CHAPTER 6

THE FUTURE OPTICAL INTERNET

ANDREA FUMAGALLI, JAVIER ARACIL, AND LUCA VALCARENGHI

6.1 INTRODUCTION

As of today, the Internet is the most versatile and widespread form of (virtually) free public telecommunication service. Beside constituting the universal bridge between continental, national, regional, and local networks, the term Internet has become the synonym of email, World Wide Web (WWW) surfing, *multimedia applications,* and many other services.

The Internet finds its origins in two U.S. government funded research networks, namely ARPANET (late 1960s) and NSFNET (middle 1980s). By the late 1980s, a number of subnetworks worldwide (e.g., the European IP backbone EuropaNET and EBONE [1,2]) were connected to these two networks. While the Internet transport and network layer protocols are still based on the original TCP/IP suite, developed in 1974 [3,4], evolution of the transmission medium technologies has radically changed the so-called *Internet link layer* (i.e., the physical and link layer of the ISO-OSI model). Early ARPANET and NSFNET connections were running on lines leased from telephone companies at a transmission rate of 56 kb/s. With the advent of low-loss fiber-based links, transmission rates were upgraded to 45 Mb/s. It is only during the latest few years that the combined skyrocketing increase in the number of Internet users and the augmented number of services available on the Internet have created an unprecedented demand for additional bandwidth, whose exponential growth in the long run can only be coped with by the latest advances in optical technology. These advances include *wavelength division multiplexing,* or (*WDM*), which potentially provides a per-fiber aggregate bandwidth in the Tb/s range. The impact

This work was supported in part by the NSF under contract # ANI-0082085.

Managing IP Networks. Edited by Saleh Aidarous and Thomas Plevyak
ISBN 0-471-39299-5

of such a revolution in Internet history is so widely recognized that a name has been created to indicate the outcome of such modernization of the Internet: the *next-generation Internet* (*NGI*). Since it is commonly recognized that the NGI originates from the high transmission bandwidth provided by the optical medium, in the rest of the chapter the terms NGI and *Optical Internet* (OI) will be used interchangeably.

The aim of this chapter is to describe the basic principles of the OI and to identify its potential benefits and design challenges. The chapter consists of four sections.

Section 6.2 describes the optical layer and presents some of the enabling optical components. The concepts of first-generation optical network (FGON) (e.g., SONET/SDH) and second-generation optical network (SGON) (e.g., WDM) are introduced. In the latter case optical circuits, or lightpaths, are established between network nodes to create all-optical transparent connections, and thus to circumvent the so-called electronic bottleneck of FGON. Four alternative approaches for designing SGON are described: static and semistatic lightpath-based networks, dynamic lightpath-based networks, optical packet-switching-based networks, and optical burst switching based networks.

Section 6.3 illustrates the evolution of the Internet layering from the X.25-based solution to IP over ATM, packet over SONET (POS), and Gigabit Ethernet. Network performance, complexity, and costs of these solutions are discussed and compared with one another, with the intent to identify efficient mechanisms to transport IP packets over the optical layer. The simplified two-layer architecture known as "IP over WDM" is then presented. Flow-switching solutions, such as MPLS and multiprotocol lambda switching (MPλS or generalized MPLS) are described that allow management of the WDM layer from the IP layer using standard mechanisms and protocols.

Section 6.4 discusses the impact of traffic self-similarity on the optical network engineering process. In particular, the huge bandwidth potentially available in optical fiber poses new challenges, e.g., efficient network dimensioning, which are of no concern in the conventional relatively "low speed" Internet scenario. A thorough analysis at the TCP connection level is provided in this section with the scope of determining efficient ways to manage the optical bandwidth dynamically.

Section 6.5 identifies some of the key challenges encountered in the realization of the OI. The challenges reviewed in this last section include efficient traffic engineering in the OI, adequate network resilience schemes at both the IP layer and WDM layer, and coordination between resilience schemes available at both the IP and the WDM layer.

6.2 OPTICAL NETWORK TECHNOLOGIES

Since its dawn in the 1960s [5], fiber optics technology has undergone a continuous evolution, with innovative devices becoming available on the market on a yearly basis. As illustrated in this section, the advent of these devices has fostered the development of optical networks and their evolution from first- to second-generation architectures. An overview of the key enabling optical components is first presented, followed by a description of a number of alternative network architectures based on such components. The description of the components is kept at a high level, giving the reader a global picture, rather than providing a comprehensive description of each single component. The reader who may be interested in details of specific optical components is referred to a number of comprehensive books on the subject [5–11].

6.2.1 Optical Components

Discovery, in the 1960s, that an inexpensive thin wire of glass is capable of propagating a huge quantity of data with small signal attenuation paved the way to a long—still lasting—revolution in the way telecommunications and data networks are designed.

Fiber optics is a transmission medium made of silica (SiO_2) in which the optical signal remains confined and guided by total internal reflection. Internal reflection is achieved by means of a two-region section of the fiber: the core and the cladding. The two regions are differently doped (by means of injected impurities), leading to different refractive indices. As a result, the light beam in the core that propagates with an angle of incidence higher than the critical angle—with respect to the boundary surface between the two regions—is completely reflected and confined within the core. Fiber optics can be either *multimode*—when more than one propagation mode is possible—or *single-mode*—when only one mode propagates in the fiber. The former type of fiber is characterized by modal dispersion and low installation cost. The latter does not present modal dispersion and is more expensive to install than the former, due to the accurate signal coupling required between fibers, transmitters and receivers. Typically, three spectral regions, referred to as *windows*, are defined that have low signal attenuation in the fiber. The *first window* is in the 800–900-nm interval range—optical frequencies are conventionally characterized in terms of wavelengths—and it is generally used to transmit multimode signals. The *second* and *third windows* are, respectively, in the 1240–1340-nm and 1500–1650-nm regions, and are typically used to transmit single-mode signals. The bandwidth potentially available in each window is about 20 THz, but the actual bandwidth available for data transmission is considerably less and limited by a number of factors. Among these factors, one can enumerate the limited bandwidth of other optical components, such as transmitters, receivers, and optical amplifiers, and the limited electronic processing speed at the network nodes, the so-called "electronic bottleneck" of the first-generation optical networks.

An optical signal is generated by either a *light-emitting diode* (LED) or a *laser diode* (*laser*). The main difference between LEDs and lasers is the spectral density of the emitted light beam. The LED signal has a large spectrum and is used in conjunction with multimode fiber. LEDs are relatively inexpensive and used in networks with low data rates (e.g., FDDI), short distances, and limited power budget (e.g., LANs, fiber to the home, fiber to the curb, coarse WDM (CWDM), access networks). Instead, the laser emits a light beam with power concentrated in a narrow bandwidth. Its cost grows with its selectivity in bandwidth. Lasers are currently expensive and used in conjunction with single-mode fiber to deploy DWDM systems, in which the wavelength or channel density is high. Lasers can be designed either to operate at one fixed wavelength or to be tunable. In the latter case there is a trade-off between the laser tuning speed and its tuning range, i.e., fast tunable lasers have a small tuning range and vice versa. At the receiver node, *photodetectors* are used to convert the optical signal back into the electronic domain. When the received optical signal is weak, *avalanche photodiodes* (APDs) can be used, in which the reception of a photon generates an avalanche of photons that are easier to detect.

Optical amplifiers became commercially available in the early 1990s to regenerate the optical signal without requiring the two phase O-E-O conversion plus electrical signal amplification that was previously utilized. Without any doubt optical amplifiers constitute one of the most important milestones in the history of optical communications. By avoiding O-E-O conversion, an optical signal can be regenerated directly in the optical domain, thus circumventing the limited bandwidth of electronic circuitry. For example, a single

optical amplifier can amplify a number of WDM channels simultaneously, with an aggregate throughput that by far exceeds the electronic maximum bandwidth.

Optical amplifiers can be divided into four categories: semiconductor optical amplifiers (SOAs), x-doped fiber amplifiers (xDFAs), linear optical amplifiers (LOAs), and nonlinear amplifiers (e.g., Raman amplifiers (RAs)). SOAs exploit the same principle used by lasers. An SOA is maintained under the lasing threshold in order to utilize its amplification capabilities without inducing lasing. SOAs can work in both the second and the third spectral window, offering a broadband gain characteristic and a total bandwidth of about 100 nm. Gain fluctuation, polarization dependency, high coupling loss with the fiber, and their inherent nonlinearity represent their main drawbacks. These amplifiers are suitable for single channel amplification. Doped fiber amplifiers consist of a segment of fiber optics that is doped with rare-earth chemical elements. Their behavior is based on the principle of *stimulated emission,* by which the electrons of the doped fiber, stimulated by the arriving signal photons, emit light. As a result, the incoming optical signal is amplified as it propagates through the doped fiber and absorbs the power of a *pump* signal launched into the fiber. Widely used, the *erbium-doped fiber amplifier* (*EDFA*) [12] is employed in the *short wavelength band* (*S* = 1450–1530 nm), in the *conventional wavelength band* (*C* = 1530–1570 nm), and in the *long wavelength band* (*L* = 1570–1620 nm) of the third spectral window [13]. For example, with a total available bandwidth of about 70 nm, this amplifier can amplify 80 bidirectional channels with 100-GHz spacing (0.8 nm), each transmitting at 10 Gb/s. Interesting properties of EDFA are high gain (e.g., 18 dB fiber-to-fiber), no crosstalk among the amplified channels, small noise figure, and low coupling loss. Its drawbacks are gain fluctuations, which are a function of the channel (wavelength) position in the spectrum, and large physical dimensions. Another doped fiber amplifier is the Praseodymium-Doped Fiber Amplifier (PDFA) that is employed in the second spectral window at 1300 nm. LOAs consist of the integration of an amplifier and a vertical cavity surface emitting laser (VCSEL) [14]. The circulating optical power of the VCSEL overlaps with the amplifier waveguide and permits the optical amplifier maintaining a constant gain to be linearized. They operate across the C wavelength band. LOAs have been shown to have a small gain transient (i.e., low dependence on the number of wavelengths amplified), almost constant BER in the presence of a varying number of amplified wavelengths, and small interchannel crosstalk. These properties make them suitable for metropolitan and access optical networks, where optical amplifiers may need to operate in the presence of dynamic lightpaths that operate at different data rates and are frequently set up and torn down. Nonlinear amplifiers resort to nonlinear effects of the fiber and require high power pump signal(s). In the most common nonlinear amplifier, the RA, the power of the lower wavelength (higher energy) pump is partially transferred to the higher wavelength (lower energy) data signal via excitation of fiber vibrational modes [15]. The pump's wavelength and power determine, respectively, the data wavelengths that are amplified and the amplification gain. The 3-dB gain bandwidth for a single pump RA is about 5 THz (at 1545 nm). To obtain broader gain spectra, multiple pump signals can be used in the same fiber. RAs yield high gain, are polarization independent, have a low noise figure, allow tight channel spacing at high transmission rate (e.g., 40 Gb/s with 100-GHz spacing), and can operate inside and outside the C- and L-bands of the fiber third window.

Another important optical component is the *wavelength* (or frequency) *converter*. A wavelength converter translates the wavelength of an optical signal to a desired value. In its simplest implementation, a wavelength converter consists of an SOA [6]. In this real-

ization the wavelength converter exploits the SOAs gain saturation effect. Two signals at two different wavelengths are launched into the SOA input. One is the data signal whose wavelength is to be converted. The other is a constant signal at high power that is transmitted on the wavelength that the other information-bearing signal must be converted to. The combined power of both signals is chosen to saturate the amplifier. Consequently, the gain experienced by the incoming constant signal is a function of the power on the data signal: intervals with high power in the incoming data signal generate low power intervals in the outgoing (originally constant) signal, and vice versa. As a result, the bit pattern of the information-bearing signal (filtered out at the SOA output) is converted to the other wavelength with a reverse bit coding. With current technology, wavelength converters are expensive components.

The components described so far constitute the building blocks of a point-to-point transmission system. In order to provide networking functionalities, optical nodes must be designed that are able to switch and route the optical signal. Both *optical add/drop multiplexers* (OADMs) and *optical cross connects* (OXCs) belong to this category. An OADM node is capable of extracting one arriving optical signal from the network and replacing it with a newly injected one. While doing so, the other optical signals in the fiber are let through the OADM node unaltered. For example, in a WDM OADM, one wavelength is dropped and added, while the other wavelengths in the fiber are optically routed through the node. An OXC is an optical node that can perform routing functions on optical signals. OXCs can be built in different ways (see [5,7,8]) but their functioning principles remain the same. Depending on their routing capabilities, OXCs are usually subdivided in *fiber optical cross connects* (F-OXC), *wavelength translating optical cross connects* (WT-OXC), and *wavelength routing optical cross connects* (WR-OXC) (see Figures 6.1, 6.2, 6.3, respectively). The F-OXC is able to switch the signals of one entire fiber from the input port to the desired output port. Individual wavelengths can be added and dropped at the F-OXC. A particular implementation of this OXC is based on *microelectromechanical switch* (MEMS) [13,16–18]. The WT-OXCs and the WR-OXCs are able to switch both single fibers and single wavelengths from the input to the desired output. In the WT-OXC (WR-OXC) the wavelength of the incoming optical signal can (cannot) be converted. Recently *waveband cross connect* (WBXC) have been proposed [19]. WBXCs permit transparent switching an aggregated set of wavelengths (*wavebands*) that share the same links along part of their paths. Switching wavebands reduces the complexity of OXCs [20] but increases the complexity of routing and wavelength assignment algorithms. The configu-

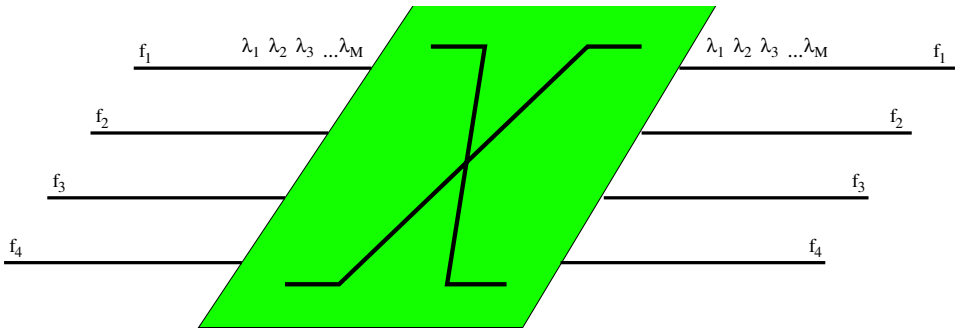

Figure 6.1 F-OXC.

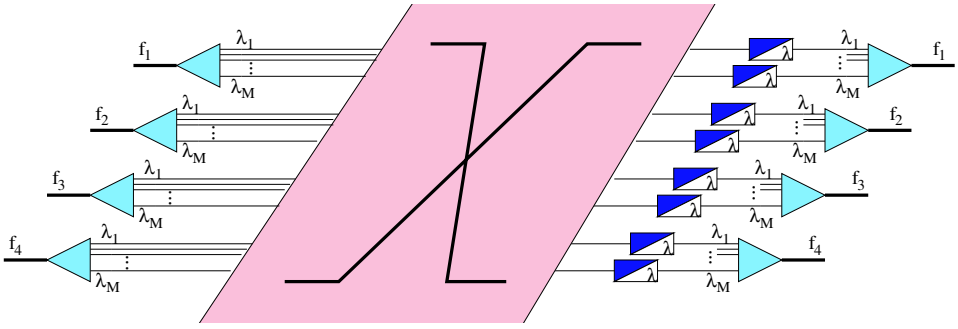

Figure 6.2 WT-OXC.

ration of an OXC can be either fixed or dynamically changed by means of an (electronic) controller (in the latter case, the component is more commonly referred to as an "optical switch").

As already mentioned, the system obtained by interconnecting the aforementioned optical components is not a mere point-to-point transport medium anymore. It presents characteristics and functions that are typical of the network layer of the ISO/OSI model. Thus, the all-optical networks (or second-generation optical networks) offer networking, multiplexing, and transport capabilities altogether. These functions have been standardized by the ITU-T to form a layered model, referred to as the Optical Layer (OL).

6.2.2 The Optical Layer

In its Recommendation G.872 [21], the ITU-T describes the functional architecture of the *optical transport network* (OTN) commonly referred to as the OL. By means of WDM, the OL provides simple routing functions that create optical circuits, or *lightpaths*, across the network. A lightpath is a point-to-point all-optical connection between physical nodes that need not be adjacent [22]. Conceptually, a lightpath is a service that belongs to the network layer of the ISO/OSI model.

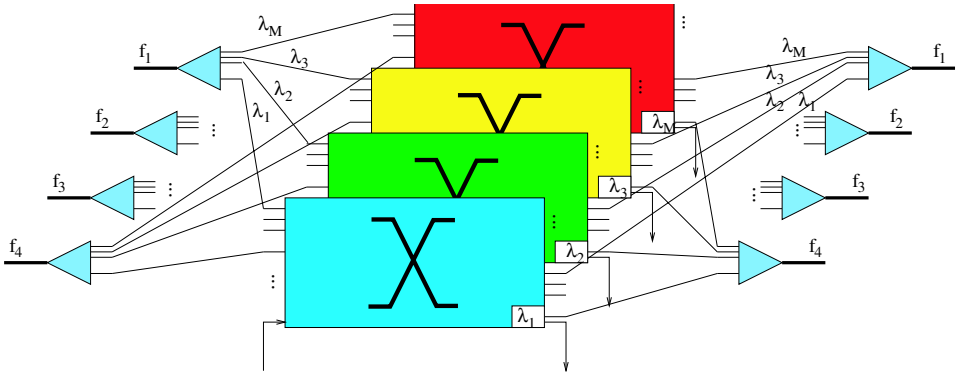

Figure 6.3 WR-OXC.

As part of the OL, three sublayers have been standardized by ITU-T:

1. Lightpath or Optical Channel (OCh)
2. Optical Multiplex Section (OMS)
3. Optical Transmission Section (OTS).

In [21] the OL is described from a network level viewpoint, taking into account an optical network layered structure, client characteristic information, client/server layer associations, networking topology, and layered functionalities that provide optical signal transmission, multiplexing, routing, supervision, performance assessment, and network survivability.

The layered structure of the optical transport network, depicted in Figure 6.4, shows the OCh section, OMS, and OTS sublayers.

6.2.2.1 Optical Channel Section This sublayer provides end-to-end networking in the form of optical channels (lightpaths) for the transparent transmission of a client's data in varying desired formats, e.g., SDH STM-N, PDH 565 Mb/s, cell-based ATM, digital wrapper, IP/MPLS. To provide end-to-end networking, the following capabilities are included in the OCh sublayer:

1. OCh connection rearrangement for flexible network routing.
2. OCh overhead processing that ensures integrity of the OCh-adapted information.
3. OCh supervisory functions that enable network level operations and management functions, such as connection provisioning, exchange of QoS parameters, and network survivability (end-to-end protection and restoration).

Network components related to this sublayer include optical line terminals (OLTs) (the interface between the electronic layer and the optical layer), optical transmitters (laser, diodes), and optical receivers (photodiodes).

