6

Architecture of Optical Transport Networks (OTNs)

This chapter examines the optical transport network (OTN), as defined by the ITU-T. The emphasis is on ITU-T Recommendations G.709, G.872, and several IETF efforts dealing with the OTN. We begin with a look at the concept of a digital wrapper, followed by an introduction to control planes and in-band/out-of-band signaling. We then examine the current digital hierarchy, as defined by the ITU-T, and the North American standards groups. Next, an analysis is made of several mapping and multiplexing arrangements; both subjects are discussed in relation to the SONET/SDH counterparts. The chapter concludes with a study of the OTN layered architecture and how the digital wrapper is being defined by the standards bodies.

THE DIGITAL WRAPPER

User traffic, such as voice, video, and data, is transported through networks by encapsulating the traffic units (the user bits) inside other traffic units. This operation is often called encapsulation, in which the user payload is placed inside the other traffic and is not examined by any node (except for possible errors introduced by the network) until the traffic reaches the final destination. This idea is shown in Figure 6–1.

Many people call encapsulation framing. This term is typically used when referring to the operations that occur between the third layer (such

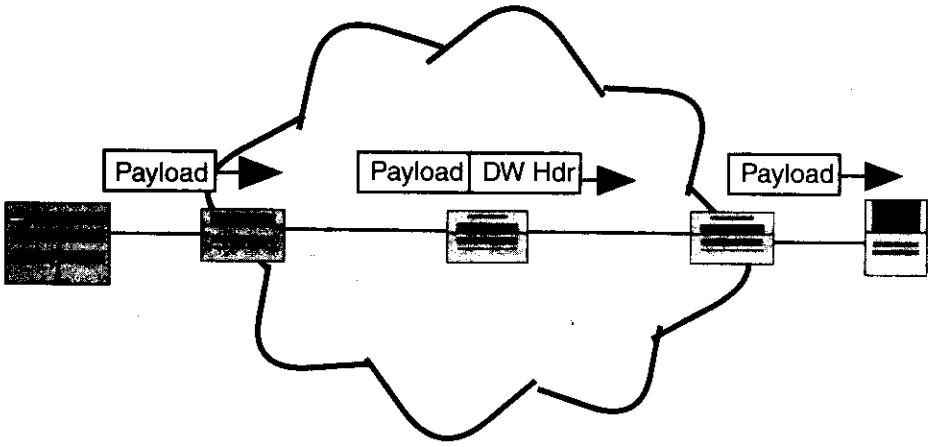


Figure 6–1 Digital wrapper header.

as IP), and second/or first layers of the layered model. As examples, SONET has a framing convention for carrying IP traffic in the layer 1 frames; Ethernet has another convention for encapsulating IP traffic into the layer 2 frames.

The idea of the digital wrapper stems from the concept of encapsulation: wrapping up user traffic, placing headers (maybe trailers) around it, and shipping it through the network. One of the key fields in the digital wrapper is a forward error correction (FEC) value. Its job is to correct bits that have been damaged due to transmission impairments. In a typical OC-192 system, FEC can provide better than 10^{-20} bit error rate (BER) performance with a new fiber system [NORT98]. Using FEC on extended spans can improve the link budget (the accumulated Db values on the end-to-end span) by 2 dB.

Another important field in the digital wrapper header is the protocol ID. It is used to identify the type of traffic that is being transported through the network, such as IP datagrams, OSPF routing packets, and ICMP diagnostic packets.

Many different encapsulation standards exist, and few of them are compatible. As examples, ATM encapsulation is not the same as Frame Relay encapsulation; SONET encapsulation is not the same as Ethernet encapsulation.

So, the idea of the digital wrapper is not new. But the intent is to standardize a digital wrapper for use in the new optical networks.

CONTROL PLANES

A control plane is a set of software and/or hardware in a node that is used to control several vital operations of the network, such as bandwidth allocations, route discovery, and error recovery.

Obviously, the control plane is important. One example of a control plane that has been in existence for some time is the SS7 protocol stack. (In SS7, a control plane is usually called a signaling plane.) Its job is to control the data plane of the telephone network.

The term data plane does not mean the traffic is only data; it might be voice or video traffic. The terms "user plane" and "transport planes" are also used to describe the data plane.

Other examples of control planes are the routing protocols (OSPF, IS-IS, BGP) used in data networks. They enable IP (in the data plane) to forward traffic correctly.

Figure 6–2 shows the relationships of the control plane to the data plane. The control messages are exchanged between nodes to perform a wide variety of operations. For optical networks, some of the more important tasks for the control plane include:

- Exchanging status messages, such as alarms and diagnostics.
- Providing timing messages to keep nodes' clocks in sync with each other.
- Using messages to download information on which wavelengths will be used between two nodes.

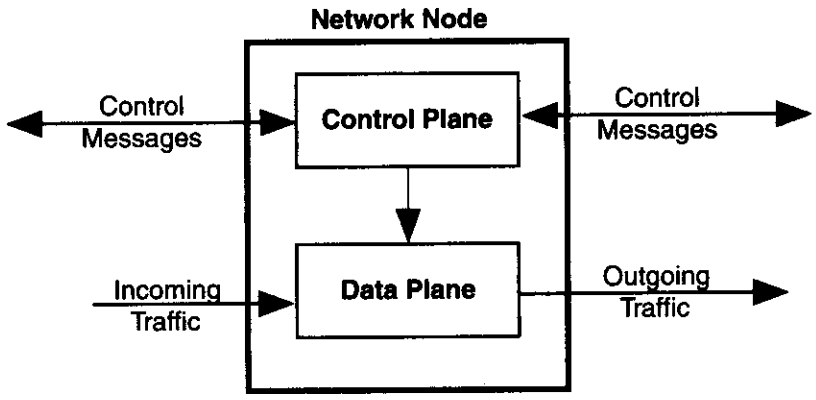


Figure 6–2 The control plane and the data plane.

- Exchanging hello messages to make certain nodes are up and running.
- Setting up and tearing down data plane connections (maybe, depending on the nature of the control plane).
- Building forwarding (cross-connect) tables to allow the data plane to relay traffic from its input link (output port) to its output link (output port).

Second generation digital transport networks such as SONET and SDH have used network management protocols in the data plane (SNMP, CMIP, TL1, or proprietary protocols residing in the DCC bytes) for protection and restoration schemes. This approach has several problems [AWDU01]:

- It leads to relatively slow convergence following a failure. The only way to expedite service recovery in such environments is to preprovision dedicated protection channels.
- It complicates the task of interworking equipment from different manufacturers, and thus different networks.
- It precludes the use of distributed dynamic routing control capabilities.

The approach for third generation digital transport networks is to define a separate, dedicated control plane that can operate in any of the following fashions:

- The control plane messages can be exchanged on a separate physical fiber link from those of the user traffic.
- Alternatively, the control messages can be sent on the same fiber link as the link used for the user traffic, as well as on the same wavelength (possible, but not encouraged).
- Control messages can also be sent on a separate wavelength on the same fiber that is transporting user traffic on the other wavelengths of that fiber.
- Control messages can be sent and received on separate nodes from those that carry the data traffic (somewhat complex in coordinating nodal activities, but permitted).

We will hold the thoughts about the control plane for later chapters; in the discussions on multiprotocol lambda switching (Chapter 10) and link management protocols (Chapter 11), they will be examined in more detail in relation to these specific protocols. Now, we examine where the messages of the control plane flow between the network nodes.

IN-BAND AND OUT-OF-BAND CONTROL SIGNALING

Many systems carry the control plane messages on the same physical link/circuit as the user traffic. For example, the IP-based protocols use this approach where, say, OSPF control plane traffic and IP user plane traffic use the same link. This approach is an in-band control plane, and it uses in-band signaling.

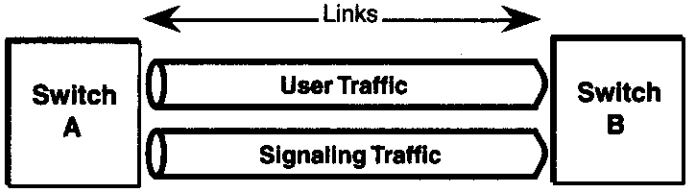
The 3G transport networks use a separate channel for signaling information. This approach is called out-of-band signaling and it is preferred to in-band signaling because it is more efficient and robust. Later discussions will amplify and reinforce this general statement.

Two types of out-of-band signaling exist. With the first type, known as physical out-of-band signaling, a separate physical channel is used for signaling. The second type is called physical in-band/logical out-of-band signaling. With this approach, signaling and user traffic share the same physical channel, but part of the channel capacity is reserved only for signaling traffic; the remainder of the bandwidth is reserved for user traffic, such as IP payloads.

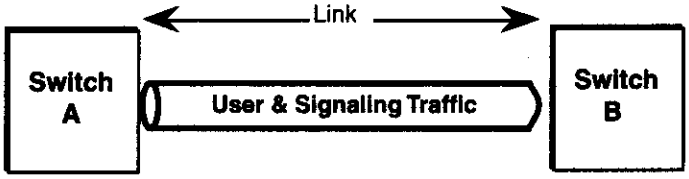
Figure 6–3 shows the differences between these two methods of out-of-band signaling. In Figure 6–3 (a), two physical links are used between two nodes, which are optical nodes in this example. In Figure 6–3 (b), one physical link is used between two nodes, with the signaling traffic allotted reserved bandwidth on the link. Figure 6–3 (b) is an example of the well-known ISDN, using a TDM control slot on the link called the B channel.

Generally, it is not practical or cost-effective to use two separate links (two separate wire-pairs, two separate fibers, etc.) for each residential customer (or a few customers), which would entail installing more cable in the local distribution plant. Consequently, the ISDN approach represents a compromise; the physical channel is shared, but some bandwidth is dedicated to the logical signaling channel.

In addition, ISDN was designed to support a limited set of users on a local loop. It was later adapted for use in high-capacity backbone networks, and it does not provide for redundant links that are used, say, in



(a) Physical Out-of-band Signaling



(b) Physical In-band/Logical Out-of-band Signaling

Figure 6–3 Comparison of signaling systems.

the core of a 1st, 2nd, or 3rd generation transport network. If the link fails, the user (or a few users) is denied service, but the failure does not affect a large population.

SS7 is an example of an out-of-band control plane. SS7 is usually deployed as a separate network within the overall telephone network architecture for the purpose of establishing and terminating telephone calls. If a user link fails, the signaling link is still operable and can continue to support other user calls.

But what happens if the signaling link fails? As shown in Figure 6–4, SS7 and 3G transport networks can be designed to support more than one signaling link; if one link fails, another link is available to take over without the loss of any signaling traffic. Since the optical backbone

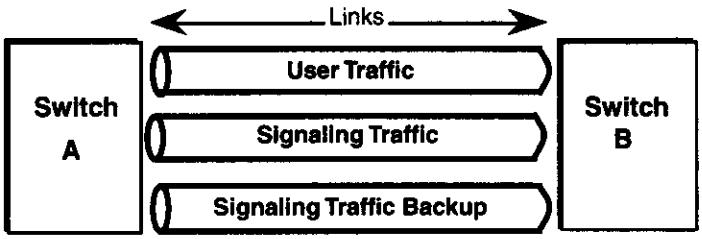


Figure 6–4 Redundant signaling links to provide robustness.

(and, of course, SS7) may support many users (in fact, millions of users), physical signaling with redundant links is quite important.

We can take this discussion one step further and pose another question: "What happens if an optical node fails?" The 3G transport network is sufficiently concerned with downtime (node unavailability), that the network also has redundant nodes, as well as redundant links between the nodes. These topologies are explained in more detail in subsequent chapters.

An In-band Signal on an O/O/O PXC

If an optical node is not capable of O/E/O operations, it can resort to the tried-and-true technique used in telephone systems for many years: the presence or absence (power on/power off) of the physical signal. In fact, many of the restoration mechanisms of optical networks are based on this notion. The presence or absence of a physical optical signal can be used to indicate the absence or presence of problems respectively on the link.

IMPORTANCE OF MULTIPLEXING AND MULTIPLEXING HIERARCHIES

As noted several times in this book, one of the key elements of a digital transport network is multiplexing, the aggregation of lower bandwidth traffic (called tributaries or containers) into higher level tributaries of greater bandwidth. The idea is shown in Figure 6–5, in which lower bandwidth DS0s are multiplexed into higher bandwidth DS1s, which, in turn, are multiplexed into still higher bandwidth DS3s.

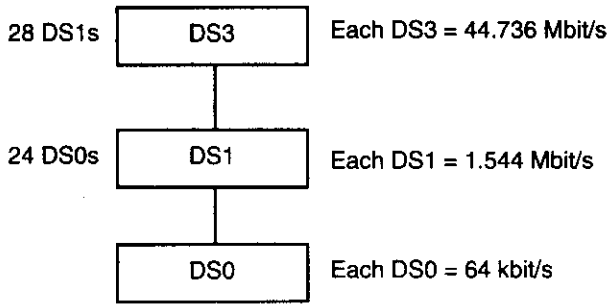
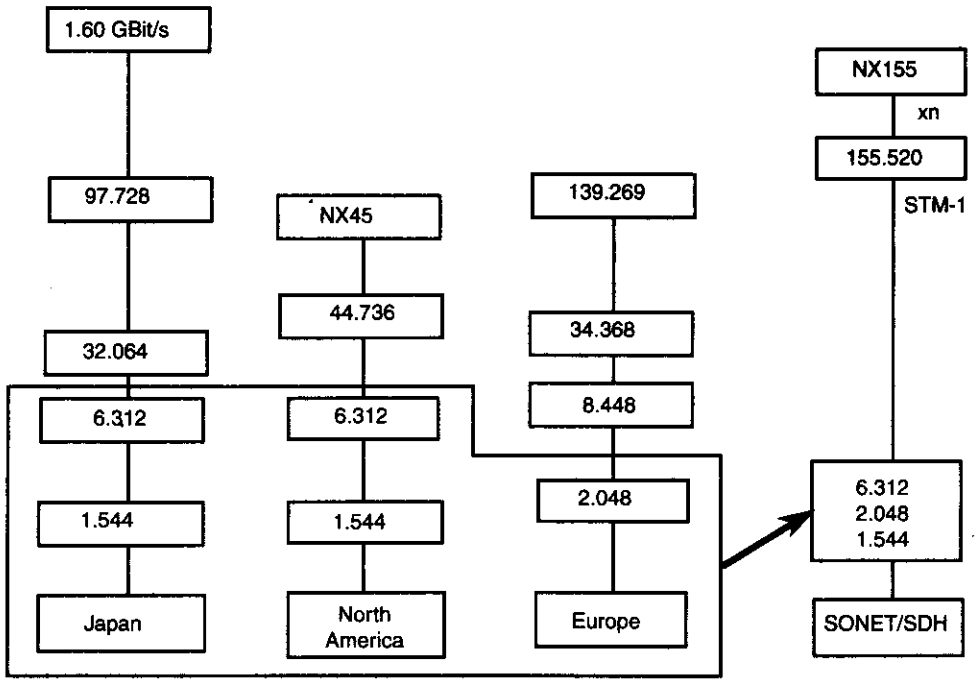


Figure 6–5 Multiplexing levels revisited.

The attractive aspect of this operation is that it allows lower bandwidth payloads to support modest bandwidth needs of certain users, and higher bandwidth payloads to support users that have such needs. At the same time, these tributaries/containers can be combined (multiplexed) for transport across high bandwidth media (optical fiber) at far less expense than transporting individual low bandwidth tributaries. When necessary, the higher level multiplexed tributaries can be demultiplexed (separated) and sent to the appropriate user.

CURRENT DIGITAL TRANSPORT HIERARCHY

SONET and SDH were developed to address several problems that exist with digital transport networks. The one of interest here is the fact that the first generation digital transport multiplexing hierarchy varied (and still varies, but to a lesser extent than before) in the different geographical regions of the world. This disparate approach is complex and makes



Note: Unless noted otherwise, speeds in Mbit/s.

Figure 6–6 SONET/SDH.

the interworking of the systems difficult and expensive. Moreover, it means that companies that build hardware and software for carrier systems must implement multiple commercial platforms for what could be one technology. SONET and SDH provide a standard from which vendors can build compatible multiplexing transport hierarchies. The SONET/SDH multiplexing hierarchy is shown in a general way in Figure 6–6, a slight re-rendering of Figure 2–5 in Chapter 2.

While SONET and SDH do not ensure equipment compatibility, they do provide a basis for vendors to build worldwide standards. Moreover, as shown with the shaded area in Figure 6–6, SONET and SDH are backwards compatible, in that they support the transport carriers' first generation transport systems in North America, Europe, and Japan. This feature is important because it allows different digital signals and hierarchies to operate with a common transport system, which is SONET/SDH.

SONET MULTIPLEXING HIERARCHY

Figure 6–7 shows the details of the SONET multiplexing and mapping hierarchy. The convention is to show the flow of operations going from the right side of the page to the left side. The boxes on the right-most side indicate the user payloads that are multiplexed and mapped into the higher levels of virtual tributaries (VTs), VT groups, and synchronous transport signal (STS) signals. The notation xN indicates the level of multiplexing; that is, how many lower level signals are multiplexed into the next higher level signal. The notation of STS-3c indicates that the lower level tributaries have been joined (concatenated) into higher level signals. This concatenation allows the full payload to be treated as one entity and not as individual tributaries. STS-3c is an optional mapping scheme. In SONET, four VT1.5s are multiplexed together to create a VTG (Virtual Tributary Group of 6.912 Mbit/s). Notice that this hierarchy also supports some of the ITU-T signals (e.g., E1 at 34.368 Mbit/s).

SDH MULTIPLEXING HIERARCHY

Figure 6–8 shows the original SDH multiplexing hierarchy as published in the ITU-T G.707 and G.708 recommendations. The basic multiplexing scheme starts on the right side of the figure and progresses to the left side of the figure. This structure is similar to the SONET structure

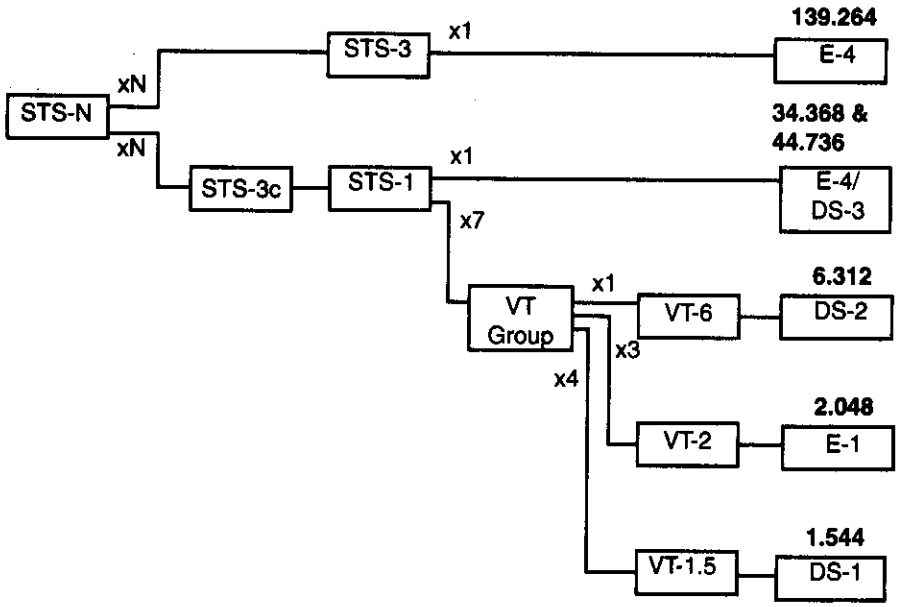


Figure 6–7 SONET multiplexing and mapping hierarchy.

explained earlier. At the lowest level, containers (C) are input to VCs. The purpose of this function is to create a uniform VC payload envelope. Various containers (ranging from 1.544 Mbit/s to 139.264 Mbit/s) are covered by the original SDH standard.

Next, VCs are aligned with TUs. This alignment entails bit stuffing to bring all inputs to a common bit transfer rate. Next, the VCs are aligned to TUs, and pointer processing operations are implemented to denote the position of the tributaries in the payload envelope.

These initial functions allow the payload to be multiplexed into TUGs. As Figure 6–8 illustrates, the Xn indicates the multiplexing integer used to multiplex the TUs to the TUGs. The next step is the multiplexing of the TUGs to higher level VCs, and TUG 2 and 3 are multiplexed into VC-3 and VC-4. These VCs are aligned with bit stuffing for the administrative units (AUs) which are multiplexed into the AU group (AUG). This payload then is multiplexed with an even N integer into the synchronous transport module (STM).

The SONET 1.544 DS1 multiplexing structure, as recommended in ITU-T G.707, is also shown in Figure 6–8. For the purpose of transfer-

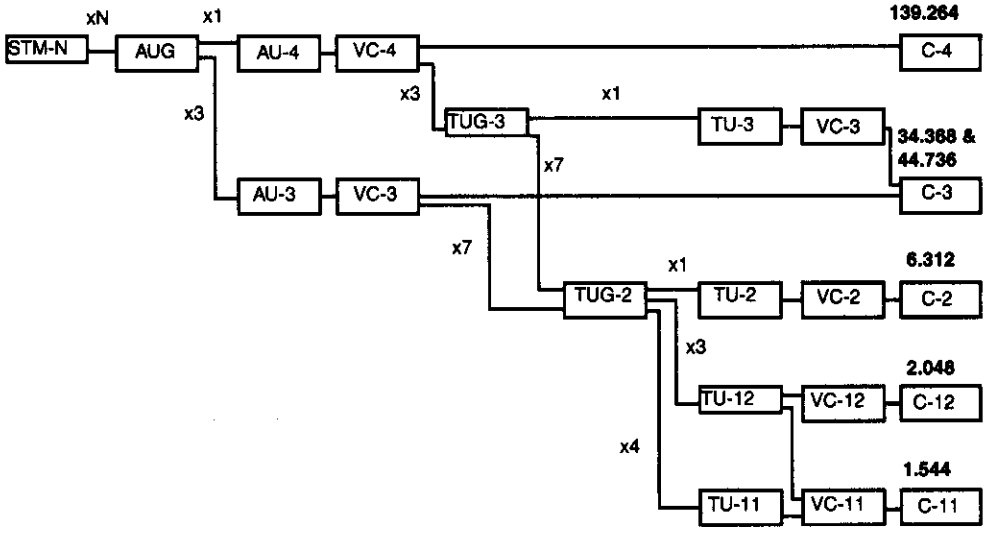


Figure 6–8 The SDH (original) multiplexing hierarchy.

ring information, almost all digital signals of conventional North American multiplexing hierarchy can be employed. Figure 6–8 depicts the SDH terminology as it applies to the North American DS1 (1.544 Mbit/s) signal. This same approach is used to represent the DS1C (3.152 Mbit/s), DS2 (6.312 Mbit/s) and DS3 (44.736 Mbit/s) signals.

Each DS1 consists of 1.544 Mbit/s and is referred to as a container (C). Each container (a C-11) becomes a virtual container (a VC-11) by the addition of path overhead bits (POH) and some stuffed bits in predefined positions. This procedure is called mapping. As noted earlier, the process referred to as aligning relates to the procedure of assembling the virtual container into a tributary unit (a TU-11) in which a pointer is added to indicate the position of the first byte of the virtual container in the tributary unit frame. (Recall that the SONET term for TU-11 is VT1.5.) Thereafter, the process is the same as the other payload mapping and multiplexing operations.

Revised SDH Transport Hierarchy

Figure 6–9 shows the revised SDH digital hierarchy. As the SDH technology matured, and as vendors and network operators gained more

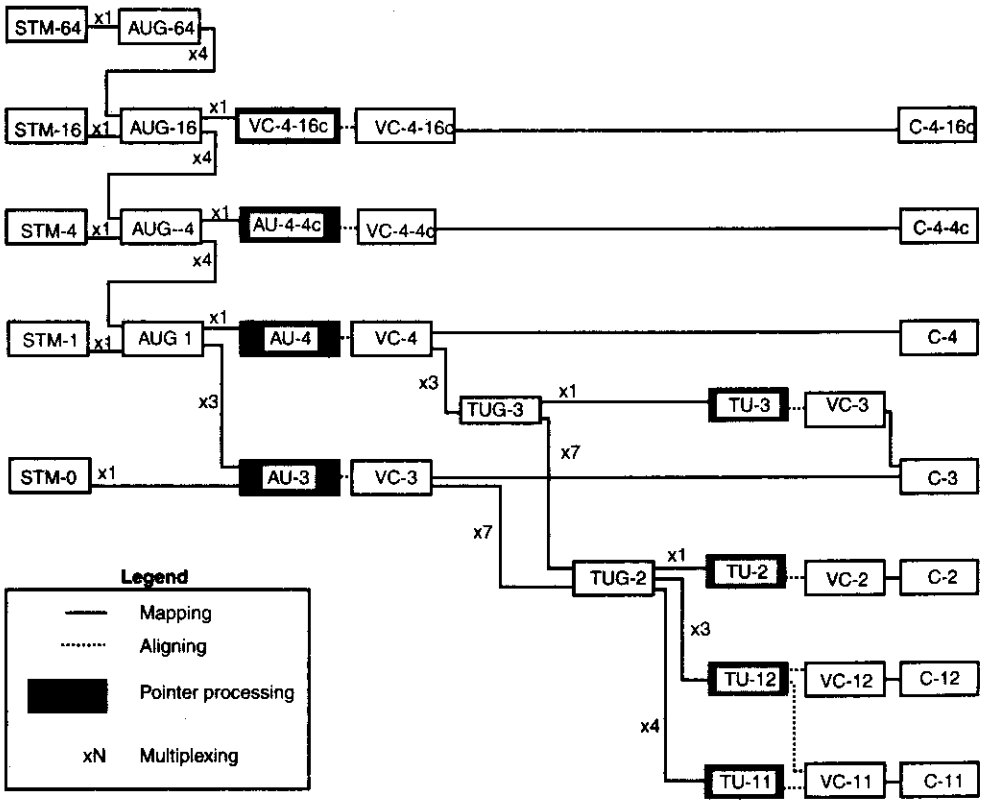


Figure 6–9 The revised SDH digital hierarchy.

experience with SDH, it was recognized that the transport hierarchy model in Figure 6–9 needed to be expanded to show changes to SDH, notably (a) how concatenation fits into the hierarchy, (b) with the resulting AUs and AUGs, and (c) how the resulting STMs (on the left side of Figure 6–9) relate to the AUs and AUGs.

Also included in this figure is a legend that explains where the following operations take place: (a) mapping, (b) alignment, (c) pointer processing, and (d) multiplexing. These operations were defined in the original SDH recommendations, but, to keep matters simple, they were not included in Figure 6–8. Notice also the STM-N frames on the left side of the figure.

