, the second failure is located at the opposite side of the central hub, i.e., the first failure is mirrored at the hub in order to obtain the location of the second failure. The third failure is placed in the middle of the first and second failures. The location of the fourth failure is found by mirroring the third failure at the hub. This procedure is repeated incrementally until all multiple failures are placed. In our numerical investigations we focus on failure scenarios which do not split the network into several disjoint subnetworks, i.e., full connectivity is not affected by the various failures. Recall from above, to guarantee full-connectivity no more than one link or one node failure must occur on each ring segment between two neighboring ring-and-star homed nodes. In the following, we set $N = 64$, $D = 8$, $S = 1$, and $f = 4$, with the other parameters set to their default values.

Let us start with link and node failures on the ring subnetwork while the star subnetwork is completely intact. Fig. 8.7 depicts the impact of ring link failures and their location on the throughput-delay performance of the network, including verifying simulation results for the two scenarios without failure and with four failures.

We consider four different locations of link failures. More precisely, the link failure(s) is (are) 0, 1, 2, or 3 hops away from the corresponding next ring-and-star homed node, where one hop denotes the distance between two adjacent nodes on the ring. We observe that the simulation gives a slightly larger mean delay than the analysis at small traffic loads, while at medium to high traffic loads the results of simulation and analysis match very well. This is due to the fact that the simulation takes the access and scheduling delays of the star

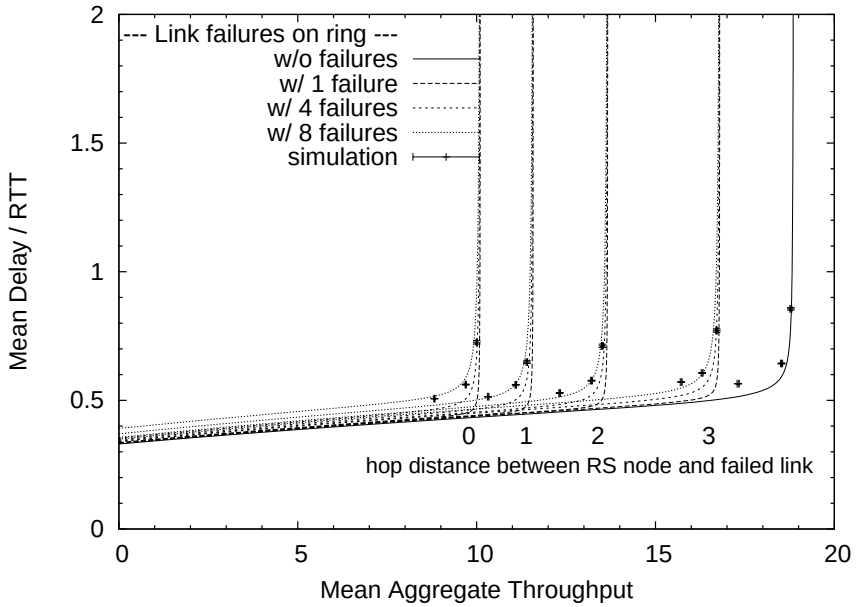


Figure 8.7: Mean delay vs. mean aggregate throughput for link failures with different locations on the ring subnetwork ($N = 64$, $D = 8$, $S = 1$, $f = 4$).

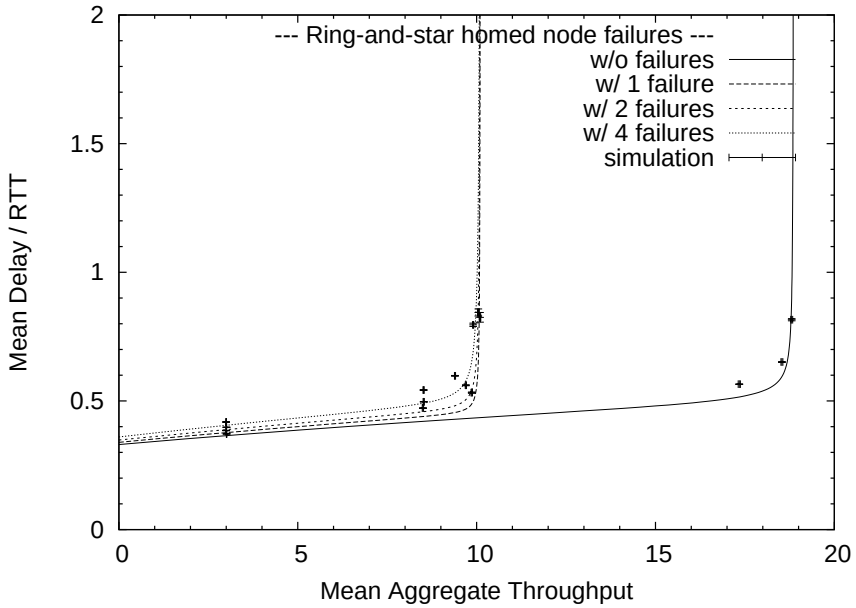


Figure 8.8: Mean delay vs. mean aggregate throughput for ring-and-star homed node failures on the ring subnetwork ($N = 64$, $D = 8$, $S = 1$, $f = 4$).

subnetwork into account, as opposed to the analysis. The access and scheduling delays result in a larger delay at small traffic loads, but both become negligible compared to the queuing delays encountered at the various transmit and transit queues as the traffic load increases. Interestingly, Fig. 8.7 shows that the location of the link failures has a significantly larger impact on the throughput-delay performance of the network than the number of link failures. We observe that for a given failure location the protectoration technique is able to make the network very resilient against multiple link failures on the ring subnetwork such that the throughput-delay performance is only slightly deteriorated with an increasing number of link failures. The performance loss, however, strongly depends on the location of the link failure(s). Apparently, link failures which are closer to ring-and-star homed nodes have a significantly more detrimental impact on the throughput-delay performance than link failures that are further away in terms of hops. This is because of two main effects. First, a link closer to a given ring-and-star homed node carries the traffic of more ring homed nodes that is sent to the ring-and-star homed node for being proxy stripped. If this link fails, more ring homed nodes are affected and need to steer the traffic in the opposite direction towards the neighboring ring-and-star homed node. Second, with a link failure close to a given ring-and-star homed node the opposite direction towards the neighboring ring-and-star homed node is larger compared to link failures that occur in the middle of two neighboring ring-and-star homed nodes. As a result, with a link failure close to a given ring-and-star homed node, wrapped and steered traffic traverses more intermediate ring homed nodes on the backup path. Both effects, more affected nodes and longer backup paths, lead to an increased congestion on the ring segment and thus an increased mean delay and a decreased mean aggregate throughput. Note that this performance loss can be alleviated by increasing the number of ring-and-star homed nodes (see Fig. 8.6). In doing so, for a given N each ring segment between two adjacent ring-and-star homed nodes contains fewer ring homed nodes, resulting in a decreased number of affected nodes and a decreased backup path length.

Next, we consider node failures. To guarantee full connectivity among the remaining functional nodes, no more than one node failure must occur between each pair of adjacent ring-and-star homed nodes. Note that the location of failed ring homed nodes has a similar impact on the throughput-delay performance of the network as the location of failed ring links, which have been examined above. In the following, we concentrate on ring-and-star homed node failures. More precisely, to maintain full connectivity among all nodes a failed ring-and-star homed node is assumed to be unable to communicate via the ring subnetwork while the transmission and reception via the star subnetwork remains fully intact (below we will investigate the complementary case where ring-and-star homed nodes are able to transmit and receive only via the ring subnetwork while they are disconnected from the star subnetwork). Fig. 8.8 depicts the mean delay vs. mean aggregate throughput with and without ring-and-star homed node failures on the ring subnetwork. We observe that the maximum achievable mean aggregate throughput decreases significantly if one ring-and-star homed node fails, but does not further decrease in the presence of additional ring-and-star homed node failures. Comparing Figs. 8.8 and 8.7 we observe that the maximum achievable mean aggregate throughput with failed ring-and-star homed nodes is only very slightly smaller than the maximum mean aggregate throughput with link failures zero hops away from ring-and-star homed nodes. In both cases, ring homed nodes next to a failed node or failed link have to wrap and steer traffic towards the neighboring ring-and-star homed node; with a failed link this wrapping and steering takes place on one side of the ring-and-star homed node, with a failed ring-and-star homed node on both sides of the node. As discussed above,

the protected traffic experiences thereby increased queuing delays at the intermediate ring homed nodes due to congestion. Recall that for a given N the congestion could be alleviated by limiting the number of ring homed nodes in each ring segment by increasing the number of ring-and-star homed nodes. Consequently, the detrimental impact of the ring subnetwork on the throughput-delay performance of protected traffic is mitigated, at the expense of more nodes being attached to the star subnetwork.

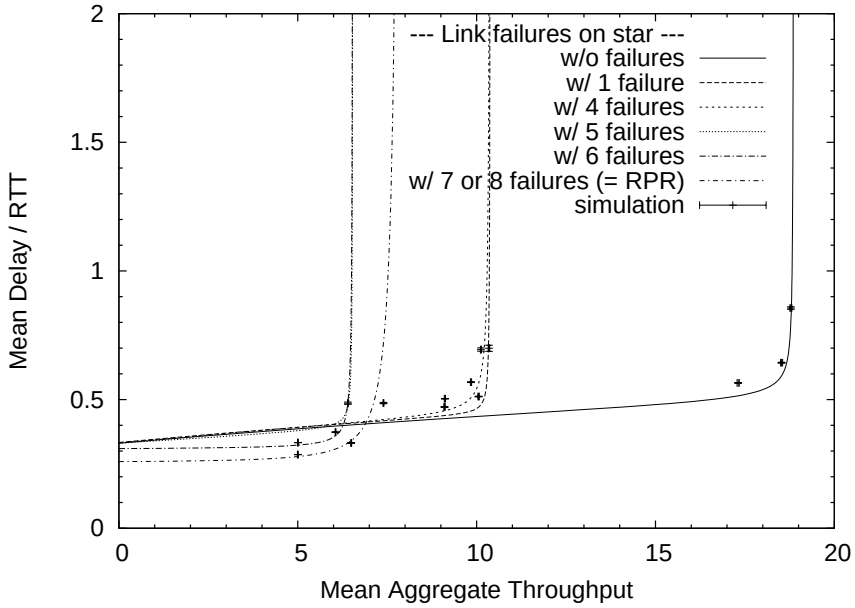


Figure 8.9: Mean delay vs. mean aggregate throughput for link failures on the star subnetwork ($N = 64$, $D = 8$, $S = 1$, $f = 4$).

Operation with Star Failures

After investigating link and node failures on the ring subnetwork, we now turn our attention to failures on the star subnetwork. We consider failure scenarios where a single, multiple, and all ring-and-star homed nodes are disconnected from the star subnetwork due to link or component (splitter, combiner, amplifier, (de)partitioner, AWG, PSC) failures. In the following, we focus on failed links that connect the ring-and-star homed nodes to the star subnetwork, but note that component failures have a similar impact on the ring-and-star homed nodes in terms of connectivity and throughput-delay performance. As shown in Fig. 8.9, with one ring-and-star homed node disconnected from the star subnetwork the maximum achievable mean aggregate throughput decreases significantly. While the throughput-delay performance remains rather unchanged for four link failures, five disconnected ring-and-star homed nodes further significantly decrease the maximum achievable mean aggregate throughput. Interestingly, with six link failures, i.e., only one pair of ring-and-star homed nodes is interconnected via the star subnetwork, the mean delay at light traffic loads is smaller than in other scenarios with fewer link failures. Note that the mean delay is further decreased with a slightly increased maximum achievable mean aggregate throughput if seven or eight links fail, i.e.,

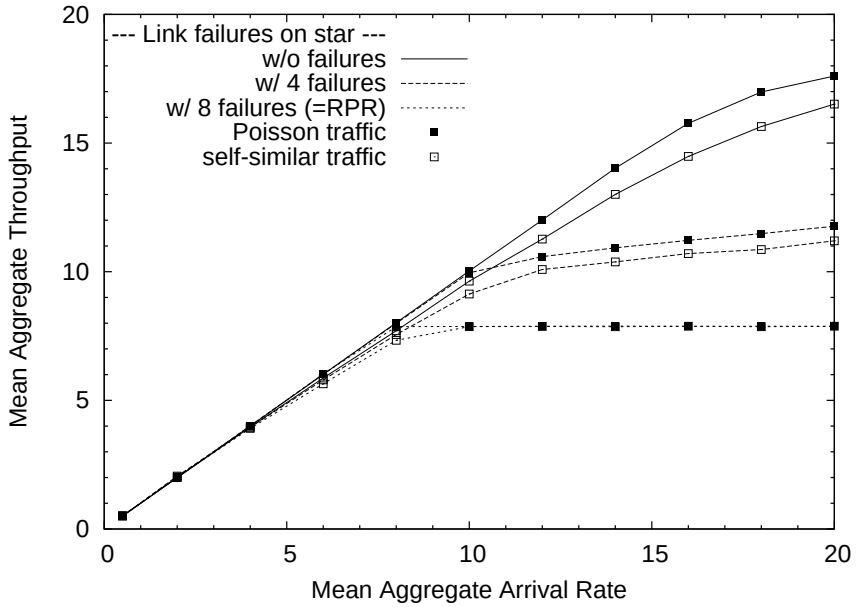


Figure 8.10: Mean aggregate throughput vs. mean aggregate arrival rate with $N = 64$, $D = 8$, $S = 1$ and $f = 4$ for Poisson and self-similar traffic without and with link failures on the star subnetwork.

no traffic is sent across the star subnetwork at all. Without any star subnetwork, the network is reduced to a conventional bidirectional RPR ring. Thus, we observe that an RPR ring achieves a slightly better throughput-delay performance than our hybrid ring-star network if the number of link failures on the star subnetwork is very large. This is mainly due to the fact that with fewer nodes connected to the star subnetwork the ring subnetwork gets more congested towards the proxy stripping ring-and-star homed hot-spot nodes, resulting in an increased delay and a decreased throughput.

Note, however, that our hybrid ring-star network without any failures is able to achieve a dramatically larger maximum mean aggregate throughput (close to 2.5 times larger) than RPR, as depicted in Fig. 8.9. Furthermore, as shown in Figs. 8.7 and 8.8 even in the presence of multiple link and/or node failures on the ring subnetwork our hybrid ring-star network not only outperforms RPR in terms of maximum achievable mean aggregate throughput but also guarantees full connectivity among all nodes, as opposed to RPR whose connectivity is entirely lost if more than a single link or node fails.

8.4.2 Self-similar Traffic

In this section we examine the performance of the protectoration network for self-similar traffic and compare it with the performance for Poisson traffic using simulations. In addition, we consider the network operation with more realistic finite buffers in this section, in contrast to the infinite buffers considered in the preceding section. More specifically, we consider the same network parameter settings as in the investigation of the failures in the star subnetwork in the preceding section, i.e., $N = 64$ nodes, $D = 8$ AWG and PSC ports, $S = 1$ combiner input port/splitter output port, star subnetwork speed-up factor $f = 4$, and all other parameters at

their default values. At each node we generate self-similar packet traffic with Hurst parameter 0.75 for each of the $N-1$ destination nodes by aggregating ON/OFF processes with Pareto distributed on-duration and exponentially distributed off-duration [189]. We set the capacity of all buffers to 96 kbyte. With this choice the transmit buffer for the star subnetwork has approximately the capacity required to hold the packets that can be transmitted within the bandwidth-propagation delay product of the star subnetwork over the PSC. (To see this note that the bandwidth-propagation delay product is $4 \cdot 2.5$ Gbit/s $\cdot \tau_{PSC}$, which corresponds to 124.3 frames of $400 \cdot 4$ Bytes each. Noting that at most one packet can be transmitted per frame over the PSC and the average packet size is approximately half a frame, we set the buffer capacity to be equivalent to 60 frames, i.e., $60 \cdot 400 \cdot 4 = 96$ kbyte.)

In Figs. 8.10–8.12 we plot the mean aggregate throughput, the relative packet loss, and the mean delay as functions of the mean aggregate arrival rate, which is given in the same units as the mean aggregate throughput. The 95% confidence intervals for the Poisson traffic simulations are generally too small to be seen in the plots, except for a few intervals for small loss probabilities in Fig. 8.11.

