CHAPTER 2

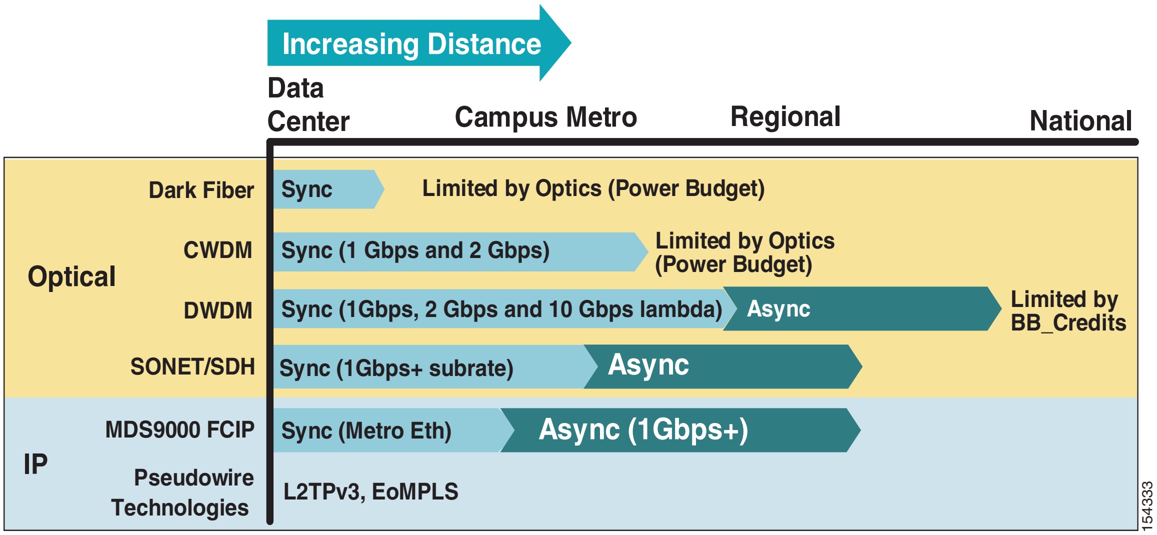
Data Center Transport Technologies

A wide variety of transport options for interconnecting the data centers provide various features and allow many different distances. Achievable distances depend on many factors such as the power budget of the optics, the lambda used for the transmission, the type of fiber, buffer-to-buffer credits, and so forth.

Before discussing some of the available technologies, it is important to consider the features of the LAN and SAN switches that provide higher availability for the data center interconnect. The required convergence time from the application that is going to use these features is also important.

Figure 2-1 shows the various transport technologies and distances.

Figure 2-1 Transport Technologies and Distances



Redundancy and Client Protection Technologies

EtherChanneling on the LAN switches and port channeling on the Cisco MDS Fibre Channel switches are two typical technologies that are used to provide availability and increased bandwidth from redundant fibers, pseudowires, or lambda.

EtherChannels allow you to bundle multiple ports for redundancy and/or increased bandwidth. Each switch connects to the other switch, with up to eight links bundled together as a single port with eight times the throughput capacity (if these are gigabit ports, an 8-Gigabit port results).

The following are benefits of channeling:

- Sub-second convergence for link failures—If you lose any of the links in the channel, the switch detects the failure and distributes the traffic on the remaining links.
- Increased bandwidth—Each port channel link has as much bandwidth as the sum of the bundled links.
- All links are active.

You can configure EtherChannels manually, or you can use Port Aggregation Protocol (PAgP) or Link Aggregation Control Protocol (LACP) to form EtherChannels. The EtherChannel protocols allow ports with similar characteristics to form an EtherChannel through dynamic negotiation with connected network devices. PAgP is a Cisco-proprietary protocol and LACP is defined in IEEE 802.3ad.

EtherChannel load balancing can use the following:

- MAC addresses or IP addresses
- Layer 4 port numbers
- Either source or destination, or both source and destination addresses or ports

The selected mode applies to all EtherChannels configured on the switch. EtherChannel load balancing can also use the Layer 4 port information. An EtherChannel can be configured to be an IEEE 802.1q trunk, thus carrying multiple VLANs.