KEY INDEXES AND OTHER TERMS

Before we proceed into a more detailed discussion of the G.709 and G.872 operations, it is important to introduce and review several terms that are used to explain bit rates, multiplexing levels, wavelengths supported, and information on enhanced functionality. Keep the entries in Box 6–1 in mind as you read the remainder of this chapter. These entries and associated acronyms are explained where appropriate.

Box 6–1 Key Indexes and Terms

Three terms dealing with the physical signal are: (a) 1R: signal is reamplified, (b) 2R: signal is reamplified and reshaped, (c) 3R: signal is reamplified, reshaped, and retimed.

Several index values are used to explain the multiplexing level and the associated bit rate. They are as follows (see Appendix 2 of [BELL01] for more details), and the terms will make more sense as you proceed through the chapter:

The index k is used to represent a supported bit rate, described as OPUk, ODUk, and OTUk. The value of k=1 represents an approximate bit rate of 2.5 Gbit/s, k=2 represents an approximate bit rate of 10 Gbit/s, k=3 an approximate bit rate of 40 Gbit/s and k=4 an approximate bit rate of 160 Gbit/s (under study by the ITU-T).

The exact bit-rate values in kbits/s for the k units are as follows:

- OPU: k=1: 2 488 320.000, k=2: 9 995 276.962, k=3: 40 150 519.322
- ODU: k=1: 2 498 775.126, k=2: 10 037 273.924, k=3: 40 319 218.983
- OTU: k=1: 2 666 057.143, k=2: 10 709 225.316, k=3: 43 018 413.559

The index m is used to represent the bit rate or set of bit rates supported on the interface. The valid values for m are (1, 2, 3, 12, 23, 123).

The index n is used to represent the order of the OTM, OTS, OMS, OPS, OCG, and OMU. This index represents the maximum number of wavelengths that can be supported at the lowest bit rate supported on the wavelength. It is possible that a reduced number of higher bit-rate wavelengths are supported. The case n=0 represents a single channel without a specific wavelength assigned to the channel.

The index r, if present, is used to indicate a reduced functionality of OTM, OCG, OCC, and OCh.

THE NEW OPTICAL TRANSPORT AND DIGITAL TRANSPORT HIERARCHY

Figure 6–10 shows the new digital transport hierarchy, known as the OTN multiplexing hierarchy [G.87200]. It is evident that the multiplexing architecture of OTN is different from the SONET and SDH schemes. All the multiplexing hierarchies still use the right-to-left multiplexing flow, but OTN has a new set of terms and concepts to describe a third generation digital transport network. The remainder of this chapter examines the entities of the OTN, and you will find Figure 6–10 to be a useful reference as your read the remainder of this chapter.

ODUk Mapping and Multiplexing

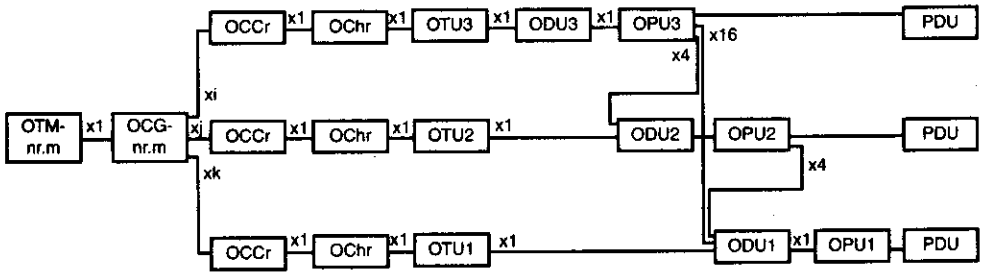
Since G.709 defines at the optical channel layer three distinct client payload bit rates, an optical channel data unit (ODU) frame has been defined for each of these bit rates. An ODUk refers to a bit rate k framing signal, where k = 1 (for 2.5 Gbit/s), 2 (for 10 Gbit/s) or 3 (for 40 Gbit/s).

As shown in Figure 6–10, in optical channel data unit (ODUk) mapping, the client signal is mapped into the optical channel payload unit (OPUk). The OPUk is mapped into an ODUk and the ODUk is mapped into an optical channel transport unit (OTUk). The OTUk is mapped into an OChr and the OChr is then modulated onto an OCCr.

Therefore, these levels of ODUk multiplexing can be defined:

ODU1 multiplexing:

- Four ODU1 are multiplexed into one OPU2, which is mapped into one ODU2.



Note: The multiplexing capabilities of OTM, OCC, and OCG can vary, depending on the values of nr, r, and m.

Figure 6–10 The OTN multiplexing hierarchy.

- Sixteen ODU1 are multiplexed into one OPU3, which is mapped into one ODU3.

ODU2 multiplexing:

- Four ODU2 are multiplexed into one OPU3, which is mapped into one ODU3.

THE OTN LAYERED MODEL

Like most communications systems in place today, optical networks are described by a layered model. This approach is also used to describe parts of the SONET/SDH network, as shown in Figure 6–11. Notice that more than one term is associated with two of the layers. SONET uses the term line and SDH uses the term multiplex for one of the layers. For one

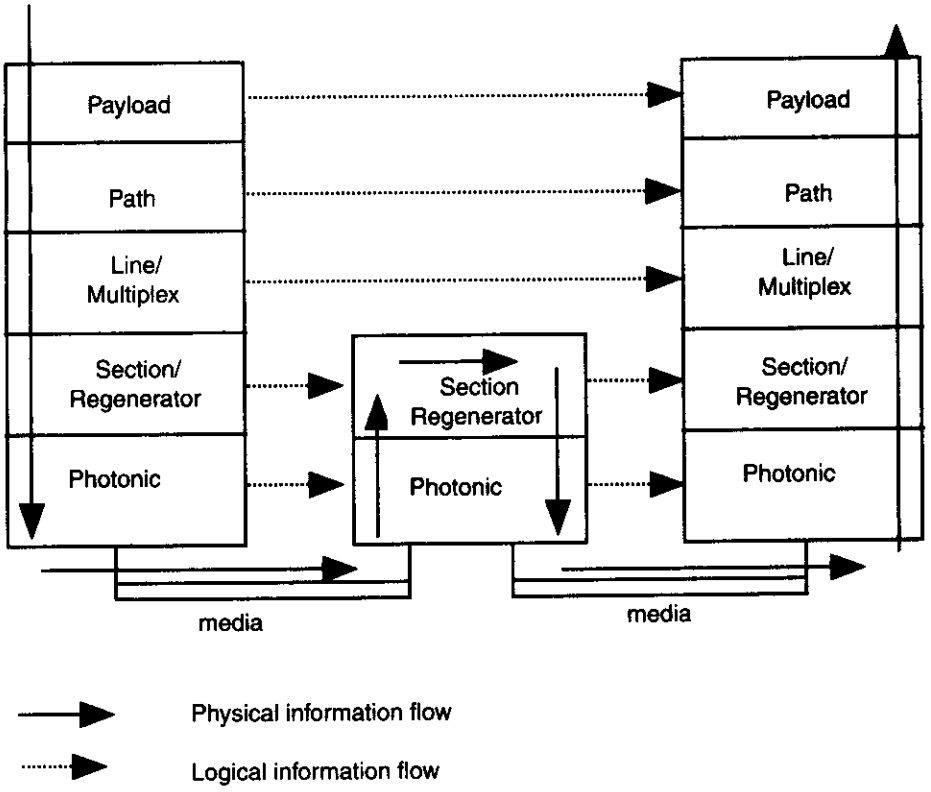


Figure 6–11 The SONET/SDH layered model.

other layer, SONET uses the term section, and SDH uses the term regenerator. The layers have different names, but they are performing the same functions. The functions of the SONET/SDH layers are described in Chapter 5 (with the headers used at these layers).

The ITU-T has published a layered model for next generation optical networks; the general scheme is shown in Figure 6–12 [G.70901] and [G.87200]. The model is designed with several layers. The bottom three layers and a multiplexing operation (shown in the figure as optical channel multiplexing) are collectively called the optical transport hierarchy (OTH). The functions of the layers are:

- *Optical channel (OCh) layer:* Provides end-to-end optical channels between two optical nodes, supporting user (client) payloads of different formats, such as ATM, STM-N, etc. Services include routing, monitoring, provisioning, and backup and recovery features.
- *Optical multiplex section (OMS) layer:* Provides for the support of WDM signals, and manages each signal as an optical channel. Services include wave division multiplexing, and multiplex section backup and recovery.
- *Optical transmission section (OTS) layer:* Provides the transmission of the physical optical signal, based on the specific type of

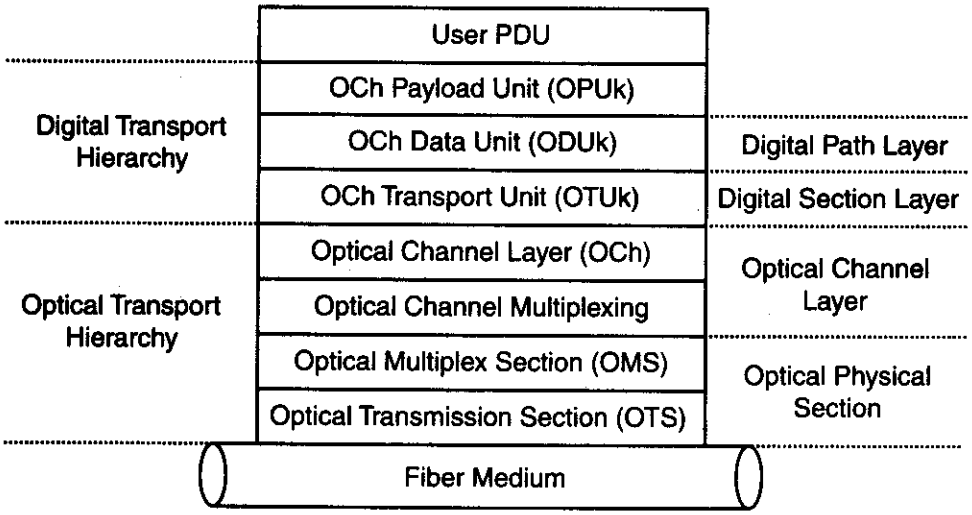


Figure 6–12 The optical network layered architecture.

fiber, such as dispersion-unshifted fiber (USF), dispersion shifted fiber (DSF), etc. Services include the correct signal generation and reception at the section level.

Each optical channel is an optical carrier that supports an optical transport unit (OTU). The bit rates for the OTU are defined by k. An OTUk (where k = 1,2,3) is composed of the entities of OTUk, ODUk, and PDUk. The upper layers are collectively called the digital transport hierarchy, also known as the *digital wrapper* layer. The functions of these layers are:

- *Optical channel payload unit (OPUk):* Provides support to map (digitally wrap) clients' signals (i.e., STM-N signals, IP packets, ATM cells, or Ethernet frames into a structured frame).
- *Optical channel data unit (ODUk) layer:* Provides client-independent connectivity, connection protection and monitoring. The layer is also called the digital path layer.
- *Optical channel transport unit (OTUk):* Provides FEC capabilities and optical section protection and monitoring capabilities. This layer is also called the digital section layer.

The layered model in Figure 6–12 is more elaborate than the SONET/SDH layered model in Figure 6–11 that has been used since the mid-1980s. The reasons for expanding the model is to establish a framework for an optical network to support (a) STDM, data-driven traffic (which is beyond the TDM voice-driven SONET/SDH network), (b) high bandwidth DWDM Tbit/s rates, (c) and enchanted OAM&P without concern for the granularity of the client payload (the user protocol data unit [PDU]) in Figure 6–12.

Another View

Let's look at these entities in a different way. Figure 6–13 shows the user payload (client traffic) encapsulated into an optical payload unit of order k (OPUk). This unit is an octet-based frame consisting of 4 rows and 3810 columns. The first two columns are overhead and the remaining 3808 bytes is for payload. The overhead content varies, depending upon the type of client traffic.

This traffic is encapsulated into an ODUk. This unit is also an octet-based frame, consisting of 4 rows and 3824 columns. Thus, it adds 14 additional columns of ODU overhead, consisting of general communications

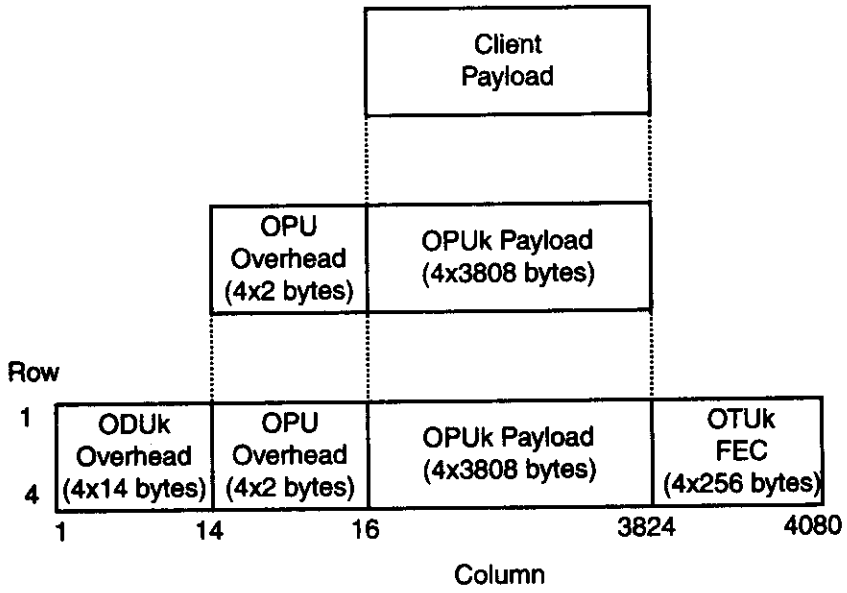


Figure 6–13 OPU, PDU, and OTU.

channels, frame alignment, monitoring and maintenance signals, as well as protection channel bytes.

The OTUk is created from the bytes discussed above, and an additional 256 columns of an FEC field.

Full Functionality Stack: OTM-n.m

Figure 6–14 shows the arrangement for a full functionality stack, designated as OTM-n.m (Optical Transport Module). The OTM-n.m is the information structure used to support OTN interfaces. Up to n OCCs can be multiplexed into an OCG-n.m using wavelength division multiplexing. The OCC tributary slots of the OCG-n.m can be of different sizes, depending on the value of the index m (m = 1, 2, 3, 12, 23, or 123). The OCG-n.m is transported via the OTM-n.m.

Reduced Functionality Stack: OTM-nr.m and OTM-0.r

The OTM also defines stacks that have reduced functionality. This arrangement is used to support optical physical section (OPS) layer connections in the OTN. Two options are permitted, with modifications to the protocol stack. These options are shown in Figure 6–15 and Figure 6–16:

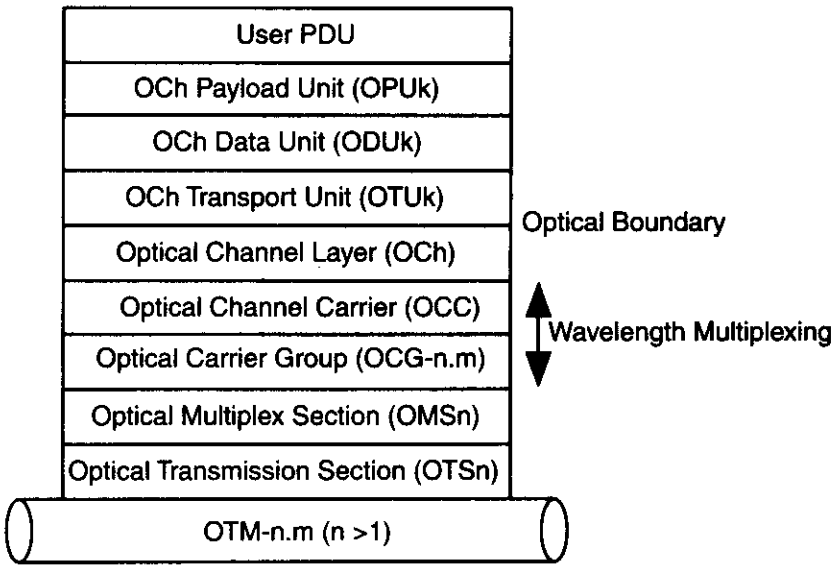


Figure 6–14 Full functional stack.

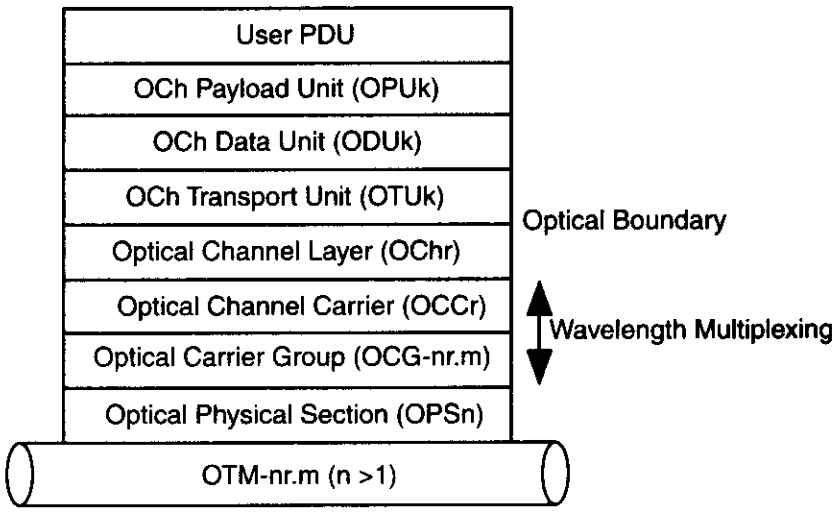


Figure 6–15 OTM-nr.m.

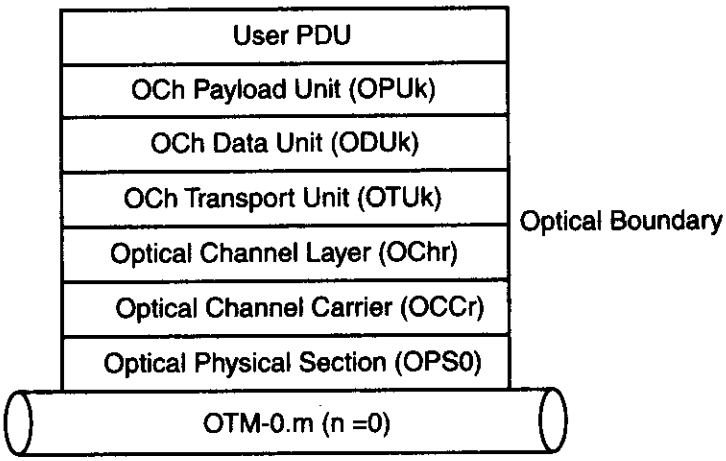


Figure 6–16 OTM-0.

- *OTM-nr.m:* consists of up to n multiplexed optical channels (see Figure 6–15). Up to OCCr are multiplexed into an OCG-nr.m using WDM. The OCCr tributaries of the OCG-r.m can range from m = 1, 2, 3, 12, 23, or 123. The OCG-nr.m is transported via the OTM-nr.m.
- *OTM-0:* consists of a single optical channel without a specific wavelength assigned (see Figure 6–16). This reduced functionality does not support WDM. Only one OCCr tributary slot is provided; thus an OCG-0r.m stack is defined. The OCCr tributary can range from m = 1, 2, 3. The OCCr is transported via the OTM-0.m.

ENCAPSULATION AND DECAPSULATION OPERATIONS

Again, as in other layered protocol models, the OTN defines the relationship of the layers to the traffic units. Figure 6–17 shows this relationship. The arrows on the left side of the figure denote the processing of the traffic going down the layers at the transmit side with the arrow that points down. Conversely, the arrow that points up denotes the processing of the processing of the traffic at the receiving side. For obvious reasons, the operations on the transmit side are called encapsulation, and the operations on the receive side are called decapsulation.

The traffic units on the left side of the figure are called signal types. Here is a description of these signal types. You will also find Figure 6–10 to be helpful during this discussion:

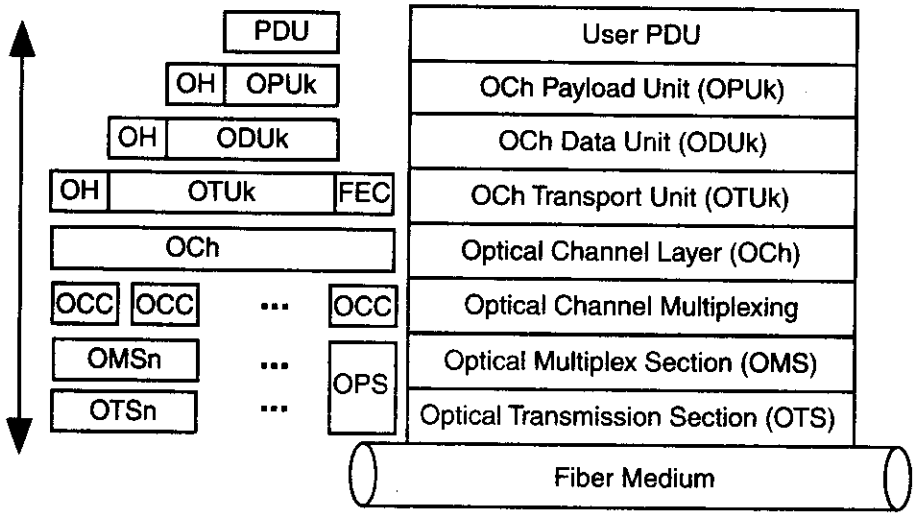


Figure 6–17 Encapsulation and decapsulation.

- *OPUk signal:* Optical channel payload unit (OPU) is defined as a structured signal of order k (k = 1, 2, 3) and is called the OPUk signal. The OPUk frame structure is organized in an octet-based block frame structure with 4 rows and 3810 columns. The two main areas of the OPUk frame (4 x 3810 Bytes) are the OPUk Overhead area (column 15 and 16) and the OPUk payload area (columns 17 to 3824).
- *ODUk signal:* Optical channel data unit (ODU) is defined as a structured signal of order k (k = 1, 2, 3) and is called the ODUk signal. The ODUk frame structure is organized in an octet-based block frame structure with 4 rows and 3824 columns. The two main areas of the ODUk frame are the ODUk Overhead area (columns 1 to 14, with column 1 dedicated to frame alignment and OTUk specific alignment) and the OPUk area (columns 15 to 3824, which are dedicated to the OPUk area).
- *OTUk signal:* Optical channel transport unit (OTU) of order k (k = 1, 2, 3) defines the conditioning for transport over an optical channel network connection. This signal is called the OTUk signal. The OTUk (k = 1,2,3) frame structure is based on the ODUk frame structure and extends it with a forward error correction (FEC). Scrambling is performed after FEC computation and insertion into the OTUk signal. In the OTUk signal, 256 columns are added to

the ODUk frame for the FEC and the reserved overhead bytes in row 1, columns 9 to 14 of the ODUk overhead, are used for OTUk specific overhead, resulting in an octet-based block frame structure with 4 rows and 4080 columns (4 x 4080 bytes).

GENERIC FRAMING PROCEDURE (GFP)

The generic framing procedure (GFP) is used to encapsulate any type of traffic over the optical channel. It follows the construction of the digital wrapper. Its implementation is intended to avoid the use of other encapsulation and framing conventions, such as SDH/SONET and Ethernet. GFP is defined to transport any client layer (defined as OTN ODU of unit-k) over fixed rate optical channels.

One of GFP's principal attributes is that it sets forth the rules for conveying idle frames through the optical network, which most other framing procedures do not define. Knowing if the signals in the networks represent traffic (non-idle) or no traffic (idle) is very important. It allows the synchronous optical network to continue its operations even though it may be transporting asynchronous traffic.

The frame format for GFP is shown in Figure 6–18. The four-byte header is a two-byte PDU length indicator (PLI) and a two-byte header error control (HEC) field. The frame check sequence field (FCS) is optional. The idle frame format includes a null PLI and the HEC field.

GFP defines also two frame-oriented mechanisms. The first is frame multiplexing in which frame multiplexing is performed on a frame-by-frame basis. When no frames are waiting, idle frames are inserted. The second is called the GFP frame delineation; framing of the payload is

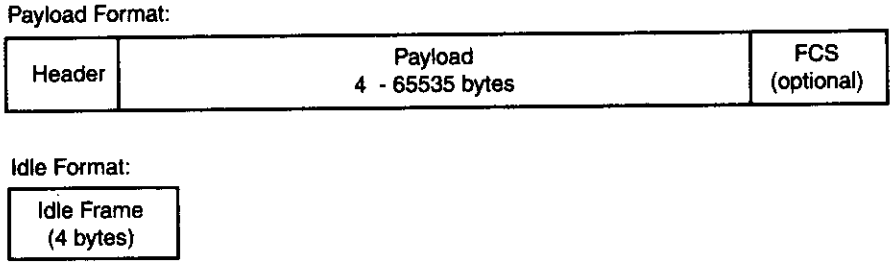


Figure 6–18 The GFP frame format [BELL01].

based on the detection of a correct HEC. After which, PLI is used to find the start of the next frame.

The GFP frames constitute the OPUk payload. The corresponding OPUk overhead is defined by the payload structure identifier (PSI), which includes the following fields: (a) PT: payload type (1-byte) and (b) RES: reserved (254-bytes).

The GFP OPUk (k = 1, 2, 3) capacities are defined so that they can include the following client bit rates: (a) GFP (OPU1): 2,488,320 kbit/s, (b) GFP (OPU2), 9,995,276 kbit/s, and (c): GFP (OPU3): 40,150,519 kbit/s.

The attractive aspect of GFP is by aligning the variable-length byte structure of every GFP frame with the byte structure of the OPUk, there are no restrictions on the maximum frame length. Therefore, a frame may cross the OPUk frame boundary.

SUMMARY

The third generation digital multiplexing scheme is an enhanced revision of the second generation SONET/SDH hierarchy. It has recently been defined, and vendors and network operators are planning the transition to the technology.

We are not finished with the 3G OTN; it is revisited in more detail later, after MPLS and GMPLS have been introduced (if you are in need of this material now, go to Chapter 10, and see "The Next Horizon: GMPLS Extensions for G.709").

7

Wavelength Division Multiplexing (WDM)

Wavelength division multiplexing (or Wave division multiplexing) was introduced in Chapter 1, and has been mentioned several times thereafter. This chapter explains WDM in more detail, including examples of WDM bandwidth allocations, WDM components, and topologies of WDM networks. The chapter concludes with an example of how parts of the WDM spectrum are managed with a wavelength plan.