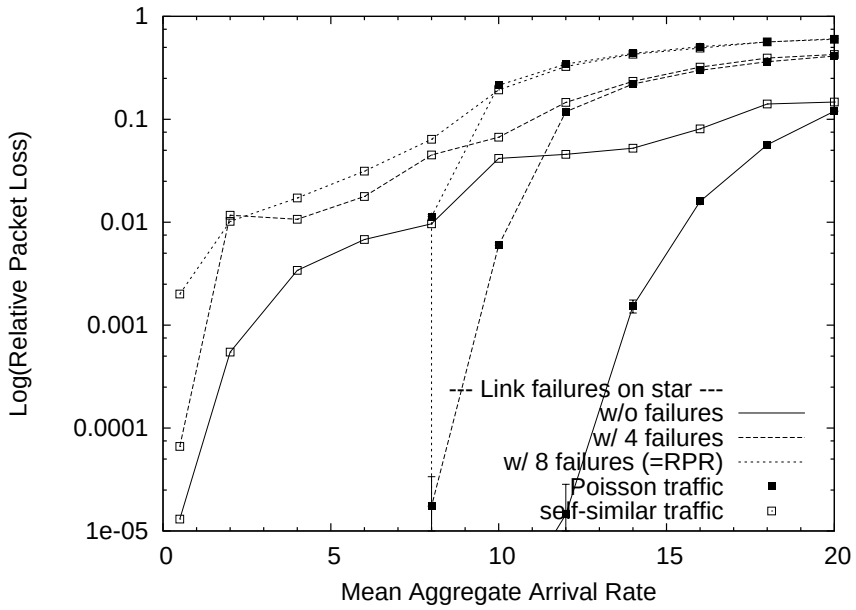


Figure 8.11: Relative packet loss vs. mean aggregate arrival rate with $N = 64$, $D = 8$, $S = 1$, $f = 4$ for Poisson and self-similar traffic without and with link failures on the star subnetwork.

We observe from Figs. 8.10 and 8.11 that for self-similar traffic, the packet loss is generally larger and the throughput smaller than for Poisson traffic, as is to be expected for the more bursty self-similar traffic. These differences become less pronounced as the network saturates, observe for instance that for the network with eight failures the differences disappear when the arrival rate exceeds 10 packet generations in steady state or equivalently $10 \cdot 2.5$ Gbit/s. This is because in the saturated network the buffers tend to be constantly filled to capacity. We observe from Fig. 8.12 that the self-similar traffic experiences slightly larger mean delays than the Poisson traffic in scenarios where the network is relatively lightly loaded, i.e., in the network with eight (four) failures for a mean arrival rate smaller than approximately five

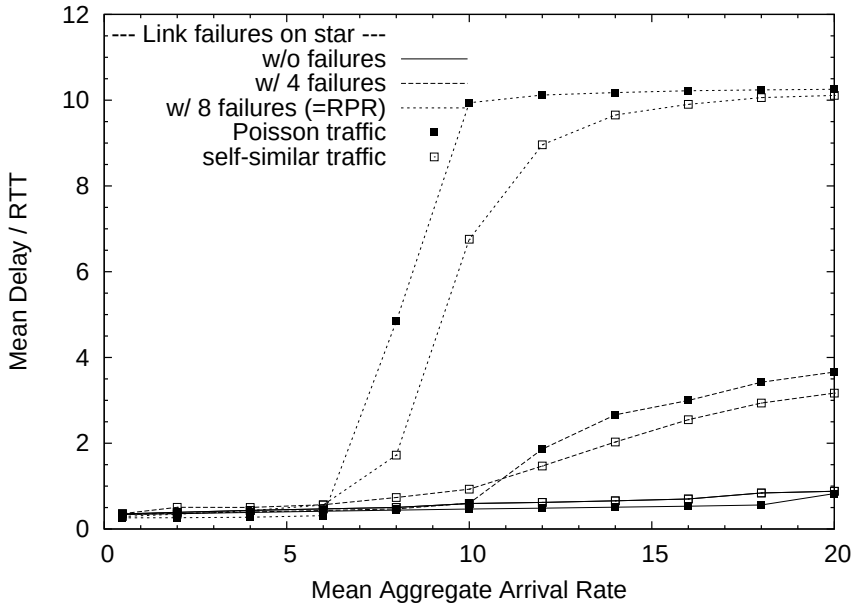


Figure 8.12: Mean delay vs. mean aggregate arrival rate with $N = 64$, $D = 8$, $S = 1$ and $f = 4$ for Poisson and self-similar traffic without and with link failures on the star subnetwork.

(eleven), and for the network without failures for the entire considered range of arrival rates. On the other hand, the self-similar traffic experiences somewhat smaller mean delays than the Poisson traffic when the network is relatively heavily loaded, i.e., in the network with eight (four) failures for arrival rates larger than approximately five (eleven). This is because with light traffic loads, the buffer occupancy levels tend to be fairly low, thus bursts of generated packets can typically be held in the buffers and experience larger delays. With heavy loads, on the other hand, the buffers tend to be constantly filled to capacity, especially with the relatively smooth Poisson traffic arrivals. The bursty self-similar traffic arrivals, on the other hand, tend to result in occasional 'dips' in the buffer occupancy levels and consequently somewhat smaller mean delays. Overall we observe that the differences in the throughput-delay performance between Poisson and self-similar traffic are relatively small. Also, from comparison with the analytical and simulation results for the scenario with Poisson traffic and infinite buffers in Fig. 8.9 we observe that the analysis for the infinite buffer and Poisson traffic scenario predicts the principal behavior of the network for finite buffers and self-similar traffic with reasonable accuracy.

8.5 Conclusions

We have developed and evaluated the *protectoration* fault recovery technique which extends our RINGOSTAR architecture to make it robust against link and node failures. For the fast and efficient recovery from failures, this technique uniquely combines the wrapping and steering methods of the conventional RPR ring network to exploit their respective strengths (fast recovery with wrapping, bandwidth efficiency with steering) to achieve a fast and bandwidth efficient recovery. In contrast to the conventional RPR network which can recover to full

network connectivity only after a *single failure*, our protectoration technique keeps all nodes connected for *multiple failures*.

Our analytical and simulation results demonstrate that the RINGOSTAR network without any failures achieves a significantly higher throughput-delay performance than the failure-free RPR network. The performance is close to that achieved with an idealized star subnetwork as analyzed in Chapter 7, i.e., the performance deterioration due to protocol overheads is rather small. The performance evaluation for failure scenarios shows that for a large number of failures on the star subnetwork, the throughput-delay performance of the protectoration network degenerates to the performance of the conventional RPR network. The impact of the failures on the ring subnetwork on the throughput-delay performance depends largely on the position of the failure (i.e., the distance from the ring-and-star homed node where the ring subnetwork is connected to the star subnetwork) and is largely independent from the number of failures.

Chapter 9

QoS Support & Fairness Control

AN important requirement for future metro solutions is Quality of Service (QoS) support (see Section 2.3.2). On one hand, metro customers demand circuit-like transmission service for transporting mission critical traffic such as video and voice. On the other hand there is an increasing demand for inexpensive data services transported without bandwidth or delay guarantees in a best-effort manner. To support such demands RINGOSTAR's MAC protocol has to be extended by corresponding mechanisms. Recall from Chapter 5 that or strategy for developing RINGOSTAR is to combine the individual strengths of RPR and the WDM star network, while mitigating their specific limitations. Clearly, one strength of RPR is its sophisticated QoS mechanism. Therefore, in this chapter, we present extensions to RPR's QoS mechanisms that migrate RPR's QoS support to RINGOSTAR.

An issue closely tied to QoS support is fairness control. Best-effort traffic from different source nodes competes for the bandwidth left over by higher priority traffic. To ensure that this bandwidth is shared among the different traffic sourcesin a fair manner a fairness control mechanism is required. Therefore, we adapt an advanced fairness control protocol for RPR termed Distributed Virtual-time Scheduling in Rings (DVSR) to support our hybrid ring-star architecture. We evaluate the performance of RINGOSTAR's fairness control protocol for self-similar traffic by means of computer simulation.

9.1 QoS Support

In this section, we first discuss the mechanisms enabling QoS support in RPR. We then discuss how these mechanisms can be adapted to RINGOSTAR. Since QoS support (as well as recovery from link and node failures, see Section 8.2.1) relies on RPR's topology discovery protocol we present an overview of the latter in this section as well. Both QoS support and the topology discovery are specified in the IEEE 802.17 Resilient Packet Ring standard [37]. The following discussion of QoS support and topology discovery closely follows [155].

9.1.1 QoS Support in RPR

RPR provides three different traffic classes A, B, and C, which have already been mentioned in the discussion of RPR's queue structure in Section 6.1. Class A provides circuit-like service with guaranteed bandwidth, zero packet loss, as well as low delay and jitter. Class A is further divided into class A0 and A1. Class B is a high-quality class with guaranteed band-

width, zero loss and predicable delay and jitter. Furthermore, and in contrast to class A, the guaranteed grant of bandwidth can be exceeded. All class B traffic up the committed information rate (CIR) belongs to class B-CIR. Class B traffic exceeding the granted bandwidth, excess information rate (EIR) traffic, is called class B-EIR traffic and transported without any guarantees like class C. Finally, class C is a best-effort service class, i.e., packets may be lost and delay increases if bandwidth gets scarce. Class B-EIR and C are fairness eligible (FE) and controlled by the fairness algorithm (see Section 9.2 below). Table 9.1 summarizes the the features RPR's traffic classes.

Class A	*Circuit like service:* Guaranteed bandwidth, low delay and jitter.
Class B	*High quality service:* Guaranteed bandwidth, bounded delay and jitter (B-CIR), granted bandwidth can be exceeded (B-EIR).
Class C	*Best-effort service:* No bandwidth guarantee, possibly packet loss, unpredictable delay and jitter.

Table 9.1: Features of RPR's traffic classes.

To ensure that the bandwidth guaranteed to class A0, A1, and B-CIR is always available, the guaranteed amount of bandwidth is preallocated for these classes. Bandwidth preallocated for class A0 is called 'reserved' and may not be used by other classes. All remaining bandwidth is called 'unreserved rate'. Bandwidth preallocated to classes A1 and B-EIR is called 'reclaimable'. Any remaining bandwidth currently not used by these classes may be used for transporting class B-EIR and C traffic. Additionally to reclaimable bandwidth not in use, class B-EIR and C compete for the remaining bandwidth that has not been preallocated. Preallocations for class A and B traffic are broadcasted during network initialization using using the topology discovery protocol. Table 9.2 summarizes the transmission privileges for each traffic class.

	Preallocation	Transit Priority	Fairness Eligible
Class A0	Reserved	PTQ	No
Class A1	Reclaimable	PTQ	No
Class B-CIR	Reclaimable	STQ	No
Class B-EIR	None	STQ	Yes
Class C	None	STQ	Yes

Table 9.2: Transmission privileges of RPR's traffic classes.

As illustrated in Fig. 9.2, each RPR node implements three queues and several traffic shapers (for each ring direction). A packet arriving from the client is put in one of the three queues, according to the packets priority A, B, or C (and according to the direction in which the packet is sent). Traffic stored in the individual queues then needs to pass one or two traffic shapers, depending on the traffic class. Each traffic shaper is implemented as a token bucket that ensures that the passed traffic does not exceed the specified average rate and reduces fluctuations of the traffic. There are individual traffic shapers for classes A0, A1 (optional), B-CIR, and for the FE traffic classes B-EIR and C. Furthermore, a downstream shaper controls that the unreserved rate is not exceeded. The rate of the A0, A1, and B-CIR shapers are preconfigured while the FE shaper is dynamically controlled by the fairness control algorithm. Packets taken from any of the queues for class A, B, or C are put into the node's stage queue

of the corresponding ring direction, as shown in Fig. 6.7 (top). Class A traffic has priority over class B which has priority over class C traffic. Therefore, class B traffic is only added to the stage queue if the queue for class A traffic is empty or if the shaper for class A currently does not pass a packet because class A has temporally reached the granted tranmission rate limit. Similarly, class C traffic is only sent if the class B queue is empty or if the class B shaper currently does not pass a packet. Finally, note that the previous discussion is simplified in that there is only one traffic shaper for each individual traffic class A0, A1, B-CIR, and FE traffic. However, the fairness algorithm discussed in Section 9.2.3 below operates at the granularity of flows between individual source-destination node pairs. In this case, at least the single FE shaper needs to be replaced by one individual FE traffic shaper per destination node. The queue arbitration algorithm is illustrated in the pseudo-code in Fig. 9.1.

```
while (true) { // endless loop
    if (packet_in_queue_A) {
        if (shaper_A0_passes_packet) {
            put_packet_in_stage_queue(); // class A0
            continue; // reenter loop
        }
        else (shaper_A1_passes_packet && downstream_shaper_passes_packet) {
            put_packet_in_stage_queue(); // class A1
            continue; // reenter loop
        }
    }
    if (packet_in_queue_B) {
        if (shaper_B_CIR_passes_packet && downstream_shaper_passes_packet) {
            put_packet_in_stage_queue(); // class B-CIR
            continue; // reenter loop
        }
        if (shaper_FE_passes_packet && downstream_shaper_passes_packet) {
            put_packet_in_stage_queue(); // class B-EIR
            continue; // reenter loop
        }
    }
    if (packet_in_queue_C) {
        if (shaper_FE_passes_packet && downstream_shaper_passes_packet) {
            put_packet_in_stage_queue(); // class C
            continue; // reenter loop
        }
    }
}
```

Figure 9.1: Pseudo-code for traffic shaper arbitration in RPR.

Recall from Section 6.1 that in RPR's dual-queue transit path high-priority class A frames are stored in the PTQ while class B and C frames are stored in the STQ. Forwarding from the PTQ has priority over forwarding from the STQ and most types of locally generated traffic. Therefore, the PTQ is usually empty and the delay of class A frames is approximately equal to the propagation delay of the optical signal from source to destination. Some small delay jitter is introduced due to the fact that class A packets arriving at a node's PTQ sometimes need to be stored until another packet that is currently sent has left the node. Due to the small delay jitter and the fact that the bandwidth is reserved in advance and must not be exceeded, class A traffic can be considered as RPR's equivalent to a circuit in SONET/SDH.

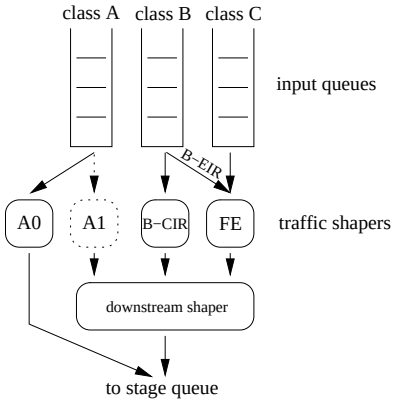


Figure 9.2: Client input queues and traffic shapers of RPR node (for one ring direction).

Class B traffic is stored in the STQ and does therefore experience larger delays. However, the delay is still bounded by the propagation delay plus the maximum waiting time in the STQss between source and destination. To see this, recall that the PTQ is usually empty and that STQ traffic gets priority over locally added traffic if the STQ length exceeds a certain threshold. Different from class A, the amount of class B traffic may exceed the reserved rate. This excess traffic (class B-EIR) is then treated as fairness eligible and shares the unreserved ring bandwidth with class C traffic. Therefore, class B provides a bandwidth guarantee similar to class A, but also allows clients to exceed the guaranteed rate, with excess traffic be treated fair among all competing FE traffic being an advantage over real circuits. Furthermore, the network utilization is improved compared to circuits as bandwidth preallocated but currently not claimed by class B can be used for transporting best-effort class C traffic. Note that in single queue mode the PTQ should not fill up with too many frames in order to forward class A transit traffic without introducing additional delays. This is achieved by letting transit traffic have priority over all locally added traffic, as illustrated in the pseudo-code for RPR's queue arbitration mechanism ('PTQ only') in Section 6.1.