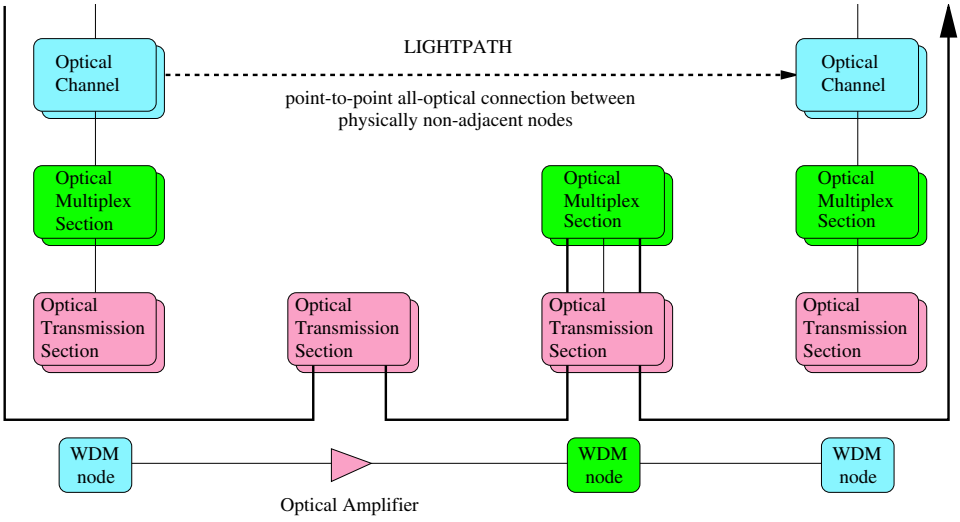

Figure 6.4 The ITU-T G.872 optical layer [7].

6.2.2.2 Optical Multiplex Section This sublayer provides the functionalities required to manage a multiwavelength optical signal. (Notice that a "multiwavelength" signal includes as a special case the single-wavelength fiber.) The capabilities of this layer include:

- OMS overhead processing that ensures integrity of the multiwavelength OMS-adapted information.
- OMS supervisory functions that enable section level operations and management functions, such as multiplex section survivability (line protection and restoration).

Network components related to this sublayer include OADMs, OXCs (F-OXC, WT-OXC, WR-OXC), wavelength converters, optical switches, passive optical couplers/splitters (e.g., passive hub).

6.2.2.3 Optical Transmission Section This sublayer provides the functionalities required to handle the transmission of the optical signal on various types of optical medium. The capabilities of this sublayer include:

- OTS overhead processing that ensures integrity of the OTS-adapted information.
- OTS supervisory functions that enable section level operations and management functions, such as survivability of the transmission section.

The network components related to this sublayer are primarily optical amplifiers and regenerators. The regeneration process can be of three types: 1R, when signal amplification and equalization (frequency, dispersion) are performed; 2R, when 1R, digital reshaping of the signal, and noise suppression take place; and 3R, when 2R regeneration and pulse retiming take place. Regeneration can be achieved using either optical amplifier (OA) (1R) or O-E-O conversion and electronic processing of the signal (1R, 2R, 3R).

6.2.3 All-Optical Network Architectures

Optical networks are often divided into two classes: first-generation optical networks and second-generation optical networks. In first-generation optical networks the optical signal undergoes O-E-O conversion at each node that it encounters on the route from source to destination. First-generation solutions include conventional telephone standards, e.g., SDH, SONET, and more recently proposed standards, e.g., Gigabit Ethernet. In second-generation optical networks (commonly referred to as "all-optical networks") the optical signal is optically routed at the network node, and does not require O-E-O conversion until it is received at the destination. A few promising commercial products are becoming available with all-optical features, e.g., WDM rings with OADM.

It is important to understand the potential advantages of all-optical networks when compared to first-generation optical networks [13]. All-optical networks provide transparency of the optical signal, hence, they are capable of supporting multiple protocols on the same optical transport infrastructure. Optical transparency eases the migration from one protocol to another because no major changes in the physical transport network are necessary. In all-optical networks, services such as protection and restoration switching can be provided directly at the optical layer, and made available to all higher protocols running in the network, including those protocols that do not have built-in survivability

features. In all-optical networks, a significant cost reduction can be potentially achieved by reducing the amount of electronic circuitry and line terminals required in the network (the so-called electronic bottleneck of first-generation optical networks).

The significant potential advantages of all-optical networks make them one of the most promising networking solution for the next-generation Internet. This section will therefore focus on various all-optical network architectures that can be implemented using the optical components presented in Section 6.2.1. Readers interested in a survey of (the more conventional) first-generation optical networks are referred to [5,7,23].

All-optical networks can be classified in fours categories: static and semistatic lightpath networks, dynamic lightpath networks, optical packet-switching networks, and optical burst switching networks.

6.2.3.1 Static and Semistatic Lightpath Networks Static and semistatic lightpath networks are the simplest SGON architecture. These networks, sometime referred to as "wavelength routed networks," provide point-to-point wavelength paths (lightpaths) between network nodes—e.g., IP routers—that need not be physically adjacent—e.g., directly connected by a cable. In a way that is similar to ATM or frame relay permanent virtual circuits, such static lightpaths can be used to transport data between gigabit routers that perform packet or cell forwarding in the electronic domain.

Wavelength routed networks circumvent the electronic bottleneck due to O-E-O conversion at every network node of FGON. By means of a lightpath, the connection from source to destination remains in the optical domain along its entire path. The only places where the transmitted signal is converted from and to electronics are, respectively, the source and the destination. For this reason, this approach is sometimes referred to as "single-hop" networking. The creation of lightpaths permits to build a desired logical topology (where the lightpaths constitute the links) on top of the physical topology (where the cables constitute the links). By creating a logical topology, network connectivity can be increased arbitrarily. In addition, the virtual topology can be redesigned or updated without requiring the huge investments that are necessary to modify the existing cabling of the physical topology.

The key components utilized in static and semistatic lightpath networks are OADMs and OXCs. Depending on their characteristics, OADMs and OXCs can be either configured during their production (static network) or reconfigured multiple times during the lifetime of the network (semistatic network). Manual reconfiguration is typically required in the semistatic network, to provision new lightpaths on a per-month or per-year basis.

The fundamental problem to be addressed in wavelength routed networks is the *routing and wavelength assignment* (RWA) problem. The RWA problem consists of determining for each required lightpath, a path across the physical topology and a wavelength to be used to establish the lightpath. Clearly, the cost of providing a desired logical topology in a given physical topology is affected by the algorithm used to solve the RWA problem. For example, a cost-effective solution will minimize the number of wavelengths required in the fiber to provide the desired logical topology.

In general, two scenarios are defined for the RWA problem. In the first scenario *wavelength continuity* is required along the lightpath. This is the scenario in which wavelength converters are not available in the network. In this case, once chosen, the wavelength of the lightpath cannot be changed from fiber to fiber. The RWA problem with wavelength continuity constraint is NP-complete [22]. In the second scenario, the wavelength continuity constraint is removed, and the RWA problem is greatly simplified. This is the sce-

nario in which wavelength converters are available at the network nodes, and can be used to change the wavelength of a lightpath along its path. However, it must be remembered that wavelength converters are, at the moment, expensive components that significantly increase the overall network cost.

Besides the single-hop approach, *multihop* optical networks have been proposed. Multihop optical networks permit routing of a connection using a concatenation of multiple lightpaths, allowing O-E-O conversion of the optical signal at some "selected" intermediate nodes. Two advantages of multihop networking are: (1) reducing the span of the individual lightpath to better cope with transmission impairments, e.g., fiber chromatic dispersion [24]; and (2) performing electronic traffic multiplexing at the nodes where O-E-O conversion takes place [25]. By multiplexing multiple connections together at selected network nodes, the multihop network yields more bandwidth-efficient solutions than does the single-hop approach.

Figure 6.5 plots the network cost of a WDM ring network obtained using three approaches: First-generation optical network, single-hop network, and multihop network. The network cost is a function of γ, which represents the cost ratio between a wavelength mile and optical terminal (OT) cost. When $\gamma = 0$ the terminal cost is predominant. When $\gamma = 1$ the per mile wavelength cost is predominant. Network cost is normalized to the cost of the multihop network. The trade-off between first-generation optical networks and the single-hop network is clearly visible. The advantage of the multihop network over the single-hop network is documented in the figure. More details on this subject can be found in [25].

Another important aspect of static and semistatic lightpath networks is *survivability*. This problem is considered of paramount importance, because a sudden network fault can disrupt revenues of both network providers and network users at the same time. In general, a network is referred to as "survivable" if it provides some ability to restore ongoing connections in the event of a catastrophic failure of a network component, such as a fiber cut. Several approaches can be used to guarantee optical layer survivability [26]. Relevant parameters of a survivable network are fast recovery time, contained network resource redundancy, and simplicity of the recovery scheme.

A simple and fast protection scheme is represented by the self-healing WDM ring [26]. This scheme provides optical protection against any single-cable failure in the ring

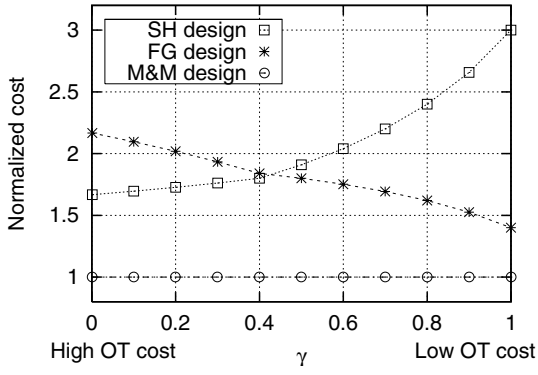

Figure 6.5 Trade-off between first-generation (FG), single-hop (SH), and multihop (MH) WDM ring design.

topology. By deflecting optical signals into a second counterpropagating fiber, the WDM ring can circumvent any single cable interruption or node fault. A recent study [27] illustrates the use of multiple WDM rings in an arbitrary topology. In this approach, a given arbitrary topology is covered with WDM rings. Lightpaths are then routed across the WDM rings to create the desired virtual topology. In this manner, if a cable fails, all lightpaths originally routed across that cable are rerouted using the WDM ring, which covers the interrupted cable. A lightpath may require a cascade of rings to reach the destination. The effect of the ring size (number of nodes) on the network cost is indicated in Figures 6.6 and 6.7. Two cost factors are considered: the total wavelength mileage required to create the virtual topology and the necessary spare resources in the WDM rings as well as the number of crossconnect ports that are required to switch a lightpath from one ring to another. As shown in Figure 6.6, the number of crossconnect ports decreases as the maximum ring size increases. Ring size is defined as the number of nodes connected to the same ring. The required wavelength mileage may be adversely affected by increasing ring size, as shown in Figure 6.7. With appropriate optimization techniques, however, it is possible to mitigate resulting wavelength mileage increase.

6.2.3.2 Dynamic Lightpath Networks With data traffic becoming predominant over voice traffic, bandwidth flexibility is becoming a key feature in high-speed networks. A way to obtain such bandwidth flexibility in the optical layer is to resort to dynamic lightpath provisioning in which lightpaths are set up and torn down on demand. We refer to this type of network as *dynamic lightpath* networks. In such networks, lightpath re-

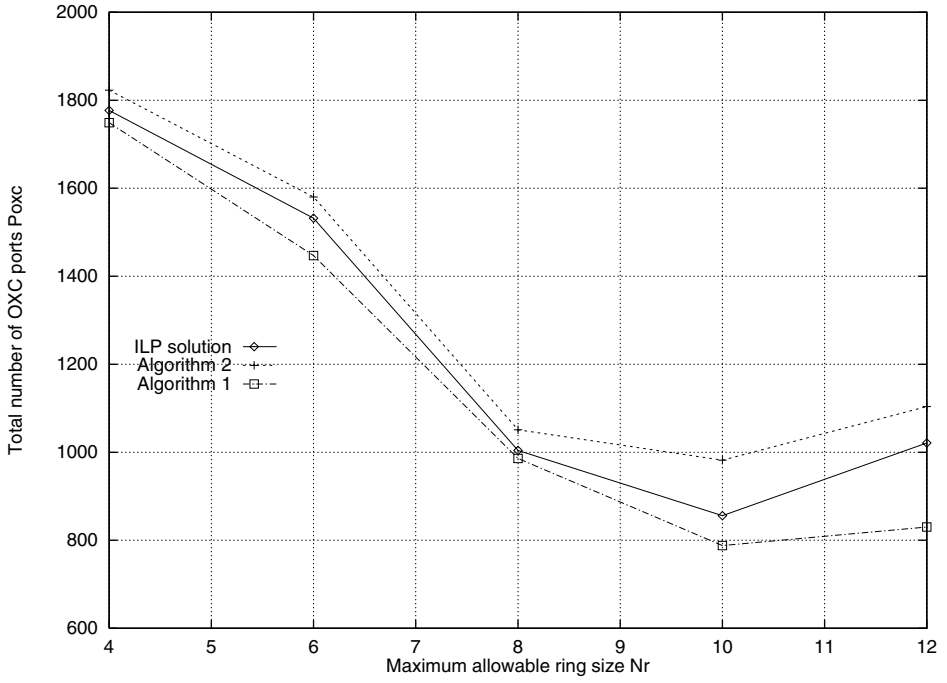

Figure 6.6 Number of OXC ports required.

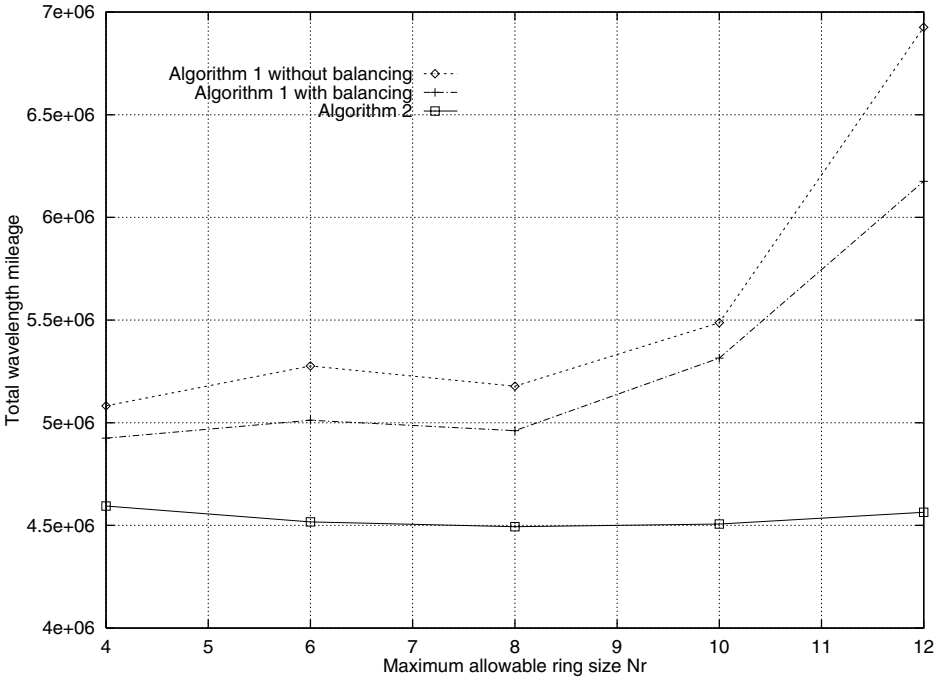

Figure 6.7 Total wavelength mileage required.

quests are expected to be generated frequently and to last for intervals of hours, minutes, and even seconds.

Similar to wavelength routed networks, dynamic lightpath networks offer the advantage of circumventing the electronic bottleneck of O-E-O at the transit nodes once the lightpath is set up. In addition, it is now possible to continuously adjust the logical network topology to best serve varying traffic patterns. The latter feature seems to be particularly suitable for today's Internet traffic. Dynamic lightpath networks are also expected to increase network utilization. When a lightpath is underutilized, it will be taken down, thus releasing some network resources that may be used more efficiently by other newly generated lightpaths.

Enabling technologies for dynamic lightpath networks include tunable transmitters, tunable receivers, and switching capabilities of both OADMs and OXCs. In order to yield satisfactory performance, the tuning and switching latencies required by these devices must be a fraction of the lightpath lifetime.

As in the case of wavelength routed networks, various algorithms for solving the RWA problem in the presence of dynamic lightpaths have been proposed. These can either require wavelength continuity or allow a lightpath to occupy distinct wavelengths on distinct network links [28]. Typical parameters that are optimized when creating lightpaths dynamically are *blocking probability*—the probability that the lightpath request is blocked due to unavailable resources—*throughput, setup delay,* and *fairness*.

Besides devising the appropriate RWA algorithm, a control protocol must be implemented to set up and tear down lightpaths. Two approaches are possible: centralized and distributed. In the centralized approach lightpath requests are sent to a controller node that

has knowledge of the actual network configuration and decides whether resources are available for setting up the requested lightpath. In the distributed approach network nodes dynamically set up lightpaths utilizing a network state update protocol that discovers the actual network configuration [28,29].

Several solutions have been proposed for the implementation of dynamic lightpath networks. The optical network for regional access using multiwavelength protocols (ONRAMP) project [30–32] consists of a WDM physical ring architecture that reconfigures the network logical topology by dynamically provisioning lightpaths between network nodes. The objective of the ONRAMP project is to explore possible architectures for the access network of the NGI. In the *LightRing* [33] a WDM ring multi-token protocol is proposed that controls the distributed set up and tear down of lightpaths. In the LightRing, access to each wavelength is regulated by a token dedicated to that channel. In a distributed way, nodes can grab a token and gain access to the token-corresponding wavelength. A performance comparison between the centralized and the distributed approach can be found in [34]. Proposals by the IETF are based on adapting the MPLS control plane to control the optical layer, e.g., OADMs and OXCs. This approach is commonly referred to as MPλS, and more recently as GMPLS [35,36]. GMPLS permits to dynamically set up and tear down lightpaths in arbitrary, mesh, and network topologies.

Survivability is another important factor in dynamic lightpath networks. Due to the dynamic behavior of the network, schemes that are flexible and able to adapt themselves to network changes are preferred over fixed protection schemes previously described. Recently, solutions for dynamic provisioning of reliable connections have been proposed [37–43]. In these schemes, upon arrival of a request for setting up (tearing down) a connection, both a primary and a backup path are established (torn down) and resources along the network are reserved (released). Path calculation is based on network status information available at the source node upon arrival of the connection request. When resources for either the working or the protection path are not available, the request is blocked. Efficiency of the scheme (e.g., maximization of resource sharing, low connection request blocking) critically depends on signaling protocol convergence time. In particular, convergence time of the network status information at the nodes must be faster than connection interarrival time. It must be noticed that providing (freeing) both primary and backup resources, when the circuit is created (torn down), may require considerable signaling that may overload the control and management channel.

Restoration schemes have the potential to yield a more dynamic and less signaling demanding solution than protection schemes. Restoration schemes search for a secondary (or restoration) path to reach the destination only upon failure of the working path *without* reserving, in advance, any network spare resource. As depicted in Figure 6.8, depending on the node that computes the secondary path, restoration schemes can be divided into two classes: centralized restoration and distributed restoration. Depending on whether or not the restoration paths are computed before the occurrence of the fault, restoration schemes can be further divided in preplanned and real time [44].