For more information, see the following URL:
http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SXF/native/configuration/guide/channel.html.

When an EtherChannel link goes down, and there are at least *min-links* up (which by default is 1), the EtherChannel stays up, and spanning tree or the routing protocols running on top of the EtherChannel do not have to reconverge. The detection speed of the link failure is immediate if the devices are connected directly via a fiber or via an optical transport technology. The detection might take longer on a *pseudo-wire*.

Fibre Channel port channeling provides the ability to aggregate multiple physical inter-switch links (ISLs) into a logical ISL (up to 16 ports). The load sharing on the link members is based on source and destination ID (SID/DID) and exchange ID(SID/DID/OXID). If one link fails, the traffic is redistributed among the remaining member links in the channel and is transparent to the end applications. The *Port Channel* feature supports both E_port and TE_port modes, creating a virtual ISL or EISL that allows transporting multiple virtual storage area networks (VSANs).

When a port channel link goes down and at least one link within the channel group is still functional, there is no topology change in the fabric.

Dark Fiber

Dark fiber is a viable method for SAN extension over data center or campus distances. The maximum attainable distance is a function of the optical characteristics (transmit power and receive sensitivity) of the LED or laser that resides in a Small Form-Factor Pluggable (SFP) or Gigabit Interface Converter (GBIC) transponder, combined with the number of fiber joins, and the attenuation of the fiber. Lower cost MultiMode Fiber (MMF) with 850 nm SX SFPs/GBICs are used in and around data center rooms. SingleMode Fiber (SMF) with 1310 nm or 1550 nm SFPs/GBICs are used over longer distances.

Pluggable Optics Characteristics

The following list provides additional information about the wavelength and the distance achieved by various GigabitEthernet, 10 GigabitEthernet, Fibre Channel 1 Gbps, and Fibre Channel 2 Gbps GBICs and SFPs. For data center connectivity, the preferred version is obviously the long wavelength or extra long wavelength version.

- 1000BASE-SX GBIC and SFP—GigabitEthernet transceiver that transmits at 850 nm on MMF. The maximum distance is 550 m on MMF with core size of 50 um and *multimodal bandwidth.distance* of 500 MHz.km.
- 1000BASE-LX/LH GBIC and SFP—GigabitEthernet transceiver that transmits at 1300 nm on either MMF or SMF. The maximum distance is 550 m on MMF fiber with core size of 62.5 um or 50 um and *multimodal bandwidth.distance* respectively of 500 MHz.km and 400 MHz.km and 10 km on SMF with 9/10 um mode field diameter, ~8.3 um core (ITU-T G.652 SMF).
- 1000BASE-ZX GBIC and SFP—GigabitEthernet transceiver that transmits at 1550 nm on SMF. The maximum distance is 70 km on regular ITU-T G.652 SMF (9/10 um mode field diameter, ~8.3 um core) and 100 km on with dispersion shifted SMF.
- 10GBASE-SR XENPAK—10 GigabitEthernet transceiver that transmits at 850 nm on MMF. The maximum distance is 300 m on 50 um core MMF with *multimodal bandwidth.distance* of 2000 MHz.km.
- 10GBASE-LX4 XENPAK—10 GigabitEthernet transceiver that transmits at 1310 nm on MMF. The maximum distance is 300 m with 50 um or 62.5 um core and *multimodal bandwidth.distance* of 500 MHz.km.
- 10BASE-LR XENPAK—10 GigabitEthernet transceiver that transmits at 1310 nm on ITU-T G.652 SMF. The maximum distance is ~10 km.
- 10BASE-ER XENPAK—10 GigabitEthernet transceiver that transmits at 1550 nm on ITU-T G.652 SMF. The maximum distance is 40 km.
- 10BASE-ER XENPAK—10 GigabitEthernet transceiver that transmits at 1550 nm on any SMF type. The maximum distance is ~80 km.
- SFP-FC-2G-SW—1 Gbps or 2 Gbps Fibre Channel transceiver that transmits at 850 nm on MMF. The maximum distance on 50 um core MMF is 500 m at 1.06 Gbps and 300 m at 2.125 Gbps.
- SFP-FC-2G-LW—1 Gbps or 2 Gbps Fibre Channel transceiver that transmits at 1310 nm on SMF. The maximum distance is 10 km on 9 um mode field diameter SMF for either speed.
- SFP-FCGE-SW—Triple-Rate Multiprotocol SFP that can be used as Gigabit Ethernet or Fibre Channel transceiver. It transmits at 810 nm on MMF. The maximum distance is 500 m on MMF with core of 50 um.
- SFP-FCGE-LW—Triple-Rate Multiprotocol SFP that can be used as Gigabit Ethernet or Fibre Channel transceiver. It transmits at 1310 nm on SMF. The maximum distance is 10 km on SMF with mode field diameter of 9 um.