THE WDM OPERATION

WDM is based on a well-known concept called frequency division multiplexing or FDM. With this technology, the bandwidth of a channel (its frequency domain) is divided into multiple channels, and each channel occupies a part of the larger frequency spectrum. In WDM networks, each channel is called a *wavelength*. This name is used because each channel operates at a different frequency and a different optical wavelength (and the higher the frequency, the shorter the signal's wavelength).

In addition to the term wavelength, the terms frequency slot, lambda, and optical channel are also used to describe the optical WDM network channels. Recall that the term lambda is often noted with the Greek letter of λ. Notwithstanding all these terms, the idea of WDM is to use the optical channels (frequency slots) to carry user traffic.

Figure 7–1 shows a simple example of a WDM link. Four fibers are connected to a WDM multiplexer, which combines or multiplexes them onto one fiber. The opposite operation occurs at the receiving multiplexer, which separates (demultiplexes) the wavelengths and sends them to an appropriate output port, perhaps to other fibers.

As Figure 7–1 shows, each wavelength is separated by an unused spectrum to prevent the signals from interfering with each other. The ITU publishes standards on this spacing. The most common spacing is referred to as a 100 GHz spacing. Others are emerging that pack the wavelengths closer together at spacings of 50 GHz and 25 GHz. The 100 GHz spacing is sometimes cited as 0.8 nm, and the 50 GHz and 25 GHz are sometimes cited as 0.4 and 0.2 nm, respectively. These formulae can be used to determine the frequency of the wavelength and the wavelength spacings:

Frequency (Hz) = speed of light in a vacuum (in meters) / wavelength (in meters)

Wavelength separation = (frequency separation × wavelength^2) / Speed of light (in meters)

The range of frequencies (and, of course, wavelengths) carried in the fiber vary. A common set of wavelengths used today are those in the 1550 nm region; they are called the C band. Table 7–1 lists the WDM

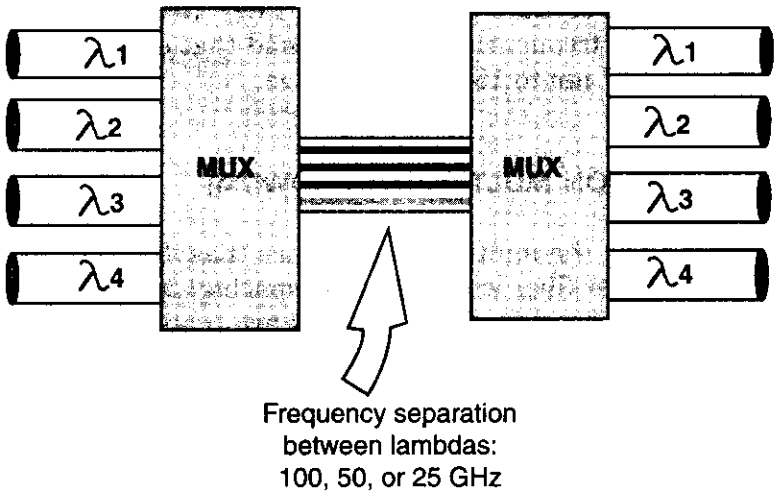


Figure 7–1 The WDM link.

Table 7–1 Examples of the C Band Wavelengths and Frequencies [LIGH01a]

Wavelength (nm)	Frequency (THz)	Wavelength (nm)	Frequency (THz)	Wavelength (nm)	Frequency (THz)
1539.766	194.7	1550.116	193.4	1560.606	192.1
1540.557	194.6	1550.918	193.3	1561.419	192.0
1541.349	194.5	1551.721	193.2	1562.233	191.0
1542.142	194.4	1552.524	193.1	1563.047	191.9
1542.936	194.3	1553.329	193.0	1563.900	191.8
1543.730	194.2	1554.134	192.9	1564.679	191.7
1544.526	194.1	1554.940	192.8	1565.496	191.6
1545.322	194.0	1555.747	192.7	1566.314	191.5
1546.119	193.9	1556.555	192.6	1567.133	191.4
1546.917	193.8	1557.363	192.5	1567.952	191.3
1547.715	193.7	1558.173	192.4	1568.773	191.2
1548.515	193.6	1558.983	192.3	1569.594	191.1
1549.315	193.5	1559.794	192.2	1570.416	190.9

wavelengths and associated frequencies ranging from 1539.766 nm to 1570.416 nm [LIGH01a]. Be aware that all vendors do not use all these wavelengths, and some vendors (and the ITU) define and use wavelengths on both sides of those shown in this table; that is, shorter and longer wavelengths. For example, the ITU G.692 Recommendation defines another band of frequencies in the L Band that operates above the C Band in the 1574.37 nm to 1608.33 nm range.

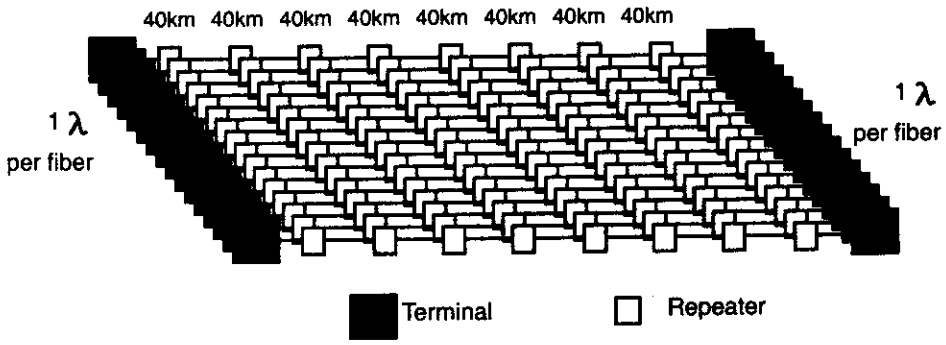
DENSE WAVE DIVISION MULTIPLEXING (DWDM)

DWDM systems allow the multiplexing of more than 160 wavelengths of 10 Gbit/s (1.6 Tbit/s per fiber with a 25 GHz spacing) by using both the C band and the L band spectra. Some vendors are proposing a spacing of 12.5 GHz. Consequently, it will be possible to transmit 320 wavelengths of 10 Gbit/s in a single fiber. A complementary method for increasing the effective capacity of a DWDM system is to include the 1480nm (S band) and 1650nm (U band) together with the deployment of fibers covering an ultra-wide waveband from 1460 to 1675nm (i.e., from the S band to the U band) [BELL01] for a total throughput of 3.2 Tbit/s per fiber.

TDM and WDM Topologies

Figure 7–2 shows the topology for a conventional optical TDM system and a WDM system. The WDM example is based on the Multiwavelength Optical NETworking (MONET) Consortium, funded partly by the Defense Advanced Research Projects Agency (DARPA) [JOHN99].

This example shows two layouts, each having a capacity of 40 Gbit/s. The optical terminals are OC-48 devices. An OC-48 link operates at 2,488.320 Mbit/s. In the TDM system, 16 fiber pairs operate at the OC-48 rate, each carrying one wavelength. In contrast, the WDM operates at the same capacity, using only a single fiber pair.



Conventional TDM System

WDM System

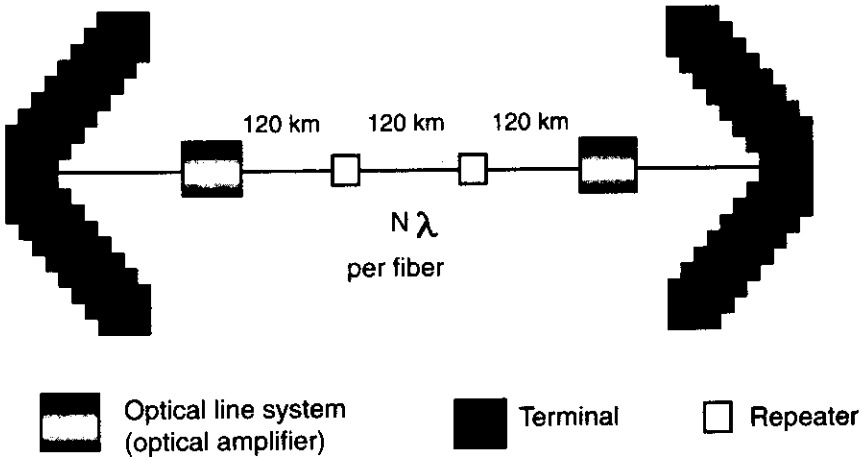


Figure 7–2 TDM and WDM optical systems.

The WDM system also needs fewer intermediate elements such as optical repeaters (also called amplifiers in the literature). Indeed, this figure illustrates a huge difference between TDM and WDM systems: the dramatic decrease in WDM networks in the number of fibers, and other components. In addition, the newer WDM technology does not need to space the amplifiers as closely together as in a TDM system.

RELATIONSHIP OF WDM TO SONET/SDH

Figure 7–3 shows the relationship of WDM to SONET and SDH [NORT98]. As previous discussions have emphasized, the idea of interworking WDM with SONET is to overlay SONET onto the photonic WDM layer.

There are parts of the network in which SONET frames are passed transparently and are not examined. One of these parts is at the multi-wavelength optical repeater (MOR). The decision to process (examine) the

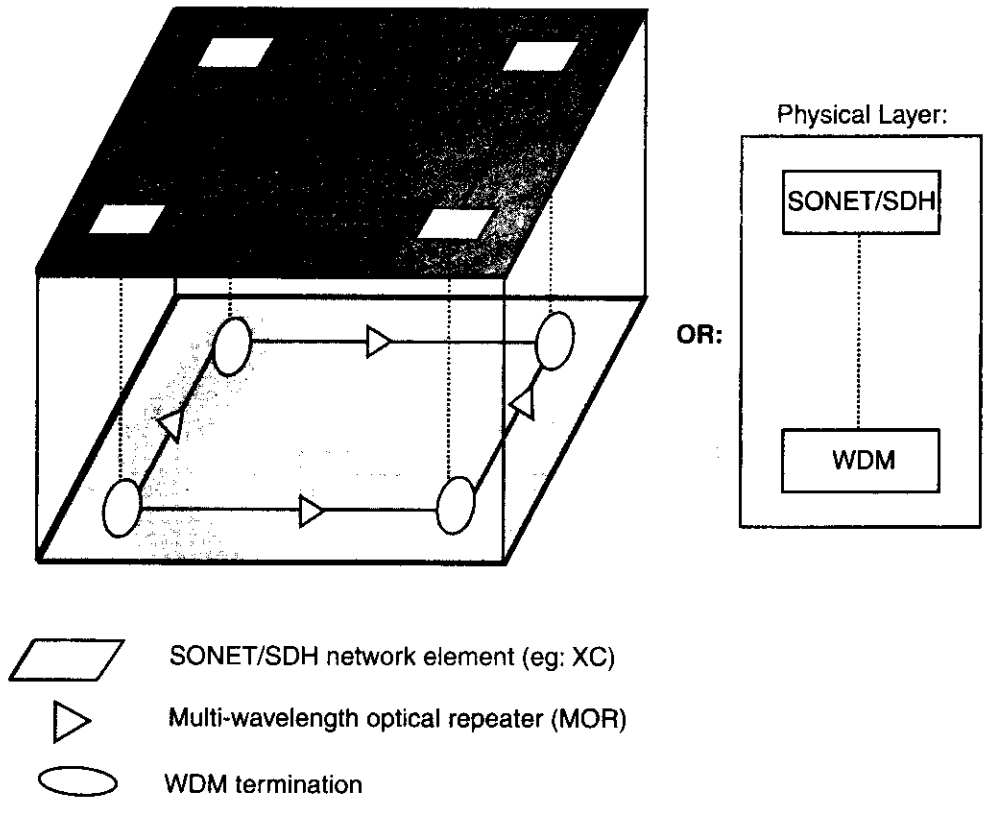


Figure 7–3 Relationship of SONET/SDH to WDM [NORT98].

SDH/SONET bits is based on the need to drop, add, or cross-connect payloads, and/or perform OAM operations, such as testing and diagnostics.

Any termination of WDM signals might remain transparent to SDH/SONET, or termination of WDM and SDH/SONET may occur at one site. This decision is based on the individual network topology and how customer payloads must be processed at each node.

Notice in Figure 7–3 that the SONET/SDH network element, such as a cross-connect, is situated at a WDM termination site. This arrangement makes good sense since an O/E/O operation at an optical node requires the termination of the optical signal. Also, the idea of these arrangements is shown in Figure 7–3 with the physical layer relationships: SDH/SONET layers operate over the WDM layer.

ERBIUM-DOPED FIBER (EDF)

With the advent of the erbium-doped fiber (EDF), the need for electronic circuitry in some of the optical components no longer exists. Moreover, EDF Amplifiers (EDFAs) are transparent to a data rate. They also provide high gain and experience low noise. The major attraction is that all the optical signal channels can be amplified simultaneously at the EDFA in a single fiber. Of course, this approach is the essence of WDM.

EDFAs play a big role in several parts of WDM optical networks. EDFAs can be found in amplifiers, optical cross-connects, wavelength add-drop multiplexers (ADM), and broadcast networks.

They are deployed as in-line amplifiers, in which they amplify an optical signal that has been attenuated by the fiber. They are used to boost optical power at the sending site as the signal enters the fiber (as well as at the receiver). They are found in optical cross-connects, and they are used to compensate for signal loss as well as wavelength ADMs for the same function. Lastly, they are now employed in optical broadcast systems to boost the power for the distribution system.

WDM AMPLIFIERS

One of the key components of WDM optical networks are optical fiber amplifiers. In the past, optical networks used optoelectronic regenerators between optical terminals. These devices converted optical signals to electrical signals and then back to optical signals. This approach required expensive high-speed electronic circuitry, and operated on one signal (one lightwave).

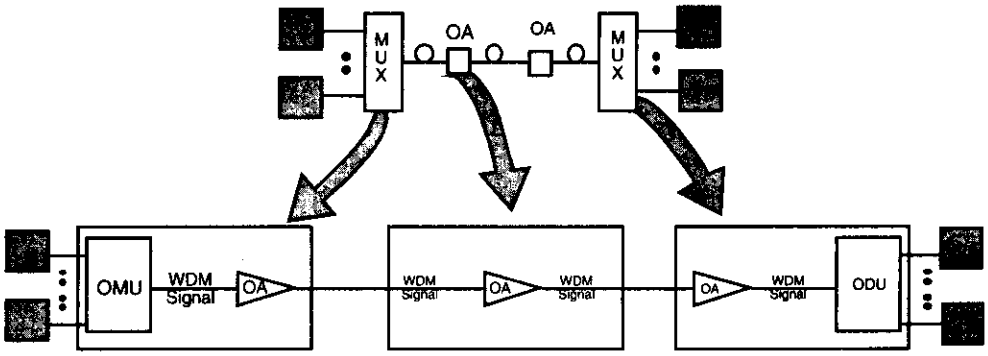


Figure 7–4 Optical amplifiers for WDM.

Figure 7–4 shows a typical schematic diagram for the use of optical amplifiers in a WDM transmission system (also called optical fiber amplifiers (OFAs)). At the sending end, multiple optical channels are combined in an optical multiplexer. This combined signal is amplified before it is launched into the first fiber span. At the receiving end, the opposite operations occur. The incoming WDM signals are amplified by a "preamplifier" and then they are demuxed and sent to their respective receivers.

As mentioned earlier, the optical amplifier is deployed as an in-line amplifier; it amplifies the optical signal that has been attenuated by the fiber. It also boosts optical power at the sending site (the optical multiplexing unit) and the receiving site (the optical demultiplexing unit).

The use of erbium-doped fiber for amplifiers in the late 1980s and early 1990s was a major milestone, leading to the development of a new generation of amplifiers for the 1500 nm wavelength window. The erbium-doped fiber amplifiers (EDFAs) are significantly less expensive than the optoelectronic regenerators. They are oblivious to the bit rate or the data format on the link, so any upgrades do not affect them. In addition, they can amplify multiple WDM wavelengths simultaneously.

Let's review once again three terms dealing with the physical signal: (a) 1R: signal is reamplified, (b) 2R: signal is reamplified and reshaped, (c) 3R: signal is reamplified, reshaped, and retimed. The OFA eliminates the need for 3R operations.

Gain Flatness

The WDM signal must be well-balanced throughout the entire fiber transmission. This means that each wavelength amplification should remain constant with respect to other wavelengths. To maintain this

consistency, the optical amplifier is gain-flattened. One of the problems with conventional erbium-doped fiber amplifiers is that they amplify different wavelengths carried on the fiber. Under certain conditions, there can be a 3 to 4 dB of gain difference per amplifier within the 1530 to 1560 nm window. One approach to handle this problem is to use an equalization filter that controls peak attenuation, center wavelength, and signal width.

ADD-DROP MULTIPLEXERS

As the capacity of optical systems increases, opportunities are created for network service providers to provide more capacity to the systems' users. These users are located in many parts of a geographical region, including business sites, industrial parks, campuses, and stand-alone offices. These diverse sites require great flexibility in bandwidth management to the customers' requirements. In WDM networks, the service provider should be able to provision bandwidth in a fast, efficient, and cost-effective manner to these sites.

One of the key "tools" to support this environment is the wavelength add-drop multiplexer (WADM). Based on earlier TDM ADMs, these devices support the management of fiber capacity by the selective adding, inserting, and removal of WDM channels at intermediate points in the network; a general example of these operations is shown in Figure 7–5.

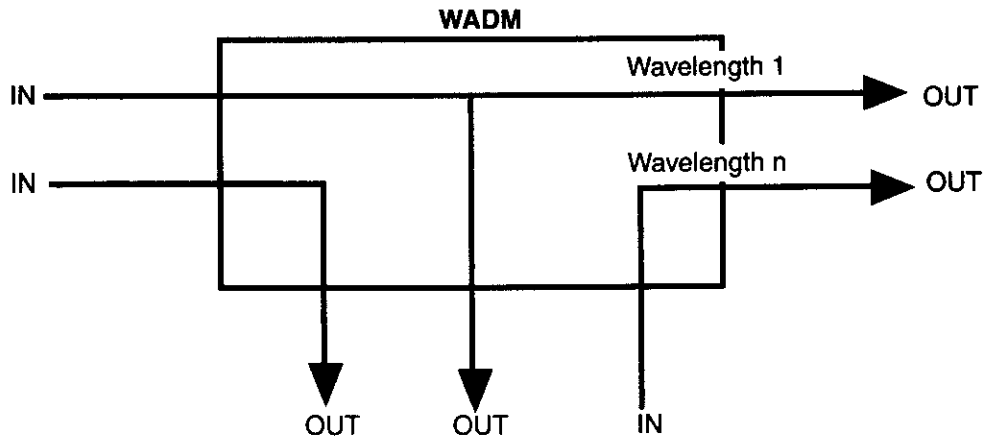


Figure 7–5 Generalized view of an optical ADM.

Metropolitan WDM networks are of keen interest in regards to WADM services. The requirements range from rearrangeable add-drop of 1–8 channels in a small business to 40 or more channels in an interoffice ring.

Due to the diverse customer mix, each WADM channel should be capable of carrying a different data rate and channel mix. The emerging WADMs support all the requirements described in this introduction.

Figure 7–5 shows a general example of an optical ADM. The machine has n inputs and a single mode fiber output with multiple wavelengths. The ADM must be able to demodulate each wavelength from the composite signal, and drop, pass through, or insert the wavelengths.

WADM Input and Output Ports

We continue the discussion of WADMs with a more detailed look at the ADM input and output ports. Figure 7–6 is a re-rendering of Figure 7–5. It shows four ways of managing a WDM channel at the ADM. The four notations are shown at the respective output port on the ADM.

- *Add:* An input channel is added to an output channel.
- *Drop:* These channels are the opposite of the adds. A channel coming into the ADM is dropped off to another node.
- *Through:* This channel is a straight pass-through the machine.
- *Drop-and-Continue or Bridge:* This channel is both a drop cross-connection and a through cross-connection from the same source. This configuration allows payload to be dropped off and also passed downstream.

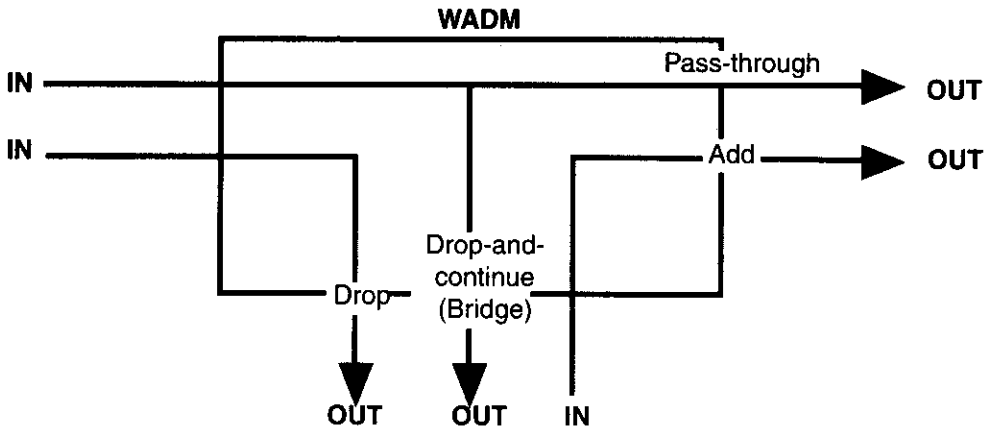


Figure 7–6 The WADM ports.

Figure 7–7 shows the WADM with in, out, add, and drop ports designated as 1–4 to later identify the relationships of the optical flows [GILE99]. The channel pathways are set up from the in and add ports to the out and drop ports. To manage these flows, a connection matrix is used, with rows corresponding to optical paths through the WADM and columns of 1s or 0s depicting the state of the flows for each WDM channel. An ideal WDM of N channels has at least 2^N possible connection states.

Figure 7–7 also shows two connection matrices for 16 WDM channels in a WADM. In the first matrix, channels 5-12 are through-channels, channels 1-4 are added, and channels 13-16 are dropped. In the second matrix, channels 2 and 11 are both dropped and added, and the other channels are passed-through the WADM.

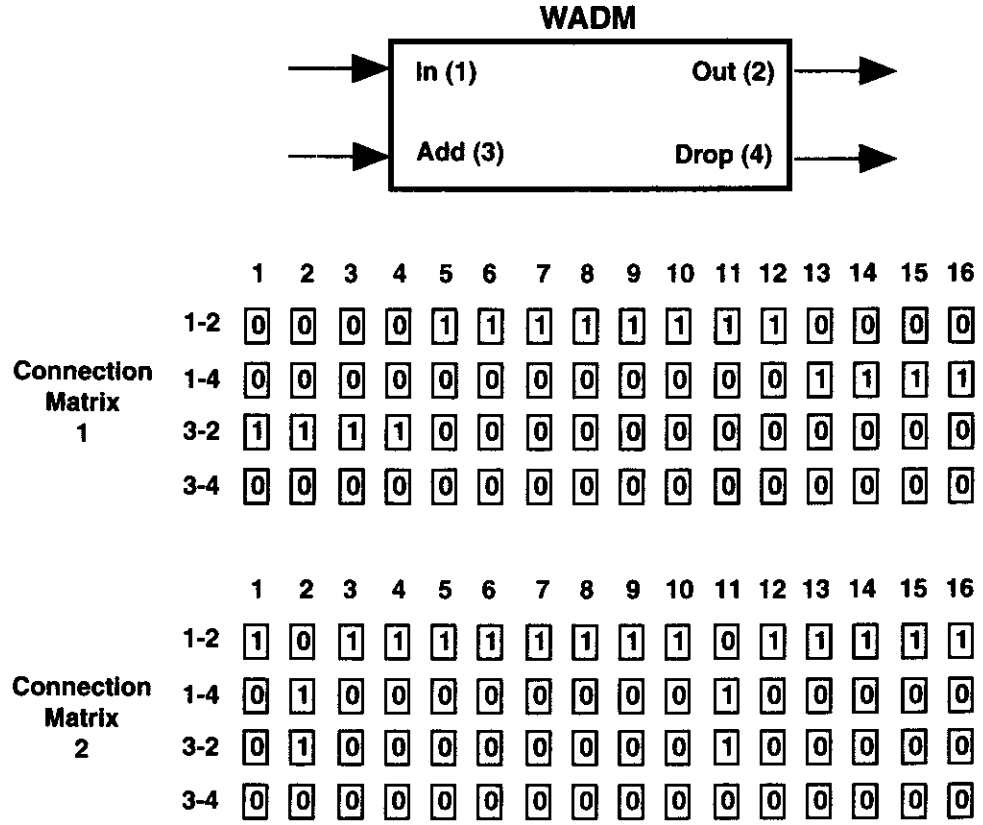


Figure 7–7 The WADM connection matrices [GILE99].

WDM CROSS-CONNECTS

The digital cross-connects (DCSs) are used to cross-connect optical links (point-to-point lines, and rings). One of their principal jobs is to map various types of input steams to output streams; in essence, they provide a central point for grooming and consolidation of user payload. The DCS can segregate high bandwidth traffic from low bandwidth traffic and send this traffic out to different ports. The DCS is also tasked with trouble isolation, loopback testing, and diagnostic requirements. It must respond to alarms and failure notifications.

The DCS may perform switching at the electrical level or (increasingly) at the optical level. The state-of-the-art optical cross-connects (OXCs) can convert an incoming signal of a specific wavelength to an outgoing signal of a different wavelength.

Figure 7–8 shows a functional diagram of an OXC. Four optical line systems (OLSs) are connected to the OXC [JACK99]. The WDM signals from two OLSs are demultiplexed and the resultant wavelengths are passed through wavelength converters. Each wavelength signal is cross-connected by the OXC, wavelength converted again, multiplexed, and sent out of a fiber interface.

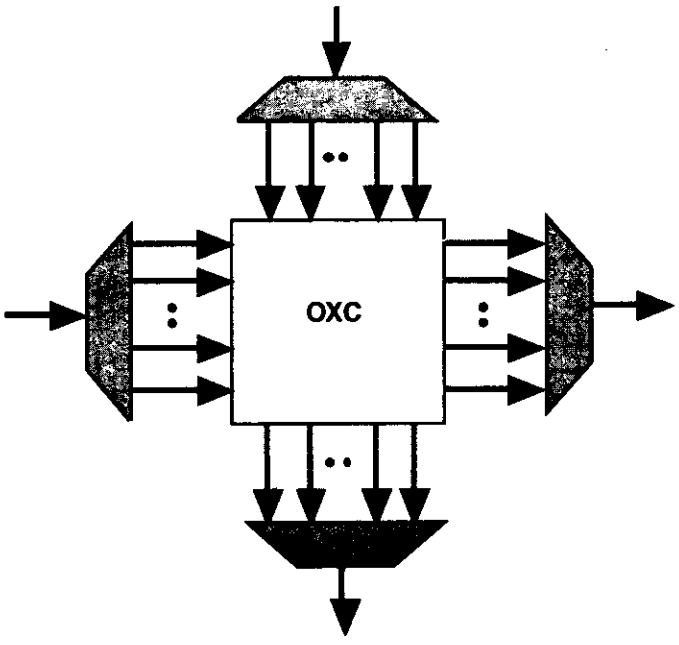


Figure 7–8 General view of a WDM OXC.