Among the restoration schemes proposed so far for the WDM layer, *alternate routing*[1] (AR) [45] and *distributed restoration algorithms* (DRA) [46] are worth mentioning. Real-time DRAs [46] best utilize network resources, as they search for the secondary path only when the primary lightpath is disrupted. This approach may require heavy sig-

[1]Sometime referred to as diverse routing.

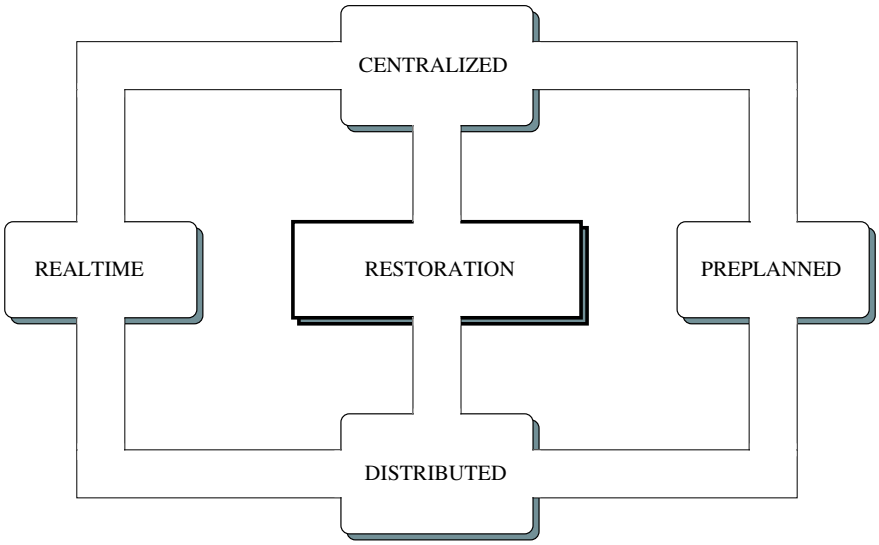

Figure 6.8 Restoration schemes.

naling when the network fault disrupts numerous connections (e.g., fiber cut), since each disrupted connection will search for a secondary path independently of others. AR schemes [45,47] are simple and provide fast recovery time by precomputing (yet not reserving) the secondary path for each active connection. Unless blocked due to lack of resources, AR schemes yield a recovery time that is merely the time necessary to set up the secondary path, which can be comparable to SONET-protection times, i.e., <50 ms [45,47].

An intermediate solution that yields relatively fast restoration time and low blocking probability is the class of *stochastic preplanned restoration* (SPR) schemes [48]. In the SPR schemes, multiple restoration paths are precomputed for each active connection. Resources along the precomputed restoration paths are not reserved. Upon failure of the working path, one of the precomputed restoration paths is randomly chosen and activated. The random selection is driven by the network status information available at the source node at the moment of failure occurrence. The selection is made with the aim of reducing the probability that the restoration path will be blocked due to lack of sufficient network resources. Figure 6.9 reports three curves that represent expected restoration blocking probability versus network load of three restoration schemes: AR, the SPR scheme with proportional weighted path choice (SPR-PW) [48] and exact solution of the integer multicommodity maximum flow (MCMF) problem with centralized control. Solution of the MCMF problem represents the theoretical optimum (minimum restoration blocking probability) for any restoration scheme.

6.2.3.3 Optical Packet-Switching Networks

Optical packet switching represents the ultimate frontier of all-optical networking. In this approach, packets are individually switched and routed in the optical domain. Statistical multiplexing of multiple data streams can be achieved in this way, while circumventing the O-E-O conversion required in conventional electronic packet-switching solutions.

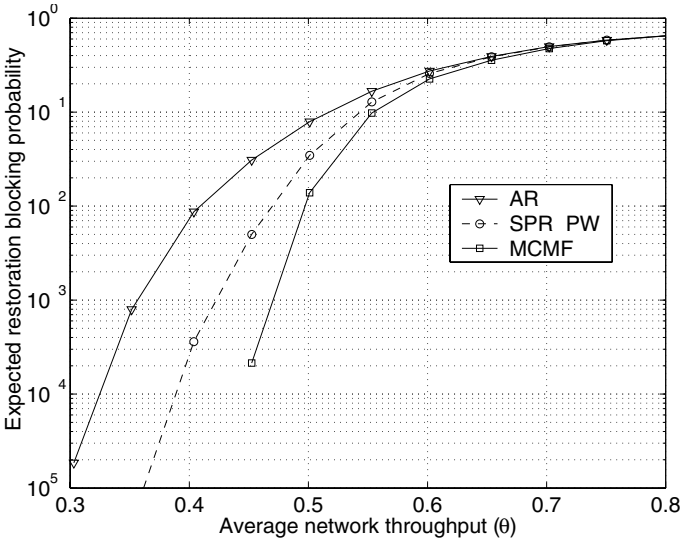

Figure 6.9 Restoration blocking probability versus network load of AR, SPR, and centralized MCMF exact solution.

Depending on how the packet header is processed at the optical node, two approaches are possible. In the *quasi-optical* approach, the packet header is converted to and processed in the electronic domain for control purposes, i.e., determining the output port intended for the arriving packet [49]. Packet payload remains in the optical domain. In the *all-optical* scenario the entire packet, header included, remains in the optical domain. Header processing can be performed by optical gates [50, 51].

With today's available technologies, neither the quasi-optical nor all-optical approach is mature, due to a number of open challenges. Some of these challenges are discussed next.

Optical packet switching imposes very stringent requirements on the optical switching node. The optical node must be able to perform some basic functions, including packet synchronization, header detection, buffering, switching, and routing in the optical domain. Since fast optical memories are not available, both synchronization and buffering require the use of cumbersome fiber delay lines (FDLs). Buffers can be implemented using *feedforward* and *feedback* delay-line structures [7, 8]. In the former case, an optical packet can be delayed (or stored) in the same FDL only once. In the latter case, an optical packet can be circulated many times in the same FDL. Header detection requires some form of packet framing and fast clock recovery of the incoming signal. The optical switch used to route the packet to the intended output port must have a switching time that is a fraction of the packet transmission time. Optical switches capable of switching in nanoseconds or subnanosecond intervals are bulky and expensive.

Some of the open issues in optical packet switching have been extensively studied over the past decade [52]. While research on all-optical header processing is still focused on enabling devices, the quasi-optical approach has been investigated in a number of network test beds, e.g., the Defense Advanced Research Projects Agency (ARPA/DARPA) sponsored contention resolution by delay lines (CORD) [53], the

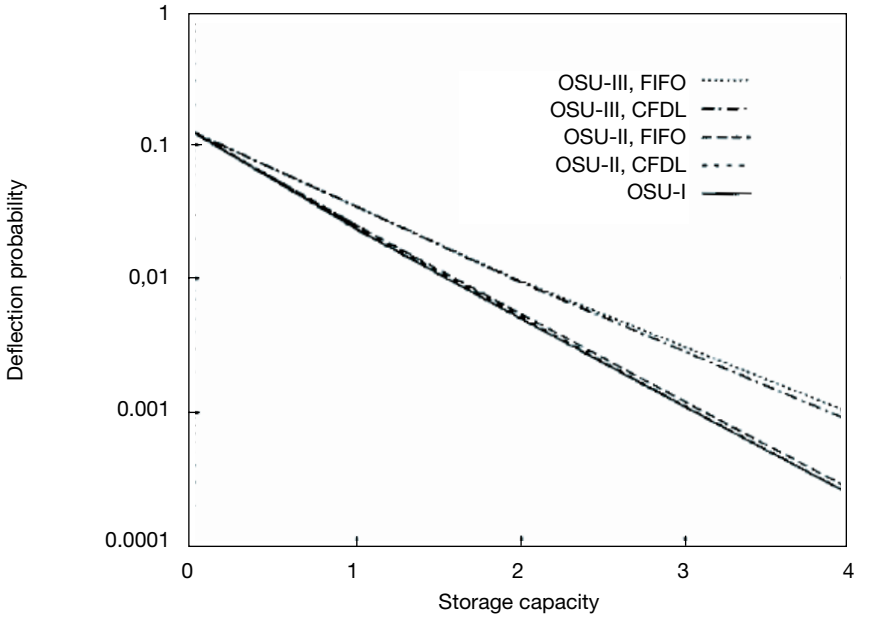

Figure 6.10 Deflection probability in an optical packet-switching node versus the number of fiber delay lines used as buffer.

European Asynchronous Transfer Mode Optical Switching (ATMOS) [54], keys to optical packet switching (KEOPS) (Advanced Communications Technologies and Services (ACTS) funded) [55], and wavelength switched-packet network (WASPNET) [56] (Engineering and Physical Sciences Research Council (EPSRC) funded). These projects have demonstrated the feasibility of optical packet switching, but not yet produced practical, off-the-shelf solutions.

An example of results obtained in these studies is found in Chlamtac and Fumagalli [57], in which a 2 × 2 switching node is proposed to optically store and forward packets of fixed size. A number of fiber delay lines are used in the switch to delay the arriving packets and resolve contention that may arise when two simultaneously arriving packets select the same output fiber. Three switch architectures are proposed and compared: *OSU-I,* a feedback architecture in which packets can circulate within the switch indefinitely; *OSU-II,* a feedforward architecture based on 3 × 3 optical switches; and *OSU-III,* a feedforward architecture based on simpler 2 × 2 optical switches. Figure 6.10 reports the packet deflection probability (i.e., packet is deflected to the wrong output fiber due to the lack of available delay lines) achieved by the three architectures versus the number of delay lines used in the switching node. Two control strategies are compared: FIFO and care packet first, don't care packet last (CFDL).[2]

[2]A packet is referred to as "don't care packet" when either of the output fibers can be used to reach its final destination using the minimum number of hops.

6.2.3.4 Optical Burst Switching Networks A possible intermediate solution between dynamic lightpath networks and optical packet-switching networks is represented by *optical burst switching* (OBS) networks [58,59]. An optical burst consists of the aggregation of multiple packets and may contain several megabytes worth of data that is assembled at the edge switch or router. Once constructed, the burst is transmitted across the network optically, by utilizing a *one-way* reservation mechanism, i.e., the burst transmission is announced by a control message that is followed by the actual burst transmission. An *offset* time may be used between the control message that announces transmission of the burst and the actual burst transmission. Offset time allows the switching node to learn about the arriving burst and prepare for the appropriate switching and routing of the incoming burst.

The aim of OBS is to achieve statistical multiplexing without requiring some of the complex functions needed in optical packet switching networks. With OBS, switching time is relaxed and optical buffers can be avoided by using contention-free burst transmission scheduling [60]. Many studies have been conducted on this recently introduced concept, also indicating the possibility to integrate OBS with MPLS and generate the so-called labeled OBS (LOBS) [58].

One solution recently proposed to transmit optical bursts in WDM rings is presented in [61]. LightRing multitoken control (see Section 6.2.3.2) is adopted to transmit data bursts. Upon reception of a token, a source with an outstanding data burst ready to transmit, checks the resources available on the token-related wavelength. If no other transmission is ongoing on that wavelength, between the source and the destination, the token is released to the downstream nodes to announce transmission of the burst (which follows immediately after the token release). Since a burst transmission can occur only when a token is acquired, all transmissions are guaranteed to be contention-free, and the efficient "tell-and-go" reservation can be used, i.e., reservation of network resources and beginning of burst transmission occur simultaneously. Figure 6.11 plots the saturation throughput achieved by the LightRing reservation protocol versus the expected burst size in a 80-km ring. The total ring bandwidth (320 Gb/s) is evenly divided over W wavelengths. Figure 6.12 plots response time which includes access and transmission time versus expected burst size. Some values of W yield response times that are only marginally affected by burst size.

6.3 PROTOCOL ARCHITECTURES, SIGNALING AND FRAMING TECHNIQUES FOR THE OPTICAL INTERNET

The phenomenal advances in optical technologies are not only driving evolution of the OI, they are also fostering progressive evolution of the TCP/IP protocol stack to cope with the new gigabit speed scenario. At its origins, the Internet was built on unreliable and low-speed links, thus presenting not only congestion problems but also physical layer errors (bit error rate (BER)). Consequently, the original TCP/IP stack is oriented toward providing a reliable best-effort datagram delivery service. Soon it became clear that facilities provided by the unreliable network layer and the end-to-end TCP would not suffice to satisfy the increasing demand for multimedia services that require guaranteed QoS from the network lower layers.

ATM was adopted as a link layer that provides on-demand bandwidth to the IP flows by means of *switched virtual circuits* (SVCs). The SVCs are meant to be set up in a dy-

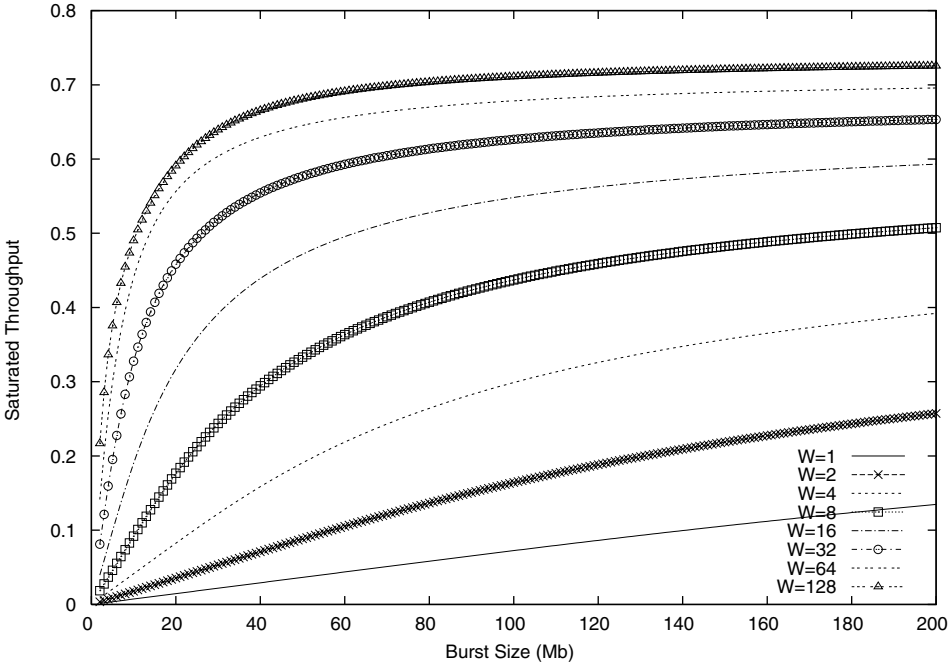

Figure 6.11 LightRing: saturation throughput versus expected burst size.

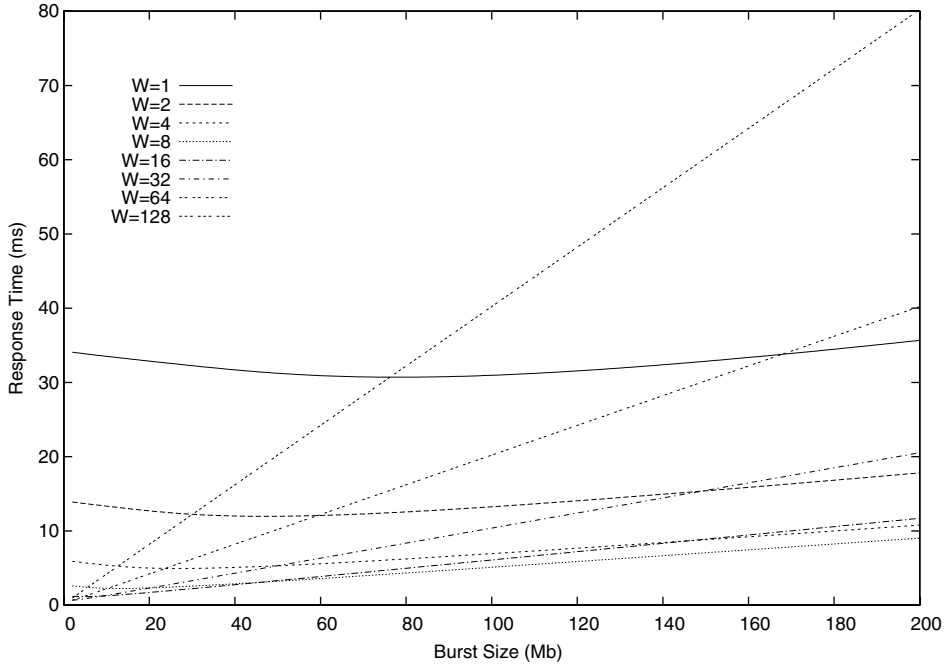

Figure 6.12 LightRing: response time versus expected burst size.

namic fashion in accordance with certain traffic flow descriptors provided by the IP. However, the complexity of provisioning dynamic bandwidth at the ATM link layer led the evolution of ATM to a mere static link layer. Nowadays, ATM *permanent virtual circuits* (PVCs) are commonly used to create links between IP routers. Such PVCs provide a constant bandwidth connection between routers. It must be pointed out that the ATM cell segmentation of the IP packet generates bandwidth inefficiency due to ATM cell headers (roughly 10%) and a significant processing burden that leads to an increase in overall network hardware cost.

In the attempt to guarantee a reliable physical layer for telephony applications, SONET and SDH standards were introduced. The SONET/SDH layer provides a reliable physical layer that the ATM layer can operate on. This includes *operations and maintenance* (OAM) features such as link protection. More recently, introduction by the ITU-T and IETF of, respectively, the OL (see 6.2.2) and the MPLS architecture [62], has led to the network layering depicted in Figure 6.13.

With the introduction of reliable protocols at the OL (e.g., second-generation optical networks) and guaranteed QoS at the IP layer (by means of MPLS), it appears that the complex multilayering architecture depicted in Figure 6.13 is no longer necessary. Some layers may be discarded to simplify the overall protocol stack and reduce network cost. In this section, we outline the evolution of OI layering by describing a number of possible realizations of IP on top of, respectively, first- and second-generation optical networks (see Figure 6.14). As part of the *first-generation optical Internet* we classify the *IP over SONET* standards and *IP over Gigabit Ethernet*. The *second-generation optical Internet* includes *IP over WDM* architectures, e.g., MPλS.

6.3.1 The First-generation Optical Internet: IP over First-Generation Optical Networks

We consider as first-generation optical internet the IP architecture based on FGONs. FGONs provide static point-to-point optical channels between physically *adjacent* IP routers. IP routers perform electronic processing of packets. The research effort in this scenario is focused on implementation of the interface between the IP and the OL that provides efficient encapsulation of IP datagrams. Both the *IP over SONET* and *IP over Gigabit Ethernet* standards are possible candidates for the interface between the IP and the optical domain.

6.3.1.1 IP over SONET The rationale of IP over SONET is to simplify the protocol stack, shown in Figure 6.14, by removing the ATM layer. By encapsulating IP datagrams directly on top of SONET, bandwidth efficiency is increased and the processing burden imposed by the ATM packet segmentation and reassembly avoided. However, a link layer protocol is still needed for packet delineation. Figure 6.15 illustrates how the IP packet is encapsulated in the SONET frame. Since SONET provides a byte-oriented stream service, the IP datagrams have to be delineated with a layer 2 header/trailer before being transmitted.