9.1.2 QoS Support in RINGOSTAR

For providing QoS support in RINGOSTAR we adapt the previously described mechanism to support the underlying proxy stripping. This is achieved by deploying an additional queue and traffic shaper structure in ring-and-star homed nodes. This structure is the same as for the two ring directions and corresponds to the star stage queue. Packets arriving from the clients which are sent via the star are stored in this structure. Remember from Section 6.3.3, that the star subnetwork features the same dual transit queue structure consisting of PTQ and STQ as the ring. Therefore, traffic using the star network as a shortcut to its destination is treated the same way as ring-only traffic in terms of forwarding priority. However, note that at a ring-and-star homed node traffic arriving from the star subnetwork that needs to be forwarded on the ring is put into the same transit queue as transit traffic arriving from the ring. The lossloss property of RPR's transit path is maintained only if the amount of both ring in-transit traffic and star-to-ring in-transit traffic remain below a certain threshold. Since at each ring-and-star homed node star-to-ring and ring transit traffic compete for the bandwidth on the outlink to the downstream ring node packets may be lost due to buffer overflow. Nevertheless, no class A and class B-CIR frames will be lost and the bandwidth

guarantee can be maintained since the required bandwidth is reserved in advance and has priority over the remaining traffic classes, namely the FE classes B-EIR and C. For these classes the fairness mechanism alleviates the congestion and thus prevent packet loss in both ring and star-to-ring transit queues. Also note that the star's pretransmission coordination protocol and retransmissions of collided frames introduce an additional delay jitter for traffic sent over the star. If the star is dimensioned properly, at most one retransmission is required providing an upper bound for the transmission delay as required for class B traffic. An MAC extension to enable circuit-like service with constant delay in a single-hop AWG based star network is discussed in [119]. This can easily be adapted to RINGOSTAR's star subnetwork to improve the support for class A traffic.

9.1.3 Topology Discovery

In this section we describe RPR's topology discovery protocol which is used to build and maintain an image of the network topology in each node and to disseminate each node's status (e.g., bandwidth reservations for class A and B-CIR traffic) around the network. The topology information enables each node to determine the shortest path to all destination nodes.

During system initialization, each node broadcasts a so-called topology discovery message all around the ring (in both directions). The message includes a time-to-live (TTL) field that is initially set to a value of 255 (maximum possible number of nodes in RPR) and decremented each time the message is forwarded to the next ring node. Therefore, each node receiving a toplogy message can determine the relative ring position of the node that issued the message. After the topology message of all nodes have propagated around the ring, each node is able to build a complete image of the network topology and to calculate the RTT to each other node. Furthermore, each topology discovery message includes status information about the node that sent the message. Information about the nodes bandwidth preallocations for class A and B-CIR traffic enables the other nodes to calculate the bandwidth remaining on each link for best-effort traffic. If a node detects a link or node failure this is also indicated in the topology discovery message so that each node is enabled to calculate alternative paths circumventing the failure. When a new node is inserted to the ring, or a node detects a failure at its links or neighboring nodes, it immediately broadcasts a topology discovery message containing the new status. A node receiving a topology discovery message inconsistent with the information in its database also sends a topology discovery message with its own status. This starts a ripple effect and all nodes are updated with the newest status information. To increase the robustness of the system each node periodically broadcasts its status in topology discovery messages, even if the status has not changed.

In RINGOSTAR, the topology message issued by a node additionally contains the information whether the node is a ring-only or a ring-and-star homed node.

9.2 Fairness Control

Traffic of class B-EIR and C, so-called fairness eligible (FE) traffic, competes for the ring bandwidth left over by class A and class B-CIR. If the aggregate amount of FE traffic exceeds the available amount of bandwidth the goal in RPR is to share the bandwidth between the competing nodes in a fair manner. If no fairness control is applied, downstream nodes suffer from starvation by upstream nodes, as illustrated in Fig. 9.3. Starvation results from the fact

that the transit path is lossless and has therefore priority over locally added traffic in case of an overload situation. To achieve fairness in RPR several fairness control mechanisms have been proposed [190, 186, 191, 192], most prominently the original RPR fairness algorithm (RPR-FA) specified in the IEEE 802.17 Resilient Packet Ring standard [37] and Distributed Virtual-time Scheduling in Rings (DVSR) [188]. It was shown in [186, 187, 188] that for a relatively simple traffic scenario with unbalanced constant-rate traffic inputs RPR-FA suffers from severe and permanent oscillations spanning nearly the entire range of the link capacity. Such oscillations hinder spatial reuse, decrease throughput, and increase delay jitter. Furthermore, a refined fairness reference model, namely Ring Ingress-Aggregated with Spatial Reuse (RIAS), which is not fully implemented by RPR-FA, has been incorporated into the RPR standard recently. DVSR has been proposed to overcome these limitations and it has been shown that it is able to mitigate oscillations and achieve nearly complete spatial reuse in accordance with the RIAS reference model. RINGOSTAR's fairness control mechanism which is discussed below is therefore derived from DVSR and not from RPR-FA. The description of RPR-FA and DVSR below closely follows [188] and [145].

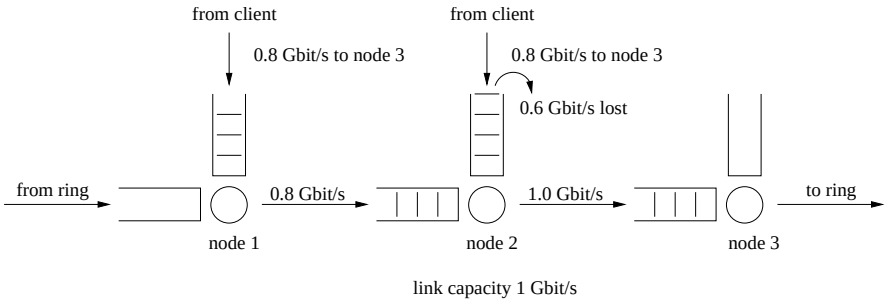


Figure 9.3: Starvation scenario: Transit traffic from upstream node 1 has priority over locally added traffic from downstream node 2 which suffers from starvation.

9.2.1 Original RPR Fairness Algorithm

In this section we describe RPR-FA as specified in the RPR standard. Note that the description omits several details specified in the standard for more clarity. Recall that the fairness mechanism only controls class B-EIR and class C traffic by adjusting the FE traffic shapers. RPR-FA can be operated in two modes. One is called aggressive mode (AM) and derived from the Spatial Reuse Protocol (SRP) [193]. This mode relies on dual-queue operation, i.e., the transit path consist of both PTQ and STQ. The other is termed conservative mode (CM) and is for single-queue operation only. CM evolved from the Alladin algorithm [194]. However, both algorithm use the same framework as described in the following.

The fairness control algorithm for the links in one direction operates independently of the links in the other direction. For simplicity we only consider one ring direction. Congestion of a node is caused by the upstream nodes. Therefore, each node periodically sends fairness control packets upstream (using the other ringlet than the fairness control operates on) to throttle these nodes. The condition whether a node is congested or not depends on the mode of operation (AM or CM) and is discussed below. A fairness control packet sent by node n contains the rate *fair_rate*[n] at which upstream nodes are allowed to maximally send on the link controlled by this node. Nodes receiving the control packet have to adapt to this rate. How *fair_rate*[n] is determined also differs for both operation modes AM or CM. To evaluate

the congestion state of a link the corresponding node n measures the amount of transit FE traffic *fw_rate* and locally added FE traffic *add_rate*. Each node has byte counters for both transit and locally added FE traffic that are periodically reset after an interval of the duration *aging_interval*. The current value of *fw_rate* and *add_rate* is calculated from the measured rates in the last two intervals. The latest measurement is weighted with a factor $1/LPCOEF$ and the previous one with $1 - 1/LPCOEF$, i.e., the measurements are low-pass filtered to reduce oscillations. A node m receiving a fairness control message forwards the message upstream. Depending on node m's own congestion status it might change the value of the fair rate in the message. If the link controlled by node m is also congested the node checks if its own *local_fair_rate*[m] is smaller than the fair rate in the control packet. If so, the value in the control packet is replaced by *local_fair_rate*[m]. If the link controlled by node m is not congested, the node checks if the local *fw_rate* is larger than the fair rate *local_fair_rate*[n] in the control packet. If not, node m assumes that the upstream nodes do not cause the congestion and replaces the fair rate in the fairness message by a null value to indicate a lack of congestion. Furthermore, a node receiving a fairness control message adjusts its local FE traffic shapers. Recall from Section 9.1.1 that each node is equipped with separate FE traffic shapers for each individual destination node. The fair rate *local_fair_rate*[n] in the control packet corresponds to a certain link. Therefore only a subset of traffic shapers is affected by the new fair rate, namely those which correspond to a destination to which the path includes this link. The node must adjust these traffic shapers so that in total an aggregate send rate equal to *local_fair_rate*[n] is not exceeded. If the fairness control packet does not contain a fair rate but a null value to indicate a lack of congestion, the allowed aggregate rate is incremented by a certain value.

Aggressive Mode

In aggressive mode, a node n considers its link congested if the STQ length exceeds a certain threshold (by default 1/8 of the STQ size) or if the total FE transmission rate *fw_rate*[n] + *add_rate*[n] exceeds the unreserved bandwidth on that link. A node that is congested sets the the fair rate *local_fair_rate*[n] in the next fairness control packet the node sends upstream equal to the rate the node adds FE traffic to the ring itself, *add_rate*[n]. If the link is not congested node n sends a null value to indicate a lack of congestion. The intention behind this mechanism is this: A node suffering from starvation forces the upstream nodes to adapt to its own transmission rate to resolve the contention. Note that the local node n has a smaller add rate than the upstream nodes since transit traffic has priority over locally added traffic. Therefore the contention is always resolved if the upstream nodes adapt to the add rate of the local node. After the contention has been resolved, node n transmits null value fairness control messages to which leads the upstream nodes to continually increase their transmission rates, slowly converging to congestion state again. However, this mechanism may lead to permanent transmission rate oscillations spanning nearly the whole unreserved link capacity. For instance, consider a simple two node example, where the upstream node sends at rate and slightly smaller than the unreserved rate and the local node has only few traffic to send. After congestion occurs the local node forces the upstream node to throttle its transmission rate to the small add rate of the local node. Then, in the following fairness cycles the local node indicates a lack of congestion and the upstream node slowly approaches its full transmission rate and the congestion state again.

Conservative Mode

In conservative mode, a node n considers its link congested if the node's access timer expires, i.e., node n has not been able to access the ring to send local FE traffic for a certain amount of time, or if the aggregate FE send rate *fw_rate*[n] + *add_rate*[n] exceeds a certain threshold *low_threshold* (80% of the unreserved capacity by default). Also, node n counts the number of active nodes, i.e., the number of nodes that have sent a transit FE packet within the last measurement interval. The fair rate that is sent in the control packet depends on several factors. First, if the node has not been congested in the previous measurement interval but is congested now, *local_add_rate*[n] is set to the unreserved capacity divided by the number of nodes that have been active in the current measurement interval, including the node itself. Second, if the node has been congested in the last two measurement intervals the fair rate depends on the amount of FE traffic *fw_rate*[n] + *add_rate*[n] sent in the last interval. If the aggregate rate is below the formerly mentioned *low_threshold* the fair rate is incremented by a certain value. If the aggregate rate is above a high threshold *high_threshold* (95% of the unreserved bandwidth by default) the fair rate is decremented. The intention is that upon first occurrence of congestion the next issued fairness control packet resolves the congestion by granting all nodes an equal share of the unreserved bandwidth. Then the fair rate is continuously adjusted to ensure a high bandwidth utilization and spatial reuse. It has been shown than RPR's fairness control in conservative mode suffers from oscillations in unbalanced traffic scenarios as well [188].

9.2.2 The RIAS Fairness Objective

After RPR-FA had been specified in the RPR standard, a new targeted performance objective called Ring Ingress-Aggregated with Spatial Reuse (RIAS) was incorporated in the standards document. The goal of RIAS is to improve spatial reuse. A formal definition of RIAS fairness can be found in [188]. Here, we provide an intuitive understanding of RIAS fairness by discussing the RIAS fair bandwidth shares in four representative scenarios presented in [195]. To simplify the discussion and without loss of generality we assume that the full link capacity is available for FE traffic and that all traffic is fairness eligible, in other words all traffic sent in the network is class C traffic. As we will see shortly, RIAS specifies fairness at the granularity of flows between a source and an individual destination node. In contrast to only considering the aggregate amount of traffic a node adds to the ring this enables improved reuse.

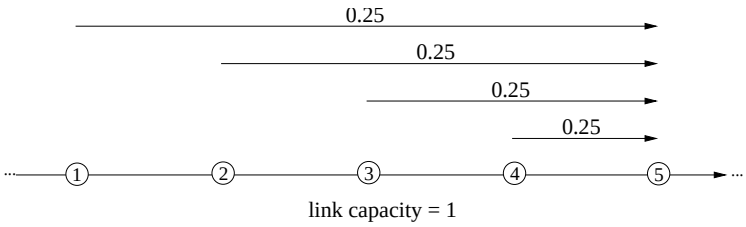


Figure 9.4: RIAS reference scenario I: Parking lot.

Consider the scenario depicted in Fig. 9.4 in which four different ring nodes send traffic to the same destination node. This could for instance be four metro customers accessing the gateway to the internet. The network operator wants to ensure that all clients receive the same share of the bandwidth, therefore all clients receive 25% of the link capacity. Note that

each client could also receive a weight and the bandwidth would be shared proportionally to these weights. However, for simplicity we assume equal weights for all customers here and in the following examples. A fairness control mechanism that establishes the bandwidth distribution illustrated in Fig. 9.4 is said to be RPR compliant. The following scenarios are optional but improve the bandwidth utilization, an important design goal of RPR.

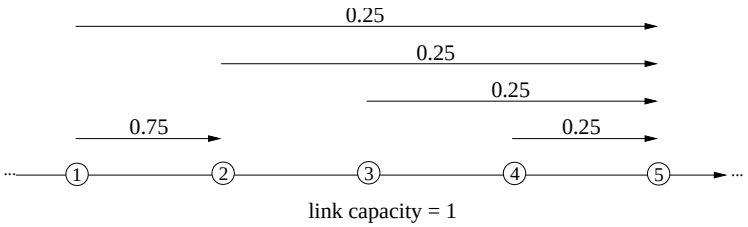


Figure 9.5: RIAS reference scenario II: Parallel parking lot.

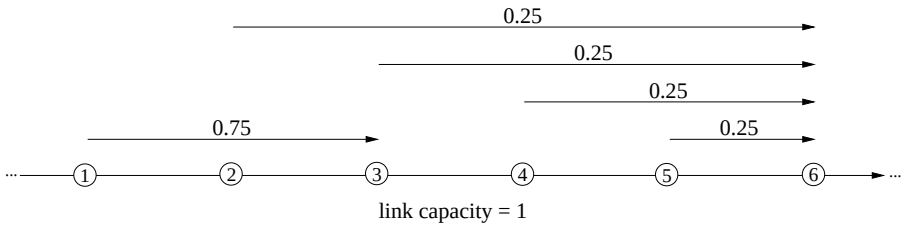


Figure 9.6: RIAS reference scenario III: Upstream parallel parking lot.

Fig. 9.5 shows previous scenario extended by an additional flow between node 1 and 2. To achieve full spatial reuse RIAS requires that this flow receives the remaining 75% capacity on the first link. More generally, a each flow should adapt to the smallest share it receives on any link it traverses, as illustrated for the flow between node 1 and 3 in Fig. 9.6.

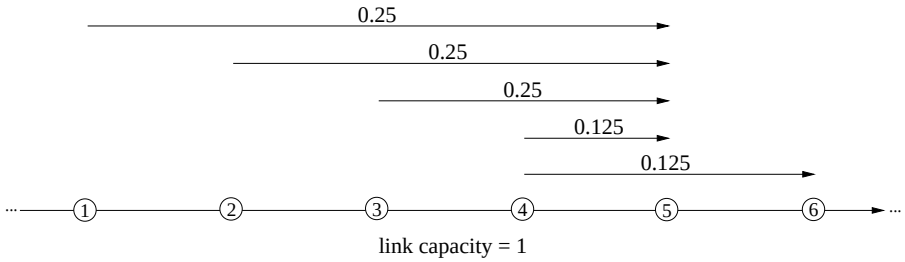


Figure 9.7: RIAS reference scenario IV: Two-exit parking lot.

Finally, if several nodes compete for the bandwidth on a certain link, all nodes should receive the same bandwidth share, independent of the number of flows of each node. Note that this is the only requirement that differentiates RIAS from the well know maxmin fairness objective [196]. An illustrative scenario is shown in Fig. 9.7, where node 4 receives 25% of the link capacity equal to all other nodes which is equally shared between the two flows originating from node 4.