WAVELENGTH CONTINUITY PROPERTY

A WDM lightwave is said to satisfy the wavelength continuity property if it is transported over the same wavelength end-to-end. For optical nodes that do not have wavelength conversion features, this property is obviously in effect.

Nonetheless, as a wavelength is transported though multiple nodes, its properties will indeed change, a subject discussed in the last part of this chapter.

EXAMPLE OF DWDM WAVELENGTH PLAN

Vendors of optical equipment have their own methods of allowing the network operator to configure the wavelengths in the network. This section discusses one method used by Nortel Networks in several of their S/DMS TransportNodes™ [NORT98]. As shown in Table 7–2, this method supports up to 32 wavelengths (using NZDF fiber). The network operator can provision as few as two wavelengths, and add more (in increments of two) as traffic increases over time. The 32 wavelengths aggregate to 320 Gbit/s per fiber span. The wavelengths are divided into two bands, as shown in the table. Each band propagates in opposite directions on the fiber span. Table 7–2 also contains a column labeled "Order of Use." This entry provides guidance on the order in which the wavelengths are deployed. This plan is designed to provide the best performance for the span, as well as to assure that different nodes in the network are in agreement about which wavelengths are actually used.

Average Versus Maximum Span Loss and Chromatic Dispersion

In planning for DWDM deployment, it is important to know the signal loss over the span in relation to the number of wavelengths deployed, and the link speed. For example, using 32 wavelengths, the optical node supports the distances between DWDM repeaters, as shown in Table 7–3.

In addition, for the deployment of OC-192 on NDSF, the net chromatic dispersion must be limited to no more than 1400 ps.nm (using a measurement at 1557 nm) at the receiver. For OC-48 NDSF deployments, the system can tolerate up to 11,400 ps/nm for chromatic dispersion (corrected to an optical reach of over 700 km). An DWDM system will experience different amounts of chromatic dispersion in each wavelength on the band. To correct these different amounts of chromatic

Table 7–2 Example of a Wavelength Plan [NORT98]

Wavelength Grid (100 GHz spacing)	Order of Use	Band	Up to 32	Up to 24	Up to 16	Up to 8
1505.60		Red	Spare	Spare	Spare	Spare
1559.79	16	Red	√			
1558.98	6	Red	√	√	√	
1558.17	15	Red	√			
1557.36	1	Red	√	√	√	√
1556.55	10	Red	√	√		
1555.75	4	Red	√	√	√	√
1554.94	14	Red	√			
1554.13	8	Red	√	√	√	
1553.33	13	Red	√			
1552.52	2	Red	√	√	√	√
1551.72	9	Red	√	√		
1550.92	3	Red	√	√	√	√
1550.12	12	Red	√	√		
1549.32	7	Red	√	√	√	
1548.51	11	Red	√	√		
1547.72	5	Red	√	√	√	
1541.30		Blue	Spare	Spare	Spare	Spare
1540.56	16	Blue	√			
1539.77	5	Blue	√	√	√	
1538.98	15	Blue	√			
1538.19	6	Blue	√	√	√	
1537.40	14	Blue	√			
1536.61	7	Blue	√	√	√	
1535.82	13	Blue	√			
1535.04	1	Blue	√	√	√	√
1534.25	10	Blue	√	√		
1533.47	4	Blue	√	√	√	√
1532.68	12	Blue	√	√		
1531.90	8	Blue	√	√	√	
1531.12	11	Blue	√	√		
1530.33	2	Blue	√	√	√	√
1529.55	9	Blue	√	√		
1528.77	3	Blue	√	√	√	√

Table 7–3 Typical Optical Reach for 32 Wavelengths [NORT98]

Fiber Type	OC-48	OC-48 and OC-48/OC-192
NDSF	700km	500km
NZDSF	700km	500km

dispersion, Nortel Networks provides red and blue band compensation modules for use in several systems that span over 400 km in length.

Nortel Networks uses the following scheme to plan average and maximum span loss values. Let's use an OC-48, six-span, eight-wavelength application, with an average span loss value (budget) of up to 27 dB per span, where each of two spans introduce 29 dB of loss (or 2 dB in excess of the recommended budget). The total loss of 4 dB can be accommodated, providing that the total system loss budget is reduced to this formula:

Adjusted System Loss = Number of Spans × [Average Loss – Excess Loss Value]
Where: Excess Loss Value equals:
0 dB for total excess loss ≤ 2 dB,
1 dB for total excess loss > 2 dB and ≤ 4 dB

For this example, with an initial total system loss budget of 162 dB (6 spans × 27 dB per span), the adjusted system loss budget is:

6 spans × (27 dB – 1 dB) = 156 dB

HIGHER DISPERSION FOR DWDM

The subject of dispersion is explained in Chapter 3. If you would like to review the basic concepts of dispersion, see the section in Chapter 3 titled "Chromatic Dispersion." There has been considerable attention paid to channel spacing in DWDM systems, that is, the bandwidth between the wavelengths. Of course, it is desirable to reduce this spacing as much as possible, since the spacing represents unused bandwidth, and more channels can be placed in a single fiber. However, tight channel spacing can lead to interference between the wavelengths, and is evident on fibers with low dispersion levels, such as NZDF. Recall from Chapter 3 that

NZDF operates over a portion of the third wavelength window, with the chromatic dispersion small enough to support individual channel rates of 10 Gbit/s over distances of over 250 km.

A new type of fiber solves some of these problems. It is called Advanced NZDF (A-NZDF), and has been developed in conformance with the ITU G.655 Recommendation. This new class of fiber exhibits a maximum dispersion of 10 psec/nm-km at the far end of the C band. It tends to suppress the non-linealities that are introduced by the tighter channel spacing. It is expected that this improved DWDM operation will allow the dispersion to be high enough to minimize non-linearities and low enough to minimize the need for dispersion compensation. It will be an important tool in the deployment of 40 Gbit/s systems.

For the reader who wants more information on this new fiber class, you should read ITU G.655, and [RYAN01] provides an excellent explanation of the performance of A-NZDF, as well as some interesting cost comparisons of A-NZDF and other NZDF technologies.

TUNABLE DWDM LASERS

The WDM wavelengths are spaced very closely together to allow many wavelengths to be placed on one fiber. However, it is quite costly to build a system in which there is a dedicated laser for each wavelength. The solution is a tunable laser, one that can be tuned over a wide range of wavelengths. Thus, this laser can be modified to suit several wavelength channels.

SUMMARY

WDM represents a significant advance in optical fiber technology. It has allowed optical networks to increase their capacity by many orders of magnitude. WDM relies on a very old concept called frequency division multiplexing, but WDM deals with the optical spectrum.

With the progress made in the last few years with dense WDM, it is now possible to operate one optical fiber in the terabit per second range. With the advent of tunable lasers, and the emerging standards on dynamic wavelength configuration and spectrum management, optical networks will be deployed that support near-instantaneous "bandwidth on demand," as well as "QOS on demand."

8

Network Topologies and Protection Schemes

This chapter examines the way in which optical networks are put together, and the shape they take. The emphasis is on point-to-point, ring, and mesh topologies, which reflect the preferred methods for the network layout. The subject of protection is also discussed, which deals with how an optical network provider can exploit network topologies to give the network user robust connections, that is, protect the customer's traffic from link or node failures.

THE NON-NEGOTIABLE REQUIREMENT: ROBUST NETWORKS

In some of my writings, I have quoted the typical network manager dictum, "Whatever it costs, whatever the effort, keep the network up and running." Certainly those people who "foot the bill" must weigh the costs of building such a network. Nonetheless, some networks cannot afford to lose the ability to transport their customers' traffic. Optical networks are the backbone of most communications networks around the world, so it is critical that they be robust.

Of course, an optical network unto itself cannot offer any more fail-safe features than a copper-wire network. Optical fiber is inherently less

error-prone, but it does not offer any more backup facilities than any other technology.

It does not matter if a network is copper- or fiber-based, any large backbone network must have ways to recover from problems and to keep these problems transparent to the network users. To gain a sense of how important this idea is, consider a telephone exchange in the metropolitan Boston area.

The telephone service provider supports over five million people in this particular area. In so doing, several powerful switches are employed to route the telephone calls to and from the calling parties and the called parties. Each switch has a backup switch in case one switch goes down. Each switch pair has ~500 communications links. Each link supports about 20 calls per second; therefore, 20 × 60s × 60min = 72,000 calls per hour per link.

This particular exchange supports the following calls: 500 links × 4 switch pairs × 72,000 = 144,000,000 calls per hour or over 40,000 calls per second.

The failure of parts of this network will have very serious consequences, leading to a considerable amount of lost revenue, lost productivity, and psychological stress.

With these thoughts of "robustness" in mind, we can now explore how optical networks provide protection services to its customers.

DIVERSITY IN THE NETWORK: WHICH CONTROL PLANE?

In several IETF working papers, the term *diversity* refers to the relationship between lightpaths, such as fibers, or wavelengths on the fibers, called *optical links* in this discussion. Two optical links are said to be diverse if they have no single point of failure. Traditionally, this diversity has been implemented by the telephone and leased line service providers.

Data service providers (using IP and other data-oriented techniques) have not had to deal with this important issue; they have relied on the private line providers for these services. Thus, IP has operated over many different kinds of carrier transport networks, without concern about the activities of the diversity operations.

This relationship holds true for third generation networks, but the idea supported by many people in the industry is to bring IP (and MPLS) into the picture to assist in achieving link diversity, routing, and protection switching.

This "integration" must be approached with considerable care. It has not yet been demonstrated that the complex ID addressing scheme (with its many rules on address aggregation, OSPF/BGP routing, private address reuse, etc.) should be integrated with (or even make known to) the optical control plane. Indeed, I am not sure this IP addressing pox should be permeated down to yet another part of a network architecture.

My view is that the IP, MPLS, and optical control planes should be designed to allow their coupling together, or to allow them to operate independently. For a review of this idea, take a look at the section titled, "Management of the Planes" in Chapter 10, and more information is provided on this issue later in the book (see "Plane Coupling and Decoupling" in Chapter 12).

LINE AND PATH PROTECTION SWITCHING

We need to define some terms at the onset of this discussion. Two forms of protection switching are employed in an optical network: line protection switching and path protection switching. Line switching performs recovery on the entire optical link (the line). Path protection switching performs recovery on selected tributaries on the fiber.

For example, an STS-1 tributary on an OC-48 line that is carrying, say, ATM traffic from city A to city B might undergo recovery operations and not affect another STS-1 tributary that is carrying ATM traffic from city A to city C. Path protection switching provides more granularity of control than does line protection switching.

Path protection switching also works hand-in-hand with tributary provisioning (setting up and tearing down specific customer tributaries on the fiber), so it is a common mode of protection switching. Of course, if a fiber link is down, it makes no difference if line protection is used, so line protection switching is also quite common.

TYPES OF TOPOLOGIES

The next parts of this chapter explain in more detail the various types of topologies found in optical networks. To start this discussion, Table 8–1 shows the various choices for a topology. The basic topology can be a point-to-point, a ring arrangement, or a meshed layout. Within these three topologies, there can be two or four fibers connecting the optical nodes, the optical signals can flow in one or both directions, and the

Table 8–1 Network Topologies and Attributes

Topology	Fibers	Signal Direction	Role of Fiber	Protection Type
Point-to-point	2 or 4	Uni- or Bidirectional	Working or Protection	Line or Path
Ring	2 or 4	Uni- or Bidirectional	Working or Protection	Line or Path
Meshed	2 or 4	Uni- or Bidirectional	Working or Protection	Line or Path

method of recovery can be performed at a line or a path level. As the table reveals, all three topologies can provide the same kinds of protection attributes. These attributes of the topologies are explained next.

WORKING AND PROTECTION FIBERS

The term working fiber (or working copy) refers to a fiber (or a wavelength) that is carrying user payload. The term protection fiber (or protection copy) refers to a fiber (or a wavelength) that is acting as a backup to the working copy. A protection copy may also be carrying user payload, but in the event of problems, this payload (usually of a low priority) is removed from the protection copy, and the other (usually of a higher priority) payload is placed onto the protection copy.

Ideally, one would like to have the working and protection fiber in different paths (different feeders) between the network elements. This option is not always possible, but, as seen in Figure 8–1 (a), it may be possible to place two fiber cable sheaths in the same conduit structure and then separate them physically within the conduit.

Many systems and their conduits are laid out in a grid structure. It may be possible (as in Figure 8–1 (b)) to place the working and protection fibers in separate conduits for at least part of the feeder connection.

Yet another possible alternative (see Figure 8–1 (c)) is to use different feeders for the working and protection fibers. Finally, Figure 8–1 (d) shows another possible way to separate the cables. Since some feeder routes in densely populated areas may intersect or be situated close together, it may be feasible to use two dual paths to two separate central offices, and have the ring connected through these offices.
The list below should be helpful as you read this chapter:

- 1+1 unidirectional: A dedicated point-to-point link between nodes to provide line or path protection.

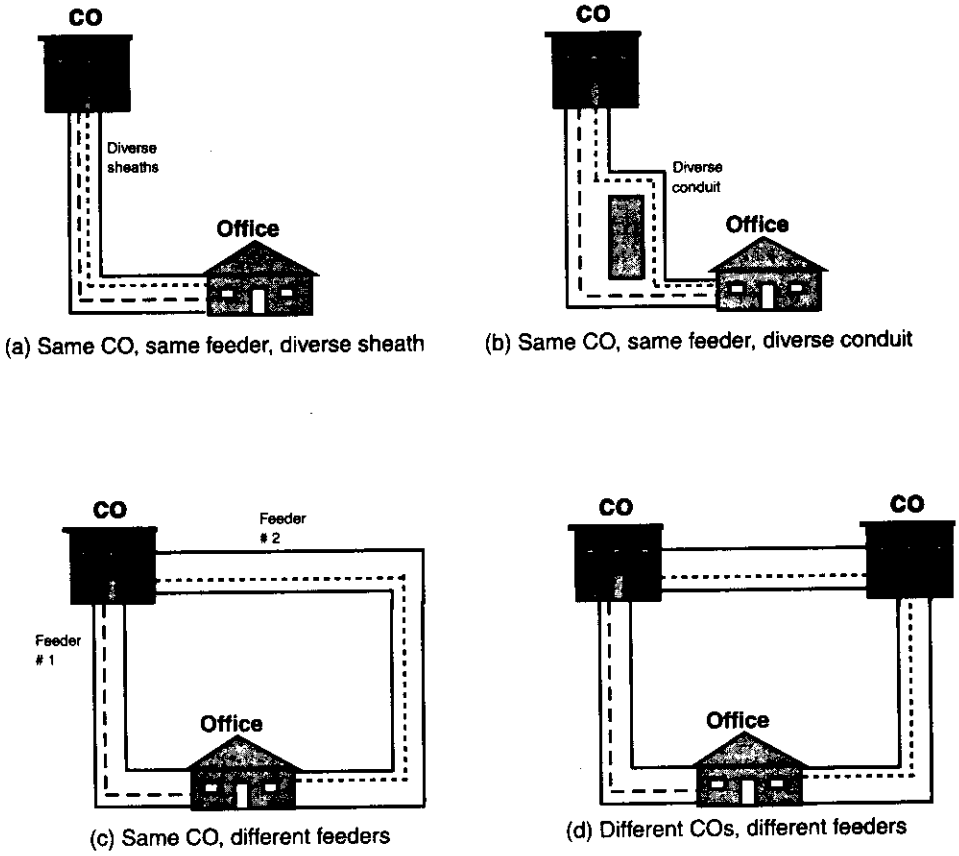


Figure 8–1 Topology diversity in the network.

- 1+1 bi-directional: A dedicated point-to-point link between nodes for line or path protection. Alarm information provided in SONET/SDH OAM bytes.
- 1:N: A protection link shared by N working links.
- 2 fiber bi-directional line switched ring (2F-BLSR): Provides line protection with one protection fiber.
- 4 fiber bi-directional line switched ring (4F-BLSR): Provides line protection with two protection fibers.
- Uni-directional path switched ring (UPSR): Provide path protection with an alternative fiber in the ring.

POINT-TO-POINT TOPOLOGY

The topology for the optical network chosen depends on the network manager's objectives pertaining to bandwidth availability, bandwidth efficiency, survivability/robustness, cost containment, and simplicity.

The point-to-point topology shown in Figure 8–2 is a common topology. The entire optical payload is terminated at each end of the fiber span between two access nodes. Two fibers connect the two optical access nodes. The term bi-directional fiber describes this arrangement: One fiber transports the signals in one direction and the other fiber transports the signals in the other direction. It is certainly possible to deploy a point-to-point system wherein only one fiber connects the nodes, but unidirectional point-to-point topologies provide no means to recover from errors on the fiber or fiber interfaces.

Point-to-point topologies are usually employed in a system that needs only a single system and single route solution. The topology is simple, but not designed to be completely survivable. The reliability of a

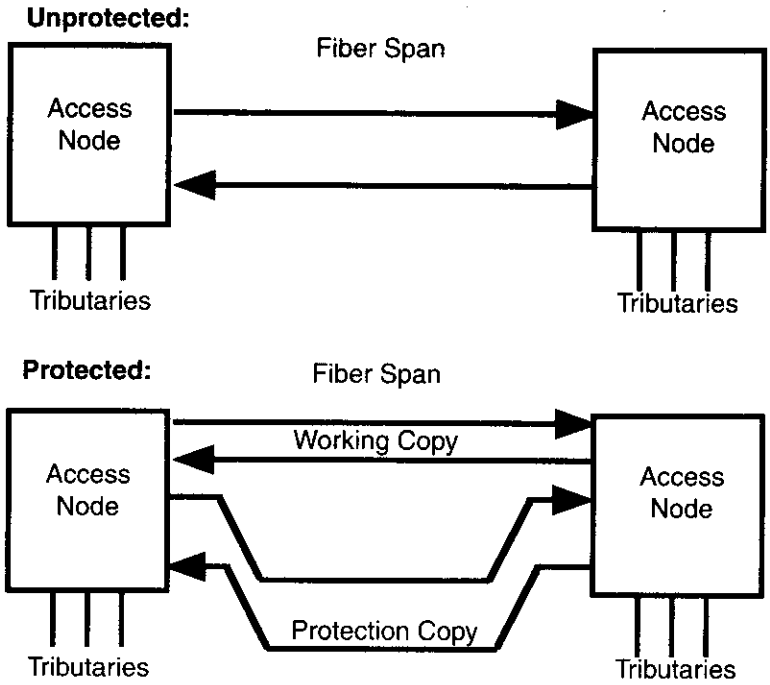


Figure 8–2 Point-to-point topology.

point-to-point system can be enhanced considerably through the user of a geographically diverse protection path, also shown in Figure 8–2. Since diverse routing may exceed the normal reach of the fiber, one or more regenerators or optical amplifiers can be employed on the span to reconstitute or boost the signal.

1:N Protection Channel Sharing

A common topology employed in many systems today is called the 1:N protection channel sharing, as shown in Figure 8–3. This topology employs multi-network elements (access nodes) in a point-to-point configuration. Its attraction is that it conserves fiber pairs by allowing multiple systems to share a common protection channel. In a network where there is rapidly growing traffic demand, the 1:N topology can either defer or avoid the deployment of new fiber cable.

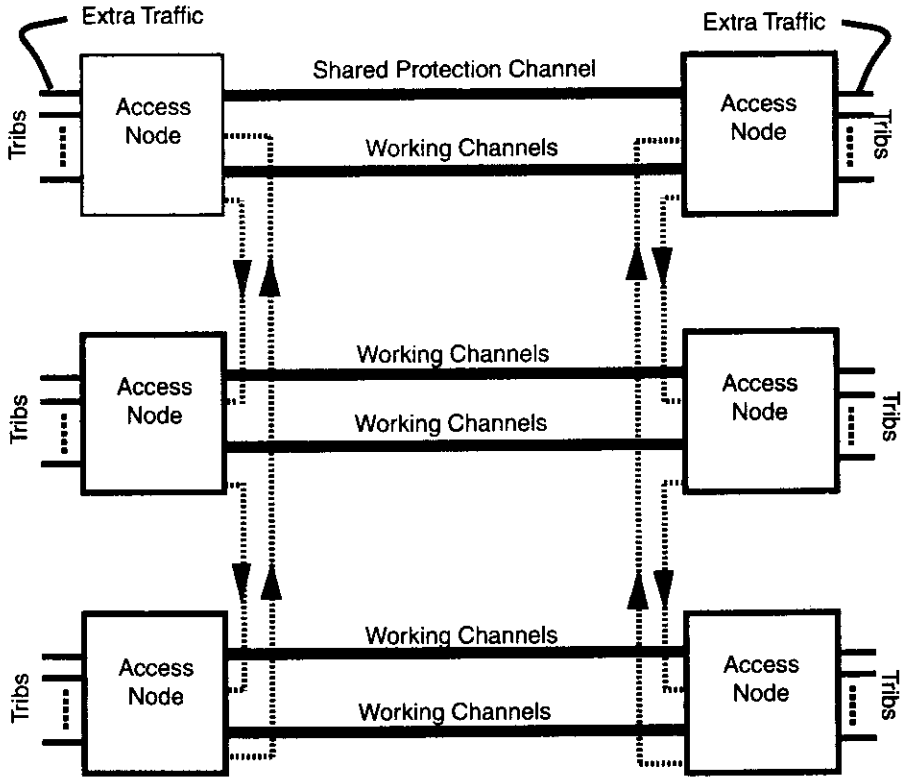


Figure 8–3 Point-to-point protection channel sharing.

If an error occurs on any of the working channels, the affected traffic on this channel is automatically diverted to the protection channel. The diversion operation occurs through the installment of inter-shelf protection loop line cards with a dedicated shelf at each end of the span.

Based on a customer's needs, the 1:N topology (which might be empty most of the time) can be used to carry extra traffic. Extra traffic is a term that refers to the exploitation of the normally idle protection fiber. While this bandwidth is available to ensure survivability, it is also capable of supporting ongoing user traffic. This approach also allows the introduction of new services without the costly implementation of additional fiber. Of course, the extra traffic is "unprotected," but additional measures may be undertaken to lessen or completely eliminate the impact of a switch to these protection facilities during an impairment to the system.

Three options can be implemented to protect the extra traffic:

- Diverse routing: The extra traffic can be protected by a redundant diverse route to prevent a single source of failure causing a service outage. The redundant route can also be provided via an extra traffic channel.
- Service discounting: The network service provider can offer an extra traffic-based service at a substantial discount with the understanding that service interruptions may occur.
- Degradable services: Multiple protected and unprotected service channels can be provisioned for certain data services. The data services would then be considered to be of lesser priority than other traffic. Due to the asynchronous nature of data traffic, the loss of a unprotected extra traffic channel will certainly result in degraded performance but may not result in a total service interruption.

Optical Channel Concatenation

Increasingly, as optical networks evolve, the legacy SONET/SDH equipment will probably be bypassed if these nodes cannot support the port speeds that are emerging today. To accommodate the high-speed optical links, a common practice is to cluster multiple backbone routers to meet the bandwidth demands. The resulting topology are many parallel links in a backbone network that are not protected with the conventional SONET/SDH ring protection operations.

Of course, one solution is to use protection channel sharing topology shown in Figure 8–3, and this is exactly what vendors are doing today. However, until recently, these schemes have been proprietary. Some use the K1 and K2 bytes of the SONET/SDH header to control the operations, some use the D bytes, and so on. Clearly, it will be beneficial to have a standardized scheme, and this section is devoted to this topic.

To give you an idea of some of the operations involved in this scheme, let's take a look at some work going on with [LEE01]. This IETF working group has set forth a scheme for managing parallel links, called optical channel concatenation. This operation uses a new framing method defined in [T1X1.501].

A new term is coined for this operation: the superchannel. It refers to a concatenation of multiple links, and perhaps wavelengths within the links. The superchannel can appear as one interface on a router, and thus be advertised as a single IP address by, say, OSPF. It is the job of the two nodes connecting the multiple parallel links to manage the individual subchannels (the ports and the wavelengths on the ports). Figure 8–4 shows how four parallel links are concatenated together, and treated as a superchannel by nodes A and B.

Figure 8–5 shows the messages that are sent between nodes A and B if all subchannels are operating without any problems. Each message is a bit map that reflects the state of each optical port. Alternatively, the bit

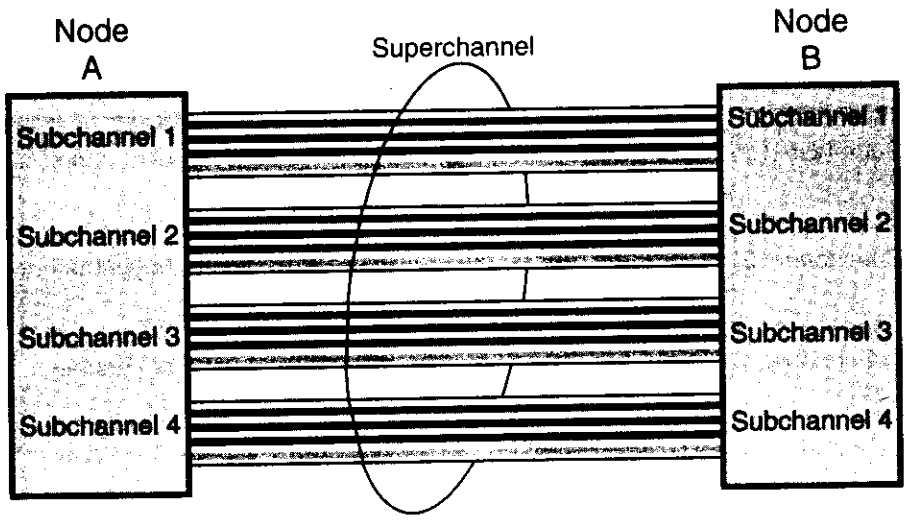


Figure 8–4 Concatenated optical channels.