Current standards propose the use of HDLC-framed PPP for layer 2 framing as described in RFC 1662/2615 [63, 64]. Figure 6.16 shows the IP datagram HDLC-framed PPP encapsulation for POS, as recommended by the standard. The figure also shows the packet delineation flag that requires byte stuffing.

Two potential drawbacks may limit the application of IP over SONET architectures [65].

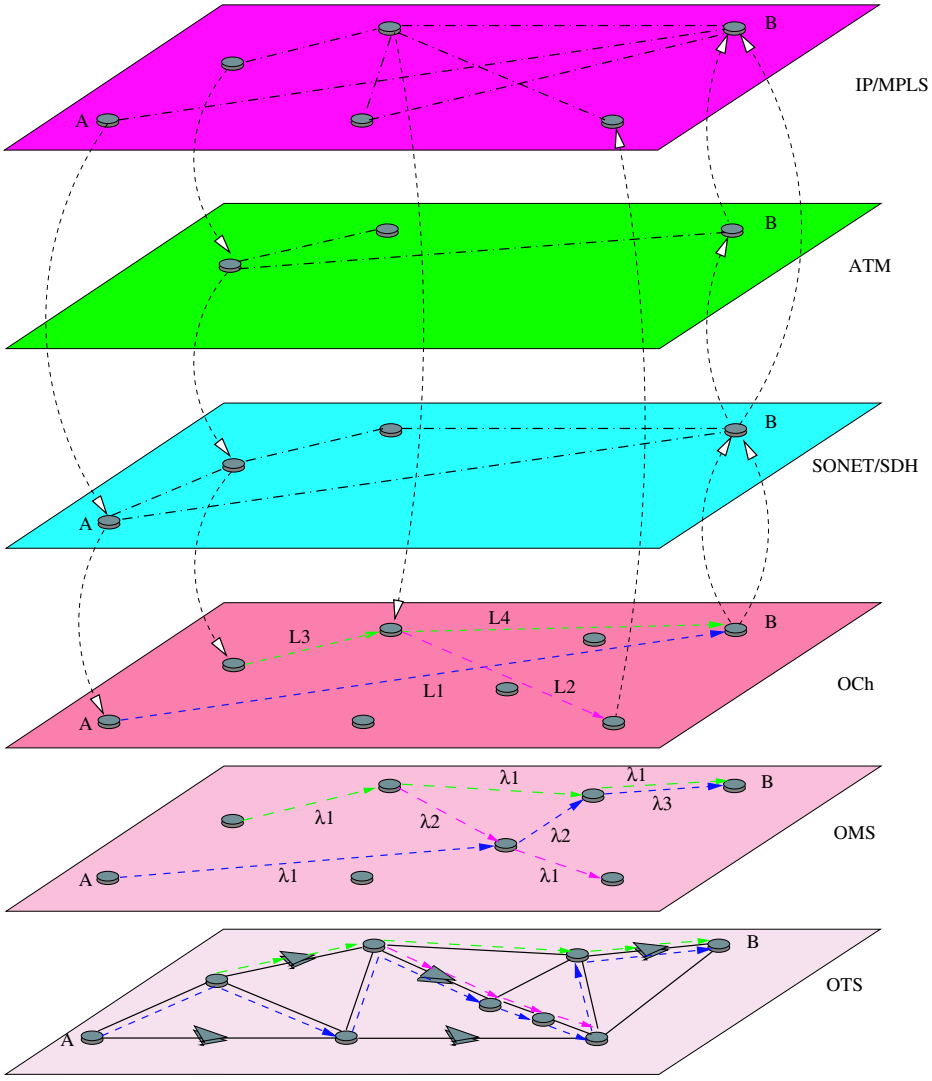

Figure 6.13 Network layering.

First of all, the use of flag 0x7E requires that the escape sequence 0x7D 0x7E is introduced in the PPP payload. Escaping a flag at gigabit rates becomes a performance bottleneck and may limit maximum transmission rate of the protocol. In addition, a malicious user may try to artificially increase the datagram length by arbitrarily inserting flag 0x7E. In doing so, the malicious user may trick the scheduling mechanisms[3] at the IP layer that sets the precedence of a certain packet over others based on the datagram size.

Second, the data scrambling provided by SONET has been shown to be insufficient. The SONET scrambler uses a simple 7-bit pseudorandom sequence,[4] which is XORed

[3]WFQ is an example of one well-known and widely used scheduling mechanism.
[4]$1 + x^6 + x^7$.

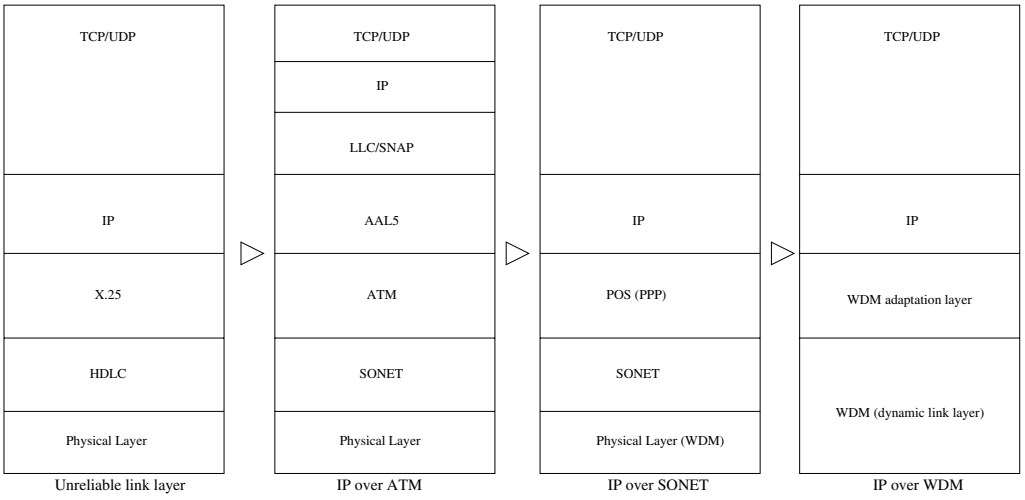

Figure 6.14 Evolution of the Internet protocol stack.

with the data in order to suppress long sequences of zeros or ones. Such long sequences of zeros and ones must be removed because they may deactivate the link due to *loss of frame* (LoF) or *loss of signal* (LoS) SONET alarms. When transmitting datagrams, the probability of still having a long sequence of zeros or ones after SONET scrambling is not negligible. An additional scrambling mechanism is thus required. At high speeds, scrambling becomes a major issue because it introduces some processing delay that can make user-perceived throughput decrease.

Due to the aforementioned drawbacks, the so-called POS standards are difficult to scale beyond OC-48 (2.5 Gb/s). To circumvent this speed limitation, other proposals for IP over SONET have been introduced, such as the *PPP over SDL* standard (RFC 2823) proposed by Lucent [66]. Figure 6.17 shows the IP frame encapsulated with PPP over

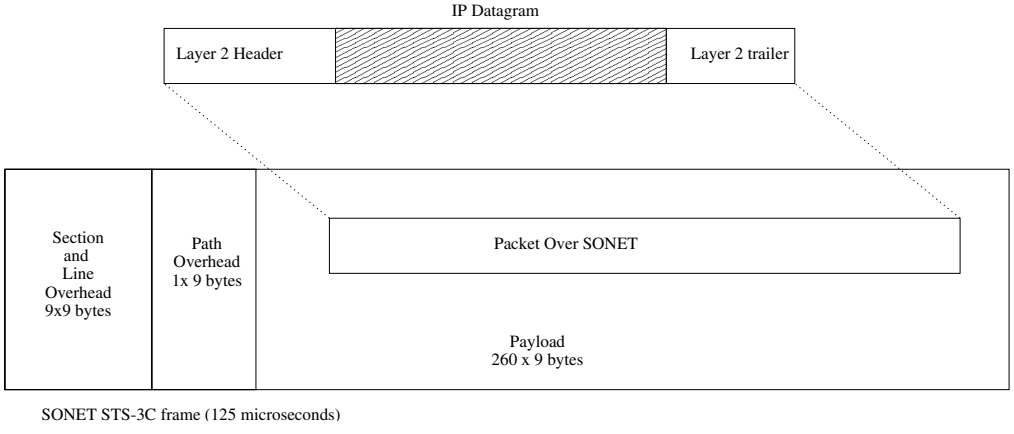

Figure 6.15 Packet encapsulation in IP over SONET.

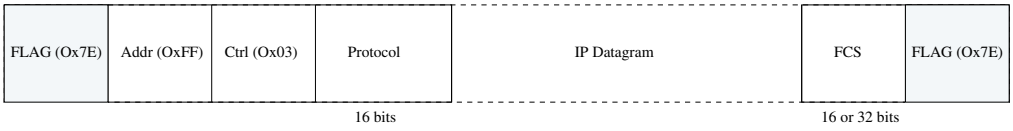

Figure 6.16 HDLC-framed PPP encapsulation.

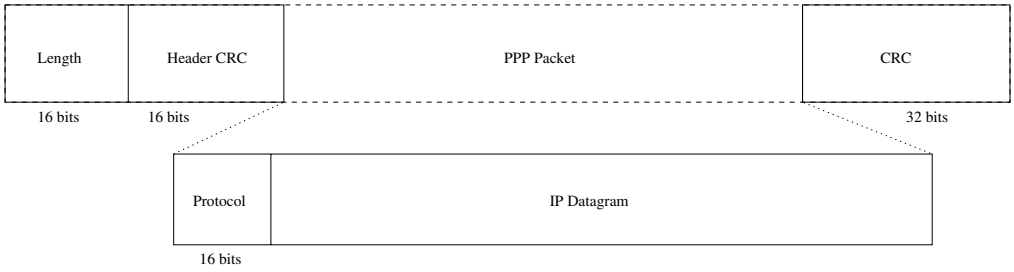

Figure 6.17 PPP SDL encapsulation.

SDL. In PPP over SDL link synchronization is achieved by means of an algorithm that is similar to I.432 ATM HEC delineation. Instead of searching for a flag (0x7E), as it is done in POS, the receiver calculates CRC over a number of bytes until it "locks" to a frame. When this happens, packet delineation is achieved and the receiver enters the SYNC state. In addition, data scrambling is performed with a self-synchronous $x^{43} + 1$ scrambler or an optional set–reset scrambler independent of user data. This makes it difficult for the malicious user to break SONET security.

Finally, it must be noted that the performance of IP over SONET, both in POS and PPP over SDL, is highly dependent upon IP packet size. Assuming that no byte stuffing (escaping 0x7E flags) is required and a 16-bit frame check sequence is adopted, the POS overhead is 7 bytes. For example, for a SONET layer offered rate of 2404 Mb/s (OC-48) the user-perceived rate on top of TCP/UDP is 2035 Mb/s for an IP packet size of 300 bytes.

6.3.1.2 IP over Gigabit Ethernet

Gigabit Ethernet is another candidate link layer for the transport of IP packets at Gb/s rates, especially in the access network [67]. There are two possible configurations for a Gigabit Ethernet: hub and point-to-point link. The latter is particularly attractive to provide access links to Gb/s users or links between gigabit routers. It must be noted that Gigabit Ethernet provides the service of an Ethernet card at Gb/s rates, i.e., asynchronous packet transport with no QoS discrimination.

In order to scale Gigabit Ethernet beyond 1 Gb/s [68] a number of solutions are being considered that are based on the principle of *inverse multiplexing*. For instance, a 10-Gb/s stream (OC-192) could be demultiplexed into four OC-48 streams (2.5 Gb/s) or eight 1.25-Gb/s streams, which could be transported by means of distinct wavelengths using 1-Gb/s Gigabit Ethernet as the link layer.

More recently, 10 Gigabit Ethernet[5] uses IEEE P802.3ae to create an new standard that

[5]http://www.10gea.org

still uses the IEEE 802.3 Ethernet MAC protocol with the same frame format and minimum and maximum frame size. The 802.3ae specification contains many technical innovations such as the definition of two different physical layer (PHY) types: the LAN and WAN PHY. The latter uses 64B66B encoding and SONET framing, and may be used for distances up to 40 km using single-mode fiber (1550 nm). The standard was approved at the June 2002 IEEE Standards Board meeting.

Finally, a new Internet Draft was issued recently [69] with a proposal for virtual concatenation of SONET envelopes. The aim is to provide an efficient way to carry Ethernet tributaries using SONET payloads. By allowing concatenation of envelopes at a given SONET hierarchy level, SONET bandwidth allocation granularity is improved to match the bandwidth required by Ethernet. For example, multiple VT1.5 payloads can be concatenated to produce a VT1.5-nv channel. By concatenating multiple VT1.5, it is then possible to reserve bandwidth from a 3.2-Mb/s (n = 2) channel to a 102.64-Mb/s (n = 64) channel. The latter can be used to accommodate a 10/100-Mb/s Ethernet transmission.

6.3.2 The Second Generation Optical Internet: IP over Second-Generation Optical Networks

The second-generation optical Internet, also commonly referred to as *IP over WDM,* represents a step forward in offering a high-speed efficient approach to provisioning IP services on top of OL. As described in Section 6.2.3.2, second-generation optical networks may provide dynamic resource allocation at the OL. A higher degree of flexibility is thus achieved when compared to the first-generation optical Internet, since optical bandwidth can be more efficiently handled by the client IP layer. Consequently, some of the intermediate layers between the IP and OL shown in Figure 6.13, e.g., ATM and SONET, are no longer necessary. In summary, two major benefits may derive from the use of IP over WDM: reduced network complexity and overhead and higher bandwidth flexibility.

Two possible approaches for IP over WDM are currently being debated: the overlay model and the peer model [70].

The overlay model provides two distinct control planes, one at the IP layer, the other at the OL. The overlay model is pushing the evolution of OI toward *flow switching* solutions, such as MPLS, which serve the purpose of simplifying IP routing and performing traffic engineering. Scheduling mechanisms are used to provide the transmitted IP packets with differentiated QoS on top of either static or dynamic WDM layers.

The peer model assumes that the IP layer and the OL are managed by a single control unit which has complete visibility of all network resources. The peer model is driving Internet evolution toward the so-called MPλS, which allows setup and tear down of optical circuits (lightpaths) on demand, to achieve efficient use of OL resources while circumventing the FGON electronic bottleneck.

6.3.2.1 IP over WDM Overlay Models: Flow Switching and Gigabit Routers

The IP over WDM overlay model assumes that the IP and WDM layers are completely decoupled. The optical network is viewed as an *autonomous system* (subnetwork) (AS) that provides connectivity to *edge IP gigabit routers,* as shown in Figure 6.18. The edge routers act as border domain gateways, and provide connectivity edge-to-edge. Such connectivity is reported to the rest of the Internet by means of BGP messages. Within the op-

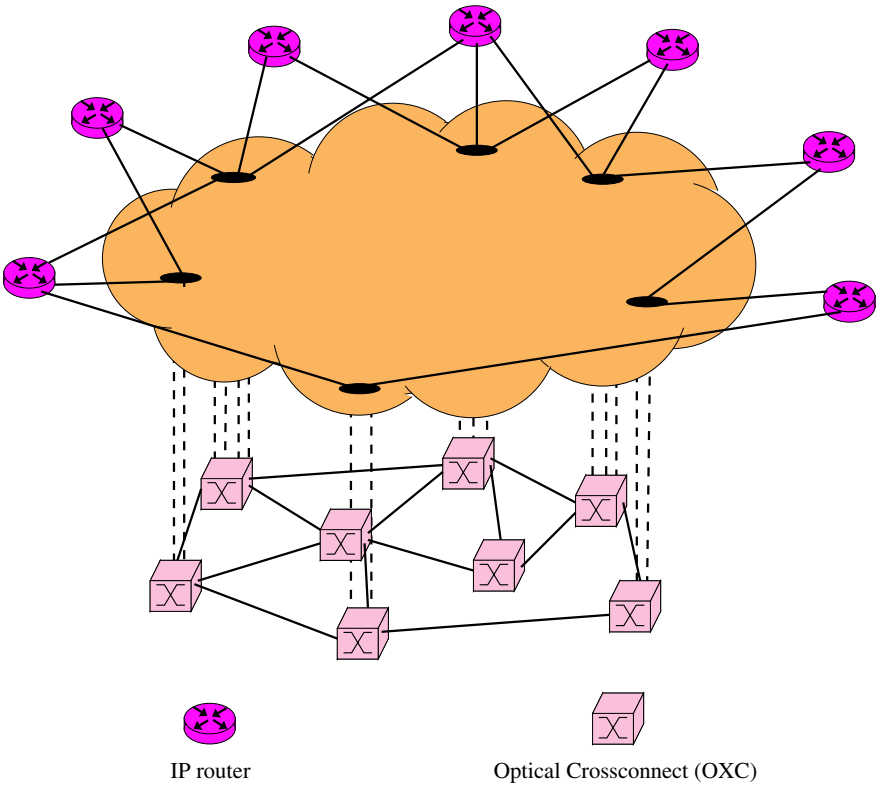

Figure 6.18 Overlay model.

tical subnetwork a different routing protocol (Interior Gateway Protocol (IGP)) may be used (e.g., OSPF), which is similar to what is done in conventional ASs.

Besides end-to-end static or semistatic lightpaths, the optical network can provide flow switching mechanisms, e.g., lightpaths on-demand or burst switching. In the latter case the IP layer acts as a *client* of the OL, which is capable of providing "on the fly" resource allocation upon request. A clear advantage of the overlay model is the fact that future technology advances that will take place at the OL will not require any change of the IP layer.

The two key elements of this architecture are the concept of flow switching and gigabit routers.

Flow Switching The idea of QoS discrimination at the link was originally introduced with the IP over ATM architecture. A network of ATM switches may provide point-to-point on-demand circuits (SVCs) in order to transport IP flows or flow bundles. As already mentioned, the use of ATM translates the problem of high-speed packet scheduling from one layer (the IP) to another (the ATM). Even though the use of fixed-size cells serves the purpose of simplifying the task, the switching and scheduling mechanism are far too elaborate for a practical implementation. As a result, current ATM implementations provide a simple static link layer to the IP.

In flow switching, the link layer is required to provide both flexible and efficient bandwidth allocation schemes and flow recognition and characterization mechanisms. The edge router must determine the flows required for a given traffic, either by automatic detection or explicit user indication. The scheduler must decide the amount of network resources that must be assigned to each flow, namely, buffer capacity and bandwidth. A number of proposals have been generated to address the flow assignment problem. In these proposals, either the IP destination address, subnetwork identifier, or HTTP file size are used to determine which flow is to be used [71, 72].