9.2.3 Distributed Virtual-Time Scheduling in Rings

Distributed Virtual-time Scheduling in Rings (DVSR) is an improved fairness control mechanism for RPR and packet rings in general that implements the RIAS fairness objective discussed above. It was shown in [188] that DVSR achieves nearly full spatial reuse while oscillations are mitigated. Furthermore, DVSR features more rapid convergence times towards the the targeted fair rates compared to conventional RPR fairness control mechanisms. As a proof-of-concept, DVSR has been implemented on a network processor that emulates an eight node ring. In the following we discuss DVSR's operation.

Different from RPR, packets arriving at the transit queue(s) and transmit queues are FIFO queued at each node. Note, however, that after the fair rates are established the queues are rarely fill up and the queue arbitration priority is of little impact. One fairness control packet circulates upstream on each of the two rings. Each fairness control packet consists of N fields that contain the fair rates of all ring links. Each node monitors both fairness control packets and writes its locally computed fair rates in the corresponding fields of the fairness control packets. To calculate the fair link rates, each node measures the number of bytes $l[k]$ arriving from node k, including the station itself, during the time interval T between the previous and the actual arrival of the control packet. Each node performs separate measurements for either ring. The fair rate F of a given link is equal to the max-min fair share among all measured link rates l_k/T with respect to the link capacity C currently available for fairness-eligible traffic. The pseudo-code in Fig. 9.8 below illustrates the calculation of the fair rate F.

```
// sort number of bytes sent from each node
sort_ascending(l[k]); // l[0] <= ... <= l[N-1]
// count number of active nodes and total amount of bytes sent
active_nodes = 0;
total_bytes = 0;
for (k = 0; i < N; k++) {
    if (l[k] > 0)
        active_nodes++;
    total_bytes += l[k];
}
// if link capacity not exhausted F is equal to largest flow rate
if (total_bytes / T < C)
    return (l[N-1] / T);
// if link capacity exhausted fair rate F is the max-min fair rate
remaining_capacity = C;
F = remaining_capacity / active_nodes;
k = 0;
while (l[k] / (C*T) < F && l[N-1] / (C*T) >= F) {
    if (l[k] > 0) {
        remaining_capacity -= l[k] / (C*T);
        active_nodes--;
        F = remaining_capacity / active_nodes;
    }
    k++;
}
return (F);
```

Figure 9.8: Pseudo-code for calculation of fair rate F.

A node receiving a fairness control packet adapts its local traffic shapers to the new fair

rates. Similar to RPR-FA, the node evaluates which destinations are affected for each fair rate in the fairness control packet and adapts the FE traffic shapers for each destination so that the granted share is not exceeded on any link. An efficient means on how to chose the rates of the individual traffic shapers for full spatial reuse in presented in the next section.

9.2.4 Fairness Control in RINGOSTAR

In this section we describe how DVSR is adapted to the RINGOSTAR architecture and access protocol. Each of the two fairness control packet is extended by $DS/2$ additional fields. The first N fields contain the fair rates of all ring links and the remaining $DS/2$ fields contain the fair rates of the star links, where one control packet carries the rates of the even numbered and the other one the rates of the odd numbered star links. Ring only nodes perform the same per node traffic measurement as in DVSR. Proxy stripping nodes additionally count the number of bytes arriving from the star for each node and use the time window of the fairness control packet that carries the fair rate of the corresponding proxy stripping node to calculate traffic rates from the byte counts. Analogously to the ring links, the fair rate F of a given star link is equal to the max-min fair share among all measured star link rates l_k/T with respect to the star link capacity C_s currently available for fairness-eligible traffic. The $(N-1)$ FE traffic shapers that limit the traffic rates are implemented as token buckets whose refill rates are set to the current fair rates of the corresponding destinations. Using the same two time windows as in the calculation of the ring and star link fair rates above, each node i counts the bytes ρ_{ij} sent to destination j during the time window. Thus, there are two sets of $(N-1)$ byte counters, one for each time window. Each time a fairness control packet arrives, a given node calculates the fair rate of each ingress flow as follows. According to the RIAS objective, the total capacity available to a given node on a certain link equals the fair rate F which is shared among all its ingress flows crossing that link. Based on the measured ingress rates ρ_{ij}/T of these flows and the available capacity F, the max–min fair share f is calculated for each crossed link. The refill rate of each token bucket is set to the minimum fair share f of these links.

9.2.5 Simulation Results

In the following, we investigate the proposed fairness control protocol for RINGOSTAR by means of simulation. We set $N = 16$, $D = 4$, and $S = 1$. We consider uniform self-similar traffic with Hurst parameter 0.75, where each node does not send any traffic to itself and sends a generated data packet to the remaining $(N-1)$ nodes with equal probability $1/(N-1)$. We consider best-effort traffic class C and assume that no bandwidth is reserved for traffic class A and 10% of the ring bandwidth are left for traffic class B, i.e., class C traffic must not use more than 90% of the ring bandwidth. Each node is assumed to have continuously data to send on the ring which operates at 2.5 Gbit/s.

Fig. 9.9 shows the RIAS fair throughput for each source-destination node pair. The throughput varies for different node pairs due to the network symmetry. Specifically, there are three types of nodes: Proxy stripping nodes (0, 4, 8, 12), nodes in the middle of two proxy stripping nodes (2, 6, 10, 12), and neighboring nodes of proxy stripping nodes (remaining nodes). All nodes of a given type achieve identical throughput to all destinations whose distance from the corresponding source node is the same. Proxy stripping nodes achieve a higher-than-average throughput to all other proxy stripping nodes due to the single-hop links

of the star subnetwork. Nodes within the same ring segment between two adjacent proxy stripping nodes achieve a higher-than-average throughput if they communicate with each other. Traffic between nodes of different ring segments is bottlenecked by the ring links next to the intermediate proxy stripping node(s), resulting in a lower-than-average throughput.

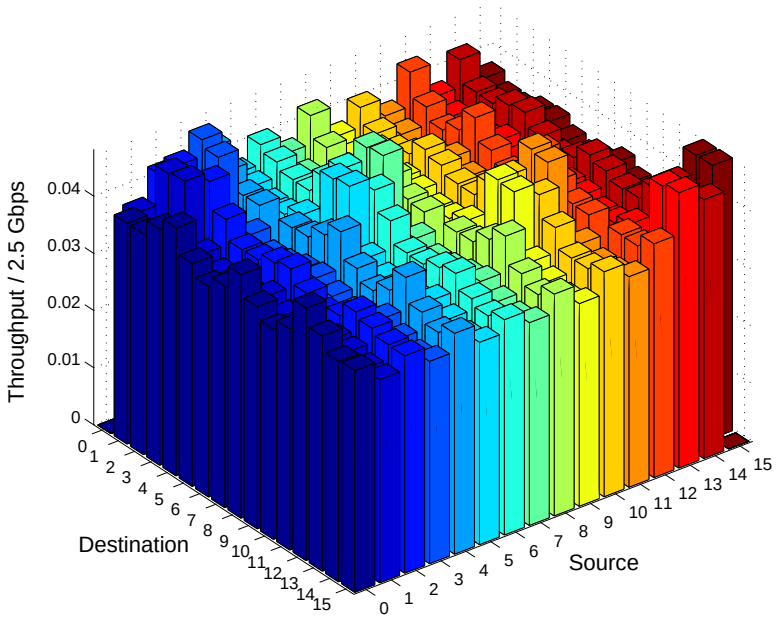


Figure 9.9: RIAS fair throughput (given in 2.5 Gbit/s) between each pair of nodes for uniform self-similar traffic ($N = 16, D = 4, S = 1$).

The dynamics of the fairness control are illustrated in Fig. 9.10 which shows the throughput of four different flows versus time which is given in multiples of the RTT of the ring. All four flows cross the ring link (0,1) from node 0 to node 1, where node 0 is assumed to be a proxy stripping node. Initially, only flow (0,1) from source node 0 to destination node 1 is active, achieving a normalized throughput of 0.9. Next, flow (15,1) is activated at 25 RTTs. After a convergence time of approximately 10 RTTs both flows equally share the available bandwidth on link (0,1). Note that before the new fair rates are established, flow (15,1) fills up the transit queue of node 0, resulting in a throttled rate of flow (0,1) and a throughput peak of flow (15,1). After 50 RTTs flow (12,1) is activated. Flow (12,1) is first sent from node 12 to 0 via the star subnetwork and then uses link (0,1) to reach node 1. We observe that it takes about 10 RTTs to converge to the new fair rates after flow (12,1) has been activated. Finally, flow (7,1) is activated after 75 RTTs. The flow uses the star subnetwork as shortcut from node 8 to node 0, similarly to flow (12,1). Since the fair rate of link (0,1) is transmitted upstream it takes longer for node 7 to receive changes of the fair rate of link (0,1) than for node 12. Note that some packets collide at node 0 and have to be retransmitted since now two flows use the star subnetwork as shortcut, resulting in an increased delay. However, the convergence time remains at approximately 10 RTTs. In summary, the sending rates adapt precisely to the theoretically expected rates in about 10 RTTs and do not suffer from severe oscillations afterwards.

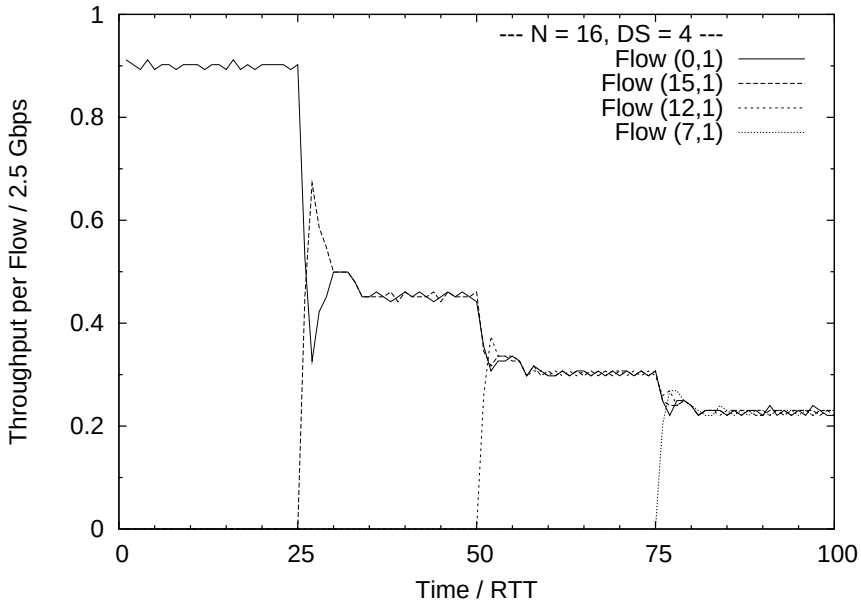


Figure 9.10: Convergence of transmission rates of flows between nodes (0,1), (15,1), (12,1), and (7,1) to their RIAS fair rates vs. time given in ring RTTs ($N = 16, D = 4, S = 1$).

9.3 Conclusions

In this chapter, we have provided an overview of RPR's QoS classes and the corresponding MAC mechanisms as well as on RPR's topology discovery protocol. We discussed how to extend RPR's QoS support to RINGOSTAR enabling circuit-like, high-quality, and best-effort service in our hybrid ring-star architecture. Furthermore, we discussed fairness control in RPR, i.e., RPR's original fairness algorithm RPR-FA that suffers from bandwidth oscillations and limited spatial reuse and the DVSR protocol that mitigates oscillations and implements the RIAS fairness objective for improved spatial reuse. From the latter we derived a fairness control mechanism for RINGOSTAR which we investigated by means of discrete event simulation for static and dynamically changing self-similar traffic demands. Our simulation results show that the proposed fairness control mechanism converges to the RIAS fair rates within approximately ten rotations of the fairness control packet(s), i.e., after approximately ten ring RTTs.

Chapter 10

Conclusions

METRO networks connect increasingly higher-speed access networks with the huge bandwidth pipes of the long-haul backbone. Furthermore, the amount of intra-metro traffic is also growing, for instance due to the deployment of proxy caches for web-traffic at the metro level. Current MANs almost exclusively rely on SONET/SDH technology which has originally been designed for TDM voice traffic and carry the ever increasing amount of bursty Internet data traffic only inefficiently. More specifically, todays SONET/SDH based metro networks suffer from *limited capacity scalability*, *poor bandwidth utilization*, *high provisioning times*, and *large system complexity*. Collectively, these problems are often referred to as the *metro gap*. To overcome these problems, future metro solutions must meet a number of requirements (see Table 2.3).

We identified three main developments targeting to overcome the metro gap, namely (*i*) improving existing SONET/SDH systems (DoS), (*ii*) the development of the RPR standard, and (*iii*) research on packet-switched WDM metro rings. Furthermore, a number of metro WDM networks with a star toplogy have been proposed. A comparison of these approaches with respect to the metro requirements (a summary can be found in Table 5.1) revealed that the combination of a packet-switched ring network like RPR with a single-hop WDM star architecture is an interesting approach for future metro systems. The intention is to use the WDM star as an evolutionary performance upgrade for existing metro packet rings.

In this work we developed and comprehensively evaluated such a hybrid ring-and-star architecture termed RINGOSTAR (see Fig. 6.5) and showed that this network is characterized by the following features:

- **Evolutionary upgrade:** Existing packet-rings, e.g., RPR, can be upgraded with the star subnetwork in an evolutionary way using dark fiber abundantly available in metro areas.
- **'Low-first-cost':** Only a *subset* of the ring nodes, i.e., the proxy nodes, need to be upgraded while the remaining nodes require only few or no modifications.
- **Huge capacity:** 'Proxy stripping' increases the network capacity significantly.
- **Scalability and 'pay-as-you-grow':** The network capacity can be scaled by increasing the number of ring nodes connected to the star subnetwork.
- **Robustness:** The 'protectoration' technique provides fast recovery from multiple link or node failures and improves resilience against such failures compared to a packet-switched ring network.

- **Instantaneous bandwidth provisioning:** No circuits need to be rolled out as opposed to SONET/SDH, connections can be provisioned instantaneously as long as there is sufficient capacity.
- **Efficiency for asymmetric traffic:** Capacity degradation due to asymmetric, e.g., hot-spot traffic is reduced compared to a packet-switched ring network. The network can be architecturally adapted to static asymmetric demands as opposed to packet-rings.
- **High utilization for bursty traffic:** Bursty data traffic, e.g., IP traffic, is transported more efficiently than in SONET/SDH due to statistical multiplexing.
- **QoS support and fairness control:** Three different service classes are provided ranging from SONET/SDH-like circuit emulation down to best-effort service with fairness control.

Of course, these features do not come for free. Compared to a packet switched ring network like RPR, the price one has to pay for the improved performance is mostly related to the additionally required optical fiber and the optical components from which the star is built. However, note, dark fiber is abundantly available in metropolitan areas and therefore relatively inexpensive. Different from the ring subnetwork, no OEO conversion is performed in the single-hop star. Therefore, optical amplifiers and dispersion compensation might be required to maintain a good enough optical signal quality over the whole transmission distance which can span more than 100 km in large rings (note that the star subnetwork has been shown to be able to cover such distances in [127]). To deploy the central AWG and/or PSC and the splitter/combiner pairs appropriate plant space must be found. Furthermore, proxy stripping, protectoration, and especially the star's reservation protocol require additionally electronic processing capacity which might turn out to be an issue if implemented on regular RPR nodes. Form a performance perspective, the star subnetwork can introduce additional delay and jitter due to retransmissions if not dimensioned properly.

10.1 RINGOSTAR vs. Metro Requirements

After providing an overview of RINGOSTAR's key features we now discuss in more detail how our proposed architecture performs with respect to the metro requirements defined in Section 2.3.2. This discussion illustrates how both underlying architectures complement each other to an improved hybrid network combining the individual strengths of one toplogy while overcoming weaknesses of the other.

Multi-Protocol-Support: Packets or streams originating from arbitrary protocols are encapsulated into RPR frames. These frames are then transmitted via RPR's asynchronous access protocol on the ring subnetwork. For transmissions over the synchronized star subnetwork ring-and-star homed nodes aggregate multiple such frames into the fixed size time slots. Therefore, the same protocols that can be transported in RINGOSTAR as in RPR.