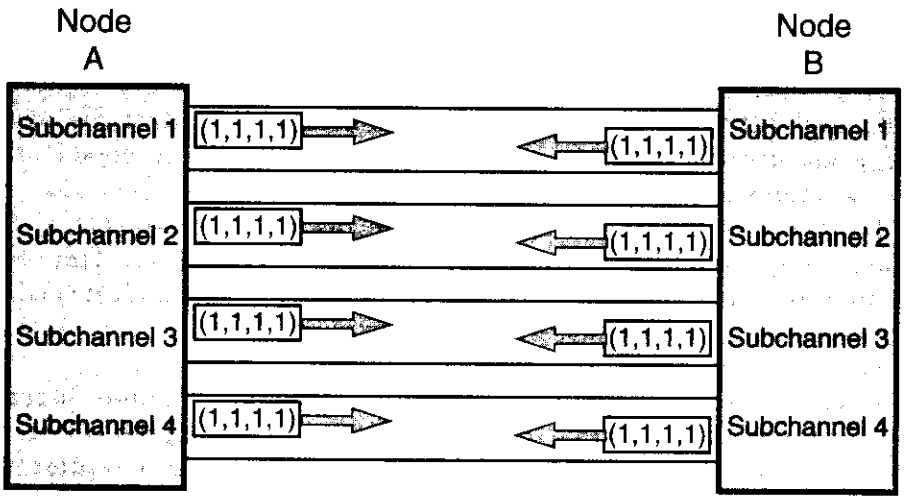


Figure 8–5 All subchannels are up and running.

map can reflect the state of the port interfaces as well as the state of each wavelength that is being received across each port. The four 1s in the message signify that all four subchannels are operating satisfactorily.

In the event that problems arise on one or more subchannels, the affected node uses the subchannels that are still operating to report on the problem. Figure 8–6 shows that subchannel 1 has failed. Both nodes

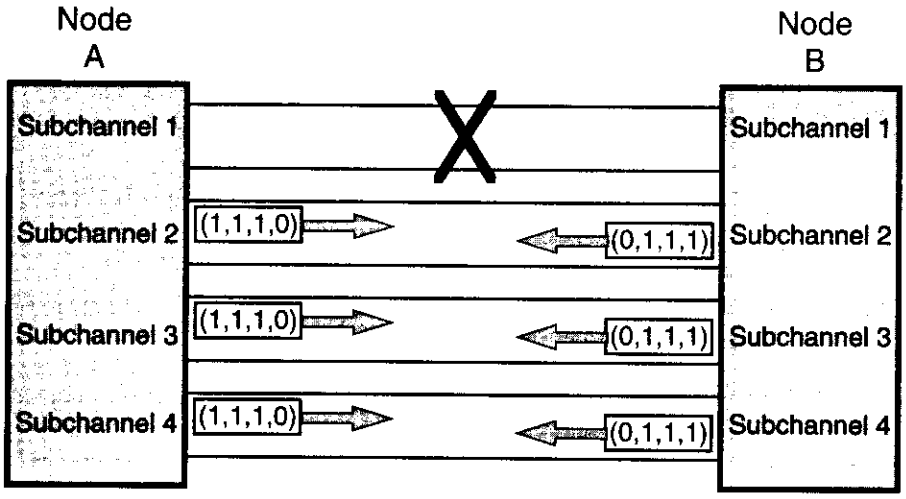


Figure 8–6 Reporting on faulty subchannels.

notice the failure, typically by not receiving any traffic from the partner node. In turn, the nodes alter the bit map in the message to identify the specific problem, in this example, at subchannel 1. The actions taken as a result of these alarms will vary, depending on the network provider's implementations. The nodes might place the user payload from subchannel 1 onto protection copies in the other subchannels, or the nodes might be configured to route this traffic to other nodes in the network. Whatever the action taken, the idea is to protect the user's traffic from these failures.

BI-DIRECTIONAL LINE-SWITCHED RING (BLSR)

The bi-directional line-switched ring (BLSR) connects adjacent nodes through a single pair of optical fibers. One fiber is used for the working copy (for user traffic) and the other fiber is used for protection. Many systems use more than two optical fibers, typically one pair for traffic, and the other pair for protection; this arrangement is covered later in the chapter.

This architecture provides a survivable closed loop architecture which can recover from either cable failure or node failure. As shown in Figure 8–7, the working traffic travels in one direction on the ring and the protection path is provided in the opposite direction. This approach is called BLSR 1:1 span protection switching: there is one protection copy for each working copy.

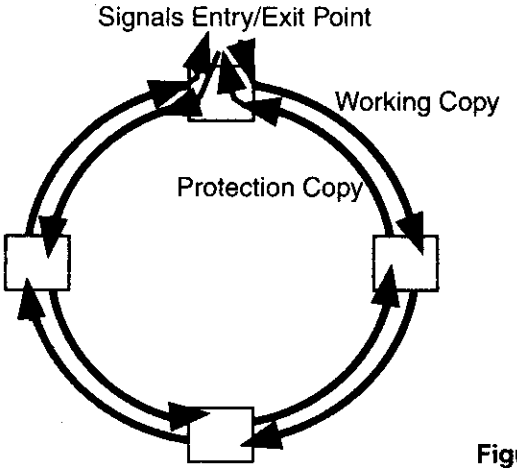


Figure 8–7 BLSR, two fibers.

Traffic may originate or terminate on the same BLSR or may be transferred to adjacent nodes into other rings. Each individual DS1, DS3, or STS1 signal travels around the ring in one direction. However, a duplicate signal passes in the opposite direction on the protection fiber. A path selector performs an ongoing monitoring function on the two fibers at each end of the path. In the event of a node failure or a fiber failure, the path selector automatically switches to the protection signal. The path selector is able to detect signal degradation as well as path failure and can transmit and receive: (a) path alarm indication signals (AIS), (b) path loss of pointer (LOP) signals, (c) signal degrade (SD) signals, and (d) excessive path layer bit interleaved parity (BIP) errors.

Figure 8–8 shows how the BLSR recovers from a fiber cut. The signal entry point is at node A and the exit point is at node B. Upon detecting the loss of a signal between nodes A and B, B sends an alarm signal to A on the working copy through nodes D and C (this operation is not shown in the figure).

This alarm alerts node A to the problem. Node A is already employing the protection path to send the traffic through the other nodes reaching the proper signal exit point at node B.

It should be emphasized that node B will notice a path failure because it receives no traffic on the working path between A and B; therefore, it does not have to rely on sending the alarm to A because it already has another signal from A on the protection path. The path selector simply selects the signal off the protection path at the signal exit point (node B).

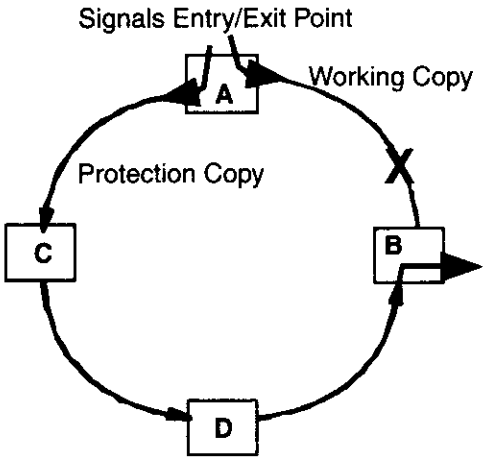


Figure 8–8 BLSR protection switching.

Notice the "cable cut" on the fiber between nodes A and B. It is now an easy task for node B to select the signal coming from node D. After all, node B is receiving nothing from node A.

Of course, the arrangement is very robust, and the recovery is instantaneous. The drawback to this topology is the consumption of resources to run the same traffic over duplicate fibers.

PROTECTION SWITCHING ON FOUR-FIBER BLSR

Figure 8–9 shows another common ring arrangement. There are four fibers in the ring, set up as two pairs. One pair is the working pair and the other is the protection pair. One fiber of the pair transmits in one direction, and the other fiber transmits in the other direction. The topology is used because it obviously provides for more capacity, and if the entire installation is performed at the same time, the benefits of the additional fibers and interfaces outweigh the costs.

Figure 8–10 shows how a four-fiber BLSR recovers from a loss of connectivity between two fiber nodes. The nodes that are affected directly (nodes A and B) can now divert (loop back; shown in the figure as LB) the traffic

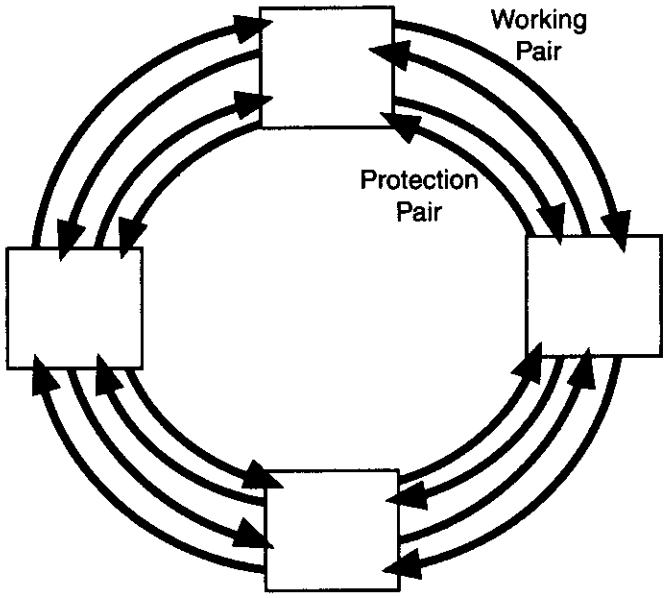


Figure 8–9 The four-fiber BLSR.

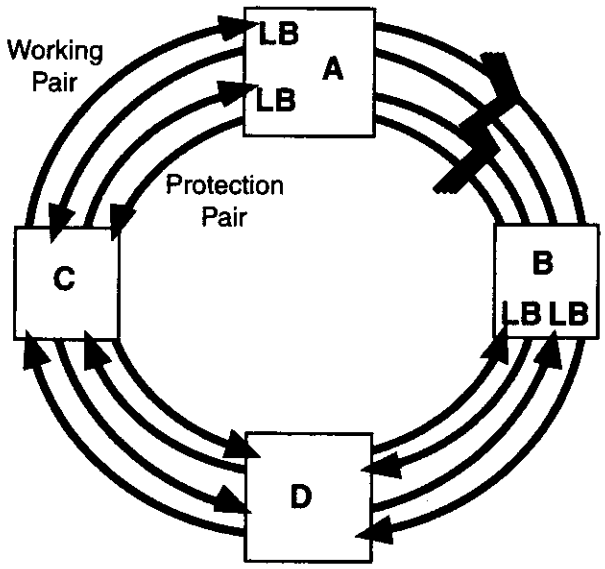


Figure 8–10 Fiber cut on working and protection fibers.

back to the other fibers in the working and protection pair. Nodes C and D are not involved in the protection switch operation, but they can be made aware of the problem by receiving diagnostic messages from nodes A and B.

The next example in Figure 8–11 shows how the ring topology can recover from a dual fault, one fault on one pair and another fault on the other pair. To keep the figure simple, the bi-directional arrows indicate that two fibers are sending the traffic in both directions on the ring. In this arrangement, the protection and working copies are utilized, based on where the errors occur:

- Between A and B: Use protection pair
- Between B and D: Use working pair
- Between C and D: Use protection pair
- Between C and A: Use working pair

The last example of ring fault recovery is shown in Figure 8–12. In this situation, node B has failed, which renders useless the links between B's neighbors, nodes A and D. These two nodes detect the failure by not receiving signals on their links to node B (or by not receiving responses to their link management protocol (LMP) hello messages, a topic for Chapter 11). In either case, nodes A and D perform the recovery by looping the signals back onto the mate of the protection and working copies.

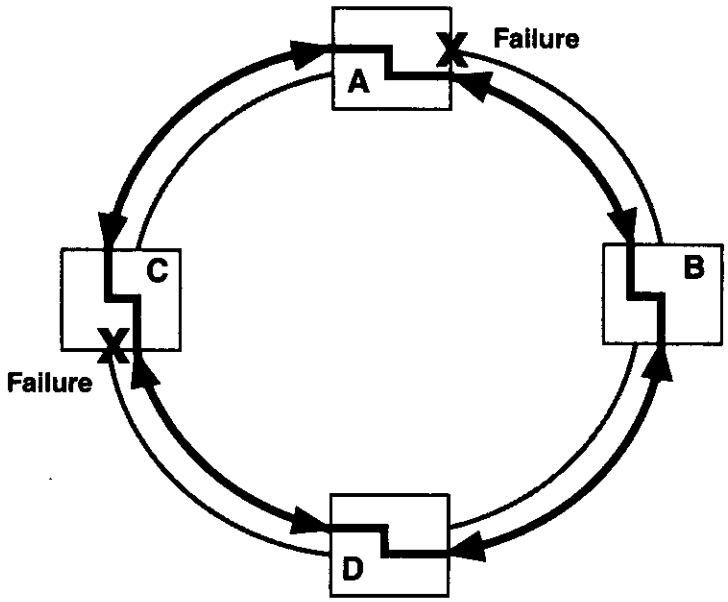


Figure 8–11 Example of a dual fault (bi-directional arrows denote two fibers).

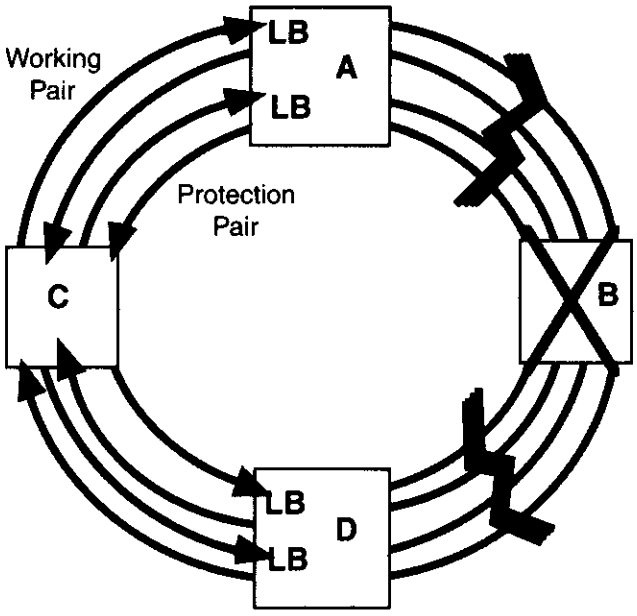
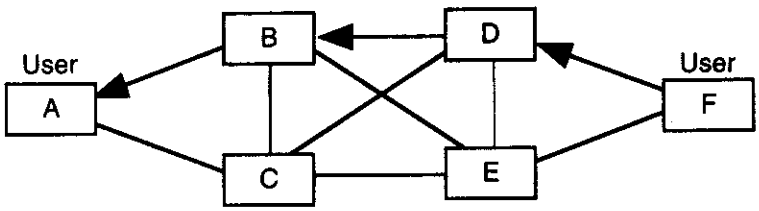


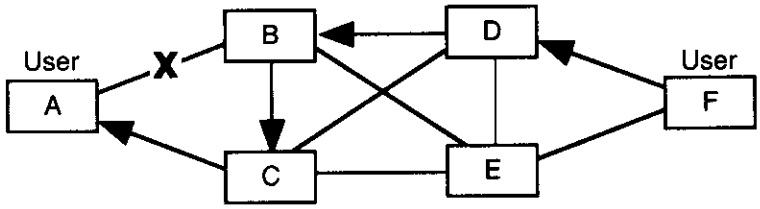
Figure 8–12 Node failure.

MESHED TOPOLOGIES

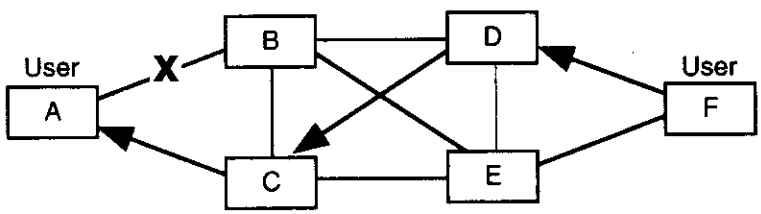
Optical rings are prevalent topologies in large-scale optical networks. But I would be remiss if meshed topologies are not discussed in this chapter, since they are a bedrock feature of Signaling System #7 (SS7), the control plane for the world's telephone system and many internets. Figure 8–13 shows a typical meshed topology. The user nodes (such as telephone central offices) are labeled A and F. The network nodes (B, C, D, and E) are called signaling transfer points (STPs), which are large switches that are responsible principally for keeping the backbone



(a) Normal Traffic Flow from F to A



(b) Diverted Traffic Flow from F to A



(c) STP B Informs STP D to Divert Traffic

Figure 8–13 Mesh topology.

network operational. The STPs in this example are fully connected, in that each node has a direct optical link with all other node. This topology is called a fully meshed network.

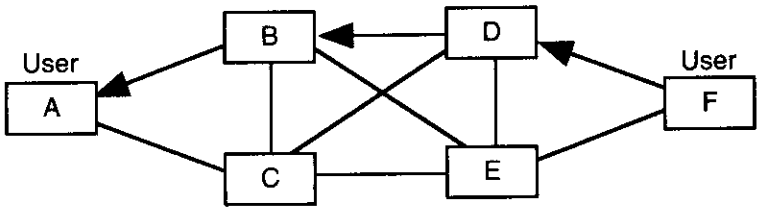
SS7 uses the term link set to define multiple links between two adjacent nodes. For optical networks, a link set could be separate fibers (most likely), or different wavelengths on the same fiber. In addition, SS7 uses a shorthand notation to describe the location of the link sets. For example, the notation link set A-B means a link set between nodes A and B.

During normal operations, routing traffic goes through the most direct route. In Figure 8–13 (a), the messages are sent across links between nodes F, D, B, and A. If a failure occurs between certain nodal pairs (B and C, for example), no routing change occurs. The reasons are that A can still reach its two switches (B and C). Also, under most conditions, upon receiving traffic from its attached switches (in this example, A) B would not send this traffic to C, but to a more direct route of B-D or B-E. Therefore, if link set B-C fails, only nodes B and C are aware of the failure. The B-C link set is always a lower priority than its alternatives.

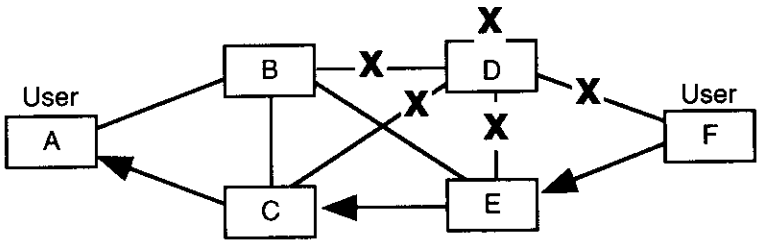
It is possible for this topology to break down to such an extent that traffic is not routable for a particular STP. For example, at switch B, if link sets B-D, B-E and B-C are unavailable, B cannot relay traffic. But this topology provides 100% redundancy. Any single point of failure does not bring down the system, because the traffic can be diverted around the failure. So, even if B's links are down (or, for that matter, if B is down), traffic is diverted, and the network remains fully operational. Anyway, multiple link failures are rare; such an event would mean that all link sets and links within each link set would have failed.

In Figure 8–13 (a), traffic is routed between signaling points F and A through link sets FD-DB-BA. In Figure 8–13 (b), a failure occurs on the link (or link set) between B and A. These nodes take this link set out of operation and B diverts the traffic to C on link set B-C. However, a better route may exist between F and A. If the B-A failure persists for greater than a set period, B will send messages to D and E to inform them that traffic for A should be sent on link sets D-C and E-C, respectively. This new configuration is shown in Figure 8–13 (c), with link set D-C the primary link set from D.

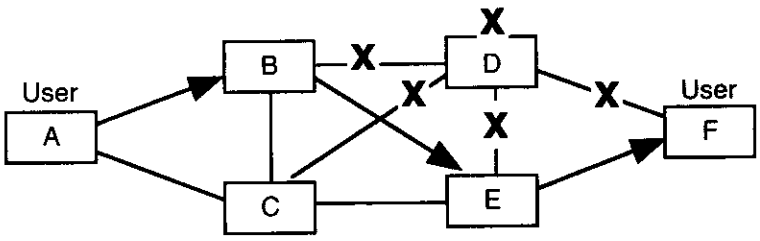
For this next example, it is assumed that node D goes down, and/or all of D's link sets become unavailable. In Figure 8–14 (a), under normal conditions, traffic is moving from F to A through link sets F-D, D-B, B-A. In Figure 8–14 (b), node D goes down and any links that interface into D are unavailable. For traffic flowing from F to A, it is diverted to E,



(a) Normal Traffic Flow from F to A



(b) Node D is Down



(c) Traffic in Other Direction

Figure 8–14 Node failure recovery.

through C, and then to A. In Figure 8–14 (c), for traffic flowing from A to F that was going through B, it is sent to B, through E, and then to F.

Once again, it is obvious how the meshed backbone topology provides for 100% recovery from any single point of failure. And some multiple points of failure are also recoverable. For example, assume that nodes B and E are declared unavailable. Traffic from A to F is diverted through link sets A-C, C-D, D-F, and traffic from F to A is diverted through link sets F-D, D-C, C-A.

It is evident that SS7 networks are highly reliable, but then they must be, since they are the control plane for the world's telephone networks.

PASSIVE OPTICAL NETWORKS (PONs)

We continue the discussion of optical network topologies with a look at passive optical networks (PONs). PONs are really not passive. They are actively sending and receiving optical signals. They are called passive because the outside plant has no electronics to power or maintain the components. As a consequence, PONs eliminate expensive power-based amplifiers, rectifiers, and, of course, batteries.

The active part of the PON is between the two ends. In most situations, these two ends are the service provider's node (say, a telephone central office), and the user node. The user node is typically a remote pedestal, often called a remote digital terminal (RDT).

Figure 8–15 shows a PON topology. The fiber can be forked out to multiple sites with the user of splitters. This approach saves a lot of money by multiplexing many user payloads on fewer fibers than in a conventional point-to-point topology. Of course, since multiple users must share the fiber, the multiplexing operation must be capable of efficient bandwidth management. The PON employs ATM for this important job.

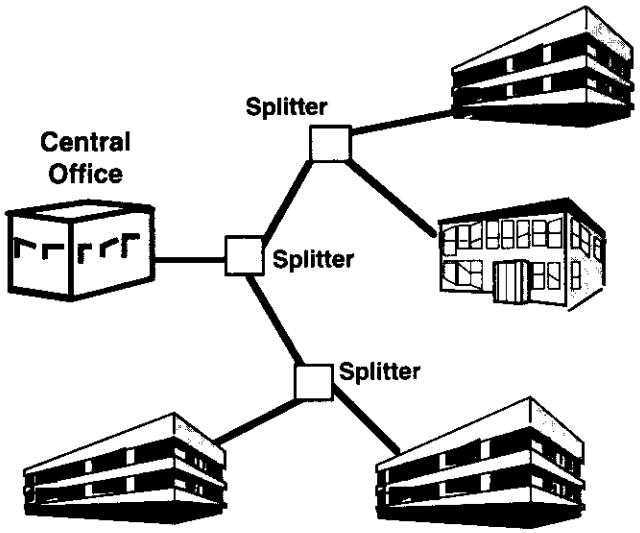


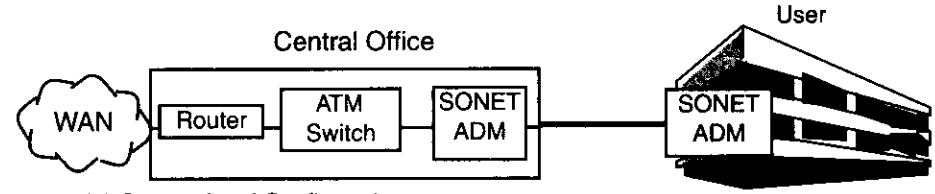
Figure 8–15 PONs.

OPTICAL ETHERNETS AND ETHERNET PONs

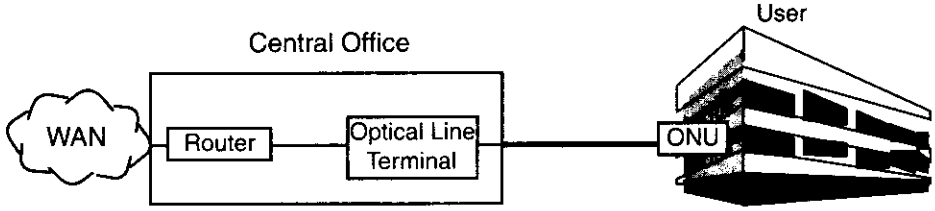
The use of Ethernet and the PON technology in the local loop reduces the cost substantially of providing high-capacity links to the customer. The reason is that expensive components such as SONET ADM, and ATM switches can be eliminated, and replaced with less expensive substitutes, as shown in Figure 8–16. At the customer premises, the SONET ADM is replaced with a simple and inexpensive optical network unit (ONU). At the Central Office (CO), the SONET ADM and the ATM switch(es) is(are) replaced with an optical line terminal (OLN). Of course, this arrangement does not offer the rich functionality of ATM and SONET, but that is precisely the point: Many user interfaces do not need all the powerful attributes of SONET and ATM.

Although Figure 8–16 shows a point-to-point topology, we learned earlier that the PON link can be split into multiple fibers with splitters or combined in a single link with splitters/couplers. Traffic is supported both upstream and downstream with the IEEE 802.3 (modified Ethernet) frame. The use of the Ethernet 1,518-byte frame makes better use of the link (compared to the ATM 53-byte cell).

Strictly speaking, the technology is not pure (original) Ethernet, in that the traffic upstream from the user to the CO uses dedicated TDM slots, and does not rely on the contention and collision detection aspects



(a) Conventional Configuration



(b) PON Ethernet Configuration

Figure 8–16 Conventional ATM/SONET and ethernet PON configurations.

of 802.3. An enhanced Ethernet feature called rate limiting allows the network operator to place transmission limits on each port on the link, that is, on each subscriber. Thus, the shared media is managed in a structured manner. In any case, the idea of the Ethernet PON is to use switches to terminate the Ethernet collision domains.

ETHERNET IN THE WIDE AREA BACKBONE?

Ethernet in optical networks is a big issue and is becoming a big industry. But as of this writing, there is very little interest in deploying Ethernet in a wide area transport network due to the limited distances resulting from Ethernet's collision window. If the CSMA/CD protocol is not used, it is not Ethernet, as defined by IEEE 802.3. So, Ethernet over optical is beyond the subject of this book, but if you want more information, I recommend [KOLE00] and [PODZ00].

METRO OPTICAL NETWORKING

Fiber optics has made remarkable progress in finding its way into the local loops, also know as metro optical networking. Originally implemented within the network, it quickly proved itself as a cost-effective digital transport technology for the local access loop as well. Indeed, much of its success is owed to those interfaces to the subscriber.

Optical networking plays a big role in residential broadband. It provides a robust, relatively inexpensive mechanism for deploying fiber in the distribution plant. Its extensive operations, administration and maintenance (OAM) capabilities make it an attractive technology for the service providers.