As will be shown later, most IP flows are actually short in size and duration, a fact that makes per-flow resource allocation and flow recognition a complicated matter. As a result, most flow-switching solutions existing nowadays are *tag-switching* solutions, e.g., the Cisco Tag Switching [73] and the MPLS [74]. Tag-switching solutions assign a tag to an incoming packet at the network edge so that subsequent routers perform the routing based on tag value. Tag switching is efficient and flexible. For instance, tag routing tables can be drastically simplified by addressing a set of destination subnetworks using a single tag. *Traffic engineering* of aggregated flows is possible which circumvents the problem of per-flow traffic engineering and related short time scale. For example, traffic engineering can be done at the *subnetwork granularity*. In this scenario all the connections terminating at a particular set of destination subnetworks are assigned the same tag. Suppose, for instance, that a large fraction of TCP connections from a European country are directed to U.S. servers. A tag is assigned to the corresponding IP packets in order for the edge routers to mark the packets as "directed to the US." Routers subsequently encountered by the transmitted packets will eventually route the packets to the transcontinental link by simply inspecting their tag.

On the other hand, separate resources may be assigned to the multiplexed TCP connections, resulting in segregation of flows directed to a given destination. A network operator could, for instance, assign different tags to different ISPs. Each tag is then assigned a desired capacity, as indicated by the corresponding peering agreement. Finally, a class of service (CoS) field in the MPLS tag allows one to define classes of service at the flow granularity and treat the various flows accordingly. For a detailed description of MPLS, the reader is referred to Chapter 3.

IP Gigabit Routers The edge routers of the optical subnetwork are often referred to as gigabit routers. Design of IP gigabit routers is oriented toward providing more efficient routing table lookups so that packets and MPLS flows can be handled with minimum processing time. Gigabit Routers can be classified into three major families: centralized, decentralized, and parallel (see Figure 6.19).

The centralized router uses a single routing engine. The approach is simple, but the single routing engine may become a bottleneck. The decentralized router replicates the routing engines in the multiple line cards. If a certain route is not found in the line card routing table, the master engine is requested to handle the packet. Packets may be routed out of order due to possible delay variations that originate during update of the line card tables. The parallel router provides a parallel architecture at both the master and line card levels. The parallel architecture improves table lookup time at the cost of elaborate schemes for consistency control at high speed.

In general, the design of efficient lookup algorithms is of fundamental importance in realizing gigabit routing. Such algorithms fall into a search algorithms area that has been extensively studied over the last years. A large number of search algorithms have been

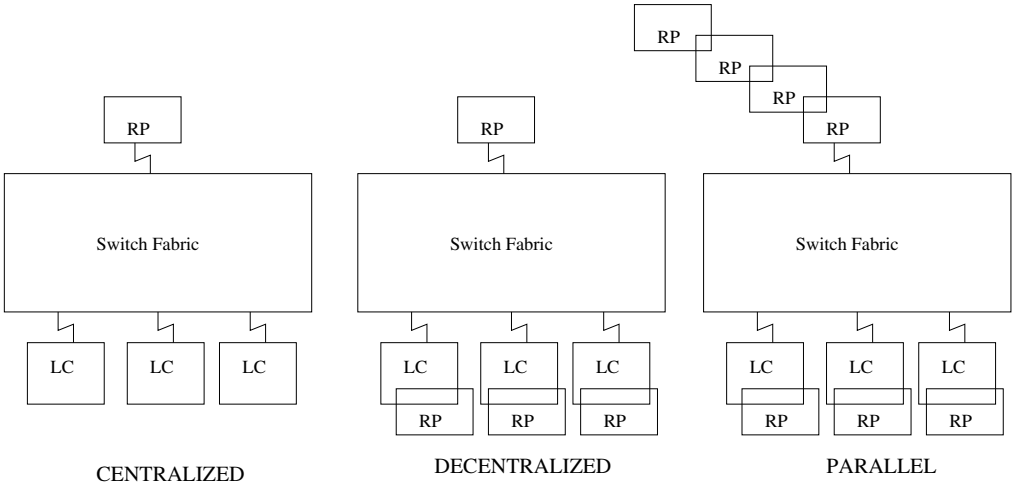

Figure 6.19 Gigabit router families.

proposed that are based on tree structures. They provide specific search optimization mechanisms at each tree level, e.g., trees with hashing tables. Tree levels can be compressed or segregated to improve the lookup time upon specific input traffic patterns. Furthermore, use of caches and *content addressable memories* (CAMs) also provide fast response times for the most frequently accessed (hot) routes. The quality of a particular algorithm is measured by the search complexity as the number of entries in the table grows. A selection of the most popular algorithms [75–80], together with the associated complexity, is shown in Table 6.1.

6.3.2.2 IP over WDM Peer Models: Multiprotocol Lambda Switching and WDM Gigabit Routers The introduction of MPLS tag switching and its advantages has paved the way to more recent proposals that suggest the use of the MPLS control plane for controlling the OL. Use of MPLS Traffic Engineering Control to configure OADMs and OXCs has been proposed by IETF to solve the problem of finding a distributed management protocol, the so-called MPλS [81], that dynamically sets up and tears

Table 6.1 Search Complexity for Several Lookup Algorithms

Algorithm	Complexity
Binary search	$O(\log N + W)$
Patricia-trie	$O(W^2)$
Dynamic prefix tree	$O(W)$
LC-tries	$O(W/k)$
Multiresolution compressed tree	$O(\log_k N + 1)$
Hash table	$O(\log W)$
Content addressable memory	$O(W/k)$
Tree + hash table	$O(W/k)$

Note: W: bits per address, *k:* constant factor.

down lightpaths. A further step is represented by the proposed generalization of the MPLS control plane to manage legacy equipment, e.g., SONET crossconnects and add-drop multiplexers. This most recent proposal is referred to as GMPLS [35].

In the MPλS framework, IP and WDM layers are not independently overlayed. On the contrary, their devices are now considered as peers within the same physical network topology. As a result, IP routers are able to directly set up lightpaths in the physical network topology as they deem appropriate. For this reason, this model is known as the *peer* model (see Figure 6.20).

In simple words, the network configuration is similar to the standard MPLS network with the difference that the core MPLS routers are replaced by OXCs equipped with MPLS control plane. These OXC's are commonly referred to as Lambda Switch Routers (λSRs) (see Figure 6.21). The MPLS control plane can thus define an LSP that actually consists of a lightpath. At the MPλS network edge there are MPλS edge routers capable of aggregating IP input packets to be transmitted over the same lightpaths (the functionality is the same one provided by conventional MPLS edge routers in MPLS networks). However, the novelty is twofold: the created LSP, or lightpath, has a predetermined capacity (that is the transmission rate determined by the lightpath transmitter and receiver) and the traffic carried by the lightpath is completely orthogonal with respect to other lightpaths.

Many similarities exists between LSRs and OXCs [81]. Both LSRs and OXCs decouple the *control plane* from the *data plane*. The LSR data plane is based on label swapping to forward a packet from source to destination, whereas, the OXC data plane utilizes switching matrices. LSR performs label switching by mapping the pair (input port, input label) onto the pair (output port, output label). OXC maps (input port, input optical channel) onto (output port, output optical channel). The previous relations are determined by the control plane and are locally activated by a switch controller. The LSR control plane is used to discover, distribute, and maintain relevant state information associated with the MPLS network and instantiate and maintain LSPs. Similarly, the OXC control plane is

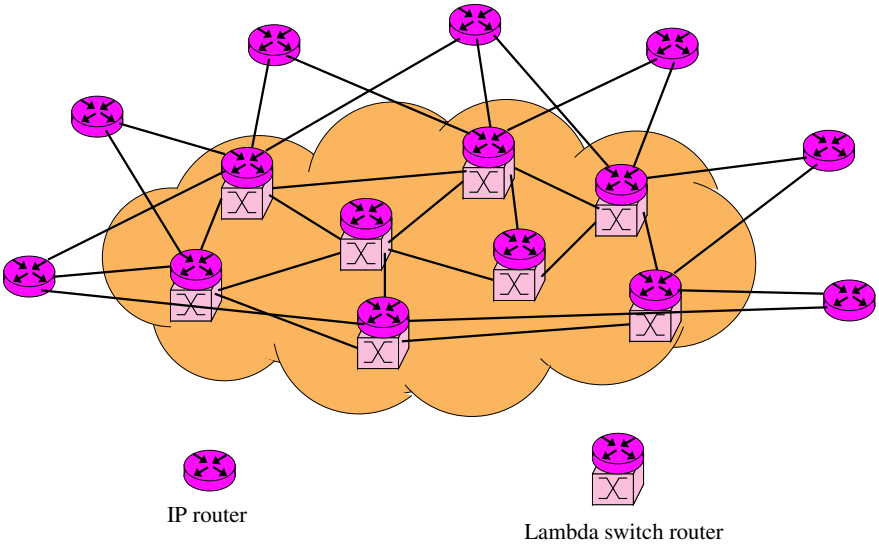

Figure 6.20 Peer model.

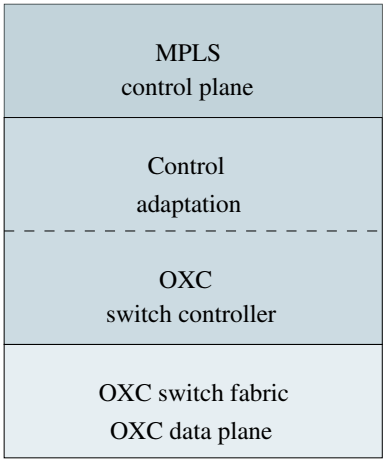

Figure 6.21 Lambda switch router.

used to discover, distribute, and maintain relevant state information associated with the OL and establish and maintain lightpaths.

The main difference between LSRs and OXCs is that OXCs are not able to perform packet level processing (e.g., aggregation and grooming), whereas, LSRs are electronic datagram devices that can perform packet level operations in the data plane. Another difference is that while in the LSR the forwarding information (tag) is carried by each individual packet. In the OXC the tag is carried, implicitly, by the lightpath's wavelength. Finally, in-band signaling is only possible in MPLS networks, whereas, in MPλS networks a separated control channel is necessary to perform out-of-band signaling.

In summary, the peer model approach is based on just one instance of routing and signaling protocols for both the electronic and optical domains. IP routing protocols (e.g., IGP such as OSPF) are used to calculate the routes and a link state advertisement (LSA) protocol distributes network state information to the network nodes. Signaling and reservation of network resources for lightpath (LSP) setup are based on MPLS signaling protocols, such as adaptations of RSVP-TE and CR-LDP [35, 82]. For network resilience, the peer model exploits the same schemes used by MPLS [26], e.g., end-to-end path restoration, path and line protection.

The MPλS peer model presents advantages and disadvantages when compared to the overlay model. It offers a framework for optical bandwidth management and provides real-time lightpath provisioning. It facilitates the coordination between IP network elements and optical network elements by resorting to the same control protocol plane used in data (MPLS) networks. It yields seamless interconnection of IP and optical networks, in accordance with the notion of second-generation optical Internet. However, it requires routing information that is specific to optical networks to be known by the IP routers. As a result, it requires a joint evolution of both the OL and the IP control plane. It is therefore expected that, from a practical point of view, the peer model is going to be less suitable than the overlay model for near-term deployment [70].

Finally, it is worth mentioning that although most of the solutions proposed so far to combine the IP layer and the OL are based on the concept of switching packets electroni-

cally and routing aggregated flows optically by means of lightpaths, some initial attempts have been made in the direction of optical packet switching. For example, WDM gigabit routers [83] combine optical switching and WDM transmission techniques to yield WDM optical packet switching. As in the case of IP gigabit routers, the main functions to be performed are packet routing (decide the packet route based on available network topology information), packet forwarding (routing table lookup and assignment of the packet to an output port), and packet switching. While efficient routing and switching mechanisms can be inherited from IP gigabit routers, the real bottleneck for WDM gigabit routers arises at the forwarding level. The high aggregate transmission rate of the wavelengths sets challenging objectives for routing table look-up. As mentioned in Section 6.2.3.3, an additional practical problem is represented by realization of inexpensive optical memories.

6.4 TRAFFIC ENGINEERING IN THE OPTICAL INTERNET

One of the open issues of OI is how to make the phenomenal amount of bandwidth supplied by the OL available to Internet applications. Available technologies in SGON provide only static lightpath configurations in which Gb/s connections are setup by management procedures. Such static lightpaths are used to transport a *high-level traffic multiplex*.

An open problem in SGON is to identify simple mechanisms that offer multiple granularities in bandwidth allotment at the optical layer (on-demand lightpaths, burst-switching, optical packet switching). In order to effectively use any of the aforementioned mechanisms, a good understanding of Internet traffic characteristics at the IP *flow* level is required.

In particular, at the coarsest level of aggregation (static lightpath), it is necessary to understand traffic characteristics obtained when a large number of flows are multiplexed. At a finer aggregation levels (burst switching, optical packet switching) a detailed characterization of IP flows is mandatory.

In this section we first describe some basic concepts of traffic *self-similarity* which is a property of high-level multiplexed Internet traffic. Then, we present a *flow level analysis* based on collected traffic traces. In conclusion, we analyze the impact of such properties on the specific scenario of the OI.

6.4.1 Self-Similarity

It is widely recognized that the multiplexing of many Internet traffic sources differs significantly from other well-known types of multiplexed traffic, such as multiplexed voice sources [84, 85]. Indeed, in contrast with voice traffic, that is Poisson modeled, Internet traffic presents self-similar characteristics.

A stationary stochastic process in discrete time $X = \{X_t, t \geq 0\} = \{X_1, X_2, \ldots\}$ is called "asymptotically second-order self-similar" with "Hurst parameter H" if for all $k \geq 1$ [86]

$$\lim_{m\to\infty} \rho^{(m)}(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \tag{6.1}$$

where, $\rho^{(m)}(k)$ is the lag k autocorrelation of the aggregated process $S^{(m)} = \{X_t^{(m)}\} = \{X_1^{(m)}, X_2^{(m)}, \ldots\}$,

$$S_t^{(m)} = \frac{1}{m}(X_{tm-m+1} + \cdots + X_{tm}), \qquad t \geq 1 \tag{6.2}$$

For $1/2 < H < 1$, this means that correlation $\rho(k)$ decays to zero so slowly that [87],

$$\sum_{k} \rho(k) = \infty \tag{6.3}$$

In simple words, process X has long memory or *long-range dependence*. In the specific case of the Internet traffic, such long-range dependence is observed in the packet-counting process, X_t (Figure 6.22), defined as the number of data bytes transmitted over a fixed time interval—δ ms [84,85].

As a consequence of the slow decay of the autocorrelation function, the overflow probability at intermediate router queues heavily increases when compared to a process with independent increments (Poisson). In Erramilli et al. [88], an experimental queueing analysis with long-range dependent traffic compares an original Internet traffic trace with a shuffled version of the same, i.e., with destroyed correlations. Results show a dramatic impact on server performance due to long-range dependence.

As a visual example, the packet counting process in time scales of 1, 10, and 1000 ms obtained from both a collected Internet traffic trace and a *Poisson process* is shown in Figure 6.23. We note that while the Poisson process (independent increments) tends to smooth out as the time scale of observation increases, the real traffic sample does not. Since X_t shows the long-range dependence of the Internet traffic trace, variance of the aggregate process X_t^m does not decay with the inverse of the number of samples aggregated (m).

The effect of *slowly decaying variance* can be observed by plotting the variance of the aggregated process S_t^m versus m (aggregation level) in log-log scales (Figure 6.24). For a Poisson process (independent increments) variance decays with the inverse of the number of samples (aggregation level), as predicted by the Central Limit Theorem. Instead, due to the effect of long-range dependence, decay of Internet traffic trace variance is slower than the independent case. As a result, Internet traffic trace shows significant burstiness at any time scale, not only at small time scales. We note that bursts of traffic are observed even at time scales of 1 s (see Figure 6.23). Such large bursts cause buffer overflow situations that are not captured by Poisson input traffic models.

6.4.1.1 Causes for Long-Range Dependence

While there is considerable debate about long-range dependence causes [84, 85, 89–91], Willinger et al. [91] showed

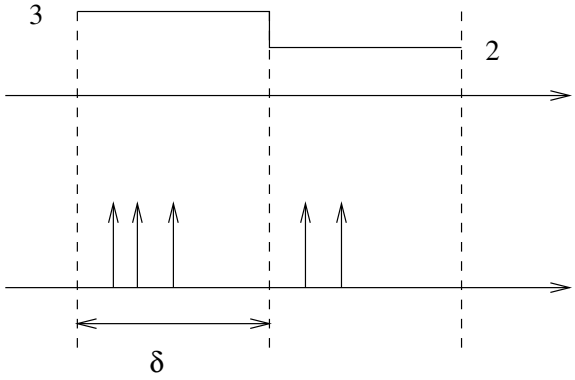

Figure 6.22 Packet-counting process (X_t).

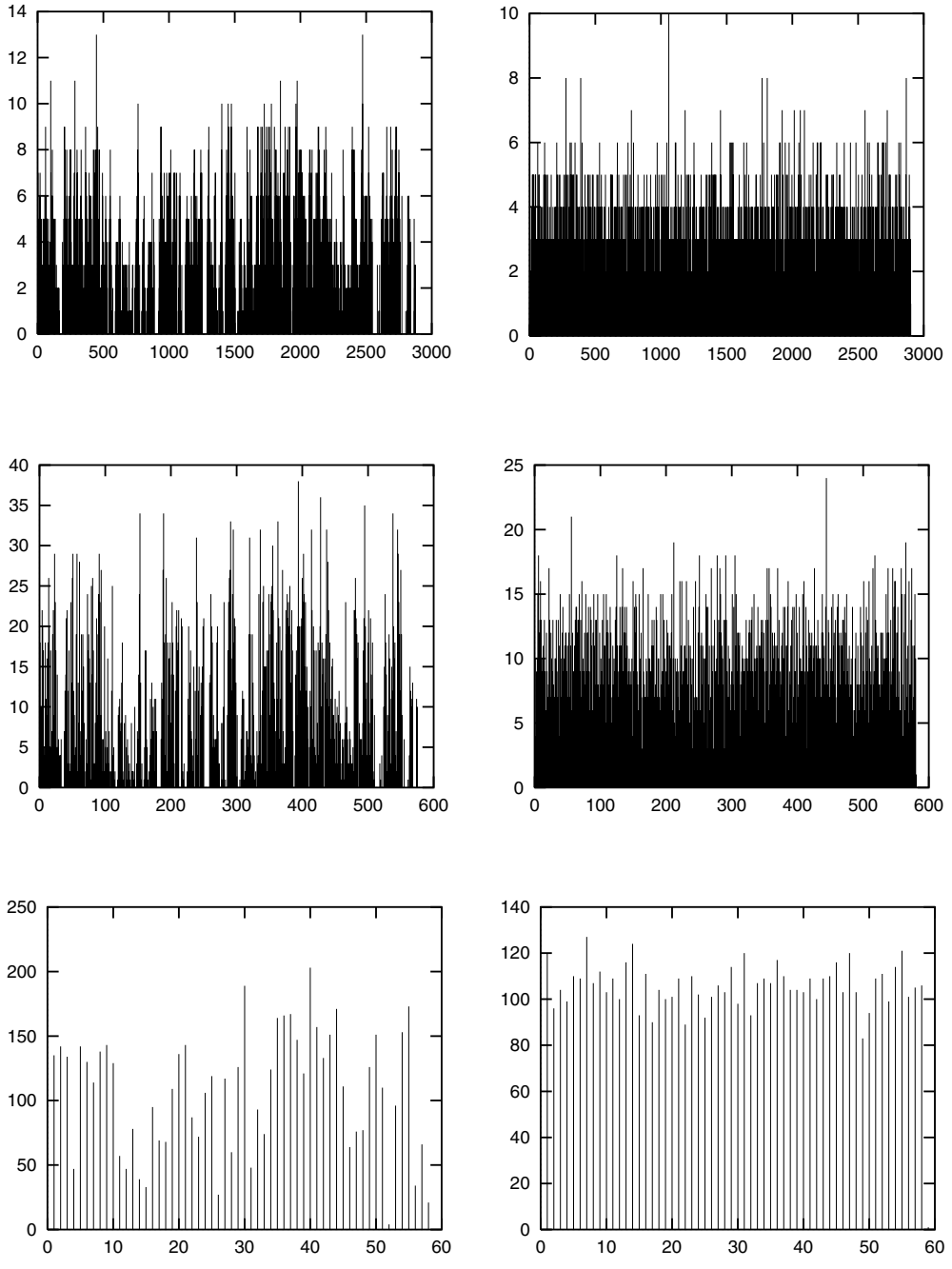

Figure 6.23 Packet-counting process over several time scales (S^m). Internet traffic trace (right-hand side) and Poisson process (left-hand side).