Optical Transparency: All ring-and-star homed nodes are connected optically transparent via the central AWG and/or PSC. Communication between this subset of nodes bypasses all other nodes, enables modulation format transparency, and provisioning of wavelength channels. If the ring-and-star homed nodes are positioned in geographic regions with high traffic demands, RINGOSTAR enables optical transparency between customers in these regions.

For this purpose, a slice of the optical spectrum below or above the star subnetwork's TDM channels would be reserved for wavelength channels between customers.

Differentiated SLAs and QoS Levels: As detailed in Section 9.1, RINGOSTAR inherits RPR's sophisticated QoS mechanism supporting three different traffic classes ranging from circuit emulation down to best-effort service.

Fast Provisioning: As in any packet-switched network, no time consuming circuit roll-outs are required like in SONET/SDH. Instead, bandwidth can be provisioned instantly as long as there is sufficient capacity available on all links between source and destination to reserve the requested amount of bandwidth.

Sub-rate Provisioning: The packet-switching paradigm, or more precisely statistical multiplexing, allows provisioning of connections at arbitrary data-rates. Each connection naturally claims the bandwidth it needs. However, flow control, as implemented by RINGOSTAR's traffic shapers, is required to ensure that no connection claims more than the granted amount of bandwidth.

High Bandwidth Utilization: Statistical multiplexing in RINGOSTAR leads to an improved bandwidth utilization for lowly aggregated and therefore bursty metro data traffic compared to circuit-switched technologies like SONET/SDH. While this is no improvement compared to neither of the two underlying architectures which also feature statistical multiplexing, RINGOSTAR in its role as a performance upgrade for packet ring networks significantly reduces the forwarding overhead of the ring nodes and therefore makes more efficient use of the available bandwidth. This results in a huge network capacity as for instance illustrated in Fig. 7.10.

Scalability: RINGOSTAR's proxy stripping mechanism not only increases the network capacity but also makes it scalable by choosing the appropriate number of ring nodes connected to the star subnetwork resulting in a certain capacity. This is a big improvement compared to ring networks like RPR where the capacity is fixed and implicitly also provides a means to combat the problem of low throughput per node for a high number of ring stations inherent to any ring network. The geographical scalability, however, is still limited as all stations need to be connected to form a ring.

Different Traffic Patterns: Asymmetric traffic patterns lead to performance degradation in ring networks as illustrated in Fig. 7.12 for RPR in an hot-spot traffic scenario. In RINGOSTAR this problem can be overcome by appropriately dimensioning the star subnetwork, e.g., by equipping ring-and-star homed nodes located in areas with higher-than-average traffic demands with multiple star interfaces. In dynamically changing traffic scenarios overdimensioning the whole star subnetwork is an effective means to avoid performance degradations for asymmetric traffic patterns. Fig. 7.16 shows that for uniform traffic the star subnetwork should be run at approximately twice the data rate of the ring subnetwork. However, each new generation of optical transceivers usually operates at four times the data rate of the previous generation. For instance, if the ring is operated at 2.5 Gbit/s the star would run at 10 Gbit/s, leaving approximately 100% 'headroom' to cope with asymmetries in the traffic distribution.

Survivability: As in RPR, no spare bandwidth is reserved for recovery from link or node

failures which would be wasted during failure free operation. On the downside this prohibits full bandwidth recovery in case of failures. However, the protectoration mechanism has been shown to result in relatively low performance degradations for many types of failures. (Note that the amount of capacity remaining in a given failure scenario depends on the position of the failed component relative to the star links, i.e., there are less and more important links.) Furthermore, the robustness against multiple failures is improved compared to RPR. While in RPR more than one failed link or node separates the ring into multiple disjoint segments, RINGOSTAR maintains full connectivity between all nodes for many multiple failure scenarios. Although it has not been analyzed in detail, the recovery timescales should be the same as in RPR, namely sub-50 ms recovery, since protectoration relies on RPR's wrapping mechanism upon first detection of a failure. Finally, RINGOSTAR inherits RPR's different survivability classes (which are the same as the QoS traffic classes A, B, and C), i.e., in case of a capacity bottleneck due to a failure packets corresponding to the lowest survivability class are dropped first.

Cost-efficiency: When installed as a performance upgrade of an existing packet ring network the cost for the first deployment is comparably low. The star links can be implemented using dark fiber which is abundantly available in metro areas and only a subset of the ring nodes need to be upgraded with additional transceivers while the existing ring network is fully reused. In case of a deployment 'from scratch' where both subnetworks need to be installed the cost is obviously higher. However, it should be considered that a packet ring can be relatively easily implemented on top of a SONET/SDH ring, which exists in most metro areas. Due to scalability in terms of both the number of nodes and the network capacity RINGOSTAR supports the metro operator's 'pay-as-you-grow' strategy. The node complexity for ring-only nodes is the same as in RPR. Ring-and-star homed nodes require additional transceivers and electronic processing capacity to run the star subnetwork's access protocol. In terms of future proofness, the hybrid ring-star topology can be regarded as a first step in a migration process towards a meshed topology when extended by additional links and more generic routing strategies later. Overall, RINGOSTAR fulfills the metro requirements better than the two underlying architectures individually. This is also illustrated by Table 10.1 that summarizes the previous discussion. Features of the two underlying subnetworks, RPR and the WDM star, that contribute to strengths of RINGOSTAR are printed in *italics*.

10.2 Summary of Contributions

In the following we summarize the major contributions of new knowledge in the order of appearance in this document.

- **Comprehensive survey on packet-switched metro ring systems:** While a survey on published work is no original contribution the corresponding publication can be considered as a helpful service to the research community (published in *IEEE Communications Surveys and Tutorials* [9]).
- **Detailed performance comparison of WDM ring and star networks:** Two representative architectures have been compared in terms of throughput, delay, and packet loss by means of computer simulation for various traffic scenarios including self-similar and hot-spot traffic (published in the *IEEE Journal on Selected Areas in Communica-*

Requirement	DoS	RPR	WDM Star	RINGOSTAR
Multi-Protocol-Support	Yes	Yes	Yes	Yes
Optical Transparency				
Bypassing Nodes	Partly	No	**Yes**[1]	**Partly**[2]
Mod. Format Transp.	No	No	**Yes**	**Partly**[2]
Wavelength Provisioning	Difficult	No	**Yes**	**Partly**[2]
Differentiated SLAs & QoS				
Circuit Emulation	Yes[3]	**Yes**	Yes	**Yes**
Best-Effort	No	**Yes**	Yes	**Yes**
Additional Classes	No	**Yes**	No[4]	**Yes**
Fast Provisioning	No	Yes	Yes	Yes
Sub-rate Provisioning	Yes	Yes	Yes	Yes
High Bandwidth Utilization				
Statistical Multiplexing	No	Yes	Yes	Yes
Forwarding Overhead	Moderate	High	**None**	**Low**
Scalability				
Number of Nodes	Limited	**Good**	Moderate	**Good**
Capacity	Limited	Limited	Moderate	**Good**
Geographically	Limited	Limited	Good	Limited
Different Traffic Patterns				
Uniform	Good	Good	Good	Good
Hot-spot	Good	Moderate	**Good**	**Good**
Dynamically Changing	Limited[5]	Moderate	**Good**	**Good**
Survivability				
Unused Spare Bandwidth	Yes	No	No recovery,	No
Full Bandwidth Recovery	Yes	No	failure	No
Sub-50 ms recovery	Yes	**Yes**	only affects	**Yes**[6]
Multiple Failure Recovery	Limited	Limited	network	**Yes**
Multiple Surv. Classes	Yes	**Yes**	locally	**Yes**
Cost-efficiency				
Low-First-Cost	Yes	**Yes**	No	**Yes**
Pay-As-You-Grow	No	Partly[7]	**Yes**	**Yes**
Node Complexity	High	Moderate	Moderate	Moderate
Migration Towards Mesh	No	No	Limited	**Yes**

Table 10.1: Comparison of RINGOSTAR to the underlying RPR and WDM star architectures. The highlights illustrate strengths inherited from RPR and/or the WDM star. ([1]single hop, [2]between ring-and-star homed nodes, [3]real circuits, [4]potentially possible, [5]using LCAS, [6]not evaluated, [7]no. nodes)

tions (JSAC) [10]).

- **Proposal and performance evaluation of the 'RINGOSTAR' architecture**: An innovative hybrid metro architecture consisting of a ring and a star subnetwork has been proposed. The star subnetwork relies on an innovative access protocol, also being suitable for stand-alone star networks, to reduce the transmission delay. RINGOSTAR's performance has been evaluated in terms of throughput, delay, and packet loss by means of mathematical analysis and verifying computer simulations for various traffic scenarios including self-similar and hot-spot traffic (published in the *IEEE/OSA Journal of Lightwave Technology (JLT)* [11]).

- **Proposal and performance evaluation of the 'proxy stripping' technique:** By means of mathematical analysis and verifying computer simulations of RPR with and without proxy stripping for various configurations and traffic scenarios we have shown that this technique underlying RINGOSTAR significantly increases the capacity of packet-switched ring networks (published in the *OSA Journal of Optical Networking (JON)* [12]).

- **Proposal and performance evaluation of the 'protectoration' technique:** The protectoration technique combines RPR's wrapping and steering mechanism with RINGOSTAR resulting in a hybrid resilience concept implementing both protection and restauration. A comprehensive performance evaluation of protectoration by means of mathematical analysis and verifying computer simulations for various traffic and failure scenarios has shown that this technique generally improves the robustness of packet-switched ring networks and specifically enables recovery from multiple link or node failures (published in the *IEEE/OSA Journal of Lightwave Technology (JLT), Special Issue on Optical Networks* [13]).

- **Proposal of QoS and fairness control mechanisms for RINGOSTAR:** To enable QoS support we have adapted RPR's QoS concept to RINGOSTAR. An adaption of the DVSR fairness control protocol to RINGOSTAR enables fairness for competing best-effort traffic in case of overload situations. We have evaluated the dynamic behavior of the fairness control mechanism by means of computer simulation for self-similar traffic (in part published in *IEEE Communications Magazine* [14]).

10.3 Future Research

As the discussion of RINGOSTAR with respect to the metro requirements has shown, the architecture is relatively complete in terms of features and its performance has been evaluated in large detail. Of course, several aspects can still be refined. For instance, the star subnetwork could be extended by a periodic bandwidth reservation mechanism to reduce delay and jitter for high priority traffic. Furthermore, hot-spot traffic scenarios where the hot-spot is not a ring-and-star homed node could be considered and transceiver tuning times could be included in performance analysis to make the results more realistic.

However, in the author's opinion, rather than focusing on further details of the architecture, access protocol, or performance evaluation, research on RINGOSTAR would benefit most from being pushed closer to a system manufacturer and/or network operator environment to address the following questions: (*i*) How can RINGOSTAR be implemented techni-

cally in an efficient way? (*ii*) What are the cost factors for RINGOSTAR in a typical metro environment?

Concerning the first question, a testbed implementation of RINGOSTAR would be interesting as a proof-of-concept. In the first version of the testbed the star subnetwork would not necessarily have to rely on an AWG/PSC based network. Instead, the all-optical star subnetwork could be replaced by technically simpler and less expensive GbE connections and a central Ethernet switch. In this simplified setting all features of RINGOSTAR such as proxy stripping, protectoration, QoS support, and fairness control could be implemented and tested in software before moving to the probably relatively costly hardware implementation of the star subnetwork.

To answer the second question, the cost for upgrading an existing SONET/SDH or RPR network with RINGOSTAR must be estimated. The cost for the required components could be derived from the testbed implementation or approximated theoretically. Another cost major factor would be the dark fiber required to build the star subnetwork. Of course, there are many other capital or operational expenditures that must be considered including deployment, rack space for the node modules and the central AWG/PSC, power consumption, or maintenance. The resulting costs must be gauged against the performance benefits resulting the RINGOSTAR upgrade which are evaluated in detail in this work.

Appendix A

Publications

In the following we provive an overview of the author's publications related to this work.

- M. Herzog, M. Maier, and M. Reisslein, "Metropolitan Area Packet-Switched WDM Networks: A Survey on Ring Systems", *IEEE Communications Surveys*, vol. 6, no. 2, pp. 2–20, May 2004.

- H.-S. Yang, M. Herzog, M. Maier, and M. Reisslein, "Metro WDM Networks: Performance Comparison of Slotted Ring and AWG Star Networks", *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 22, no. 8, pp. 1460–1473, October 2004.

- M. Herzog, M. Maier, and A. Wolisz, "RINGOSTAR: An Evolutionary AWG Based WDM Dark-Fiber Upgrade of Optical Ring Networks", *IEEE/OSA Journal of Lightwave Technology (JLT)*, vol. 23, no. 4, pp. 1637–1651, April 2005.

- M. Herzog, S. Adams, and M. Maier, "PROXY STRIPPING: A Performance-Enhancing Technique for Optical Metropolitan Area Ring Networks", *OSA Journal of Optical Networking (JON)*, vol. 4, no. 7, pp. 400–431, July 2005.

- M. Maier, M. Herzog, M. Scheutzow, and M. Reisslein, "PROTECTORATION: A Fast and Efficient Multiple-Failure Recovery Technique for Resilient Packet Ring (RPR) Using Dark Fiber", *IEEE/OSA Journal of Lightwave Technology (JLT), Special Issue on Optical Networks*, vol. 23, no. 10, pp. 2816–2838, October 2005.

- M. Herzog and M. Maier, "RINGOSTAR: An Evolutionary Performance-Enhancing WDM Upgrade of IEEE 802.17 Resilient Packet Ring (RPR)", *IEEE Communications Magazine*, vol. 44, no. 2, pp. S11–S17, February 2006.

Appendix B

Acronyms

2D	two-dimensional
3D	three-dimensional
3R	reamplifying, reshaping, retiming
ADM	add-drop multiplexer
AM	aggressive mode
APD	avalanche photodiode
APS	Automatic Protection Switching
ASE	amplified spontaneous emission
ASK	amplitude shift keying
ASTN	Automatic Switched Transport Network
ATM	Asynchronous Transfer Mode
ATMR	Asynchronous Transfer Mode Ring
AWG	arrayed-waveguide grating
BER	bit error rate
CBR	constant bit rate
CC	control channel
CDN	content distribution network
CIR	committed information rate
CM	conservative mode
CSMA/CA	carrier sense multiple access with collision avoidance
CSMA/CD	carrier sense multiple access with collision detection

DCF	dispersion compensation fiber
DoS	Data over SONET/SDH
DQBR	Distributed Queue Bidirectional Ring
DQDB	IEEE 802.6 Distributed Queue Dual Bus
DSL	digital subscriber loop
DVSR	Distributed Virtual-time Scheduling in Rings
DWADM	dynamic wavelength add-drop multiplexer
EDFA	erbium doped fiber amplifier
EFM	Ethernet in the First Mile
EIR	excess information rate
EPON	Ethernet passive optical network
ESCON	Fiber Distributed Data Interface
FBG	fiber Bragg grating
FCC	Federal Communications Commission
FCFS	first-come-first-served
FDDI	Fibre Distributed Data Interface
FDL	fiber delay line
FE	fairness eligible
FIFO	first-in-first-out
FR	fixed-tuned receiver
FSK	frequency shift keying
FSR	free spectral range
FT	fixed-tuned transmitter
FTP	file transfer protocol
FWM	four-wave mixing
GbE	Gigabit Ethernet
GFP-F	Frame-Mapped GFP
GFP	Generic Framing Procedure
GFP-T	Transparent GFP

GMPLS	Generalized Multiprotocol Label Switching
GUI	graphical user interface
HOL	head-of-line
HORNET	Hybrid Optoelectronic Ring NETwork
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPoRPR	IP over Resilient Packet Ring
ITU-T	International Telecommunication Union-Telecommunication Standardization Sector
LAN	local area network
LCAS	Link Capacity Adjustment Scheme
LC	liquid crystal
LD	laser diode
LED	light emitting diode
LQ	longest queue
MAC	medium access control
MAN	metropolitan area network
M-ATMR	multi-channel ATMR
MAWSON	Metropolitan Area Wavelength Switched Optical Network
MEF	Metro Ethernet Forum
MEMS	micro-electro-mechanical system
MMF	multi-mode fiber
MMR-MS	MMR Multiple SAT
MMR	Multi-MetaRing
MMR-SS	MMR Single SAT
MPλS	multi-protocol wavelength switching
MQW	multiple-quantum-well
MTIT	multitoken interarrival time

MTU	maximum transfer unit
OADM	optical add-drop multiplexer
OA&M	operation, administration, and management
OBS	optical burst switching
OEO	optical-electronic-optical
OOO	all-optical
OPS	optical packet switching
OXC	optical crossconnect
PCM	pulse code modulation
PD	photodiode
PMD	polarisation-mode dispersion
POP	point of presence
PoS	Packet over SONET/SDH
PSC	passive star coupler
PSR	photonic slot routing
PtP	point-to-point
PTQ	primary transit queue
PXT	pre-cross-connected trail
QoS	Quality of Service
RAM	random access memory
RIAS	Ring Ingress-Aggregated with Spatial Reuse
RINGO	RING Optical network
ROADM	reconfigurable add-drop multiplexer
RPR-FA	RPR fairness algorithm
RPR	IEEE 802.17 Resilient Packet Ring
RPRWG	RPR Working Group
RTT	round-trip time
SAR-OD	segmentation and reassembly on demand
SBS	Stimulated Brillouin Scattering

SCM	sub-carrier multiplexing
SDL	switched delay line
SLA	service level agreement
SMARTNet	Scalable Multi-channel Adaptable Ring Terabit Network
SMF	single-mode fiber
SNR	signal-to-noise ratio
SOA	semiconductor optical amplifier
SONET/SDH	synchronous optical network/synchronous digital hierarchy
SPM	self-phase modulation
SR^3	SRR with Reservations
SRP	Spatial Reuse Protocol
SRR	Synchronous Round Robin
SRS	Stimulated Raman Scattering
STQ	secondary transit queue
TCP/IP	Transmission Control Protocol/Internet Protocol
TDMA	time division multiple access
TDM	time division multiplexing
TR	tunable receiver
TTL	time-to-live
TT	tunable transmitter
UMTS	Universal Mobile Telecommunication System
UNI	User Network Interface
VBR	variable bit rate
VCSEL	vertical-cavity surface-emitting laser
VC	Virtual Concatenation
VLAN	virtual LAN
VOQ	virtual output queue
WAN	wide area network
WDM	wavelength division multiplexing

WG	working group
WLAN	wireless local area network
WRS	wavelength-routing switch
WSXC	wavelength selective crossconnect
XPM	cross-phase modulation

Bibliography

[1] B. Mukherjee. *Optical WDM Networks*. Springer, 1st edition, January 2006.