Figure 8–17 shows how an optical network is being deployed to support the residential broadband technology. The most common approach is to employ rings. These rings tie together the customers to the CO with optical-based remote digital terminals (RDTs). The user payload can be added or dropped off at the customer sites through conventional add-drop operations. The optical signals are provisionable, allowing a wide range of data rates. These "typical" rates are a function of the bandwidth requirements of the users attached to the ring.

These optical networks provide a variety of options to allow the service provider to adapt the residential broadband media to specific requirements. One of these requirements could be the use of rings, as

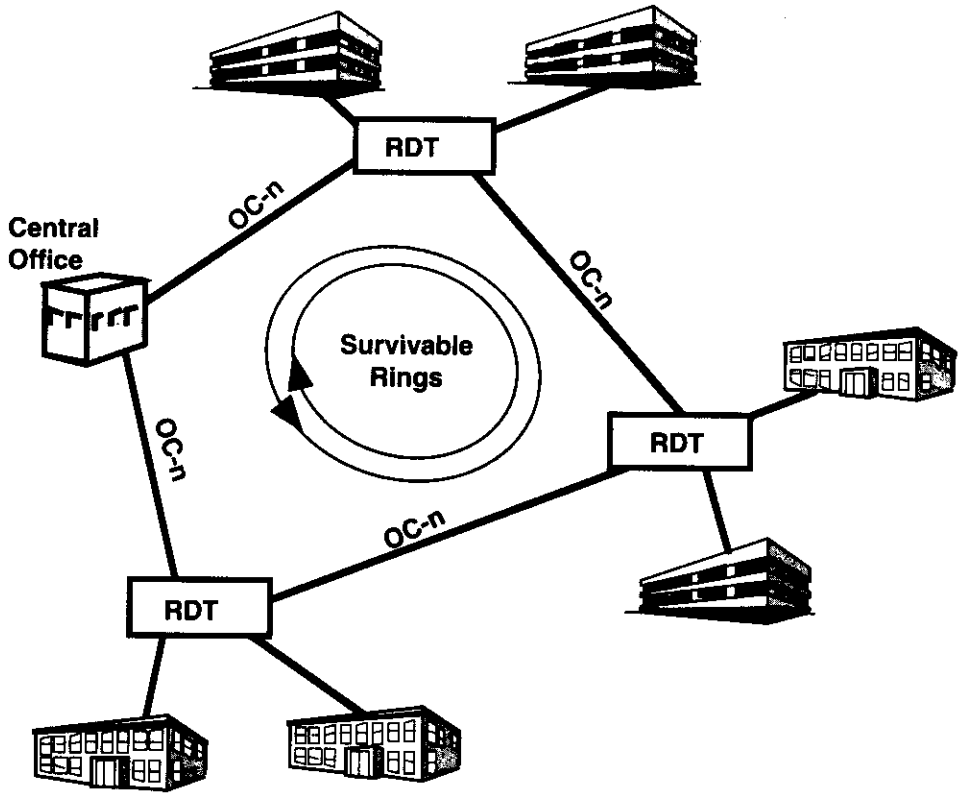


Figure 8–17 Metro optical networking.

illustrated in Figure 8–17, but other requirements may dictate other solutions, such as point-to-point topologies.

SUMMARY

There is no question that the use of backup links and backup nodes is an expensive undertaking. Yet, for large-scale backbone networks such as the telephone system or the Internet, there is no alternative. Downtime, due to a link or nodal failure, is not only inconvenient to the network customers, it might be catastrophic. Thus, the deployment of redundant facilities on rings, point-to-point systems, or even meshed networks is well-accepted and supported by both the network provider and the network customer.

9

MPLS and Optical Networks

Label switching and multiprotocol label switching (MPLS) are considered by many to be key components in third generation transport networks. Therefore, this chapter is devoted to these subjects and how they fit into optical networks. The major features of MPLS are explained with emphasis on the operations of label switching routers (LSRs), label assignments, and label swapping.

As a note to the introduction, this chapter is closely associated with several subsequent chapters in this book. For purposes of organization and clarity, it is necessary to break down the subject matter into different chapters. To lend continuity to the discussions, frequent references are made to related material in these chapters. You can ignore these references; I have written all material to be read serially from start to finish. Still, you might find it helpful to look ahead (or backwards) to the referenced material.

WHAT IS LABEL SWITCHING?

The basic concept of label switching is very simple. To show why, let's assume a user's traffic (say, an email message) is relayed from the user's computer to the recipient's computer. In traditional internets (those that do not use label switching), the method to relay this email is similar to postal mail: A destination address is examined by the relaying entity (for

our work, a router; for the postal service, a mail person). This address determines how the router or mail person forwards the data packet or the mail envelope to the final recipient.

Label switching is different. Instead of using a destination address to make the routing decision, a number (a label) is associated with the packet. In the postal service analogy, a label value is placed on the envelope and is thereafter used in place of the postal address to route the mail to the recipient.

In computer networks, a label is placed in a label header in the packet and is used in place of an address (an Internet Protocol [IP] address, usually). The label is used by the router to direct the traffic to its destination.

Reasons for Using Label Switching

Let's look at the reasons label switching is of such keen interest in the industry. They can be summarized as follows:

- Speed, delay, and jitter: Label switching is considerably faster than traditional IP forwarding. This speed translates into less delay in transporting traffic through the network. It also translates into less variable delay (jitter), an important consideration for applications that cannot tolerate a lot of jitter, such as voice and video.
- Scalability: MPLS allows a large number of IP addresses to be associated with one or a few labels. This approach reduces further the size of address (actually label) tables, and allows a router to support more users; that is, to scale to a large user population.
- Resource consumption: Label switching networks do not need a lot of the network's resources to execute the control mechanisms to establish label switching paths (LSPs) for users' traffic.
- Route control: Most IP-based networks use the concept of destination-based routing, wherein the destination IP address in the IP datagram determines the route through a network. Destination-routing is not always an efficient operation, and MPLS offers methods to use more efficient route control techniques, thus providing a higher level of service to the user.
- Traffic engineering: As part of route control, many of the MPLS operations are designed to allow the network provider to engineer

the links and nodes in the network to support different kinds of traffic, as well as constrain the traffic to specific parts of the network. This idea is important in optimizing expensive network resources.

- Labels and Lambdas: If label switching is used in optical networks, it is possible to correlate (map) a label (or labels) to wavelengths, then use a PXC O/O/O switch for forwarding the traffic, thus reducing further the delay and jitter of user payload processing.

THE FORWARDING EQUIVALENCE CLASS (FEC)

The term FEC is applied to label switching operations. FEC is used to describe an association of discrete packets with a destination address, usually the final recipient of the traffic, such as a host machine. FEC implementations may also associate an FEC value with the destination address and a class of traffic. The class of traffic is associated (typically) with a destination TCP/UDP port number, and/or the protocol ID (PID) field in the IP datagram header.

Why is FEC used? First, it allows the grouping of packets into classes. From this grouping, the FEC value in a packet can be used to set priorities for the handling of the packets, giving higher priority to certain FECs. FECs can be used to support efficient QOS operations. For example, FECs can be associated with high-priority, real-time voice traffic, or low-priority newsgroup traffic, and so on.

Scalability and Granularity: Labels and Wavelengths

The network administrator has control over how big the MPLS forwarding tables become by implementing FEC coarse granularity. If only the IP destination address is used for the FEC, the tables can probably be kept to a manageable size. Yet this "coarse granularity" does not provide a way to support classes of traffic and QOS operations. On the other hand, a network supporting "fine granularity" by using port numbers and PIDs will have more traffic classifications, more FECs, more labels, and a larger forwarding table. This network will not scale as easily to a large user base. Fortunately, label switching networks need not be one or the other. A combination of coarse and fine granularity FECs is permissible.

Nonetheless, the issue is important in relation to how many labels are correlated to optical wavelengths, a topic discussed later in this chapter, and in Chapter 10 (see "Considerations for Interworking Layer 1 Lambdas and Layer 2 Labels"), and Chapter 12 (see "Granularity of Labels vs. Wavelength Support").

TYPES OF MPLS NODES

Figure 9–1 shows the three types of MPLS nodes. They perform the following functions:

- Ingress LSR: Receives native-mode user traffic (for example, IP datagrams), and classifies it into an FEC. It then generates an MPLS header and assigns it an initial label. The IP datagram is encapsulated into the MPLS packet, with the MPLS header attached to the datagram. If it is integrated with a QOS operation (say, DiffServ), the ingress LSR will condition the traffic (such as using different queues for different priorities of the traffic) in accordance with the DiffServ rules.
- Transit, interior, or core LSR: Receives the packet and uses the MPLS header to make forwarding decisions. It will also perform label swapping (exchanging label values). It is not concerned with processing the IP header, only the label header.
- Egress LSR: Performs the decapsulation operations (i.e., it removes the MPLS header).

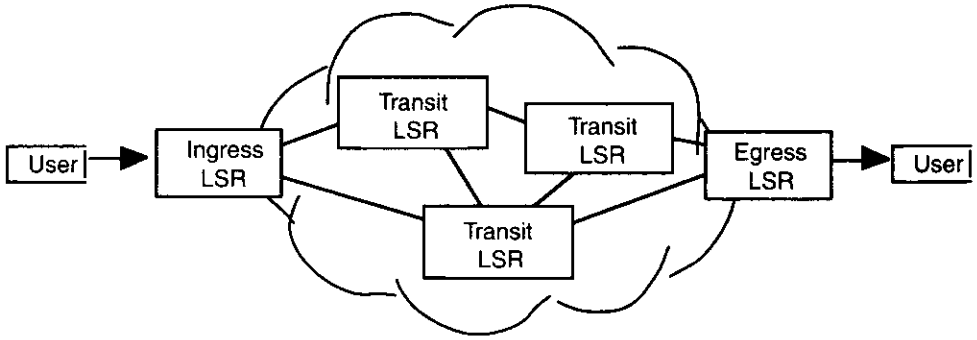


Figure 9–1 The MPLS nodes.

LABEL DISTRIBUTION AND BINDING

To use labels between the LSRs, MPLS executes a control plane to perform (a) the advertising of a range of label values that an LSR wants to use, (b) the advertising of associated IP addresses that are to be associated with the labels, and (c) perhaps the advertising of QOS performance parameters and suggested routes for the user's label switching path through the network. The process of agreeing to these parameters, and then building label switching tables in the LSRs, is called binding.

Methods for Label Distribution

MPLS does not stipulate a specific label distribution protocol.[1] Since several protocols are currently in operation that can support label distribution, it makes sense to use what is available. Nevertheless, the IETF has developed a specific label distribution protocol to complement MPLS that is called the label distribution protocol (LDP).

Another protocol, the constraint-based LDP (CR-LDP), is an extension to LDP. It allows the network manager to set up explicitly routed Label Switched Paths (LSPs). CR-LDP operates independently of any IGP. It is used for delay-sensitive traffic and emulates a circuit-switched network. CR-LDP is also designed to support traffic engineering operations by allowing the network administrator to dictate how and where the users' label switched paths flow through the network.

RSVP can also be used for label distribution; this extension is called RSVP-TE. By using the RSVP PATH and RESV messages (with extensions), it supports label binding and distribution operations. Extensions to BGP are yet another method for advertising and distributing labels. Most of the attention in the industry is focused on extended LDP and extended RSVP for label distribution.

Additionally, an extension to MPLS, called generalized MPLS (GMPLS), has been published that provides information on using MPLS (and extended RSVP or extended LDP) for optical networks. GMPLS is discussed later in this chapter. The other protocols are explained in the companion book to this series on MPLS [BLAC02].

---

[1]Some papers call a label distribution protocol a signaling protocol. If this term is used, be aware that it does not refer to conventional signaling protocols, such as ISDN's Q.931 and SS7's ISUP.

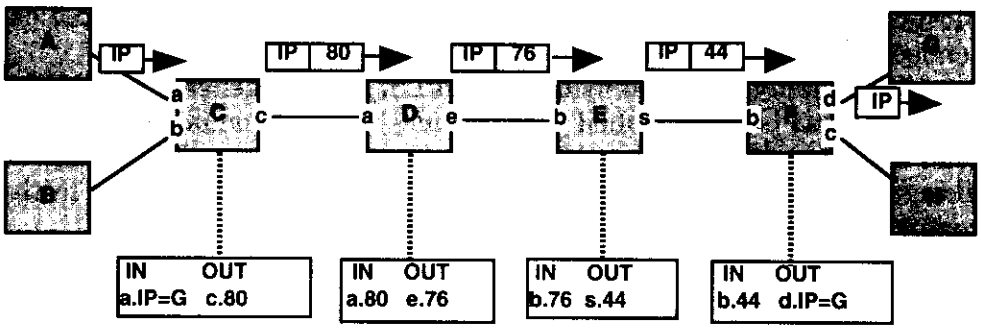
LABEL SWAPPING AND TRAFFIC FORWARDING

In Figure 9–2, nodes A, B, G, and H are user machines and are not configured with MPLS. Node C is the ingress LSR, nodes D and E are transit (interior) LSRs, and node F is the egress LSR.

The example in Figure 9–2 uses generic addresses. For example, the address for node G is "G," which could be an IP address or some other address, such as IPX, a telephone number, etc.

LSR C receives an IP datagram from user node A on interface a. This datagram is destined for node G. LSR C analyzes the FEC fields, correlates the FEC with label 80, encapsulates the datagram behind a label header, and sends the packet to output interface c. The OUT entry in LSR C's table directs it to place label 80 onto the label header in the packet. This operation at LSR C is called a label push.

Hereafter, LSRs D and E process only the label header, and their swapping tables are used to (at LSR D) swap label 80 for label 76, and (at LSR E) swap label 76 for label 44. Notice that the swapping tables use the ingress and egress interfaces at each LSR to correlate the labels to the ingress and egress communications links. Egress LSR F is configured to recognize label 44 on interface b as its own local label; that is, there are no more hops, and the end of the LSP has been reached. Notice that the OUT entry in F's table directs LSR F to send this datagram to G on interface d; this implies removing the label from the packet. This label removal is part of an operation called a label pop.



Notes: Nodes A, B, G, and H are not aware of MPLS in this example. Nodes C, D, E, and F are MPLS-aware.

Figure 9–2 Label swapping and forwarding.

MPLS SUPPORT OF VIRTUAL PRIVATE NETWORKS (VPNs)

Before proceeding with more examples of label operations, the subject of label switching and virtual private networks (VPNs) is discussed. Figure 9–3 shows how MPLS could be used to support a large base of VPN customers with a very simple arrangement.

Certain assumptions must be made for this operation to work well. First, the customers are at the same ends of the MPLS end-to-end path (the label switching path, or LSP). Second, they have the same QOS requirements, and FEC parameters. But these two requirements should not be unusual. For example, many customers may be running VoIP, or Web retrievals, and so forth, from one site to another.

This example shows the idea of label stacking: the placing of more than one label in the MPLS header. Three sets of customers are supported by the VPN in this example (A-D, B-E, and C-F), but there could be hundreds or even thousands or customers. Label stacking allows designated LSRs to exchange information with each other and act as ingress and egress nodes to a large domain of networks and other LSRs. This concept allows certain labels to be processed by a node while others are ignored. The point is that, by label stacking, the VPN backbone can accommodate all the traffic with one set of labels for the LSP in the backbone. The customers' labels are pushed down and are not examined through the MPLS tunnel.

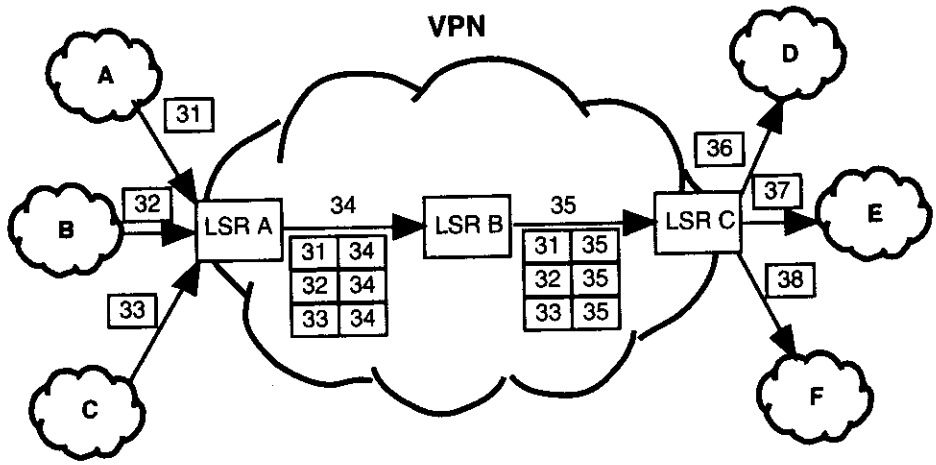


Figure 9–3 Label stacking in a VPN.

For example, labels 31, 32, and 33 from users A, B, and C, respectively, are pushed down into the label header, and label 34 is used between the network LSR A and LSR B to process the packets. LSR B swaps label 34 for label 35, but does not examine labels 31, 32, and 33. When the packets arrive at LSR C, its label switching table reveals that label 35 represents the end of the label switching path. So, LSR C pops label 35, thus revealing labels 31, 32, and 33. Upon examination of these labels, LSR C knows (a) to swap them for labels 36, 37, and 38, respectively, and (b) to pass the packets to end users on networks D, E, and F.

MPLS TRAFFIC ENGINEERING (TE)

As noted, traffic engineering (TE) deals with the performance of a network in supporting the network's customers and their QOS needs. The focus of TE for MPLS networks is (a) the measurement of traffic and (b) the control of traffic. The latter operation deals with operations to ensure the network has the resources to support the users' QOS requirements.

An Internet working group has published RFC 2702. This informational RFC defines in a general way the requirements for traffic engineering over MPLS [AWDU99]. The next part of this chapter provides a summary of [AWDU99], with my comments added to the discussion.

Traffic Oriented or Resource Oriented Performance

Traffic engineering in an MPLS environment establishes objectives with regard to two performance functions: (a) traffic oriented objectives and (b) resource oriented objectives.

Traffic oriented performance supports the QOS operations of user traffic. In a single class, best effort Internet service model, the key traffic oriented performance objectives include: minimizing traffic loss, minimizing delay, maximizing of throughput, and enforcement of service level agreements (SLAs). Resource-oriented performance objectives deal with the network resources, such as communications links, routers, and servers—those entities that contribute to the realization of traffic oriented objectives.

Efficient management of these resources is vital to the attainment of resource oriented performance objectives. Available bandwidth is the bottom line; without bandwidth, any number of TE operations is worthless, and the efficient management of the available bandwidth is the essence of TE.

Traffic Trunks, Traffic Flows, and Label Switched Paths

An important aspect of MPLS TE is the distinctions between traffic trunks, traffic flows, and label switched paths (LSPs). A traffic trunk is an aggregation of traffic flows of the same class which are placed inside an LSP. A traffic trunk can have characteristics associated with it (addresses, port numbers). A traffic trunk can be routed, because it is an aspect of the LSP. Therefore, the path through which the traffic trunk flows can be changed.

MPLS TE concerns itself with mapping traffic trunks onto the physical links of a network through label switched paths. As explained in Chapters 10 and 12, third generation transport networks extend the MPLS traffic engineering of label switched paths to optical switched paths (OSPs) by correlating labels to wavelengths (If you want to look ahead, see "Interworking the Three Control Planes" in Chapter 10, and "Correlating the Wavelength OSP with the MPLS LSP" in Chapter 12).

LDP, CR-LDP, RSVP-TE, and OSPF (Extensions) for TE Support

Four protocols have been developed or extended to provide the signaling capabilities for MPLS. Some of them were introduced earlier in this chapter. They are explained in detail in a companion book to this series [BLAC02], and are summarized here:

- LDP: Designed specifically for MPLS label advertising and distribution operations.
- CR-LDP: An enhanced LDP that supports the building of defined (constrained) LSPs in an MPLS network.
- RSVP-TE: To RSVP that permit the negotiation and establishment of LSPs.
- OSPF (several extensions): Several extensions to OSPF for discovery of LSPs based on network-specific criteria.

MULTIPROTOCOL LAMBDA SWITCHING (MPλS)

The framework for interworking optical networks and MPLS is called MPλS, and is defined in [AWDU01]. As noted in the introduction to this chapter, MPLS and optical networks are a good match, but it is important that this effort has a standardized approach. Both technologies have control mechanisms (a control plane) to manage the user traffic.

These control planes (introduced here and covered in more detail in Chapters 10 and 12), are shown in a general way in Figure 9–4. The MPLS control plane is concerned with label distribution and binding an end-to-end LSP. The optical control plane is concerned with setting up wavelengths, optical coding schemes (SDH/SONET), transfer rates (in bit/s), and protection switching options (1:1, 1:N, etc.) on an OSP between two adjacent nodes.

The [AWDU01] authors hold that it is not a good idea to have different control planes when two technologies must interwork. They cite IP over ATM, with IP using OSPF, BGP, and IS-IS for its control plane and ATM using PNNI (and, to some extent, Q.2931) for its control plane.

I understand [AWDU01]'s points, but some interworking of the MPLS and optical control planes is not only necessary, but desirable (and suggested by the vertical arrows between these planes in Figure 9–4). The reasons for this statement are presented later in this chapter and in Chapters 10, 12, and 13.

For this immediate discussion, [AWDU01] describes the framework for adapting the MPLS TE control plane for optical cross-connects. It is a useful document because it sets up a model for an MPLS-based optical Internet, including an analysis of the similarities and differences between MPLS and optical network control operations.

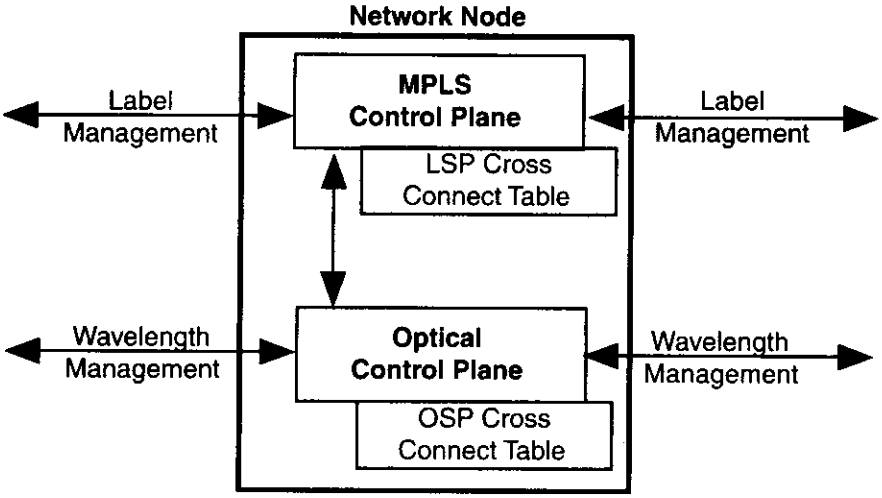


Figure 9–4 The MPLS and optical control planes.

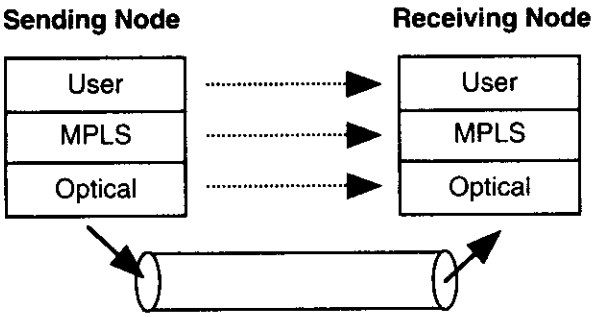


Figure 9–5 MPλS.

Relationships of OXC and MPLS Operations

A convenient way to view the relationship of MPLS and optical networking is through the layered model, as shown in Figure 9–5. The optical operations occur in layer 1; the MPLS operations occur in a combination of layers 2 and 3.[2]

The data plane of an LSR uses the label swapping operation to transfer a labeled packet from an input port to an output port. The data plane of an OXC uses a switching matrix to connect an optical channel trail from an input port to an output port. Recall that an optical trail is an optical connection between two nodes, such as an LSR or an OXC.

Traffic from the control plane of an upper layer can be sent to either the data or control plane of the adjacent lower layer on the transmit side, with a reverse operation occurring on the receive side. In fact, such an approach is common. For example, an LSR may send a control message to an adjacent LSR that sets up some timers for label management operations. This control message could go over either an optical data channel or an optical control channel.

An LSR performs label switching by first establishing a relation between an input port and an input label, and an output port and an output label. Likewise, an OXC provisions an optical channel by first establishing a relation between an input port and an input optical channel (and/or wavelength), and an output port and an output optical channel (and/or wavelength). In the LSR, the next hop label forwarding entry (NHLFE) maintains the input–output relations.

---

[2]Some of the new protocols' operations span layers 2 and 3. ATM is one example; MPLS is cited in some literature as another. Strictly speaking, MPLS is a layer 3 protocol in that it does not define the critical function of layer 2 frame delineation.

In the OXC, the switch controller reconfigures the internal interconnection fabric (called an optical switching path [OSP] cross-connect table or a wavelength forwarding information base [WFIB]) to establish the relationships.

The functions of the control plane include resource discovery, distributed routing control, and connection management. In particular, the control plane of the LSR is used to discover, distribute, and maintain relevant state information associated with the MPLS network, and to manage label switched paths (LSPs).

The control plane of the OXC is used to discover, distribute, and maintain relevant state information associated with the OTN, and to establish and maintain optical channel trails under various optical internetworking traffic engineering rules and policies.

A significant difference between current LSRs and OXCs is that, with LSRs, the forwarding information is carried explicitly as part of the labels appended to data packets, while, with OXCs, the switching information is implied from the wavelength or optical channel. The label is used by the LSP cross-connect table (the NHLFE), and the wavelength is used by the OSP cross-connect table.