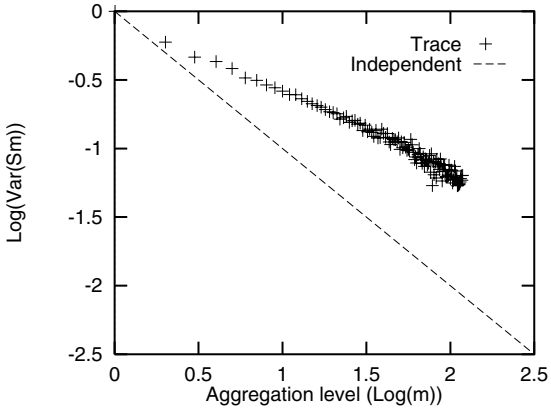

Figure 6.24 Variance-aggregation plot (Log(Var(S^m)) versus Log(m)).

that the multiplex of on–off sources with heavy-tailed on-off periods (Figure 6.25) turns out to have self-similar properties [84, 85, 89, 91]. Furthermore, Tsybakov and Georganas [86] showed that as long as the on-period of the individual connections is heavy-tailed, the resulting traffic multiplex is asymptotically second-order self-similar, even though the connection arrival process is Poisson.

A heavy-tailed random variable, R, has a distribution tail

$$P(R < r) = 1 - \left(\frac{K}{r}\right)^{-\alpha} \tag{6.4}$$

where α takes on values $1 < \alpha < 2$. The resulting random variable has finite mean but infinite variance. Parameter α in Equation (6.4) is related to the Hurst parameter H as follows [86]:

$$H = \frac{3 - \alpha}{2} \tag{6.5}$$

Values of H in the range $1/2 < H < 1$ indicate long-range dependence. Such H values correspond to α values in the range $1 < \alpha < 2$.

The Poisson-arriving heavy-tailed bursts hypothesis for long-range dependence can be verified empirically. At the Public University of Navarra the data traffic of a large

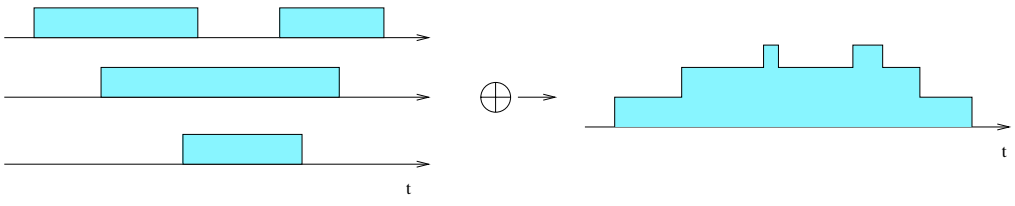

Figure 6.25 Multiplex of on–off sources.

multiplex of Internet users (1500 hosts), which share IP over ATM access links [92], has been analyzed. Traffic trace, recorded in February 2000, clearly reveals that WWW traffic is dominant. Indeed, TCP traffic percentage equals 99% in bytes transmitted, 82.8% of which (96.9% of the total number of TCP connections) are WWW connections. A detailed analysis of TCP connections recorded in the trace, focusing on *arrival process* and *connection holding time,* was performed. Hourly intervals in the morning, afternoon, and evening were observed, generating the plots reported in Figure 6.26. The plots report the *survival function* of connection interarrival time ($S(x) = P(X > x)$) in log-linear scale. The almost straight lines shown in the figure indicate that interarrival times are best modeled by an exponential random variable. Since users' traffic is independent of one another, the aggregate arrival process can be assumed to be Poisson.

Figure 6.27 and Figure 6.28 report survival functions as a function of the connection size (in bytes) and duration of the connection (in seconds) in log-log scale. The survival function of the heavy-tailed random variable defined in Equation (6.4) yields a line of slope $-\alpha$ when plotted in log-log scales. By plotting the distribution tail least-square regression line in both figures, estimated values of $\alpha = 1.2$ and $\alpha = 1.15$ were computed for size and duration, respectively. Such values are in accordance with previous studies that report values of 1.1 and 1.2 [93].

The results in Figure 6.26, 6.27, and 6.28 show that the traffic multiplex can indeed be modeled as a multiplex of Poisson arriving heavy-tailed bursts. TCP connections show a Poisson behavior in the arrival process since they originate from a multiplex of traffic originating from a large number of independent users. On the other hand, the heavy-tailed nature of connection size and duration is due to the fact that Internet file sizes can be

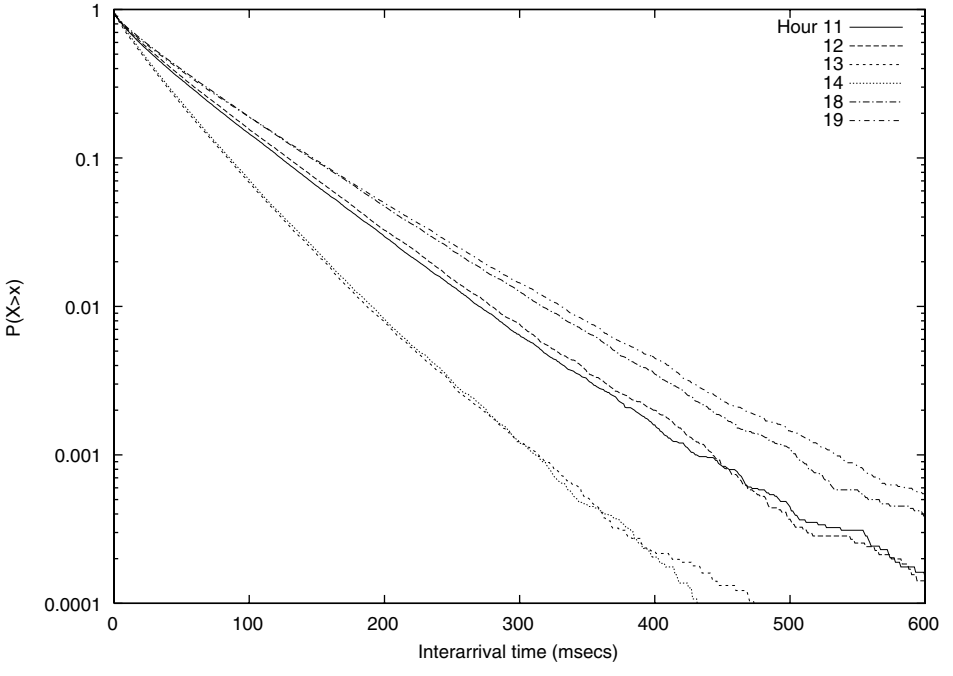

Figure 6.26 Survival function of connection interarrival time.

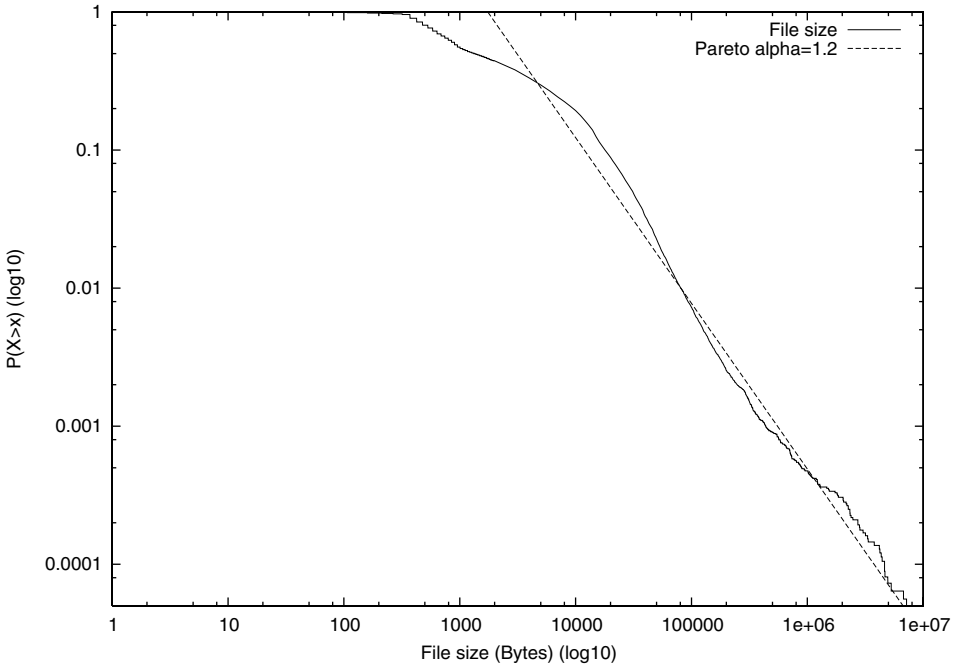

Figure 6.27 Survival function of file size.

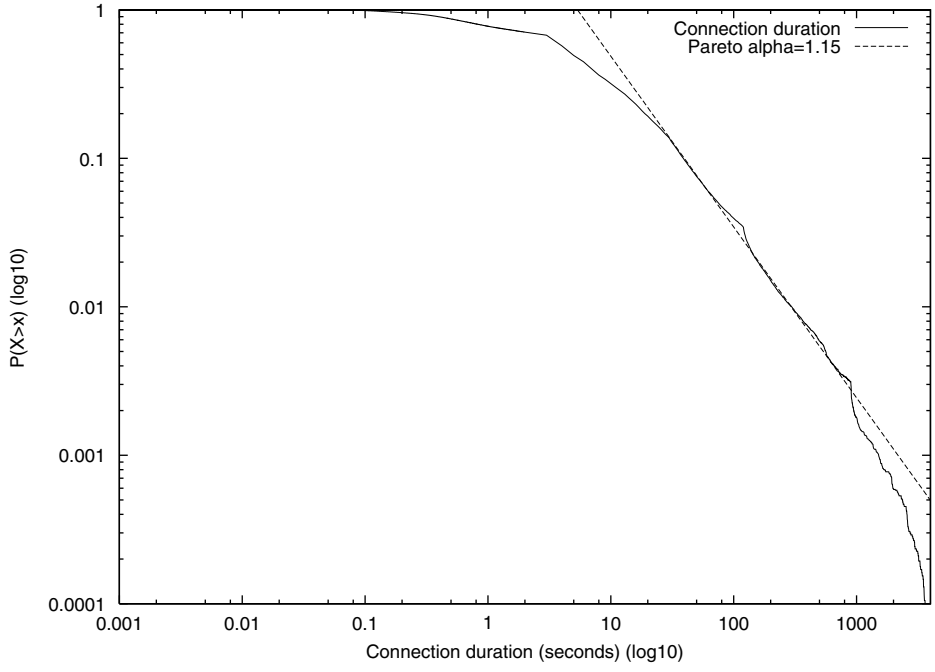

Figure 6.28 Survival function of connection duration.

modeled by a Pareto random variable (see Equation (6.4)). Pareto random variables show infinite variance whenever $1 \leq \alpha \leq 2$ and seem to be satisfactory models of the number of pages found in a book randomly chosen in a large library. The intuitive interpretation of this result is that file sizes in the Internet show a large variability, since there are a myriad of files that range from plain text to large video clips.

6.4.2 Flow Level Analysis

Considering that the most likely evolution of the OI in the near future is toward providing dynamic bandwidth allocation at the OL to support IP flows, a traffic analysis at the flow level is in order. Macroscopic analysis of traffic trace clearly indicates that the traffic sample is dominated by short TCP connections. Figure 6.29 reports the top 10 TCP ports sorted by number of connections and transmitted bytes. Note that some services generate a significant number of connections, each consisting of few bytes, such as AUTH(113), LOGIN(49), and DNS over TCP(53). However, some of the latter services are not among the top services in number of bytes transmitted. The AUTH service is normally used in conjunction with FTP, in order to allow anonymous FTP servers to authenticate the client. On the other hand, other services, such as Hotline (5501), consume a significant share of network resources (third in the amount of generated bytes) with very few connections (only 181 in a week). Hotline integrates multiple services, such as chat, file transfer, and news in the same session. As a result, the transfer of large files produces a significant amount of bytes transmitted by Hotline connections.

Traffic trace is dominated by the WWW with 80% of the total traffic in the amount of bytes, and 90% in the number of connections. We observe from Figure 6.29 that WWW uses port 80 for direct TCP connections and usually port 8080 for proxy WWW connections. The WWW is followed at a considerable distance[6] in bytes generation by the FTP (port 20 for data and port 21 for control) and Hotline, which is quite similar to FTP due to the file-transfer nature of both services. A small percentage of transmitted bytes are due to mail retrieval through POP3 (port 110), mail upload from client to server with SMTP (port 25), virtual terminal services like Telnet (port 23), and secure transactions with HTTPS} protocol (port 443). Table 6.2 presents some connection level statistics for the most popular services found in the trace. We note that WWW connections are small in size, with mean equal to 7.5 kB and 99% percentile equal to 70 kB. We also note a strong asymmetry in bytes transferred from server to client with respect to bytes from client to server, except for SMTP.

In conclusion, this analysis shows that most of traffic trace is dominated by short TCP connections due to WWW, also noted in recent studies, such as Miller et al. [94]. Such short TCP connections pose significant challenges regarding traffic engineering at the OL. As opposed to voice calls, whose call holding time justifies the use of a switched virtual circuit per call, per-flow switching in the OI is not practical except for a small percentage of connections. Thus, some flow aggregation mechanisms become necessary. While some proposals exist in the literature regarding per-port, subnetwork, or pair origin–destination address, finding an adequate flow-switching solution in the OI still remains an open issue.

[6]Note the logarithmic scale in the *y*-axis.

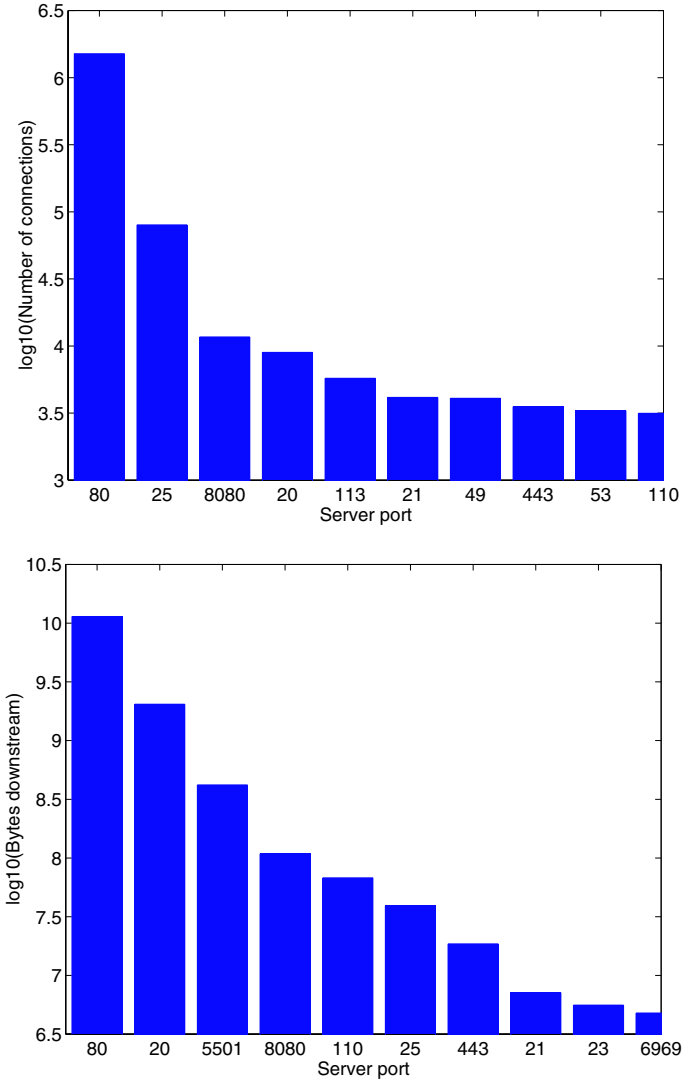

Figure 6.29 Top 10 ports (log-linear) by number of connections (top) and number of bytes transmitted (bottom).

Table 6.2 Number of Bytes and Duration of Connection

Service	Bytes per Connection Cli → Serv	Bytes per Connection Serv → Cli	Duration (s)
WWW	551	7,552	17.2
SMTP	26,394	490	40.0
POP3	69	21,494	17.1
FTPdata	0	227,603	10.6
Telnet	339	12,212	148.4

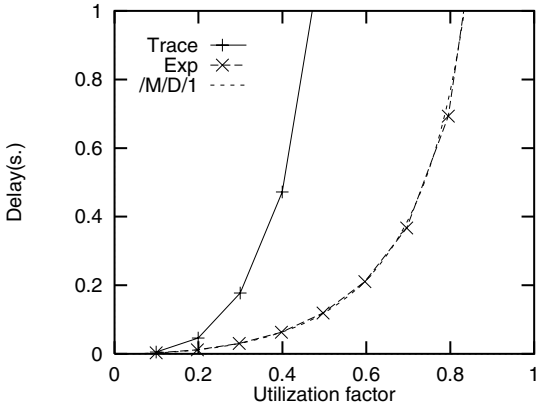

Figure 6.30 Average queue delay versus throughput for (1) Internet trace, (2) Poisson arrival of packets with exponentially distributed length, and (3) Poisson arrival of packets with deterministic length.

6.4.3 Impact on Optical Internet Traffic Engineering

From the previous section it is clear that long-range dependence of Internet traffic produces significant burstiness at any time scale. Such burstiness impacts queueing performance at intermediate routers. Figure 6.30 reports the average queue delay versus throughput curves obtained for (1) Internet trace, (2) Poisson arrival of packets with exponentially distributed length, and (3) Poisson arrival of packets with deterministic length. A dramatic performance degradation is experienced for Internet trace, which is due to the aforementioned long-range dependence effect.