[2] K. G. Coffman and A. M. Odlyzko. *Optical Fiber Telecommunications IV-B: Systems and Impairments*, chapter Growth of the Internet, pages 17–56. Academic Press, 2002.

[3] M. Lesk. How much information is there in the world? unpublished paper, available at http://www.lesk.com/mlesk/diglib.html, 1997.

[4] B. Mukherjee. WDM Optical Communication Networks: Progress and Challenges. *IEEE Journal on Selected Areas in Communications*, 18(10):1810–1824, October 2000.

[5] P.-H. Ho and H. T. Mouftah. A Framework of Scalable Optical Metropolitan Networks for Improving Survivability and Class of Service. *IEEE Network*, 16(4):29–35, July/Aug. 2002.

[6] S. Yao, S. J. B. Yoo, and B. Mukherjee. All–optical packet switching for metropolitan area networks: Opportunities and challenges. *IEEE Communications Magazine*, 39(3):142–148, March 2001.

[7] G. Barish and K. Obraczka. World Wide Web Caching: Trends and Techniques. *IEEE Communications Magazine*, 38(5):178–185, May 2000.

[8] L. G. Kazovsky, I. M. White, K. Shrikande, and M. S. Rogge. High Capacity Metropolitan Area Networks for the Next Generation Internet. In *Proc., 35th Asilomar Conference on Signals, Systems, and Computers*, volume 1, pages 3–7, 2001.

[9] M. Herzog, M. Maier, and M. Reisslein. Metropolitan Area Packet-Switched WDM Networks: A Survey on Ring Systems. *IEEE Communications Surveys and Tutorials*, 6(2):2–20, Second Quarter 2004.

[10] H.-S. Yang, M. Herzog, M. Maier, and M. Reisslein. Metro WDM Networks: Performance Comparison of Slotted Ring and AWG Star Networks. *IEEE Journal on Selected Areas in Communications*, 22(8):1460–1473, October 2004.

[11] M. Herzog, M. Maier, and A. Wolisz. RINGOSTAR: An Evolutionary AWG Based WDM Upgrade of Optical Ring Networks. *IEEE/OSA Journal of Lightwave Technology*, 23(4):1637–1651, April 2005.

[12] M. Herzog, S. Adams, and M. Maier. PROXY STRIPPING: A Performance Enhancing Technique for Optical Metropolitan Area Ring Networks. *OSA Journal of Optical Networking (JON)*, 4(7):400–431, July 2005.

[13] M. Maier, M. Herzog, M. Scheutzow, and M. Reisslein. PROTECTORATION: A Fast and Efficient Multiple–Failure Recovery Technique for Resilient Packet Ring (RPR) Using Dark Fiber. *IEEE/OSA Journal of Lightwave Technology (JLT), Special Issue on Optical Networks*, 23(10):2816–2838, October 2005.

[14] M. Herzog and M. Maier. RINGOSTAR: An Evolutionary Performance-Enhancing WDM Upgrade of IEEE 802.17 Resilient Packet Ring (RPR). *IEEE Communications Magazine*, 44(2):S11–S17, February 2006.

[15] K. Petermann. Einführung in die optische Nachrichtentechnik, 2004. Lecture Notes, Institute of Photonics, Technical University Berlin.

[16] ITU-T Rec. G.692. Optical interfaces for multichannel systems with optical amplifiers, October 1998.

[17] R. C. Alferness. *Guided-Wave Optoelectronics*, chapter Titanium-diffused Lithium Niobate Waveguide Devices. Springer, 1988.

[18] Sorrento Networks. Metropolitan optical networks: Overview and requirements. white paper, available at http://www.lightreading.com, June 23 2000.

[19] W. Grover. *Mesh-based Survivable Transport Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking*, chapter Failure Impacts, Survivability Principles, and Measures of Survivability, pages 103–172. Prentice Hall PTR, 2003.

[20] N. F. Maxemchuk, I. Ouveysi, and M. Zukerman. A quantitative measure for telecommunications networks topology design. *IEEE/ACM Transactions on Networking*, 13(4):731–742, August 2005.

[21] L. G. Kazovsky, K. Shrikhande, I. M. White, M. Rogge, and D. Wonglumson. Optical Metropolitan Area Networks. In *Proc., Optical Fiber Communication Conference and Exhibit (OFC), paper WU1*, volume 3, pages WU1–1–WU1–3, Anaheim, CA, March 2001.

[22] I. Chlamtac and A. Ganz. A Multibus Train Communication (AMTRAC) Architecture for High–Speed Fiber Optic Networks. *IEEE Journal on Selected Areas in Communications*, 6(6):903–912, July 1988.

[23] M. N. Ransom and D. R. Spears. Applications of Public Gigabit Networks. *IEEE Network*, 6(2):30–40, March 1992.

[24] P. E. Green. *Fiber Optic Networks*. Prentice Hall, 1993.

[25] R. Ramaswami. Optical Fiber Communication: From Transmission To Networking. *IEEE Communications Magazine*, 40(5):138–147, May 2002.

[26] S. Mokbel. Canada's Optical Research and Education Network: CA*net3. In *Proc., Design of Reliable Communication Networks (DRCN)*, pages 10–32, Munich, Germany, April 2000.

[27] Y. Cai, R. M. Fortenberry, and R. S. Tucker. Demonstration of Photonic Packet–Switched Ring Network with Optically Transparent Nodes. *IEEE Photonics Technology Letters*, 6(9):1139–1141, 1994.

[28] D. Guo and A. S. Acampora. Scalable Multihop WDM Passive Ring with Optimal Wavelength Assignment and Adaptive Wavelength Routing. *IEEE/OSA Journal of Lightwave Technology*, 14(6):1264–1277, June 1996.

[29] W. Goralski. *SONET/SDH*. Osborne McGraw-Hill, 3rd edition, October 2002.

[30] E. Modiano and P. J. Lin. Traffic Grooming in WDM Networks. *IEEE Communications Magazine*, 39(7):124–129, 2001.

[31] R. Jain. Optical networking: Recent developments, issues and trends. Tutorial at *IEEE Infocom 2003*, March 2003.

[32] R. D. Doverspike, S. J. Phillips, and J. R. Westbrook. Transport Network Architectures in an IP World. In *Proc., IEEE INFOCOM*, pages 305–314, Tel Aviv, Israel, March 2000.

[33] ITU–T Rec. G.7041. Generic Framing Procedure (GFP), October 2001.

[34] ITU–T Rec. G.707. Network Node Interface for the Synchronous Data Hierarchy, October 2000.

[35] ITU–T Rec. G.7042. Link Capacity Adjustment Scheme (LCAS) for Virtual Concatenation, October 2001.

[36] S. S. Gorshe and T. Wilson. Transparent Generic Framing Procedure (GFP): A Protocol for Efficient Transport of Blockcoded Data through SONET/SDH Networks. *IEEE Communications Magazine*, 40(5):88–95, May 2002.

[37] IEEE. IEEE Standard 802.17: Resilient Packet Ring. 2004. http://ieee802.org/17.

[38] D. E. Huber, W. Steinlin, and P. J. Wild. SILK: An Implementation of a Buffer Insertion Ring. *IEEE Journal on Selected Areas in Communications*, SAC–1(5):766–774, Nov. 1983.

[39] I. Cidon and Y. Ofek. Metaring – A Full–duplex Ring with Fairness and Spatial Reuse. In *Proc., IEEE INFOCOM*, pages 969–978, San Francisco, CA, June 1990.

[40] I. Cidon and Y. Ofek. MetaRing – A Full–Duplex Ring with Fairness and Spatial Reuse. *IEEE Transactions on Communications*, 41(1):110–120, January 1993.

[41] G. Kramer and G. Pesavento. Ethernet Passive Optical Network (EPON): Building a Next–Generation Optical Access Network. *IEEE Communications Magazine*, 40(2):66–73, February 2002.

[42] T. Shan, J. Yang, and C. Sheng. EPON Upstream Multiple Access Sheme. In *Proc., IEEE International Conference on Infotech and Infonet (ICII)*, volume 2, pages 273–278, Beijing, China, October/November 2001.

[43] G. Kramer, B. Mukherjee, and G. Pesavento. IPACT: A Dynamic Protocol for an Ethernet PON (EPON). *IEEE Communications Magazine*, 40(2):74–80, February 2002.

[44] A. Gumaste and I. Chlamtac. A Protocol to Implement Ethernet Over PON. In *Proc., IEEE International Conference on Communications (ICC)*, volume 2, pages 1345–1349, Anchorage, AK, May 2003.

[45] M. C. Nuss. Optical Ethernet in the Metro. In *Proc., 14th Annual Meeting of the IEEE Lasers and Electro–Optics Society (LEOS)*, volume 1, page 289, San Diego, CA, November 2001.

[46] A. Richter, H. Bock, W. Fischler, P. Leisching, P. M. Krummrich, A. Mayer, R. E. Neuhauser, J. P. Elbers, and C. Glingener. Germany–wide DWDM field trial: Transparent connection of a long haul link and a multiclient metro network. In *Proc., Optical Fiber Communications Conference and Exhibit (OFC), paper ML3*, volume 1, pages ML3/1–ML3/3, Anaheim, MD, March 2001.

[47] D. Stoll, P. Leisching, H. Bock, and A. Richter. Metropolitan DWDM: A Dynamically Configurable Ring for the KomNet Field Trial in Berlin. *IEEE Communications Magazine*, 39(2):106–113, February 2001.

[48] R. Gaudino, A. Carena, V. Ferrero, A. Pozzi, V. De Feo, P. Gigante, F. Neri, and P. Poggiolini. RINGO: A WDM Ring Optical Packet Network Demonstrator. In *Proc., 27th European Conference on Optical Communication (ECOC 2001)*, pages 620–621, Amsterdam, Netherlands, October 2001.

[49] R. Gaudino. RINGO: Demonstration of a WDM packet network architecture for metro applications. In *Proc., 4th International Conference on Transparant Optical Networks*, volume 1, pages 77–80, 2002.

[50] S. M. Gemelos, I. M. White, D. Wonglumson, K. Shrikhande, T. Ono, and L. G. Kazovsky. WDM Metropolitan Area Network Based on CSMA/CA Packet Switching. *IEEE Photonics Technology Letters*, 11(11):1512–1514, November 1999.

[51] K. V. Shrikhande, I. M. White, D.-R. Wonglumsom, S. M. Gemelos, M. S. Rogge, Y. Fukashiro, M. Avenarius, and L. G. Kazovsky. HORNET: A Packet–Over–WDM Multiple Access Metropolitan Area Ring Network. *IEEE Journal on Selected Areas in Communications*, 18(10):2004–2016, October 2000.

[52] Y. Fukashiro, K. Shrikhande, M. Avenarius, M. S. Rogge, I. M. White, D. Wonglumson, and L. G. Kazovsky. Fast and fine tuning of a GCSR laser using a digitally controlled driver. In *Proc., Optical Fiber Comminucation Conference and Exhibit (OFC), paper WM43*, volume 2, pages 338–340, Baltimore, MD, March 2000.

[53] K. Shrikhande, I. M. White, M. S. Rogge, F.-T. An, A. Srivatsa, E. S. Hu, S. S.-H. Yam, and L. G. Kazovsky. Performance Demonstration of a Fast–Tunable Transmitter and Burst–Mode Packet Receiver for *HORNET*. In *Proc., Optical Fiber Communication Conference and Exhibit (OFC), paper ThG2*, volume 4, pages ThG2–1–ThG2–3, Anaheim, CA, March 2001.

[54] D. Wonglumson, I. M. White, S. M. Gemelos, K. Shrikande, and L. G. Kazovsky. HORNET — a Packet–Switched WDM Metropolitan Area Ring Network: Optical Packet Transmission and Recovery, Queue Depth, and Packet Latency. In *Proc., IEEE*

Lasers and Electro–Optics Society (LEOS), volume 2, pages 653–654, San Francisco, CA, November 1999.

[55] D. Wonglumsom, I. M. White, K. Shrikhande, M. S. Rogge, S. M. Gemelos, F.-T. An, Y. Fukashiro, M. Avenarius, and L. G. Kazovsky. Experimental Demonstration on an Access Point for HORNET — A Packet–Over–WDM Multiple–Access MAN. *IEEE/OSA Journal of Lightwave Technology*, 18(12):1709–1717, December 2000.

[56] I. M. White, Y. Fukashiro, K. Shrikande, C. Wonglumson, M. S. Rogge, M. Avenarius, and L. G. Kazovsky. Experimental Demonstration of a Media Access Protocol for HORNET: A WDM Multiple Access Metropolitan Area Ring Network. In *Proc., Optical Fiber Comminication Conference and Exhibit (OFC), paper WD3*, volume 2, pages 50–52, Baltimore, MD, March 2000.

[57] I. M. White, M. S. Rogge, K. Shrikande, Y. Fukashiro, D. Wonglumson, F.-T. An, and L. G. Kazovsky. Experimental Demonstration of a Novel Media Access Protocol for HORNET: A Packet–Over–WDM Multiple–Access MAN Ring. *IEEE Photonics Technology Letters*, 12(9):1264–1266, September 2000.

[58] B. Mukherjee. WDM–Based Local Lightwave Networks Part I: Single–Hop Systems. *IEEE Network*, 6(3):12–27, May 1992.

[59] S. Johansson, A. Manzalini, M. Giannoccaro, R. Cadeddu, M. Giorgi, R. Clemente, R. Brandstrom, A. Gladisch, J. Chawki, L. Gillner, P. Ohlen, and E. Berglind. A Cost–Effective Approach to Introduce an Optical WDM Network in the Metropolitan Environment. *IEEE Journal on Selected Areas in Communications*, 16(7):1109–1122, September 1998.

[60] M. A. Summerfield. MAWSON: A Metropolitan Area Wavelength Switched Optical Network. In *Proc., Asia Pacific Conference on Communication (APCC)*, volume 1, pages 327–331, Sydney, Australia, November 1997.