MPLS and Optical Wavelength Correlation

A key aspect of MPLS and optical network interworking is to correlate an MPLS label value with an optical wavelength. This operation is introduced in this chapter and explained in more detail in Chapters 10 and 12. In Figure 9–6, the user sends traffic to the network (the definition on page 17 might be helpful). At the ingress node (now noted as an LSR/OXC)

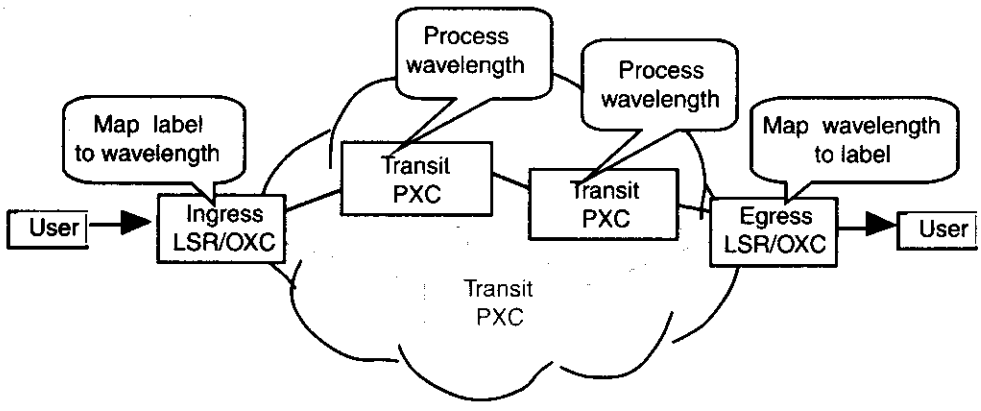


Figure 9–6 Processing the user traffic.

the MPLS label is correlated to an appropriate wavelength; that is, an appropriate channel into the network and out of the network to reach the destination user. The transit nodes, now labeled as transit PXCs, have been configured to process the wavelength to make the routing decisions.

The MPLS label is not examined at the transit PXCs. For this example, it is assumed that the user payload is to be sent through the network from the ingress LSR/OXC to the egress LSR/OXC. Thus, it is not necessary to know about the MPLS label, as long as all nodes know the relationship of the wavelength that is associated with the label, and its final destination.

An explicit LSP is one whose route is defined at its origination node, or by a control protocol such as OSPF (that discovers and sets up a path through the network). However the path is defined, once it is set up, it remains stable, unless problems occur at a node or on an optical trail. Explicit LSPs and optical channel trails exhibit certain commonalties. They are both uni-directional, point-to-point relationships. An explicit LSP provides a packet forwarding path (traffic-trunk) between an ingress LSR and an egress LSR. An optical channel trail provides an optical channel between two endpoints for the transport of user traffic.

The payload carried by both LSPs and optical trails is transparent to intermediate nodes along their respective paths. Both LSPs and optical trails can be configured to stipulate their performance and protection requirements.

Label Merging and Label Stacking. In the MPLS networks, it is possible to merge and stack labels. Label stacking was explained earlier in this chapter. Label merging is the replacement of multiple incoming labels for a particular FEC with a single outgoing label. This operation can reduce the number of labels processed by an LSR.

There are commonalties in the allocation of labels to LSPs and the allocation of wavelengths to optical trails. Two different LSPs that traverse through a given LSR port or interface cannot be allocated the same label. The exception is for LSP aggregation using label merge or label stacking. Similarly, two different optical trails that traverse through a given OXC port cannot be allocated the same wavelength.

Failure of the Optical Connection

In the event of a fiber failure or a fiber node failure, there must be a method to find a backup route. This subject is discussed in Chapter 8 regarding how optical networks use backup fibers and protection switching to recover from failures.

In an MPλS network, there must be very close coordination between the optical and label control planes, if they are indeed different software processes in the nodes. For example, if an optical connection is lost, the optical control plane must be able to inform the MPλS control plane so that neighbor LSRs can be informed of the problem. As the MPLS and the optical layers mature, it is likely that this coordination will be aided by a data link layer especially designed for this purpose; this is the subject of Chapter 11.

MPLS AND OPTICAL TE SIMILARITIES

RFC 2702 establishes the major requirements for TE support in an MPLS network. This part of the chapter summarizes this part of RFC 2702 as well as associated ideas from [AWDU01]. In reading this summary, it will be helpful to visualize the relationship of the optical layer to MPLS by substituting the MPLS term *traffic trunk* with the optical layer term *optical channel trail.*

A traffic trunk is an aggregation of traffic belonging to the same FEC which is forwarded through a common path. It is used in MPLS to allow certain attributes of the traffic transported through LSPs to be configured based on TE parameters, such as delay and throughput. The attributes that can be associated with traffic trunks include:

- Traffic parameters: Indicate the bandwidth requirements of the traffic trunk.
- Adaptivity attributes: Specify the sensitivity of the traffic trunk to changes in the state of the network, with the possibility of rerouting the traffic trunk to a different part of the network.
- Priority attributes: Impose a partial order on the set of traffic trunks and allow path selection and path placement operations to be prioritized.
- Preemption attributes: Indicate whether a traffic trunk can preempt an existing traffic-trunk from its path.
- Resilience attributes: Stipulate the survivability requirements of the traffic trunk and, in particular, the response of the system to faults that impact the path of the traffic trunk.
- Resource class affinity attributes: Further restrict route selection to specific subsets of resources. Allow inclusion and exclusion policies to be implemented.

POSSIBILITIES FOR THE MPλS NETWORK

Additional work remains to be done on a complete MPλS network, but [AWDU01] has established a coherent framework for further efforts. In concluding this chapter, we review several ideas from this IETF working draft, and my thoughts about its suppositions. As noted, other examples of the interworking of labels and wavelengths are provided in Chapters 10, 11, and 13.

If the XC is a wavelength routing switch (a PXC), then the physical fiber between a pair of PXCs can represent a single link in the OTN network topology. Individual wavelengths or channels can be analogous to labels. If there are multiple fibers between a pair of PXCs, then, as an option, these multiple fibers could be logically grouped together through a process called bundling and represented as a single link in the OTN network topology. This concept is supported in the emerging optical link management protocol, discussed in Chapter 11.

If a fiber terminates on a device that functions as both an OXC and an IP router, then the following situations may be possible:

- A subset of optical channels within the fiber may be uncommitted. That is, they are not currently in use and hence are available for allocation.
- A second subset of channels may already be committed for transit purposes. That is, they are already cross-connected by the PXC element to other out-bound optical channels and thus are not immediately available for allocation.
- Another subset of optical channels (within the same fiber) could be in use as terminal channels. That is, they are already allocated but terminate on the local PXC/router device, for example, as SONET interfaces.

In the above scenario, one way to represent the fiber in the OTN network topology is to depict it is as several links, where one of these links would represent the set of uncommitted channels that constitute the residual capacity of the fiber, while each terminal channel that terminates on the PXC/router could be represented as an individual link.

IS-IS or OSPF and possibly additional optical network specific extensions would be used to distribute information about the optical transport network topology, about available bandwidth and available channels per fiber, as well as other OTN network topology state data. This information

is then used to compute explicit routes for optical channel trails. An MPLS signaling protocol, such as RSVP extensions, is used to instantiate the optical channel trails. Using the RSVP extensions, for example, the wavelength information and/or optical fiber information can be carried in the LABEL object, which will be used to control and reconfigure the PXCs.

The use of a uniform control plane technology for both LSRs and PXCs introduces a number of interesting architectural possibilities. One such possibility is that a single MPLS traffic engineering control plane can span both routers and PXCs. In such an environment, a label switched path can traverse an intermix of routers and PXCs, or can span just routers, or just PXCs. This concept offers the potential for real bandwidth-on-demand networking, in which an IP router may dynamically request bandwidth services (a part of the bandwidth of a wavelength, for example) from the optical transport network.

Another possibility cited by [AWDU01] is that PXCs and LSRs may run different instances of the control plane which are decoupled with little or no interaction between the control plane instances. I am not convinced that the decoupling of the MPLS and optical control planes will be the best approach, and I will explain my reasons for this statement as we proceed though the remainder of this book. For now, take a look at footnote 3 for a brief explanation.[3]

To configure the mapping and switching functions, PXCs must be able to exchange control information. The favored method that is emerging in the industry is to preconfigure a dedicated control wavelength between each pair of adjacent PXCs, or between a PXC and a router, and to use this wavelength as a supervisory channel for exchange of signaling and OAM traffic. This idea is examined in Chapters 10, 11, and 12.

In the proposed control plane approach, a PXC maintains a wavelength forwarding information base (WFIB), also called an OSP cross-connect table per interface (or per fiber). This approach is used because lambdas and/or channels (labels) are specific to a particular interface (fiber), and the same lambda and/or channel (label) could be used concurrently on multiple interfaces (fibers).

[3]As I explain in Chapter 10, I favor using an integrated control plane, one in which MPLS labels and optical wavelengths are correlated. This approach facilitates the interworking for MPLS and lambdas for protection switching. It appears to me to be difficult to implement de-coupled lal ! and optical control planes. But then, we are in uncharted waters here, and I would welcome your views after you read my thoughts about this matter in Chapters 10, 12, and 14.

If the bandwidth associated with an LSP is small relative to the capacity of an optical channel trail, then inefficient utilization of network resources might result if only one LSP is mapped onto a given optical channel trail. To improve utilization of resources, therefore, it is necessary to be able to map several low bandwidth LSPs onto a relatively high-capacity optical channel. Note that since a PXC cannot perform label push/pop operations, the start/end of a nested LSP must be on a router (as nesting requires label push/pop).

Control and Data Planes Interworking

My thoughts about this aspect of [AWDU01] (and looking to the future) is that it is not desirable to have a PXC O/O/O node involved in label management; otherwise, it becomes an O/E/O node. Fine, but the questions then remain of how the PXC is going to be able to (a) interwork labels with wavelengths, and (especially) (b) how protection switching performed on fibers and/or wavelengths can be correlated to the labels that are running on the label/wavelength. To illustrate, if an optical channel fails, and it destroys a label channel, the label layer must be aware of this failure (if the optical layer cannot recover).

The key to interworking effectively the MPLS and optical planes is to carefully distinguish the interactions between not only the control planes but also between the data planes and the control planes. The subsequent chapters will lay out the procedures to accomplish these operations.

SUMMARY

The use of MPLS in the emerging 3G transport network is certainly not a given. Traffic engineering and some other features of label switching can be achieved by other means. For example, constrained routing can be set up at either layer 3 or layer 1, or both. But MPLS is gaining a lot of momentum, at least for networks that will be deployed in the future, and many vendors intend to build products that combine and/or interwork the L3, L2, and L1 control planes. In the meantime, stay tuned to tried-and-true ATM over optical.

10

Architecture of IP and MPLS-based Optical Transport Networks

One of the central issues arising in Internet and optical networks is how best to interwork the two sets of technologies. Clearly, IP is not going to go away for the foreseeable future; it is heavily embedded into PCs, palm devices, Web servers, and so on. Optical transport networks are here to stay for a while as well, and so are label switching networks using MPLS.

This chapter picks up on the introductory information on control planes in Chapters 6 and 9, and discusses one approach to the efficient and graceful interworking of IP and optical networks. In addition, the role of MPLS is explained as well. Three control planes of IP, MPLS, and the optical layer are examined, and we show how these planes can interwork to exploit all the features of IP, MPLS, and optical networks.

The chapter concludes with an explanation of generalized MPLS (GMPLS), and its use in optical networks in general and specifically in the G.709 –based OTN, discussed in Chapter 6. As noted in earlier discussions, in some circles the interworking of GMPLS in third generation optical transport networks is called MPλS.

IP, MPLS, AND OPTICAL CONTROL PLANES

This part of the chapter extends the introduction of control planes made in Chapter 9, and describes in more detail a control plane model for third generation transport networks that encompasses three control planes:

(a) IP, (b) MPLS, and (c) optical. Also, it might prove helpful to review the material in Chapter 6 (see Figure 6–2) that introduces the concepts of the control and data planes.

The Internet Control and Data Planes

The Internet control and data planes are well-understood in the industry. But until recently, these two planes have not been so named in most Internet RFCs. As shown in Figure 10–1, they have usually been named the routing layer (for the control plane) and the forwarding layer (for the data plane).

The control plane is comprised of the Internet and ISO routing protocols, shown in Figure 10–1 as OSPF, IS-IS, and BGP. The data plane is comprised of the forwarding protocol, IP.

In order to connect networks together so that they may exchange information, and in order to move traffic through these networks efficiently, a method is needed whereby a specific path (a route) is found among the many nodes (routers, servers, workstations) and routes that connect two or more network users together. The identification of a route entails a route discovery operation, which is called routing, and fits into the control plane of the Internet layered model. In its simplest terms, route discovery in the Internet control plane is the process of finding the best route between two or more nodes in a network or in multiple networks.

The MPLS Control and Data Planes

MPLS also operates with control and data planes, as depicted in Figure 10–2. The job of the control plane is to advertise labels and

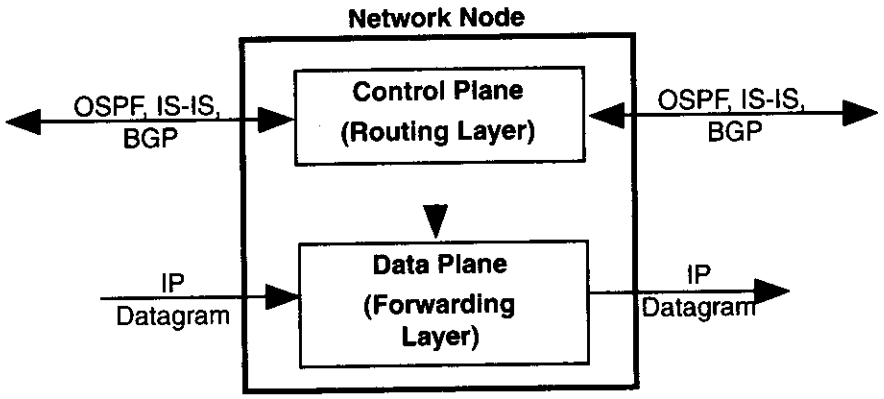


Figure 10–1 The Internet control and data planes.

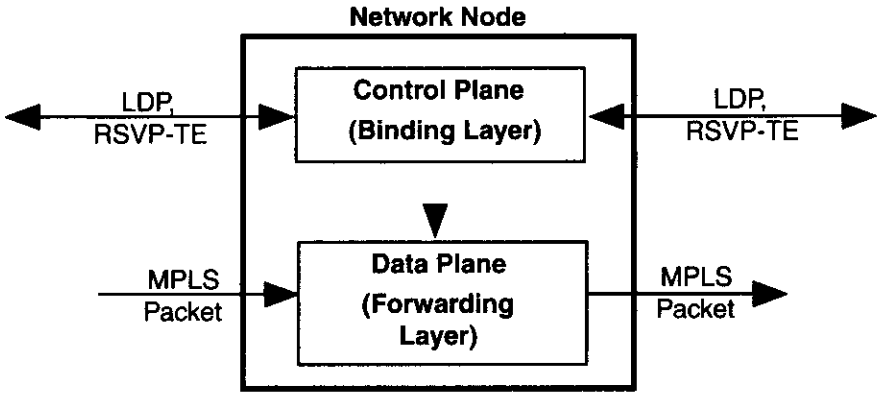


Figure 10–2 MPLS control and data planes.

addresses and to correlate them; that is, to bind labels to addresses. This idea is explained in more detail in Chapter 9.

There is more than one protocol operating at the MPLS control plane. Extensions to RSVP have been made to allow the use of this protocol to advertise, distribute, and bind labels to IP addresses. This extension is called RSVP-TE. LDP is yet another option for executing the MPLS control plane.

After MPLS nodes have exchanged labels and IP addresses, they bind the labels to addresses. Thereafter, the MPLS data plane forwards all traffic by examining the label in the MPLS label. The IP address is not examined until the traffic is delivered across the network (or networks) to the receiving user node. The label is then removed, and the IP address is used by the IP data plane to deliver the traffic to the end user.

The Optical Control and Data Planes

Figure 10–3 shows the optical control and data planes for a third generation optical transport network. The control plane (which I call the λ mapping layer) can be executed with LMP or GMPLS or a combination of the two. Whatever the specific implementation may be, the optical control plane is used to coordinate the use of wavelengths between adjacent optical nodes, as well as to insure the nodes are up and running. As explained later in this chapter, the optical control plane is also responsible for configuring optical nodes to accept different kinds of traffic, such as SDH or SONET formatted frames. It also is used to negotiate the bit transfer rate that is used between nodes.

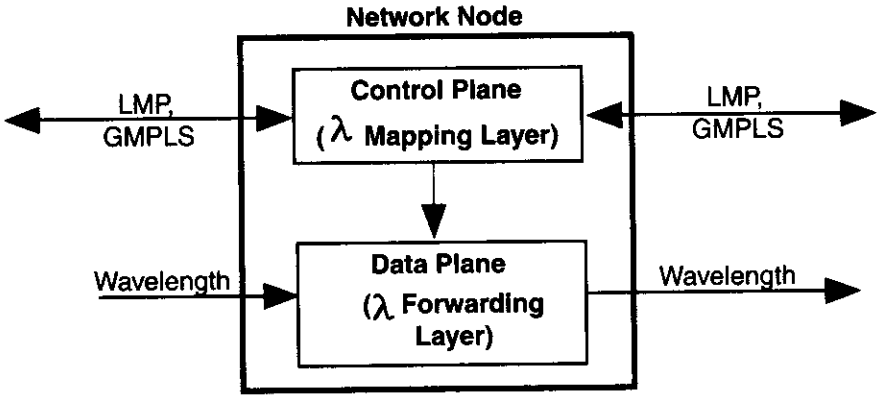


Figure 10–3 The optical control and data planes.

Thereafter, the control plane is invoked only for ongoing management operations, diagnostics, recovery, and so on. User traffic is processed in the data plane (the λ forwarding layer).

Requirements of the Optical Control Plane. The IP optical networking group defines the requirements for the optical control plane [FREE00] and has established that this plane must be able to support the following types of connections:

- A permanent optical channel set up by the network management system via network management protocols.
- A soft permanent optical channel set up by the network management system, using network-generated signaling and routing protocols to establish connections.
- A switched optical channel, which can be set up by the customer on demand using signaling and routing protocols.

INTERWORKING THE THREE CONTROL PLANES

The separate operations of the IP, MPLS, and optical control planes should be coordinated in order to take advantage of (a) the route discovery capabilities of the IP control plane, (b) the traffic engineering capabilities of the MPLS control plane, and (c) the forwarding (switching) speed of the optical data plane. Figure 10–4 illustrates how this interworking can be accomplished, from the author's perspective.

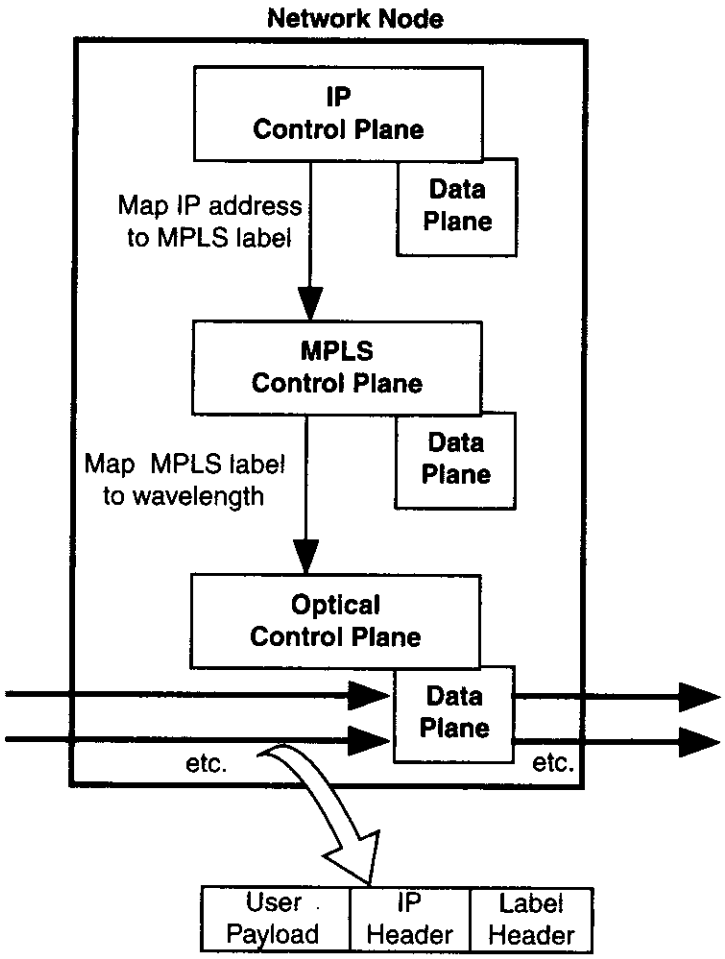


Figure 10–4 Interworking the three control planes.

Using Figure 10–4 for reference, the following three events must take place to exploit the powerful capabilities of all three control planes:

1. The IP routing protocols advertise and discover addresses as well as the routes to the nodes that are identified by the addresses.
2. The MPLS label distribution protocols distribute labels associated with the IP addresses, so that the cumbersome IP addresses do not have to be used in the network. Remember that this idea is called binding; it maps certain addressees to certain labels.

3. The MPLS labels can be mapped to specific wavelengths between adjacent optical nodes so that the nodes can resort to PXC-based O/O/O operations and not be concerned with MPLS label swapping and cumbersome O/E/O operations. Ideally, the same wavelength is used on each OSP segment of the end-to-end LSP.

Events 1 and 2 are beyond the subject of this book; they are covered extensively in other books in this series. Event 3 is certainly a relevant topic for this book and is explained in the next part of this chapter and in Chapters 12 and 14.

MANAGEMENT OF THE PLANES

While some literature suggests that the three control planes should not be coupled, this author strongly disagrees, and several IETF working groups are defining a model for the control plane interworkings. The basic idea is to implement yet another layer, as shown in Figure 10–5. This idea is not new, as networks such as ATM have used this idea for years, and revisions to SS7 also define a management plane. The function of the management plane is to coordinate the interactions of the other planes, or equally important, allow the control planes to operate independently of each other. In most implementations, such as ATM, the interactions are defined by OSI-based primitive calls, which are used by software programmers to write function calls or system calls between the

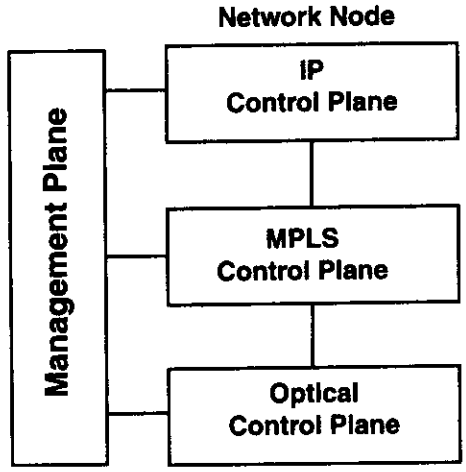


Figure 10–5 The management plane.

management plane and the control planes. For 3G transport networks, these calls have not yet been defined.

Diverse Views on Control Planes' Interworkings

At this stage in the evolution of the optical Internet, there is no common agreement on the exact manner in which the control planes will interact with each other. Consequently, there is no wide-spread agreement on how the many operations in OSPF (extensions), IS-IS, BGP (extensions), RSVP-TE, CR-LDP, LMP, GMPLS, or the OIF UNI and NNI will be executed. Indeed, there are different views on this issue. In the remainder of this chapter, and in Chapters 12 and 14, we will examine the main aspects of these issues, and I will present my views and those of several IETF working groups.

To get this exercise started, the next section provides an overview of two key Internet drafts on the subject; I recommend these papers to you if you are going to be involved in this aspect of 3G transport networks.

A FRAMEWORK FOR IP OVER OPTICAL NETWORKS

[RAJA02] sets forth, in a general way, the issues of operating IP over optical networks. The authors state that there is general consensus in the industry that the optical network control plane should utilize IP-based protocols for dynamic provisioning and restoration of lightpaths within and across optical sub-networks. This opinion is based on the view that signaling and routing mechanisms developed for IP traffic engineering applications could be re-used in optical networks. I agree in general with [RAJA02], but there is no consensus, and I refer you to [FREE00] for an opposing view. I think revisions to current IP-based protocols, and the addition of GMPLS, LMP, and the OIF UNI and NNI procedures will work, and this book uses this model, but [FREE00] provides some balanced, opposing views on the issue. Later, we will take a look at the main arguments for the [FREE00] working group.

Regarding the use of revisions to existing protocols, two major issues are discussed. The first is the adaptation and reuse of IP control plane protocols within the optical network control plane, irrespective of the types of clients that utilize the optical network. The second is the transport of IP traffic through an optical network together with the control and coordination issues that arise therein.

Two General Models

Two general models are discussed in [RAJA02]. I expect that, in future commercial implementations, these models will be merged because a combination of the two will provide the best alternative. Below is a summary of these models (I will refer you to my descriptions of these models in other parts of this book).

Domain Services Model

Under the domain services model, the optical network primarily offers high bandwidth connectivity in the form of lightpaths. Standardized signaling across the UNI is used to invoke the following four services, which are covered in this book in Chapter 13:

- Lightpath creation: This service allows a lightpath with the specified attributes to be created between a pair of termination points in the optical network, perhaps based on network administration decisions, such as security.
- Lightpath deletion: This service allows an existing lightpath to be deleted.
- Lightpath modification: This service allows certain (and limited) parameters of the lightpath to be modified.
- Lightpath status enquiry: This service allows the status of certain parameters of the lightpath to be queried by the router that created the lightpath.

A service discovery procedure may be employed as a precursor to obtaining UNI services. Service discovery allows a client to determine the static parameters of the interconnection with the optical network, including the UNI signaling protocols supported. The protocols for neighbor and service discovery are different from the UNI signaling protocol itself; for example, see LMP in Chapter 11.

Because a small set of well-defined services is offered across the UNI, the signaling protocol requirements are minimal. As you read Chapter 13, you will notice that the UNI message set is relatively sparse. Specifically, the signaling protocol is required to convey a few messages with certain attributes in a point-to-point manner between the router and the optical network. Such a protocol may be based on RSVP-TE or LDP, for example.

The optical domain services model does not deal with the type and nature of routing protocols within and across the optical network. The ODS model results in the establishment of a lightpath topology between routers at the edge of the optical network.

Unified Service Model

With the united service model, the IP and optical networks are treated as a single integrated network from a control plane view. The XCs are treated just like any other router in regard to control plane operations. Thus, in principle, there is no distinction between the UNI, NNIs and any other router-to-router interface from a routing and signaling point of view. It is assumed that this control plane is MPLS-based, as described in Chapter 9 and in the last section of this chapter.

The optical network services are obtained implicitly during end-to-end MPLS signaling. An edge router can create a lightpath with specified attributes, or delete and modify lightpaths as it creates MPLS LSPs. In this regard, the services obtained from the optical network are similar to the domain services model. These services, however, may be invoked in a more seamless manner as compared to the domain services model. For instance, when routers are attached to a single optical network, a remote router could compute an end-to-end path across the optical internetwork.