An analytical expression of an infinite-queue single-server system under self-similar input[7] is presented in Norros [95]. The survival function of packet delay in the system is given by

$$P(X > x) \sim \exp\left(-\frac{(C-m)^{2H}}{2k(H)^2 c_v^2 m^2} x^{2-2H}\right) \tag{6.6}$$

where C is the link capacity, m is the input traffic mean rate, $c_v = \sigma/m$ is the standard deviation divided by input traffic mean or marginal distribution coefficient of variation [95], and $k(H) = H^H(1-H)^{1-H}$, H being the Hurst parameter. The equation reveals that besides the utilization factor, the queueing performance depends on the input traffic long-range dependence (H parameter) and the marginal distribution variability (c_v parameter).

In optical networking, depending on the transfer mode adopted in the optical layer—dynamic lightpaths, optical bursts, etc.—the network can work at different operating points (c_v, H). For example, it has been shown that OBS serves to reduce self-similarity (H) [96]. The apparent self-similarity decrease is due to the fact that packets are aggregated to form bursts, which shifts the traffic scaling region to longer time scales. It must be observed, however, that an increase in the coefficient of variation in short time scales implies performance penalties even worse than those resulting from self-similarity [97].

[7]Fractional Gaussian noise.

As an example, assume that bursts of packets are used to enable the transfer of files swiftly across the network using in a single optical burst. Assuming file transmission at a line speed of 1 Gb/s using the same sequence of files as is detected in the collected Internet trace, i.e., using the time stat of the file transfer and size of the file, few curves are computed. The bits per second time series is depicted in Figure 6.31, and the marginal probability density function is plotted in Figure 6.32.

Clearly, the marginal distribution variability is increased and cannot be modeled as Gaussian. In order to model such a high-speed traffic stream, a family of random processes that have higher variability in the marginal distribution in comparison to a Gaussian process must be used [92]. Specifically, the *α-stable random processes* [98] provide non-Gaussian marginal distributions. Such processes are characterized by a higher variability (infinite variance in the marginal distribution) and do not have an analytical closed form for such a marginal distribution. The characteristic function is given by

$$E[e^{i\theta X}] = \begin{cases} e^{-\sigma|\theta|^{\alpha}(1-i\beta(\text{sign}\theta)\tan(\pi\alpha/2))+i\mu\theta}, & \alpha \neq 1 \\ e^{-\sigma|\theta|(1+i\beta(2/\pi)(\text{sign}\theta)\ln|\theta|)+i\mu\theta}, & \alpha = 1 \end{cases} \tag{6.7}$$

with $1 < \alpha < 2$.

We note that the resulting traffic stream can be best modeled with an α-stable random variable with the following parameters: $\alpha = 1.39$, $\beta = 1$, $\sigma = 43{,}297$, and $\mu = 149{,}147$ (bytes per second), as shown in Figure 6.32.

Most interestingly, we note that input traffic now shows independent increments due to the Poissonian nature of connection arrivals and near-infinite bandwidth, which tends to

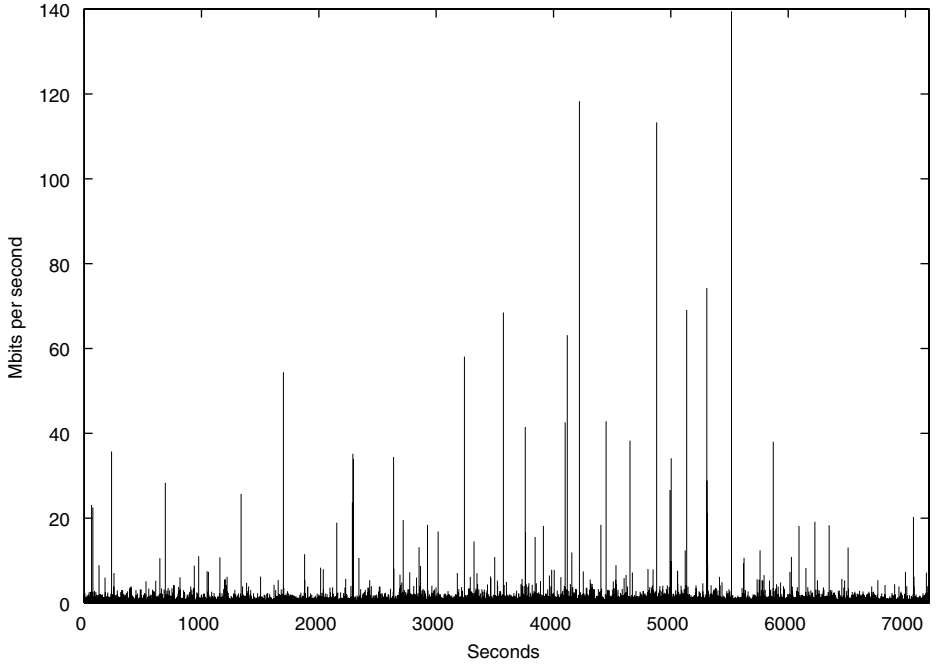

Figure 6.31 Link traffic (bytes per second) for a 1-Gb/s access link.

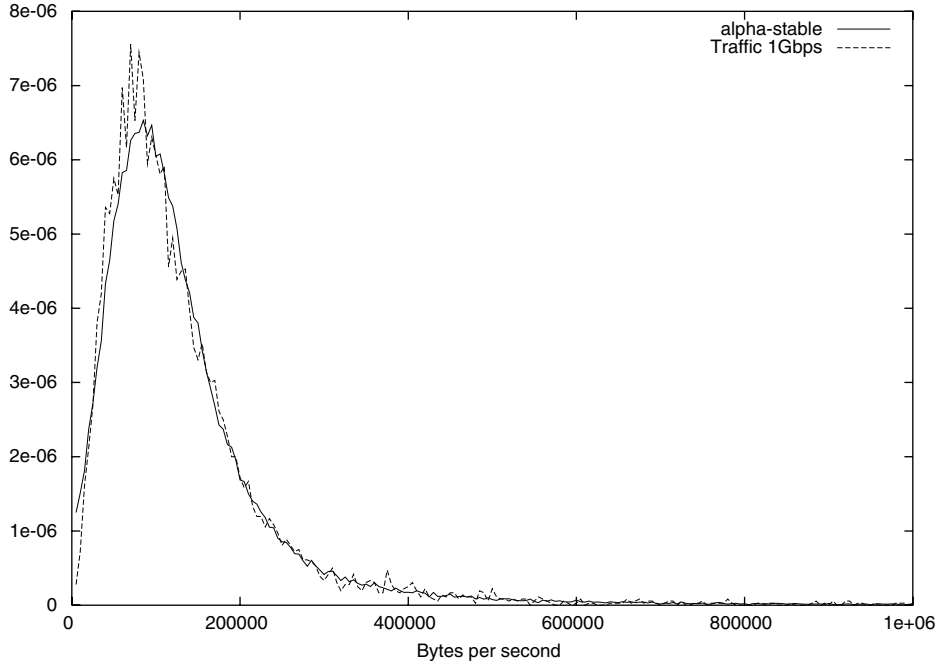

Figure 6.32 Link traffic probability density function for a 1-Gb/s access link.

make connection service times very short. The traffic marginal distribution variability is extremely large, because of the heavy-tailed nature of the size of the files transmitted over the Internet, an inherent characteristic of the current and future Internet. As a result, we note that the use of optical bursts in this scenario of "file-switching" translates dependency into marginal distribution variability. The preceding example serves to illustrate that the optical transfer mode in use, together with the traffic grooming at the edges of the optical subnetwork, may radically change the statistical characteristics of the optical network traffic load.

6.5 OPEN CHALLENGES

This final section summarizes some of the key challenges encountered in the development and deployment of the OI. These challenges include efficient traffic engineering in the OI, adequate network resilience schemes at both the IP layer and WDM layer, and coordination between resilience schemes available at both the IP and the WDM layers.

6.5.1 Traffic Engineering in the Optical Internet

Internet traffic engineering has undergone a significant evolution due to the advances in the statistical modeling of input traffic. From the oversimplified Poisson scenario to the recent discovery of the non-Poisson nature of the Internet traffic [84, 85], this field has spurred considerable research in the area of long-range dependence modeling. However,

today's network operators still face the issue of how to dimension Internet links accurately. Indeed, there are no practical dimensioning rules currently available that can be used in an effective way, such as the well-known Erlang models used in designing telephone networks.

Design of the OL poses additional challenges regarding network dimensioning. The impact of recently proposed optical transfer modes, such as the dynamic lightpath provisioning and optical burst switching, will surely affect traffic characteristics at the network intermediate and destination nodes. The high variability of input traffic may even become more significant than the long-range dependence properties of current Internet traffic, thus requiring further investigation of queueing models with α-stable input [92].

In parallel, flow switching techniques, either in an overlay model (MPLS) or peer model (MPλS), require flow recognition and bandwidth reservation mechanisms that represent a challenge due to the nature of the input traffic pattern. In this regard, a number of "intelligent techniques" are being proposed that aim at performing a priori estimates of the incoming traffic, both at the multiplexed-flow and single-flow levels, in order to proactively adapt the network resources to the incoming traffic demands [99]. Whether such intelligent techniques will satisfactorily solve the problem of resource allocation at gigabit rates still remains an open issue.

6.5.2 Resilience in IP over WDM Networks

In the two-layer IP over WDM architecture, each layer can provide its own independent resilient scheme. Restoration and protection schemes can be implemented at both the IP layer (possibly using the MPLS control plane) and the WDM layer (possibly using the MPλS control plane).

In general, resilient schemes available at the network layer, such as IP (IP/MPLS), have the capability to recover multiple faults and operate at fine traffic granularity. However, these schemes are generally slow, as they require on-line processing upon failure occurrence. Dynamic routing [100] and MPLS protection switching [101] are schemes currently considered to achieve network survivability at the IP (IP/MPLS) layer.

At the OL, both the OCh section and OMS—two of the OL sublayers—feature *dynamic restoration* and *preplanned protection*. Similar to what happens at the higher layers, restoration schemes are more efficient from a capacity viewpoint, but relatively slow (typical restoration completion time is on the order of seconds or minutes). On the other hand, protection schemes may be less efficient, but they guarantee service restoration completion times of hundreds, tens, and even fractions of milliseconds.

The main difference between OCh and OMS resilient schemes is represented by the granularity at which they operate. OCh resilient schemes protect individual lightpaths, thus allowing selective recovery of OLT failures. OMS resilient schemes work at the aggregate signal level, thus recovering all lightpaths present in the failed fiber [102]. OCh resilient schemes require more spare devices, e.g., transponders, multiplexers, demultiplexers, than their OMS counterparts, because each working lightpath must be demultiplexed and multiplexed at every node in order to be individually switched [102].

Although current IP/MPLS reliability schemes offer varying levels of restoration and protection granularity (from aggregate to single flow), they cannot efficiently provide all the necessary functionalities needed by the next-generation Internet, i.e., fast restoration for real-time services. It is thus expected that both IP and WDM layers will each provide some degree of survivability against network faults.

The availability of two resilient schemes in the same network (e.g., IP dynamic routing and OL protection) poses the question of which scheme must be used to protect which traffic. The option of using both resilient schemes for all traffic may not be the most cost-effective one. A more careful design may consist of a hybrid solution in which only part of the traffic is protected by an OL resilient scheme, while the remaining part relies only on IP restoration. Some solutions exploiting this direction have been proposed. More interest on this subject has been originated since the introduction of MPλS, in which the MPLS control plane is responsible for network restoration and protection [26, 103–105].

6.5.2.1 IP over WDM Layer Resilience Coordination The presence of different resilient schemes at different network layers requires coordination between these schemes to avoid their concurrent activation upon a single network fault [104]. In general, such coordination can be achieved in three different ways [106]: ordered response, managed response, and simultaneous response.

Ordered response is commonly achieved by resorting to *escalation strategies* that sequentially activate the different resilient schemes, starting from either the lowest or the highest network layer. Escalation strategies are governed by either explicit messaging between the different layers, or by arbitrarily setting failure detection and restoration completion times [106].

Managed response is a more sophisticated multilayer resilience strategy than the ordered response. Fault recovery procedures are managed under the supervision of a network management system that chooses the most appropriate action to take, depending on the fault nature. This scheme avoids, almost completely, multiple restoration collisions. However, due to its required large quantity of signaling, it is slow. Another potential drawback is the fact that failure of the management center may leave the entire network without any kind of resilience.

In the simultaneous response, each layer utilizes its own resilience scheme without any specific coordination with other schemes available elsewhere in the network. Consequently, multiple resilience schemes may be simultaneously activated by the same fault. This approach seems to be more economical and less complicated than the previous two. In addition, with this approach, unnecessary delays in restoring the service are avoided.

Table 6.3 summarizes some of the fault detection and restoration completion times at the IP/MPLS and WDM layers. It must be noted that MPLS resilient schemes are not yet mature and not yet experimentally tested. Similarly, OCh/OMS restoration schemes are not yet standardized. On the contrary, IP dynamic routing and OCh/OMS protection techniques are already either used in commercial networks or tested in field trials. Their characteristics are thus well known [110].

Table 6.3 Service Restoration Times of Various Resilience Techniques [107–109]

Scheme	Detection Time	Restoration Completion Time
IP dynamic routing	100 ms–180 s	1–100 s
MPLS fast (link) rerouting	0.1–100 ms	50 ms–100 ms
MPLS edge-to-edge rerouting	100 ms–80 s	1–100 s
OCh and OMS restoration	> 100 ms	≥ 50 ms
Dedicated OL protection	1–10 ms	10 μs–10 ms
Shared OL protection	1–10 ms	1–100 ms

It is also worth mentioning emerging and promising MPLS resilient schemes. Fault detection in MPLS can be achieved by exploiting different existing techniques, such as SONET/SDH-based framing alarms, OL fault-detection techniques, and IP-based KEEPALIVE, HELLO, and ICMP messages [100, 109, 111]. These detection mechanisms guarantee detection times ranging from few tenths of a millisecond to hundred of seconds. MPLS solutions may resort to control plane signaling, such as CR-LDP or modified RSVP, to perform traffic restoration, with time intervals ranging from a few dozens of milliseconds to hundreds of seconds. Due to the similarities between the MPLS and MPλS control plane, the same MPLS resilient schemes, and in particular the same signaling, can be utilized in the OL.

6.5.3 Outlook

Advent of the Internet has undoubtedly shaped the way data networks are designed and built. Opposite to the ISO-OSI layered approach, with seven well-defined, rigid, and complex protocol layers from physical to application, which even provide some redundant functionalities, the Internet protocol suite, TCP/IP, provides minimal complexity and superb reliability in the presence of faulty and noisy channels. Internet applications are built on a simple and efficient API, the BSD socket interface, which facilitates design and implementation of Internet clients and servers. As a result, a myriad of Internet applications have appeared, making the Internet evolve into a phenomenal commercial success.

In the current evolution to (gigabit speed) OI, it seems that recently proposed high-speed transfer modes, such as ATM, are being questioned in view of the latest proposed IP over WDM architectures. A parallelism to the OSI-TCP/IP evolution perhaps can be perceived at this point. While the ATM proposed a layered structure with redundant functionalities at the link, network, and transport layers, the IP over WDM architecture presents itself as a simple bandwidth-efficient alternative.

In conclusion, the IP over WDM architecture is a promising approach for realizing the OI. The IP over WDM architecture is still in its experimental phase, with a number of key issues still unresolved. Among these issues, one can find the standardization of transfer modes that support multiple granularities in bandwidth reservation, ranging up to Gb/s rates. Network resilience issues are also of fundamental importance. The challenge here is to introduce adequate resilience levels without creating unnecessary redundancies at the various protocol layers. Finally, it is expected that the OI will consist of multiple technologies and protocols, whose combination will yield network flexibility, ease of use, bandwidth efficiency, and low-cost deployment.

ACKNOWLEDGMENTS

The authors thank James Cai, Isabella Cerutti, Mikel Izal, Daniel Morato, Eduardo Magana, and Marco Tacca for providing some of the results discussed in the chapter.

REFERENCES

1. R. Sabatino and J. M. De Arce, “Implementation of the Pan-European Academic Research Network: TEN-155,” *Computer Networks: the International Journal of Distributed Informatique,* vol. 31, no. 21, pp. 2253–61, November 1999.

2. F. Greisen, "EBONE and Other Parts of the European Internet," in *Proceedings of JENC7. 7th Joint European Networking Conference (JENC7). Networking in the Information Society. TERENA.* Amsterdam, The Netherlands, pp.211/1–6, 1996.
3. A. S. Tanenbaum, *Computer Networks,* Prentice Hall, Upper Saddle River, New Jersey, third edition, 1996.
4. W. R. Stevens, *TCP/IP Illustrated,* vol. 1, Addision-Wesley, Reading, Massachusetts, 1994.
5. P. E. Green, Jr., *Fiber Optic Networks,* Prentice Hall, Englewood Cliffs, New Jersey, 1993.
6. G. P. Agrawal, *Fiber-Optic Communication Systems,* John Wiley & Sons, New York, second edition, 1997.
7. R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective,* Morgan Kaufmann Publishers, 1998.
8. R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective,* Morgan Kaufmann Publishers, second edition, 2001.
9. L. Kazovsky, S. Benedetto, and A. Willner, *Optical Fiber Communication Systems,* Artech House, Boston, Massachusetts, 1996.
10. B. Mukherjee, *Optical Communication Networks,* McGraw-Hill, New York, 1997.
11. T. E. Stern and K. Bala, *Multiwavelength Optical Networks. A Layered Approach,* Addison-Wesley, Reading, Massachusetts, 1999.
12. E. Desurvire, *Erbium-Doped Fiber Amplifier Principles and Applications,* John Wiley & Sons, New York, 1991.
13. P. Green, "Progress in Optical Networking," *IEEE Communications Magazine,* vol. 39, no. 1, pp. 54–61, January 2001.
14. D. A. Francis, S. P. DiJaili, and J. D. Walker, "A Single-chip Linear Optical Amplifier," in *Proceedings of Optical Fiber Communication Conference and Exhibit, 2001. OFC 2001,* vol. 4, pp. PD13–P1-3, 2001.
15. T. N. Nielsen, "Raman Amplifiers in WDM Systems," in *Proceedings of LEOS '99. IEEE Lasers and Electro-Optics Society 1999 12th Annual Meeting,* LEOS, vol. 2, pp. 471–472, 1999.
16. D. J. Bishop, C. R. Giles, and G. P. Austin, "The Lucent LambdaRouter: MEMS Technology of the Future Here Today," *IEEE Communications Magazine,* vol. 40, no. 3, pp. 75–79, March 2002.
17. P. B. Chu, S.-S. Lee, and S. Park, "MEMS: The Path to Large Optical Crossconnects," *IEEE Communications Magazine,* vol. 40, no. 3, pp. 80–87, March 2002.
18. P. De Dobbelaere, K. Falta, L. Fan, S. Gloeckner, and S. Patra, "Digital MEMS for Optical Switching," *IEEE Communications Magazine,* vol. 40, no. 3, pp. 88–95, March 2002.
19. R. Izmailov, S. Ganguly, T. Wang, Y. Suemura, Y. Maeno, and S. Araki, "Hybrid Hierarchical Optical Networks," *IEEE Communications Magazine,* vol. 40, no. 11, pp. 88–94, November 2002.
20. E. Ciaramella, "Introducing Wavelength Granularity to Reduce the Complexity of Optical Cross Connects," *IEEE Photonics Technology Letters,* vol. 12, no. 6, pp. 699–701, June 2000.
21. ITU, *Draft ITU-T Recommendation for Optical Transport Networks, G.872,* July 1998.
22. I. Chlamtac, A. Ganz, and G. Karmi, "Lightpath Communications: A Novel Approach to High Bandwidth Optical WAN-s," *IEEE Transactions on Communication,* vol. 40, no. 7, pp. 1171–1182, July 1992.
23. T.-H. Wu, *Fiber Optic Survivability,* Boston, Massachusetts, Artech House, 1992.
24. I. Cerutti, A. Fumagalli, and M. J. Potasek, "Effects of Chromatic Dispersion and Self-phase Modulation in Multihop Multirate WDM Rings," *IEEE Photonics Technology Letters,* vol. 14, no. 3, March 2002.
25. I. Cerutti, A. Fumagalli, M. Tacca, A. Lardies, and R. Jagannathan, "The Multi-hop Multi-rate

Wavelength Division Multiplexing Ring," *Journal of Lightwave Technology,* vol. 18, no. 12, pp. 1649–1656, December 2000.