[61] J. Fransson, M. Johansson, M. Roughan, L. Andrew, and M. A. Summerfield. Design of a Medium Access Control Protocol for a WDMA/TDMA Photonic Ring Network. In *Proc., IEEE GLOBECOM*, volume 1, pages 307–312, November 1998.

[62] M. J. Spencer and M. A. Summerfield. WRAP: A Medium Access Control Protocol for Wavelength–Routed Passive Optical Networks. *IEEE/OSA Journal of Lightwave Technology*, 18(12):1657–1676, December 2000.

[63] A. Carena, V. Ferrero, R. Gaudino, V. De Feo, F. Neri, and P. Poggiolini. RINGO: A Demonstrator of WDM Optical Packet Network on a Ring Topology. In *Proc., Optical Network Design and Modeling 2002*, February 2002.

[64] A. Bianco, M. Bonsignori, E. Leonardi, and F. Neri. Variable–Size Packets in Slotted WDM Ring Networks. In *Proc., Optical Network Design and Modelling (ONDM)*, pages 151–166, Torino, Italy, February 2002.

[65] M. Ajmone Marsan, A. Bianco, E. Leonardi, M. Meo, and F. Neri. On the Capacity of MAC Protocols for All–Optical WDM Multi–Rings with Tunable Transmitters and Fixed Receivers. In *Proc., IEEE INFOCOM*, volume 3, pages 1206–1216, March 1996.

[66] M. Ajmone Marsan, A. Bianco, E. Leonardi, M. Meo, and F. Neri. MAC Protocols and Fairness Control in WDM Multirings with Tunable Transmitters and Fixed Receivers. *IEEE Journal of Lightwave Technology*, 14(6):1230–1244, June 1996.

[67] M. Ajmone Marsan, E. Leonardi, M. Meo, and F. Neri. Modeling slotted WDM rings with discrete–time Markovian models. *Computer Networks*, 32(5):599–615, May 2000.

[68] A. Bianco, V. Distefano, A. Fumagalli, E. Leonardi, and F. Neri. A–Posteriori Access Strategies in All–Optical Slotted WDM Rings. In *Proc., IEEE GLOBECOM*, volume 1, pages 300–306, Sydney, Australia, November 1998.

[69] K. Shrikande, A. Srivatsa, I. M. White, M. S. Rogge, D. Wonglumson, S. M. Gemelos, and L. G. Kazovsky. CSMA/CA MAC Protocols for IP–HORNET: An IP over WDM Metropolitan Area Ring Network. In *Proc., IEEE GLOBECOM*, volume 2, pages 1303–1307, San Francisco, CA, Nov./Dec. 2000.

[70] W.-P. Chen and W.-S. Hwang. A Packet Pre–Classification CSMA/CA MAC Protocol for IP over WDM Ring Networks. In *Proc., IEEE International Conference on Communication Systems*, volume 2, pages 1217–1221, 2002.

[71] C.-C. Li, S.-W. Kau, and W.-S. Hwang. A CSMA/CP MAC Protocols for IP over WDM Metropolitan Area Ring Networks. In *Proc., IEEE International Conference on Communication Systems (ICCS)*, volume 2, pages 1212–1216, Singapore, November 2002.

[72] K. Bengi and H. R. van As. Efficient QoS Support in a Slotted Multihop WDM Metro Ring. *IEEE Journal on Selected Areas in Communications*, 20(1):216–227, January 2002.

[73] K. Bengi. An Analytical Model for a Slotted WDM Metro Ring with A–Posteriori Access. In *Proc., Optical Network Design and Modelling (ONDM)*, Torino, Italy, February 2002.

[74] C. S. Jelger and J. M. H. Elmirghani. A Simple MAC Protocol for WDM Metropolitan Access Ring Networks. In *Proceedings of IEEE Global Telecommunications Conference*, volume 3, pages 1500–1504, November 2001.

[75] C. S. Jelger and J. M. H. Elmirghani. Performance of a slotted MAC Protocol for WDM Metropolitan Access Ring Networks under Self–Similar traffic. In *Proc., IEEE International Conference on Communications (ICC)*, volume 5, pages 2806–2811, New York, NY, April 2002.

[76] C. S. Jelger and J. M. H. Elmirghani. Photonic Packet WDM Ring Networks Architecture and Performance. *IEEE Communications Magazine*, 40(11):110–115, November 2002.

[77] E. W. M. Wong, A. Fumagalli, and I. Chlamtac. Performance Evaluation of CROWNs: WDM Multi–Ring Topologies. In *Proc., IEEE International Conference on Communications (ICC)*, volume 2, pages 1296–1301, Seattle, WA, June 1995.

[78] I. M. White, M. S. Rogge, K. Shrikhande, and L. G. Kazovsky. Design of a control–channel–based media–access–control protocol for HORNET. *Journal of Optical Networking*, 1(12):460–473, December 2002.

[79] I. M. White, K. Shrikhande, M. S. Rogge, S. M. Gemelos, D. Wonglumsom, G. Desa, Y. Fukashiro, and L. G. Kazovsky. Architecture and Protocols for HORNET: A Novel Packet–over–WDM Multiple–Access MAN. In *Proc., IEEE GLOBECOM*, volume 2, pages 1298–1302, San Francisco, CA, Nov./Dec. 2000.

[80] I. M. White, M. S. Rogge, Y.-L. Hsueh, K. Shrikhande, and L. G. Kazovsky. Experimental Demonstration of the HORNET Survivable Bi–directional Ring Architecture. In *Proc., Optical Fiber Communication Conference and Exhibit (OFC), paper WW1*, pages 346–349, Anaheim, CA, March 2002.

[81] K. Bengi. An Optical Packet–Switched IP–over–WDM Metro Ring Network. In *Proc., 27th Annual IEEE Conference on Local Computer Networks (LCN)*, pages 43–52, Tampa, FL, November 2002.

[82] K. Bengi. Access Protocols for an Efficient Optical Packet–Switched Metropolitan Area Ring Network Supporting IP Datagrams. In *Proc., Eleventh International Conference on Computer Communications and Networks (ICCCN)*, pages 284–289, Miami, FL, October 2002.

[83] M. Boroditsky *et al.* Experimental demonstration of composite packet switching on a WDM photonic slot routing network. In *Proc., Optical Fiber Communication Conference and Exhibit (OFC), paper ThG6*, volume 4, pages ThG6–1–ThG6–3, Anaheim, CA, March 2001.

[84] A. Smiljanić, M. Boroditsky, and N. J. Frigo. Optical Packet–Switched Ring Network with Flexible Bandwidth Allocation. In *Proc., IEEE Workshop on High Performance Switching and Routing (HPSR)*, pages 83–87, Dallas, TX, May 2001.

[85] A. Smiljanić, M. Boroditsky, and N. J. Frigo. High–Capacity Packet–Switched Optical Ring Network. *IEEE Communications Letters*, 6(3):111–113, March 2002.

[86] I. Chlamtac, V. Elek, A. Fumagalli, and C. Szabó. Scalable WDM Access Network Architecture Based on Photonic Slot Routing. *IEEE/ACM Transactions on Networking*, 7(1):1–9, February 1999.

[87] V. Elek, A. Fumagalli, and G. Wedzinga. Photonic Slot Routing: A Cost–Effective Approach to Designing All–Optical Access and Metro Networks. *IEEE Communications Magazine*, 39(11):164–172, November 2001.

[88] W. Cho and B. Mukherjee. Design of MAC Protocols for DWADM–Based Metropolitan–Area Optical Ring Networks. In *Proc., IEEE GLOBECOM*, volume 3, pages 1575–1579, San Antonio, TX, November 2001.

[89] A. Fumagalli, J. Cai, and I. Chlamtac. The Multi–Token Inter–Arrival Time (MTIT) Access Protocol for Supporting IP over WDM Ring Network. In *Proc., IEEE International Conference on Communications*, volume 1, pages 586–590, Vancouver, Canada, June 1999.

[90] J. Cai, A. Fumagalli, and I. Chlamtac. The Multitoken Interarrival Time (MTIT) Access Protocol for Supporting Variable Size Packets Over WDM Ring Network. *IEEE Journal on Selected Areas in Communications*, 18(10):2094–2104, October 2000.

[91] I. Rubin and H.-K. Hua. An All–Optical Wavelength–Division Meshed–Ring Packet–Switching Network. In *Proc., IEEE INFOCOM*, volume 3, pages 969–976, Boston, MA, April 1995.

[92] I. Rubin and H.-K. Hua. SMARTNet: An All–Optical Wavelength–Division Meshed–Ring Packet–Switching Network. In *Proc., IEEE GLOBECOM*, volume 3, pages 1756–1760, Singapore, November 1995.

[93] I. Rubin and H.-K. H. Hua. Synthesis and Throughput Behavior of WDM Meshed–Ring Networks Under Nonuniform Traffic Loading. *IEEE/OSA Journal of Lightwave Technology*, 15(8):1513–1521, August 1997.

[94] I. Rubin and J. Ling. All–Optical Cross–Connect Meshed–Ring Communications Networks using a Reduced Number of Wavelengths. In *Proc., IEEE INFOCOM*, volume 2, pages 924–931, New York, NY, March 1999.

[95] M. Ajmone Marsan, A. Bianco, E. Leonardi, F. Neri, and S. Toniolo. An Almost Optimal MAC Protocol for All–Optical WDM Multi–Rings with Tunable Transmitters and Fixed Receivers. In *Proc., IEEE International Conference on Communications*, volume 1, pages 437–442, Montreal, June 1997.

[96] J. Chen, I. Cidon, and Y. Ofek. A Local Fairness Algorithm for the MetaRing, and its Performance Study. In *Proc., IEEE GLOBECOM*, volume 3, pages 1635–1641, Orlando, FL, December 1992.

[97] M. Ajmone Marsan, A. Bianco, E. Leonardi, F. Neri, and S. Toniolo. MetaRing Fairness Control Schemes in All–Optical WDM Rings. In *Proc., IEEE INFOCOM*, volume 2, pages 752–760, Kobe, Japan, April 1997.

[98] K. Bengi and H. R. van As. QoS Support and Fairness Control in a Slotted Packet–Switched WDM Metro Ring Network. In *Proc., IEEE GLOBECOM*, volume 3, pages 1494–1499, San Antonio, TX, November 2001.

[99] M. Ajmone Marsan, A. Bianco, E. Leonardi, A. Morabito, and F. Neri. SR^3: A Bandwidth–Reservation MAC Protocol for Multimedia Applications over All–Optical WDM Multi–Rings. In *Proc., IEEE INFOCOM*, volume 2, pages 761–768, Kobe, Japan, April 1997.

[100] M. Ajmone Marsan, A. Bianco, E. Leonardi, A. Morabito, and F. Neri. All–Optical WDM Multi–Rings with Differentiated QoS. *IEEE Communications Magazine*, 37(2):58–66, February 1999.

[101] A. Fumagalli, J. Cai, and I. Chlamtac. A Token Based Protocol for Integrated Packet and Circuit Switching in WDM Rings. In *Proc., IEEE GLOBECOM*, volume 4, pages 2339–2344, Sydney, Australia, November 1998.

[102] A.S.T. Lee, D. K. Hunter, D. G. Smith, and D. Marcenac. Heuristic for setting up a stack of WDM rings with wavelength reuse. *IEEE Journal of Lightwave Technology*, 18(4):521–529, April 2000.

[103] I. Rubin and J. Ling. Delay Analysis of All–Optical Packet–Switching Ring and Bus Communications Networks. In *Proc. of IEEE Globecom 2001*, volume 3, pages 1585–1589, November 2001.

[104] I. M. White, E. S.-T. Hu, Y.-L. Hsueh, K. V. Shrikhande, M. S. Rogge, and L. G. Kazovsky. Demonstration and system analysis of the HORNET. *IEEE/OSA Journal of Lightwave Technology*, 21(11):2489–2498, November 2003.

[105] I. M. White, M. S. Rogge, K. V. Shrikhande, and L. G. Kazovsky. A summary of the HORNET project: A next-generation metropolitan area network. *IEEE Journal on Selected Areas in Communications*, 21(9):1478–1494, November 2003.

[106] K. Shrikhande, A. Srivatsa, I. M. White, M. S. Rogge, D. Wonglumsom, S. M. Gemelos, and L. G. Kazovsky. CSMA/CA MAC Protocols for IP–HORNET: An IP over WDM Metropolitan Area Ring Network. In *Proceedings of Globecom '00*, volume 2, pages 1303–1307, San Francisco, CA, November 2000.

[107] O. A. Lavrova, G. Rossi, and D. J. Blumenthal. Rapid tunable transmitter with large number of ITU channels accessible in less than 5 ns. In *Proc. of ECOC '00*, volume 2, pages 169–170, Munich, Germany, September 2000.

[108] D.-R. Wonglumsom, I. M. White, K. V. Shrikhande, M. S. Rogge, S. M. Gemelos, F.-T. An, Y. Fukashiro, M. Avenarius, and L. G. Kazovsky. Experimental demonstration of an access point for HORNET—A Packet-Over-WDM Multiple-Access MAN. *IEEE/OSA Journal of Lightwave Technology*, 18(12):1709–1717, December 2000.

[109] B. Mukherjee. WDM–Based Local Lightwave Networks Part I: Single–Hop Systems. *IEEE Network*, 6(3):12–27, May 1992.

[110] M. Maier, M. Reisslein, and A. Wolisz. Towards Efficient Packet Switching Metro WDM Networks. *Optical Networks Magazine*, 3(6):44–62, Nov. 2002.

[111] M. Scheutzow, M. Maier, M. Reisslein, and A. Wolisz. Wavelength reuse for efficient packet-switched transport in an AWG-based metro WDM network. *IEEE/OSA Journal of Lightwave Technology*, 21(6):1435–1455, June 2003.

[112] M. Maier, M. Reisslein, and A. Wolisz. A Hybrid MAC Protocol for a Metro WDM Network Using Multiple Free Spectral Ranges of an Arrayed–Waveguide Grating. *Computer Networks*, 41(4):407–433, March 2003.

[113] K. Kato, A. Okada, Y. Sakai, and K. Noguchi *et al.* 10–Tbps Full–Mesh WDM Network Based On Cyclic–Frequency Arrayed–Waveguide Grating Router. In *Proc. of ECOC '00*, volume 1, pages 105–107, Munich, Germany, September 2000.

[114] A. Okada, T. Sakamoto, Y. Sakai, and K. Noguchi *et al.* All–Optical Packet Routing by an Out–of–Band Optical Label and Wavelength Conversion in a Full–Mesh Network Based on a Cyclic–Frequency AWG. In *Proc. of OFC 2001 Technical Digest, paper ThG5*, Anaheim, CA, March 2001.

[115] N. P. Caponio, A. M. Hill, F. Neri, and R. Sabella. Single–Layer Optical Platform Based on WDM/TDM Multiple Access for Large–Scale 'Switchless' Networks. *European Trans. on Telecomm.*, 11(1):73–82, Jan./Feb. 2000.

[116] A. Bianco, E. Leonardi, M. Mellia, and F. Neri. Network Controller Design for SONATA — A Large–Scale All–Optical Passive Network. *IEEE Journal on Selected Areas in Communications*, 18(10):2017–2028, October 2000.

[117] M. Maier and A. Wolisz. Demonstrating the Potential of Arrayed–Waveguide Grating Based Single–Hop WDM Networks. *Optical Networks Magazine*, 2(5):75–85, September 2001.

[118] B. Mukherjee. WDM–Based Local Lightwave Networks Part II: Multihop Systems. *IEEE Network*, 6(4):20–32, July 1992.

[119] M. Maier. *Metropolitan Area WDM Networks – An AWG Based Approach.* Norwell, MA: Kluwer Academic Publishers, 2003.

[120] M. Maier, M. Reisslein, and A. Wolisz. Towards Efficient Packet Switching Metro WDM Networks. *Optical Networks Magazine*, 3(6):44–62, Nov./Dec. 2002.

[121] C. Fan, M. Maier, and M. Reisslein. The AWG||PSC Network: A Performance Enhanced Single–Hop WDM Network with Heterogeneous Protection. In *Proc. of IEEE Infocom '03*, pages 2279–2289, San Francisco, March 2003.