It can then establish an LSP across the optical internetwork. But the edge routers must still recognize that an LSP across the optical internetwork is a lightpath, or a conduit for multiple LSPs. The concept of "forwarding adjacency" can be used to specify virtual links across optical internetworks in routing protocols such as OSPF. In essence, once a lightpath is established across an optical internetwork between two edge routers, it can be advertised as a forwarding adjacency (a virtual link) between these routers. Thus, from a data plane point of view, the lightpaths result in a virtual overlay between edge routers. The decisions as to when to create such lightpaths, and the bandwidth management for these lightpaths, are identical in both the domain services model and the unified service model.

Interconnections for IP over Optical

Given that IP/MPLS over optical can use the models described above, the transport of the IP datagrams over an optical network can occur though three kinds of interconnections: (a) peer, (b) overlay, and (c) augmented.

Peer. Under the peer model, the IP/MPLS layers act as peers of the optical transport network, so that a single control plane runs over both the IP/MPLS and the optical domains. When there is a single optical network involved, presumably a common routing protocol such as OSPF or IS-IS, with appropriate extensions, can be used to distribute topology information over the integrated IP-optical network. In the case of OSPF, opaque LSAs can be used to advertise topology state information. In the case of IS-IS, extended TLVs can be defined to propagate topology state information.

Overlay. Under the overlay model (supported by the optical domain service interconnect (ODSI)), the IP/MPLS routing, topology distribution, and signaling protocols are independent of the routing, topology distribution, and signaling protocols at the optical layer. Topology distribution, path computation, and signaling protocols are defined for the optical domain. Interactions between signaling and routing are accomplished through UNI-defined procedures.

Augmented. Under the augmented model, there are actually separate routing instances in the IP and optical domains, but information from one routing instance is passed through the other routing instance. For example, external IP addresses could be carried within the optical routing protocols to allow reachability information to be passed to IP clients.

AN OPPOSING VIEW

As noted, [FREE00] takes a different view from [RAJA02]. Here are the main arguments quoted from the [FREE00] working group:

> User plane connections are separable entities from the control plane in the OTN. In particular, user plane and control plane traffic need not be congruently routed. This is one distinguishing feature that does not apply to traditional IP connectionless (CNLS) networks. More importantly, however, the user-plane is transparent, connection-oriented (CO), and has no practical buffering. In essence, the OTN user-plane is completely agnostic regarding the type of traffic it carries (indeed, this is also true of other layer 1 technologies such as SDH/SONET, which can support ATM, IP, PDH, etc. encapsulated within fixed length frames).
>
> A key feature of a layer 1 user plane is its ability to accommodate large administrative bandwidth entities and have the flexibility to accommodate new

client layer networks as they are introduced. These client layer networks may or may not be customer application interfacing.

Given the CO and client-agnostic nature of the OTN, one cannot logically draw the conclusion that a set of control-plane protocols which were originally developed to suit a CNLS environment are the right choice for the control plane of an OTN. Indeed, the only valid justification for this is the re-use of an existing set of protocols, which can save development effort. Based on this argument, one would then have to ask why other control plane protocols developed for a CO environment should not be used.

We have not seen any evidence that such an analysis has been carried out.

Which Approaches to Use?

Once again, it is certainly possible that combinations of all the approaches cited in [RAJA02] and [FREE00] will see implementations. At this juncture, I favor the approaches cited in [RAJA02]. Let's now take a more detailed look at some of the general thoughts of these study groups. We start here with GMPLS and then continue the discussion in the remainder of this book.

GENERALIZED MPLS (GMPLS) USE IN OPTICAL NETWORKS

As noted earlier in this book, the method of distributing and binding labels between LSRs can vary. The LDP can be used, and so can extensions to RSVP. GMPLS has been developed to support MPLS operations in optical networks with the ability to use such optical technologies as time-division (e.g., SONET ADMs), wavelengths, and spatial switching (e.g., incoming port or fiber to outgoing port or fiber) [ASHW01]. This part of the chapter describes those parts of GMPLS that pertain to optical networks.

MPLS assumes that LSRs have a forwarding protocol that is capable of processing and routing signals that have packet, frame, or cell boundaries. LSRs are assumed to be OXCs (O/E/O devices). In contrast, GMPLS assumes that LSRs are PXCs (O/O/O devices) that recognize neither packet nor cell boundaries. Thus, the forwarding decision is based on time slots, wavelengths, or physical ports.

We must pause a moment here to clarify some terms used in GMPLS. Recall that this book uses the term OXC to identify an O/E/O device, the term PXC to identify an O/O/O device, and XC as a generic term to identify either an OXC or a PXC. GMPLS uses the following terms (the remainder of this chapter adopts the GMPLS terminology):

- Packet-switched capable (PXC): Processes traffic based on packet/cell/frame boundaries.
- Time-division multiplex capable (TDM): Processes traffic based on a TDM boundary, such as a SONET/SDH node.
- Lambda-switch capable (LSC): Processes traffic based on the optical wavelength.
- Fiber-switch capable (FSC): Processes traffic based on the physical interface, such as an optical fiber.

Therefore, our term of OXC describes the GMPLS PXC and TDM. Our term of PXC describes the GMPLS LSC.

Traditional MPLS LSPs are uni-directional, but GMPLS supports the establishment of bi-directional LSPs. Bi-directional LSPs have the benefit of lower setup latency and the requirement for fewer messages to support a setup operation.

Considerations for Interworking Layer 1 Lambdas and Layer 2 Labels

In Chapter 2 (see "Considerations for Interworking Layer 1 and Layer 2 Networks"), five points were made about the relationships of layer 1 circuit-switched and layer 2 label-switched networks. Recall that optical switching is considered a layer 1 switching function. To set the stage for the remainder of this chapter, we revisit the five points and modify them in relation to MPLS generally, and to GMPLS specifically.

1. Circuit-switched nodes may have thousands of physical links (ports). A key issue for optical/MPLS networks is the configuration and management of these ports, and the wavelengths on the ports. Insofar as possible, it is desirable not to execute O/E/O functions in the core network in the data plane. Therefore, the ideal lightpath through the network would have use of the same wavelength end-to-end. This is not a trivial task, since it requires all nodes in one or more networks to be able to agree on the specific wavelength. Nevertheless, GMPLS permits the negotiation of these wavelengths.

   Note: At this writing, it is not clear whether or not the optical switches are going to need switching matrices that support these thousands of ports. It appears the MEMS technology is being

pushed into the future, and very large photonic cross-connects have not reached a point of high demand.

2. The layer 1 switch ports do not have IP addresses. It is the intent to use IP addresses for all nodes and the nodes' interfaces. This requirement adds a significant task to the 3G transport network.
3. Layer 1 neighbor nodes do not need to know about their neighbor's internal port number ID; they need to know the channel ID on the port in order to recognize each piece of traffic. This condition still holds when MPLS is added to the mix, and I recommend a per-interface label assignment (as opposed to a per-platform (per-switch) arrangement) in order to more easily meld with current layer 1 practices. Examples that follow in later chapters all use a per-interface label assignment scheme.
4. Many of the circuit-switches' features are configured manually, and the operations remain static (fixed slots, etc.). This practice cannot continue if the network resources are to be dynamically and adaptively utilized. Therefore, the 3G transport network adopts a new concept: The network no longer consists of fixed pipes; it is now dynamically changing. A good analogy is offered by [XU01]: The transport network can be conceived as a large circuit switch with a dynamically-configurable backplane. Thus, the layer 1 operations must be amenable to the same kinds of bandwidth and OAM manipulation as the upper layers (such as ATM and MPLS at layer 2).
5. The switching technology on circuit-switches is based on a very fast hardware-oriented cross-connect fabric, wherein the input and output ports are very tightly synchronized. In an optical/MPLS node, the next hop should be set up by binding the MPLS label in the cross-connect fabric to the output port to that next node. Furthermore, label distribution, say with LDP, is analogous to say SS7 (ISUP) setting up connections at layer 1.

Examples of GMPLS Operations

This part of the chapter explains the GMPLS messages used in an optical network. The message that conveys this information can be sent to an XC via a variety of protocols, such as the extensions to RSVP and LDP. Consequently, we do not need to delve into the specific formats (bit positions, and so on) here, but will confine ourselves to understanding the functions of the fields in the message.

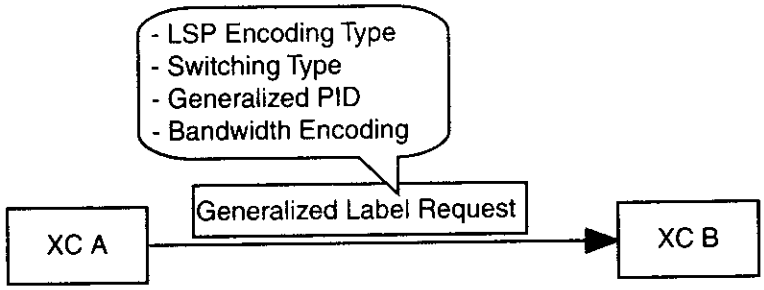


Figure 10–6 The generalized label request.

Generalized Label Request. Figure 10–6 shows one example of how GMPLS is employed. XC A sends a GMPLS message to XC B. This message contains a control label called a Generalized Label Request and its intent is to provide an XC B with sufficient information to set up resources for the connection of an LSP with XC A. As depicted in Figure 10–6, there are three required and one optional fields in this message. These fields perform the following functions:

The first field is the LSP encoding type. It identifies the encoding (format) type that is to be used with the data associated with the LSP. This field tells the receiving node about the specific framing format. Thus far, twelve types have been stipulated, as shown in Table 10–1. Note that

Table 10–1 LSP Encoding Type

Value	Type of Encoding (Format of Traffic)
1	Packet (conventional IP formats)
2	Ethernet V2
3	ANSI PDH (DS1, etc., payloads)
4	ETSI PSH (E1, etc., payloads)
5	SDH ITU-T G.707 (1996)
6	SONET ANSI T1.105 (1995)
7	Digital Wrapper
8	Lambda (photonic)
9	Fiber
10	Ethernet IEEE 802.3
11	SDH ITU-T G.707 (2000)
12	SONET ANSI T1.105 (2000)

Table 10–2 Generalized PID (G-PID)

Value	Type	Technology
0	Unknown	All
1	DS1 SF (Superframe)	ANSI-PDH
2	DS1 ESF (Extended Superframe)	ANSI-PDH
3	DS3 M23	ASNI-PDH
4	DE3 C-Bit Parity	ANSI-PDH
5	Asynchronous mapping of E4	SDH
6	Asynchronous mapping of DS3/T3	SDH
7	Asynchronous mapping of E3	SDH
8	Bit synchronous mapping of E3	SDH
9	Byte synchronous mapping of E3	SDH
10	Asynchronous mapping of DS2/T2	SDH
11	Bit synchronous mapping of DS2/T2	SDH
12	Byte synchronous mapping of DS2/T2	SDH
13	Asynchronous mapping of E1	SDH
14	Byte synchronous mapping of E1	SDH
15	Byte synchronous mapping of 31 * DS0	SDH
16	Asynchronous mapping of DS1/T1	SDH
17	Bit synchronous mapping of DS1/T1	SDH
18	Byte synchronous mapping of DS1/T1	SDH
19	Same as 12, but in a VC-12	SDH
20	Same as 13, but in a VC-12	SDH
21	Same as 14, but in a VC-12	SDH
22	DS1 SF asynchronous	SONET
23	DS1 ESF asynchronous	SONET
24	DS3 M23 asynchronous	SONET
25	DS3 C-bit parity asynchronous	SONET
26	VT	SONET
27	STS	SONET
28	POS – no scrambling, 16 bit CRC	SONET
29	POS – no scrambling, 32 bit CRC	SONET
30	POS – scrambling, 16 bit CRC	SONET
31	POS – scrambling, 32 bit CRC	SONET
32	ATM mapping	SONET/SDH
33	Ethernet	Lambda, Fiber
34	Ethernet	Lambda, Fiber
35	SONET	Lambda, Fiber
36	Digital wrapper	Lambda, Fiber
37	Lambda	Fiber

the Lambda (photonic) encoding type identifies wavelength switching (needing the services of an LSC module). The fiber encoding type identifies an FSC-capable module.

The second field is the switching type, and it informs the XC about the type of switching that is to be perfumed on a particular link. The switching capabilities are defined as (a) several variations for PSC, (b) layer-2 switching capable (for example, ATM and Frame Relay), (c) TDM switching capable, LSC switching capable, or FSC switching capable. The multiple options for PSC allow the network operator to stipulate more than one packet-switched operation.

The third field is the generalized protocol ID (G-PID), and it identifies the payload that is carried by an LSP; that is, the traffic from the client for the specific LSP. This field uses the standard Ethertype values as well as the values shown in Table 10–2.

The fourth field is a bandwidth encoding value. It defines the bandwidth for the LSP. Table 10–3 shows the recommended values for this field.

Table 10–3 Bandwidth Values

Signal Type	Bit Rate
DS0	0.064 Mbit/s
DS1	1.544 Mbit/s
E1	2.048 Mbit/s
DS2	6.312 Mbit/s
E2	8.448 Mbit/s
Ethernet	10.00 Mbit/s
E3	44.736 Mbit/s
DS3	44.736 Mbit/s
STS-1	51.84 Mbit/s
Fast Ethernet	100.00 Mbit/s
E4	139.264 Mbit/s
OC-3/STM-1	155.52 Mbit/s
OC-12/STM-4	622.08 Mbit/s
Gigabit Ethernet	1000.00 Mbit/s
OC-48/STM-1	2488.32 Mbit/s
OC-192-STM-64	9953.28 Mbit/s
10 G-Ethernet - LAN	1000.00 Mbit/s
OC-768/STM-256	39812.12 Mbit/s

Generalized Label. The generalized label request can be extended to identify not only the labels for the packets, but (a) a single wavelength within a waveband or fiber, (b) a single fiber in a bundle of fibers, (c) a single waveband within a fiber, or (d) a set of time slots within a fiber or a waveband. This extension can also identify conventional ATM or Frame Relay labels.

Port and Wavelength Labels. As Chapter 11 explains in more detail, the FSC and PSC may use multiple channels/links that are controlled by a single control channel. If so, the port/wavelength label identifies the port, fiber, or lambda that is used for this purpose.

Wavelength Label. This label groups contiguous wavelengths and identifies them with a unique waveband ID. Its function is to provide a tool for the switch to cross-connect multiple wavelengths as one unit. This label contains three fields:

- Waveband ID: The unique ID (selected by the sending node) of the wavelengths, to be used on all subsequent, related messages.
- Start label: The lowest value wavelength in the waveband.
- End label: The highest value wavelength in the waveband.

Suggested Labels for the Wavelengths

GMPLS can be used to configure the PXC's hardware. One method is called *suggested labels* and is used by an upstream PXC to notify its neighbor downstream of a label that is to be used (suggested) for a wavelength, or a set of wavelengths. Chapter 12 shows this operation in conjunction with the configuration of a PXCs MEMS's hardware.

For this discussion, the GMPLS specification defines the use of a label set to limit the label choices made between adjacent GMPLS nodes. Label sets are useful when an optical node is restricted to a set of wavelengths; obviously, not all optical nodes have the same capabilities. The other helpful aspect of negotiating labels in regards to wavelengths is that some optical nodes are O/E/O capable, and others are O/Ó/O capable, and this aspect of the node will necessarily dictate wavelength capabilities.

BI-DIRECTIONAL LSPs IN OPTICAL NETWORKS

GMPLS defines the use of bi-directional LSPs that have the same traffic engineering requirements in each direction.[1] First, to establish a bi-directional LSP when using RSVP-TE or CR-LDP, two uni-directional paths between peer LSRs must be independently established. The principal disadvantage to this approach is the time it takes to set up this bi-directional relationship. Second, setting up two uni-directional LSPs requires more messages to be exchanged than with the setting up of one symmetrical bi-directional LSP. Third, independent LSPs for a user's traffic profile can lead to different routes taken through the network for the two LSPs. This situation may not be a problem, but it does make for more complex resource allocation schemes.

Of course, since either optical node can initiate label allocations, it is possible for two peer nodes that are in conflict with each other to suggest labels at about the same time. Examples are the same label values for different fibers, or the same labels for different wavelengths on a fiber. This is not a major issue. Contention resolution of potentially conflicting virtual circuit or label bindings is well-studied, and GMPLS defines the procedures for resolving these label contention conflicts.

Link Protection

Another attractive feature of GMPLS is the provision for link protection information, including what kind of protection is needed. The following types of link protection are defined in GMPLS:

- Enhanced: Indicates that a protection scheme that is more reliable than dedicated 1+1 should be used (e.g., 4-fiber).
- Dedicated 1+1: Indicates a dedicated link layer protection scheme.

[1]My clients and I have long favored the simultaneous setting up of bi-directional virtual circuits in X.25, Frame Relay, ATM, and for MPLS: LSPs. It is good news to us that GMPLS adapts this approach for the reasons cited in the text of this chapter. Of course, one idea behind two uni-directional connections/LSPs for a user's traffic is the understanding that many users' traffic flows are asymmetric, with more traffic flow in one direction than the other. The idea is to find the bandwidth, wherever it is in the network, to support this asymmetric flow. Fine, but a multi-gigabit, fiber network makes the argument for independent uni-directional traffic paths even less tenable.

- Dedicated 1:1: Indicates that a dedicated link layer protection scheme, i.e., 1:1 protection, should be used to support the LSP.
- Shared: Indicates that a shared link layer protection scheme, such as 1:N protection, should be used to support the LSP.
- Unprotected: Indicates that the LSP should not use any link layer protection.
- Extra Traffic: Indicates that the LSP should use links that are protecting other higher priority traffic. Such LSPs may be preempted when the links carrying the higher priority traffic being protected fail.

THE NEXT HORIZON: GMPLS EXTENSIONS FOR G.709

G.709 establishes the framework for new-generation optical networks; this was explained in Chapter 6. [BELL01] sets forth the guidelines for interworking GMPLS with G.709. This part of the chapter highlights [BELL01], and I recommend this working draft to you if you need more details on the G.709/MPLS (otherwise, this part of the chapter probably contains too much detail for the casual reader).

Also, as noted earlier, it is important to repeat that the IETF working drafts are subject to change, so you should frequently go to www.ietf.org to make certain that you have the latest version.

Adapting GMPLS to control G.709 can be achieved by considering that G.709 defines two transport hierarchies: a digital hierarchy (also known as the digital wrapper) and an optical transport hierarchy. First, within the digital hierarchy (the previously defined digital wrapper), a digital path layer is defined. Then, within the optical transport hierarchy, an optical channel layer or optical path layer, including a digital OTM overhead signal (OOS; i.e., a non-associated overhead) is defined.

The generalized label request includes a technology independent part and a technology dependent part (i.e., the traffic parameters).

Technology Independent Part

The GMPLS LSP encoding type and the generalized protocol identifier (G-PID) constitute the technology independent part. Since G.709 defines two networking layers (ODUk layers and OCh layer), the LSP encoding type can reflect these two layers or can be considered as a common layer.

If an LSP encoding type is specified per networking layer or, more precisely, per group of functional networking layer (i.e., ODUk and OCh), then the signal type must not reflect these layers. This means that two LSP encoding types have to be defined: (a) one reflecting the digital hierarchy (the digital wrapper layer) through the definition of the digital path layer (i.e., the ODUk layers) and (b) the other reflecting the optical transport hierarchy through the definition of the optical path layer (i.e., the OCh layer).

The G-PID identifies the payload carried by an LSP. The G-PID, which defines the client layer of that LSP, is used by the G.709 endpoints of the LSP. The G-PID could take one of the following values at the digital path layer, in addition to the payload identifiers already defined in GMPLS:

- CBRa: asynchronous constant bit rate (i.e., STM-16/OC-48, STM-64/OC-192 and STM-256/OC-768)
- CBRb: bit synchronous constant bit rate (i.e., STM-16/OC-48, STM-64/OC-192 and STM-256/OC-768)
- ATM: constant bit rate at 2.5, 10, and 40 Gbit/s
- BSOT: non-specific client bit stream with octet timing at 2.5, 10, and 40 Gbit/s
- BSNT: non-specific client bit stream without octet timing at 2.5, 10, and 40 Gbit/s

The G-PID defined in GMPLS are then used when the client payloads are encapsulated through the GFP mapping procedure: Ethernet, ATM mapping and IP packets.

Backward Compatibility. In order to include pre-OTN developments, the G-PID at the optical channel layer can, in addition to the G.709 digital path layer (at 2.5 Gbit/s for ODU1, 10 Gbit/s for ODU2, and 40 Gbit/s for ODU3), take one of these values: (a) SDH: STM-16, STM-64, and STM-256, (b) SONET: OC-48, OC-192, and OC-768, and (c) Ethernet: 1 Gbit/s and 10 Gbit/s.

Technology Dependent Part

The technology dependent of the generalized label request (also referred to as traffic-parameters) must reflect the following G.709 features: (a) ODUk mapping, (b) ODUk multiplexing, (c) OCh multiplexing,

(d) OTM overhead signal (OOS), and (e) transparency (only for pre-OTN). As defined in GMPLS, the traffic-parameters must include the technology-specific G.709 networking signal types (i.e., the signals processed by the GMPLS control-plane). The corresponding identifiers reflect the signal types requested during the LSP setup. The following signal types must be considered: ODU1, ODU2, ODU3, and (at least one) OCh.

A second field must indicate the type of multiplexing being requested for ODUk LSP or OCh LSP. Two kinds of multiplexing are currently defined: flexible multiplexing (or simply multiplexing) and inverse multiplexing.

At the ODUk layer (i.e., digital path layer), flexible multiplexing refers to the mapping of an ODU2 into four arbitrary OPU3 tributary slots (i.e., each slot containing one ODU1) arbitrarily selected. Inverse multiplexing currently under definition at ITU-T should also be considered. The requested multiplexing type must include a default value indicating that neither ODUk flexible multiplexing nor ODUk inverse multiplexing is requested.

At the Och layer, flexible multiplexing is not defined today while inverse multiplexing means that the requested composed signal constitutes a waveband (i.e., an optical channel multiplex). A waveband, denoted as OCh[j.k] (j >= 1), is defined as a non-contiguous set of identical optical channels j x OCh, each of them associated with an OTM-x.m (x = nr or n) subinterface signal. The bit rate of each OCh constituting the waveband (i.e., the composed L-LSP) must be identical; k is unique per OCh multiplex.

Consequently, since the number of identical components included in an ODU multiplex or an OCh multiplex is arbitrary, a dedicated field indicating the requested number of components must also be defined in order to reflect individual signals constituting the requested LSP.

OTM Overhead Signal (OOS)

A dedicated field should support the following options:

- With OTM-0r.m and OTM-nr.m interfaces (reduced functionality stack), OTM overhead signal (OOS) is not supported. Therefore, with these types of interface signals, non-associated OTM overhead indication is not required.
- With OTM-n.m interfaces (full functionality stack), the OOS is supported and mapped into the Optical Supervisory Channel (OSC), which is multiplexed into the OTM-n.m using wavelength division multiplexing.

- With OTM-n.m interfaces or even with OTM-0.m and OTM-nr.m interfaces, non-standard OOS can be defined to allow for instance interoperability with pre-OTN based devices or with any optical devices that do not support G.709 OOS specification. This specific OOS enables the use of any proprietary monitoring signal exchange through any kind of supervisory channel (it can be transported by using any kind of IP-based control channel).

Transparency

Transparency is defined only for pre-OTN developments since, by definition, any signal transported over an OTN is fully transparent. This feature is used to request a pre-OTN LSP (i.e., a non-standard lambda-LSP) including transparency support. It may also be used to set up the transparency process to be applied in each intermediate LSR.

As is commonly the case today with pre-OTN capable interfaces, three kinds of transparency levels are currently defined:

- SONET/SDH Pre-OTN interfaces with RS/Section and MS/Line overhead transparency: The pre-OTN network is capable of transporting transparently STM-N/OC-N signals.
- SONET/SDH Pre-OTN interfaces terminating RS/Section overhead with MS/Line overhead transparency: The pre-OTN network is capable of transporting transparently MSn signals.
- SONET/SDH Pre-OTN interfaces terminating RS/Section and MS/Line overhead: The pre-OTN network is capable of transporting transparently HOVC/STS-SPE signals.

For pre-OTN optical channels a specific field (in the generalized label request) must indicate the transparency level requested during the L-LSP setup. However, this field is relevant only when the LSP encoding type value corresponds to the on-standard lambda layer.

G.709 Label Space

The G.709 label space must include two sub-spaces: the first reflecting the digital path layer (i.e., the ODUk layers) and the second, the optical path layer (i.e., the OCh layer).

ODUk Label Space. As noted in Chapter 6, the digital path layer (i.e., ODUk layers), G.709 defines three different client payload bit rates. An optical data unit (ODU) frame has been defined for each of these bit

rates. Recall that ODUk refers to the frame at bit rate k, where k = 1 (for 2.5 Gbit/s), 2 (for 10 Gbit/s), or 3 (for 40 Gbit/s).

In addition to the support of ODUk mapping into OTUk, the G.709 label space must support the sub-levels of ODUk flexible multiplexing (or simply ODUk multiplexing):

- ODU2 multiplexing: The mapping of an ODU2 into four arbitrary OPU3 tributary slots selected arbitrarily (i.e., each slot containing one ODU1).
- ODU3 multiplexing: Not applicable today since higher order OPU tributary slots are not defined in the current G.709 recommendation.

Also recall that the value space of the k1, k2, and k3 fields are defined as follows:

- k1: indicates a particular ODU1 in one ODU2 (k1 = 1,..,4), ODU3 (k1 = 5,..,20); k1 values from 21 to 84 are reserved for future use.
- k2: indicates a particular ODU2 in one ODU3 (k2 = 1,..,4); k2 values from 5 to 20 are reserved for future use.
- k3: k3 values (k3 = 1,..,4) are reserved for future use.

If k1, k2, and k3 values are equal to zero, the corresponding ODUk are not structured; that is, k[i]=0 (i=1,2,3) indicates that the ODU[I] is not structured and the ODU[i] is simply mapped into the OTU[I].

If k1 and k2 values are equal to zero, a particular ODUk is not structured; that is, ki=0 indicates that the ODUi is not structured. Here are some examples:

- k2=0, k1=0 indicates a full ODU3 (full 40 Gbit/s).
- k2=0, k1=3 indicates the third unstructured ODU1 in the ODU2.
- k2=2, k1=0 indicates the second unstructured ODU2 in the ODU3.
- k2=0, k1=8 indicates the fourth unstructured ODU1 in the ODU3.
- k2=4, k1=2 indicates the second ODU1 of the fourth ODU2 in the ODU3.

OCh Label Space

The OCh label space should be consistently defined as a flat value space whose values reflect the local assignment of OCh identifiers corresponding to the OTM-x.m sub-interface signals (m = 1, 2, or 3 and x = 0r,