26. A. Fumagalli and L. Valcarenghi, "IP Restoration and WDM Protection: Is There an Optimal Choice?" *IEEE Network,* November/December 2000.
27. A. Fumagalli, M. Tacca, and I. Cerutti, "Minimizing the Number of Optical Crossconnect Ports in Mesh Networks Based on Bidirectional Line-switched WDM Self-healing Ring Protection," in *Proceedings of Eight International Conference on Computer Communications and Networks, ICCCN99,* pp. 398–403, 1999.
28. H. Zang, J. P. Jue, and B. Mukherjee, "A Review of Routing and Wavelength Assignment Approaches for Wavelength-Routed Optical WDM Networks," *Optical Networks Magazine,* vol. 1, no. 1, pp. 47–60, January 2000.
29. R. Ramaswami and A. Segall, "Distributed Network Control for Optical Networks," *IEEE/ACM Transactions on Networking,* vol. 5, no. 6, pp. 936–943, December 1997.
30. M. Kuznetsov, N. M. Froberg, S. R. Henion, H. G. Rao, J. Korn, K. A. Rauschenbach, E. H. Modiano, and V. W. S. Chan, "A Next-Generation Optical Regional Access Network," *IEEE Communications Magazine,* vol. 38, no. 1, pp. 66–72, January 2000.
31. N. M. Froberg, S. R. Henion, H. . Rao, B. K. Hazzard, S. Parikh, B. R. Romkey, and M. Kuznetsov, "The NGI ONRAMP Test Bed: Reconfigurable WDM Technology for Next Generation Regional Access Networks," *Journal of Lightwave Technology,* vol. 18, no. 12, December 2000.
32. B. Ganguly and V. Chan, "A Scheduled Approach to Optical Flow Switching in the ONRAMP Optical Access Network Testbed," in *Proceedings of Optical Fiber Communication Conference and Exhibit 2002. OFC2002,* 2002.
33. J. Cai and A. Fumagalli, "Distributed On-line Wavelength Assignment in Multi-Token Based WDM Ring Networks," Tech. Rep. EE-03-00, University of Texas at Dallas, 2000.
34. J. Cai, A. Fumagalli, and C. Guan, "Centralized vs. Distributed On-demand Bandwidth Reservation Mechanisms in WDM Ring," in *Proceedings of Optical Fiber Communication Conference and Exhibit, OFC 2001,* vol. 1, pp. MH2/1–MH2/3, 2001.
35. A. Banerjee, J. Drake, J. P. Lang, B. Turner, K. Kompella, and Y. Rekhter, "Generalized Multiprotocol Label Switching: an Overview of Routing and Management Enhancements," *IEEE Communications Magazine,* vol. 39, no. 1, pp. 144–150, January 2001.
36. A. Banerjee, L. Drake, L. Lang, B. Turner, D. Awduche, L. Berger, K. Kompella, and Y. Rekhter, "Generalized Multiprotocol Label Switching: an Overview of Signaling Enhancements and Recovery Techniques," *IEEE Communications Magazine,* vol. 39, no. 7, pp. 144–151, July 2001.
37. C. Qiao and Dahai Xu, "Distributed Partial Information Management (DPIM) Schemes for Survivable Networks—Part I," in *Proceedings of INFOCOM 2002,* vol. 1, 2002.
38. S. Sengupta and R. Ramamurthy, "From Network Design to Dynamic Provisioning and Restoration in Optical Cross-Connect Mesh Networks: An Architectural and Algorithmic Overview," *IEEE Network,* vol. 15, no. 4, July/August 2001.
39. E. Bouillet, J. F. Labourdette, G. Ellinas, R. Ramamurthy, and S. Chaudhuri, "Stochastic Approaches to Compute Shared Mesh Restored Lightpaths in Optical Network Architectures," in *Proceedings of INFOCOM 2002,* vol. 1, 2002.
40. X. Su and C.-F. Su, "An Online Distributed Protection Algorithm in WDM Networks," in *ICC 2001,* 2001, vol. 5, pp. 1571–1575.
41. D. Elie-Dit-Cosaque, M. Ali, and L. Tancevski, "Informed Dynamic Shared Path Protection," in *Proceedings of OFC 2002,* pp. 492–493, 2002.
42. B. Zhou and H. T. Mouftah, "Survivable Alternate Routing for WDM Networks," in *Proceedings of OFC 2002,* 2002.

43. C.-X. Chi, D.-W. Huang, D. Lee, and X.-R. Sun, "Lazy Flooding: A new Techniuqe for Siganling in All Optical Network," in *Proceedings of OFC 2002,* pp. 551–552, 2002.
44. Luca Valcarenghi, *Survivable IP-over-WDM Networks,* Ph.D. Thesis, University of Texas at Dallas, December 2001.
45. G. Li, J. Yates, R. Doverspike, and D. Wang, "Experiments in Fast Restoration Using GMPLS in Optical/Electronic Mesh Networks," in *Proceedings of OFC 2001,* vol. PD, pp. PD34_1–PD34_3, March 2001.
46. R. R. Iraschko and W. D. Grover, "A Highly Efficient Path-Restoration Protocol for Management of Optical Network Transport Integrity," *IEEE Journal on Selected Areas in Communications,* vol. 18, pp. 779–794, May 2000.
47. J. Yates, G. Smith, P. Sebos, C. Cannon, P. Arias, J. Rice, and A. Greenberg, "IP Control of Optical Networks: Design and Experimentation," in *Proceedings of OFC 2001,* vol. 1, pp. MH5_1–MH5_3, March 2001.
48. L. Valcarenghi and A. Fumagalli, "Implementing Stochastic Preplanned Restoration with Proportional Weighted Path Choice in IP/GMPLS/WDM Networks," *Photonic Network Communications,* vol. 4, no. 3/4, July/December 2002.
49. A. Carena, M. D. Vaughn, R. Gaudino, M. Shell, and D. J. Blumenthal, "OPERA: An Optical Packet Experimental Routing Architecture with Label Swapping Capability," *Journal of Lightwave Technology,,* vol. 16, no. 12, pp. 2135–2145, December 1998.
50. P. Toliver, I. Glesk, R. J. Runser, K.-L. Deng, B. Y. Yu, and P. R. Prucnal, "Routing of 100 Gb/s Words in a Packet-switched Optical Networking Demonstration (POND) Node," *Journal of Lightwave Technology,* vol. 16, no. 12, pp. 2169–2180, December 1998.
51. P. Toliver, K.-L. Deng, I. Glesk, and P. R. Prucnal, "Simultaneous Optical Compression and Decompression of 100-Gb/s OTDM Packets Using a Single Bidirectional Optical Delay Line Lattice," *IEEE Photonics Technology Letters,* vol. 11, no. 9, pp. 1183–1185, September 1999.
52. D. K. Hunter and I. Andronovic, "Approaches to Optical Internet Packet Switching," *IEEE Communications Magazine,* vol. 38, no. 9, pp. 116–122, September 2000.
53. I. Chlamtac, A. Fumagalli, et al., "CORD: Contention Resolution by Delay Lines," *IEEE Journal on Selected Areas in Communications,* vol. 14, no. 5, pp. 1014–1029, June 1996.
54. F. Masetti et al., "High Speed, High Capacity ATM Optical Switches for Future Telecommunication Transport Networks," *IEEE Journal on Selected Areas in Communications,* vol. 14, no. 5, pp. 979–998, June 1996.
55. C. Guillemot et al., "Transparent Optical Packet Switching: The European ACTS KEOPS Project Approach," *Journal of Lightwave Technology,* vol. 16, no. 12, pp. 2117–2134, December 1998.
56. D. K. Hunter et al., "WASPNET: A Wavelength Switched Packet Network," *IEEE Communications Magazine,* vol. 37, no. 3, March 1999.
57. I. Chlamtac and A. Fumagalli, "An Optical Switch Architecture for Manhattan Networks," *IEEE Journal on Selected Areas in Communications,* vol. 11, no. 4, May 1993.
58. C. Qiao, "Labeled Optical Burst Switching for IP-over-WDM Integration," *IEEE Communications Magazine,* vol. 38, no. 9, pp. 104–114, September 2000.
59. L. Xu, H. G. Perros, and G. Rouskas, "Techniques for Optical Packet Switching and Optical Burst Switching," *IEEE Communications Magazine,* vol. 39, no. 1, pp. 136–142, January 2001.
60. P. Mehrotra, I. Baldine, D. Stevenson, and P. Franzon, "Network Processor Design for Optical Burst Switched Networks," in *Proceedings of 14th Annual IEEE International ASIC/SOC Conference,* pp. 296–300, 2001.
61. J. Cai and A. Fumagalli, "An Analytical Framework for Performance Comparison of Bandwidth Reservation Schemes in WDM Ring," in *Proceedings of INFOCOM 2002,* 2002.

62. E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," RFC 3031, January 2001.
63. A. Malis and W. Simpson, "PPP over SONET/SDH," RFC 2615, June 1999.
64. A. Malis and W. Simpson, "PPP in HDLC-like Framing," RFC 1662, July 1994.
65. J. Manchester, J. Anderson, B. Doshi, and S. Davida, "IP over SONET," *IEEE Communications Magazine,* May 1998.
66. J. Carlson, P. Langner, E. Hernandez-Valencia, and J. Manchester, "PPP over Simple Data Link (SDL) Using SONET/SDH with ATM-like Framing," RFC 2823 (Experimental), May 2000.
67. R. Seifert, *Gigabit Ethernet,* Addison-Wesley, Reading, Massachusetts, 1998.
68. D. H. Su, "Standards: The IEEE P802.3ae Project for 10 Gb/s Ethernet," *Optical Networks Magazine,* vol. 1, no. 4, October 2000.
69. N. Jones and C. Murton, "Extending PPP over SONET/SDH with Virtual Concatenation, High Order and Low Order Payloads," Internet Draft <draft-ietf-pppext-posvcholo-06.txt>, December 2001.
70. B. Rajagopalan, D. Pendarakis, D. Saha, R. S. Ramamurthy, and K. Bala, "IP over Optical Networks: Architectural Aspects," *IEEE Communication Magazine,* September 2000.
71. K. Nagami, H. Esaki, Y. Katsube, and O. Nakamura, "Flow-Aggregated, Traffic Driven Label Mapping in Label-Switching Networks," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 6, pp. 1170–1177, June 1999.
72. M. Ilvesmaki, M. Luoma, and R. Kantola, "Flow Classification in Traffic Based Multilayer IP Switching—Comparison between Conventional and Neural Approach," *Computer Communications,* vol. 21, pp. 1184–1194, 1998.
73. Y. Rekhter, B. Davie, D. Katz, E. Rosen, and G. Swallow, "Tag Switching Architecture Overview," RFC 2105, February 1998.
74. *IEEE Communications Magazine, Special Issue on MPLS,* December 1999.
75. H. H. Tzeng and T. Przygienda, "On Fast Address-Lookup Algorithms," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 6, pp. 1067–1082, June 1999.
76. S. Nilsson and G. Karlsson, "IP-Address Lookup Using LC-Tries," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 6, June 1999.
77. M. Waldvogel, G. Varghese, J. Turner, and B. Plattner, "Scalable High Speed IP Routing Lookups," in *Proceedings of ACM SIGCOMM Symposium on Communications Architectures and Protocols,* Cannes, France, pp. 25–35, 1997.
78. M. Degermark, A. Brodnik, S. Carlsson, and S. Pink, "Small Forwarding Tables for Fast Routing Lookups," in *Proceedings of ACM SIGCOMM Symposium on Communications Architectures and Protocols,* Cannes, France, pp. 3–14, 1997.
79. K. Sklower, "A Tree-Based Packet Routing Table for Berkeley Unix," in *Proceedings of Winter USENIX,* Dallas, TX, pp. 93–99, 1991.
80. W. Doeringer, G. Karjoth, and M..Nassehi, "Routing on Longest-Matching Prefixes," *IEEE/ACM Transactions on Networking,* vol. 4, no. 1, pp. 86–97, February 1996.
81. D. Awduche and Y. Rekhter, "Multiprotocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Crossconnects," *IEEE Communications Magazine,* vol. 39, no. 3, pp. 111–116, March 2001.
82. T. M. Chen and W. Wu, "Multi-protocol Lambda Switching for IP over Optical Networks," in *Proceedings of SPIE,* November 2000.
83. C. Guillemot and F. Clerot, "Optical Packet Switching for WDM IP Gigabit Routers," in *Proceedings of 24th European Conference on Optical Communication (ECOC),* vol. 1, pp. 433–434, 1998.

84. V. Paxson and S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking,* vol. 4, no. 2, pp. 226–244, April 1996.
85. W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the Self-similar Nature of Ethernet Traffic," *IEEE/ACM Transactions on Networking,* vol. 2, no. 1, pp. 1–15, January 1994.
86. B. Tsybakov and N. D. Georganas, "On Self-similar Traffic in ATM Queues: Definitions, Overflow Probability Bound and Cell Delay Distribution," *IEEE/ACM Transactions on Networking,* vol. 5, no. 30, pp. 397–409, June 1997.
87. J. Beran, *Statistics for Long-Memory Processes,* Chapman & Hall, London, 1994.
88. A. Erramilli, O. Narayan, and W. Willinger, "Experimental Queueing Analysis with Long-Range Dependent Packet Traffic," *IEEE/ACM Transactions on Networking,* vol. 4, no. 2, pp. 209–223, April 1996.
89. M. S. Taqqu, W. Willinger, and R. Sherman, "Proof of a Fundamental Result in Self-Similar Traffic Modeling," *ACM SIGCOMM Computer Communication Review,* 1997.
90. A. Veres and M. Boda, "The Chaotic Nature of TCP Congestion Control," in *Proceedings of IEEE INFOCOM 00,* Tel Aviv, Israel, 2000.
91. W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level," *IEEE/ACM Transactions on Networking,* vol. 5, no. 1, February 1997.
92. J. Aracil and D. Morato, "Characterizing Internet Load as a Multiplex of Variable Rate TCP Streams," in *Proceedings of IEEE International Conference on Computer Communications and Networks,* Las Vegas, Nevada, 2000.
93. M. E. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *IEEE/ACM Transactions on Networking,* vol. 5, no. 6, pp. 835–846, December 1997.
94. G. Miller K. Thompson and R. Wilder, "Wide-Area Internet Traffic Patterns and Characteristics," *IEEE Network,* November/December 1997.
95. I. Norros, "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks," *IEEE Journal on Selected Areas in Communications,* vol. 13, no. 6, pp. 953–962, August 1995.
96. F. Callegati et al., "On Optical Burst Switching and Self-Similar Traffic," *IEEE Communications Letters,* vol. 4, no. 3, March 2000.
97. M. Izal and M. Aracil, "On the Influence of Self-similarity on Optical Burst Switching Traffic," in *Proceedings of Globecom 2002,* Taipei, Taiwan, 2002.
98. G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes,* Chapman & Hall, New York, 1994.
99. *IEEE Journal on Selected Areas in Communications, Special Issue on Intelligent Techniques In High Speed Networks,* February 2000.
100. J. Anderson, S. Manchester, J., A. Rodriguez-Moral, and M. Veeraraghavan, "Protocols and Architectures for IP Optical Networking," *Bell Labs Technical Journal,* vol. 4, no. 1, pp. 105–124, January-March 1999.
101. T. M. Chen and T. H. Oh, "Reliable Services in MPLS," *IEEE Communications Magazine,* vol. 37, no. 12, pp. 58–62, December 1999.
102. O. Gerstel and R. Ramaswami, "Optical Layer Survivability: A Services Perspective," *IEEE Communication Magazine,* vol. 38, no. 3, pp. 104 –113, March 2000.
103. N. Ghani, S. Dixit, and Ti-Shiang Wang, "On ip-over-wdm Integration," *IEEE Communications Magazine,* vol. 38, no. 3, pp. 72–84, March 2000.
104. D. Colle, S. De Maesschalck, C. Develder, P. Van Heuven, A. Groebbens, J. Cheyns, I. Lievens, M. Pickavet, P. Lagasse, and P. Demeester, "Data-centric Optical Networks and Their

Survivability," *IEEE Journal on Selected Areas in Communications,* vol. 20, no. 1, pp. 6–20, Jan. 2002.

105. L. Sahasrabuddhe, S. Ramamurthy, and B. Mukherjee, "Fault Management in IP-over-WDM Networks: WDM Protection versus IP Restoration," *IEEE Journal on Selected Areas in Communications,* vol. 20, no. 1, pp. 21–33, January 2002.
106. D. Johnson, "Survivability Strategies for Broadband Networks," in *Proceedings of Global Telecommunications Conference, 1996. GLOBECOM '96,* London, vol. 1, pp. 452–456, November 1996.
107. S. Ramamurthy and B. Mukherjee, "Survivable WDM Mesh Networks, Part II—Restoration," in *Proceedings of IEEE International Conference on Communications, 1999. ICC '99,* Vancouver, Canada, vol. 3, pp. 2023 –2030, June 1999.
108. C. Metz, "IP Protection and Restoration," *IEEE Internet Computing,* vol. 4, no. 2, pp. 97–102, March-April 2000.
109. G. Hjálmtýsson, P. Sebos, G. Smith, and J. Yates, "Simple IP Restoration for IP/GbE/10GbE Optical Networks," in *Proceedings of Optical Fiber Communication Conference, OFC 2000,* Baltimore, Maryland, March 2000.
110. D. Cavendish, "Evolution of Optical Transport Technologies: From SONET/SDH to WDM," *IEEE Communications Magazine,* vol. 38, no. 6, pp. 164–172, June 2000.
111. N. Ghani, "Lambda-Labeling: A Framework for IP-over-WDM Using MPLS," *Optical Networks Magazine,* vol. 1, no. 2, pp. 45–58, April 2000.