[122] J. J. O. Pires, M. J. O'Mahony, N. Parnis, and E. Jones. Size limitations of a WDM ring network based on arrayed-waveguide grating OADMs. In *Proceedings of IFIP International Conference on Optical Network Design and Modelling (ONDM)*, pages 71–79, Paris, France, February 1999.

[123] J. J. O. Pires, M. J. O'Mahony, N. Parnis, and E. Jones. Scaling limitations in full–mesh WDM ring networks using arrayed–waveguide grating OADMs. *IEEE Electronics Letters*, 35(1):73–75, January 1999.

[124] N. Antoniades, K. Ennser, V. L. da Silva, and M. Yadlowsky. Computer simulation of a metro WDM ring network. In *Digest of the IEEE LEOS Summer Topical Meetings*, pages IV19–IV20, July 2000.

[125] H. I. Saleheen. Impact of closed cycle ASE accumulation on optical network performance. In *Proceedings of 14th Annual Meeting of the IEEE Lasers and Electro-Optics Society (LEOS)*, pages 32–33, November 2001.

[126] A. Yu, M. J. O'Mahony, and A. M. Hill. Transmission limitation of all-optical network based on $N \times N$ multi/demultiplexer. *Electronics Letters*, 33(12):1068–1069, June 1997.

[127] M. Herzog. Design und Untersuchung von optischen Metropolitan Area Networks (MANs) unter Berücksichtigung von neuen MAC Protokollen, September 2002. Masters Thesis, Institute of Photonics, Technical University Berlin.

[128] M. Cerisola, T. K. Fong, R. T. Hofmeister, L. G. Kazovsky, C.-L. Lu, P. Poggiolini, and D.J.M. Sabido IX. CORD — a WDM optical network: Control mechanism using

subcarrier multiplexing and novel synchronization solutions. In *Proceedings of ICC '95*, pages 261–265, Seattle, WA, June 1995.

[129] I. Chlamtac, A. Fumagalli, L.G. Kazovsky, P. Melman, W.H. Nelson, P. Poggiolini, M. Cerisola, A.N.M.M. Choudhury, T.K. Fong, R.T. Hofmeister, C.-L. Lu, A. Mekkittikul, D.J.M. Sabido IX, C.-J. Suh, and E.W.M. Wong. CORD: contention resolution by delay lines. *IEEE Journal on Selected Areas in Communications*, 14(5):1014–1029, June 1996.

[130] R. T. Hofmeister, C.-L. Lu, M.-C. Ho, P. Poggiolini, and L. G. Kazovsky. Distributed slot synchronization (DSS): A network-wide slot synchronization technique for packet-switched optical networks. *IEEE/OSA Journal of Lightwave Technology*, 16(12):2109–2116, December 1998.

[131] R. Ramaswami and K. Sivarajan. *Optical Networks: A Practical Perspective*. Morgan Kaufmann, 2002.

[132] K. Park and W. Willinger, editors. *Self–Similar Network Traffic and Performance Evaluation*. John Wiley & Sons Inc., 2000.

[133] I. Cidon and Y. Ofek. MetaRing – A Full–Duplex Ring with Fairness and Spatial Reuse. *IEEE Transactions on Communications*, 41(1):110–120, January 1993.

[134] J. Chen, I. Cidon, and Y. Ofek. A Local Fairness Algorithm for the MetaRing and its performance study. *IEEE Journal of Selected Areas in Communications*, 11:1183–1192, October 1993.

[135] K. Imai, T. Ito, H. Kasahara, and N. Morita. ATMR: Asyncronous Transfer Mode Ring Protocol. *Computer Networks and ISDN Systems*, 26:785–798, March 1994.

[136] H.-S. Yang, M. Herzog, M. Maier, and M. Reisslein. Metro WDM Networks: Performance Comparison of Ring and Star Topologies. Technical report, Arizona State University, Department of Electrical Eng., available at `http://www.fulton.asu.edu/~mre`, January 2004.

[137] C. Zhou and Y. Yang. Wide–sense Nonblocking Multicast in a Class of Regular Optical WDM Netowrks. *IEEE Trans. on Comm.*, 50(1):126–134, January 2002.

[138] M. Maier, M. Scheutzow, and M. Reisslein. The Arrayed–Waveguide Grating Based Single–Hop WDM Network: An Architecture for Efficient Multicasting. Technical report, Arizona State University, Telecommunications Research Center, available at `http://www.fulton.asu.edu/~mre`, December 2002.

[139] I. M. White, M. S. Rogge, K. Shrikhande, and L. G. Kazovsky. A Summary of the HORNET Project: A Next-Generation Metropolitan Area Network. *IEEE Journal on Selected Areas in Communications*, 21(9):1478–1494, November 2003.

[140] H.-S. Yang, M. Maier, M. Reisslein, and W. M. Carlyle. A Genetic Algorithm based Methodology for Optimizing Multi-Service Convergence in a Metro WDM Network. *IEEE/OSA Journal of Lightwave Technology*, 21(5):1114–1133, May 2003.

[141] C. Fan, M. Maier, and M. Reisslein. The AWG||PSC Network: A Performance-Enhanced Single-Hop WDM Network With Heterogeneous Protection. *IEEE/OSA Journal of Lightwave Technology*, 22(5):1242–1262, May 2004.

[142] R. Ramaswami and K. N. Sivarajan. *Optical Networks – A Practical Perspective*. Morgan Kaufmann, 2001. Second Edition.

[143] A. Hopper and R. C. Williamson. Design and Use of an Integrated Cambridge Ring. *IEEE Journal on Selected Areas in Communications*, SAC–1(5):775–784, Nov. 1983.

[144] F. E. Ross. FDDI - A Tutorial. *IEEE Communications Magazine*, 24(5):10–17, May 1986.

[145] P. Yuan, V. Gambiroza, and E. Knightly. The IEEE 802.17 Media Access Protocol for High-Speed Metropolitan-Area Resilient Packet Rings. *IEEE Network*, 18(3):8–15, May/June 2004.

[146] X. Zhang and C. Qiao. An Effective and Comprehensive Approach for Traffic Grooming and Wavelength Assignment in SONET/WDM Rings. *IEEE/ACM Transactions on Networking*, 8(5):608–617, October 2000.

[147] E. Modiano and R. Berry. Using grooming cross–connects to reduce ADM costs in SONET/WDM ring networks. In *Proc., OFC, paper WL3*, 2001.

[148] J. Eberspächer and L. Heiss. Ein breitbandiges Lokales Netz mit kombinierter Ring-Stern-Struktur. *ntzArchiv*, 10(9):247–257, 1988.

[149] M. S. Goodman. Multiwavelength Networks and New Approaches to Packet Switching. *IEEE Communications Magazine*, 27(10):27–35, October 1989.

[150] M. S. Goodman. Optical Networks: New Approaches to Interconnection and Switching. In *Conference Digest, LEOS Summer Topical on Optical Multiple Access Networks*, pages 13–14, July 1990.

[151] A. M. Hill, M. Brierley, R. M. Percival, R. Wyatt, D. Pitcher, K. M. I. Pati, I. Hall, and J.-P. Laude. Multiple-Star Wavelength-Router Network and Its Protection Strategy. *IEEE Journal on Selected Areas in Communications*, 16(7):1134–1145, Sept. 1998.

[152] W.-P. Lin, M.-S. Kao, and S. Chi. The Modified Star–Ring Architecture for High–Capacity Subcarrier Multiplexed Passive Optical Networks. *IEEE/OSA Journal of Lightwave Technology*, 19(1):32–39, January 2001.

[153] F. Kastenholz. A Core Standard For Transmission of IP Packets Over IEEE 802.17 (Resilient Packet Ring) Networks. *Internet Draft, draft–ietf–iporpr–core–00.txt*, Dec. 2002.

[154] M. Maier and M. Reisslein. AWG-Based Metro WDM Networking. *IEEE Communications Magazine*, 42(11):S19–S26, November 2004.

[155] F. Davik, M. Yilmaz, S. Gjessing, and N. Uzun. IEEE 802.17 Resilient Packet Ring Tutorial. *IEEE Communications Magazine*, 42(3):112–118, 2004.

[156] I. P. Kaminow and T. Li, editors. *Optical Fiber Telecommunications*, volume IVB, chapter 8, pages 329–403. Academic Press, 2002.

[157] W. Bux and M. Schlatter. An Approximate Method for the Performance Analysis of Buffer Insertion Rings. *IEEE Transactions on Communications*, COM-31(1):50–55, January 1983.

[158] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot. Packet-Level Traffic Measurements from the Sprint IP Backbone. *IEEE Network*, 17(6):6–16, Nov./Dec. 2003.

[159] T.-H. Wu. Emerging Technologies for Fiber Network Survivability. *IEEE Comm. Mag.*, 33(2):58–74, February 1995.

[160] D. Zhou and S. Subramaniam. Survivability in Optical Networks. *IEEE Network*, 14(6):16–23, Nov./Dec. 2000.

[161] J. Zhang and B. Mukherjee. A Review of Fault Management in WDM Mesh Networks: Basic Concepts and Research Challenges. *IEEE Network*, 18(2):41–48, March/April 2004.

[162] A. Carena, V. DeFeo, J.M. Finochietto, R. Gaudino, F. Neri, C. Piglione, and P. Poggiolini. RingO: An Experimental WDM Optical Packet Network for Metro Applications. *IEEE Journal on Selected Areas in Communications*, 22(8):1561–1571, October 2004.

[163] P.-H. Ho and H.T. Mouftah. Shared protection in mesh WDM networks. *IEEE Comm. Mag.*, 42(1):70–76, January 2004.

[164] P.-H. Ho, J. Tapolcai, and H. T. Mouftah. On achieving optimal survivable routing for shared protection in survivable next-generation Internet. *IEEE Transactions on Reliability*, 53(2):216–225, June 2004.

[165] G. Mohan and C. S. R. Murthy. Lightpath restoration in WDM optical networks. *IEEE Network*, 14(6):24–32, November/December 2000.

[166] A. Narula-Tam, E. Modiano, and A. Brzezinski. Physical Topology Design for Survivable Routing of Logical Rings in WDM-Based Networks. *IEEE Journal on Selected Areas in Communications*, 22(8):1525–1538, October 2004.

[167] C. Ou, J. Zhang, H. Zang, L.H. Sahasrabuddhe, and B. Mukherjee. New and improved approaches for shared-path protection in WDM mesh networks. *IEEE Journal of Lightwave Technology*, 22(5):1223–1232, May 2004.

[168] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee. Survivable WDM Mesh Networks. *IEEE/OSA Journal of Lightwave Technology*, 21(4):870–883, April 2003.

[169] M. Sridharan, M. V. Salapaka, and A. K. Somani. A practical approach to operating survivable WDM networks. *IEEE Journal on Selected Areas in Communications*, 20(1):202–215, January 2002.

[170] D. Xu, Y. Xiong, C. Qiao, and G. Li. Failure protection in layered networks with shared risk link groups. *IEEE Network*, 18(3):36–41, May/June 2004.

[171] J. Wang, L. Sahasrabuddhe, and B. Mukherjee. Path vs. Subpath vs. Link Restoration for Fault Management in IP-over-WDM Networks: Performance Comparisons Using GMPLS Control Signaling. *IEEE Communications Magazine*, 40(11):80–87, November 2002.

[172] Q. Zheng and G. Mohan. Protection approaches for dynamic traffic in IP/MPLS-over-WDM networks. *IEEE Communications Magazine*, 41(5):S24–S29, May 2003.

[173] O. Gerstel and G. Sasaki. Quality of Protection (QoP): A Quantitative Unifying Paradigm to Protection Service Grades. *Optical Networks Magazine*, 3(3):40–50, May/June 2002.

[174] C.V. Saradhi, M. Gurusarny, and L. Zhou. Differentiated QoS for survivable WDM optical networks. *IEEE Communications Magazine*, 42(5):S8–S14, May 2004.

[175] M. Tacca, A. Fumagalli, A. Paradisi, F. Unghvary, K. Gadhiraju, S. Lakshmanan, S. M. Rossi, A. de Campos Sachs, and D. S. Shah. Differentiated reliability in optical networks: theoretical and practical results. *IEEE Journal of Lightwave Technology*, 21(11):2576–2586, November 2003.

[176] K. Wu and L. Valcarenghi A. Fumagalli. Restoration Schemes with Differentiated Reliability. In *Proc., IEEE ICC*, pages 1968–1972, 2003.

[177] O. Gerstel and R. Ramaswami. Optical Layer Survivability: A Services Perspective. *IEEE Communications Magazine*, 38(3):104–113, March 2000.

[178] O. Gerstel and R. Ramaswami. Optical Layer Survivability—An Implementation Perspective. *IEEE J. Sel. Areas in Comm.*, 18(10):1885–1899, October 2000.

[179] Y. Ye, S. Dixit, and M. Ali. On Joint Protection/Restoration in IP-Centric DWDM-Based Optical Transport Networks. *IEEE Communications Magazine*, 38(6):174–183, June 2000.

[180] A. Fumagalli and L. Valcarenghi. IP Restoration vs. WDM Protection: Is There an Optimal Choice? *IEEE Network*, 14(6):34–41, Nov./Dec. 2000.

[181] W. Aiello, S. N. Bhatt, F. R. K. Chung, A. L. Rosenberg, and R. K. Sitaraman. Augmented Ring Networks. *IEEE Transactions on Parallel and Distributed Systems*, 12(6):598–609, June 2001.

[182] W. D. Grover and D. Stamatelakis. Cycle-Oriented Distributed Preconfiguration: Ring-like Speed with Mesh-like Capacity for Self-planning Network Restoration. In *Proc., IEEE International Conference on Communications (ICC)*, pages 537–543, 1998.

[183] G. Shen and W. D. Grover. Extending the p-Cycle Concept to Path Segment Protection for Span and Node Failure Recovery. *IEEE Journal on Selected Areas in Communications*, 21(8):1306–1319, October 2003.

[184] T. Y. Chow, F. Chudak, and A. M. Ffrench. Fast Optical Layer Mesh Protection Using Pre-Cross-Connected Trails. *IEEE/ACM Transactions on Networking*, 12(3):539–548, June 2004.

[185] G. Suwala and G. Swallow. SONET/SDH-Like Resilience for IP Networks: A Survey of Traffic Protection Mechanisms. *IEEE Network*, 18(2):20–25, March/April 2004.

[186] P. Yue, Z. Liu, and J. Liu. High Performance Fair Bandwidth Allocation Algorithm for Resilient Packet Ring. In *Proc., International Conference on Advanced Information Networking and Applications (AINA)*, pages 415–420, March 2003.

[187] X. Zhou, G. Shi, H. Fang, and L. Zeng. Fairness Algorithm Analysis in Resilient Packet Ring. In *Proc., International Conference on Communication Technology (ICCT)*, volume 1, pages 622–624, April 2003.

[188] V. Gambiroza, P. Yuan, L. Balzano, Y. Liu, S Sheafor, and E. Knightly. Design, Analysis, and Implementation of DVSR: A Fair High-Performance Protocol for Packet Rings. *IEEE/ACM Transactions on Networking*, 12(1):85–102, February 2004.

[189] K. Park and W. Willinger, editors. *Self-Similar Network Traffic and Performance Evaluation*. Wiley, 2000.

[190] F. Alharbi and N. Ansari. A Novel Fairness Algorithm for Resilient Packet Ring Networks with Low Computational and Hardware Complexity. In *Proc., IEEE LANMAN*, pages 11–14, April 2004.

[191] C.-G. Liu and J.-S. Li. Improving RPR Fairness Convergence. In *Proc., The 2004 IEEE Asia-Pacific Conference on Circuits and Systems*, pages 469–472, 2004.

[192] H. Fang, P. Wang, D. Jin, and L. Zeng. A New RPR Fairness Algorithm Based on Deficit Round Robin Scheduling Algorithm. In *Proc., 2004 International Conference on Communications, Circuits, and Systems (ICCCAS)*, pages 698–702, 2004.

[193] D. Tsiang and G. Suwala. The Cisco SRP MAC Layer Protocol. Internet RFC 2892, August 2000.

[194] A. Mekkittikul et al. Alladin Proposal for IEEE Standard 802.17, Draft 1.0, November 2001.

[195] RIAS Fairness Reference Model. http://www-ece.rice.edu/networks/RIAS/.

[196] D. Bertsekas and R. Gallager. *Data Networks*. Prentice Hall, 2nd edition, December 1997.