For a complete list of Cisco Gigabit, 10 Gigabit, Course Wave Division Multiplexing (CWDM), and Dense Wave Division Multiplexing (DWDM) transceiver modules, see the following URL: http://www.cisco.com/en/US/products/hw/modules/ps5455/products_data_sheets_list.html.

For a list of Cisco Fibre Channel transceivers, see the following URL: http://www.cisco.com/warp/public/cc/pd/ps4159/ps4358/prodlit/mds9k_ds.pdf

Note On MMF, the modal bandwidth that characterizes different fibers is a limiting factor in the maximum distance that can be achieved. The *bandwidth.distance* divided by the bandwidth used for the transmission gives the maximum distance.

CWDM

When using dark fiber with long wavelength transceivers, the maximum achievable distance is ~10 km. CWDM and DWDM allow greater distances. Before discussing CWDM and DWDM, it is important to be familiar with the ITU G.694.2 CWDM grid, and more specifically the transmission bands (most systems operate in the 1470–1610 nm range):

- O-band—Original band, which ranges from 1260 nm to 1360 nm
- E-band—Extended band, which ranges from 1360 nm to 1460 nm
- S-band—Short band, which ranges 1460 nm to 1530 nm
- C-band—Conventional band, which ranges from 1530 nm to 1565 nm
- L-band—Long band, which ranges from 1565 nm to 1625 nm
- U-band—Ultra long band, which ranges from 1625 nm to 1675 nm

CWDM allows multiple 1 Gbps or 2 Gbps channels (or colors) to share a single fiber pair. Channels are spaced at 20 nm, which means that there are 18 possible channels between 1260 nm and 1610 nm. Most systems support channels in the 1470–1610 nm range. Each channel uses a differently colored SFP or GBIC. These channels are networked with a variety of wavelength-specific add-drop multiplexers to enable an assortment of ring or point-to-point topologies. Cisco offers CWDM GBICs, SFPs, and add-drop multiplexers that work with the following wavelengths spaced at 20 nm: 1470, 1490, 1510, 1530, 1550, 1570, 1590, and 1610 nm:

- CWDM 1470-nm SFP; Gigabit Ethernet and 1 Gbps and 2 Gbps Fibre Channel, gray
- CWDM 1490-nm SFP; Gigabit Ethernet and 1 Gbps and 2 Gbps Fibre Channel, violet
- CWDM 1510-nm SFP; Gigabit Ethernet and 1 Gbps and 2 Gbps Fibre Channel, blue
- CWDM 1530-nm SFP; Gigabit Ethernet and 1 Gbps and 2-Gbps Fibre Channel, green
- CWDM 1550-nm SFP; Gigabit Ethernet and 1Gbps and 2 Gbps Fibre Channel, yellow
- CWDM 1570-nm SFP; Gigabit Ethernet and 1 Gbps and 2 Gbps Fibre Channel, orange
- CWDM 1590-nm SFP; Gigabit Ethernet and 1 Gbps and 2 Gbps Fibre Channel, red
- CWDM 1610-nm SFP; Gigabit Ethernet and 1 Gbps and 2 Gbps Fibre Channel, brown

For a complete list of Cisco Gigabit, 10 Gigabit, CWDM, and DWDM transceiver modules, see the following URL:
http://www.cisco.com/en/US/products/hw/modules/ps5455/products_data_sheets_list.html.

For a list of Cisco Fibre Channel transceivers, see the following URL:
http://www.cisco.com/warp/public/cc/pd/ps4159/ps4358/prodlit/mds9k_ds.pdf

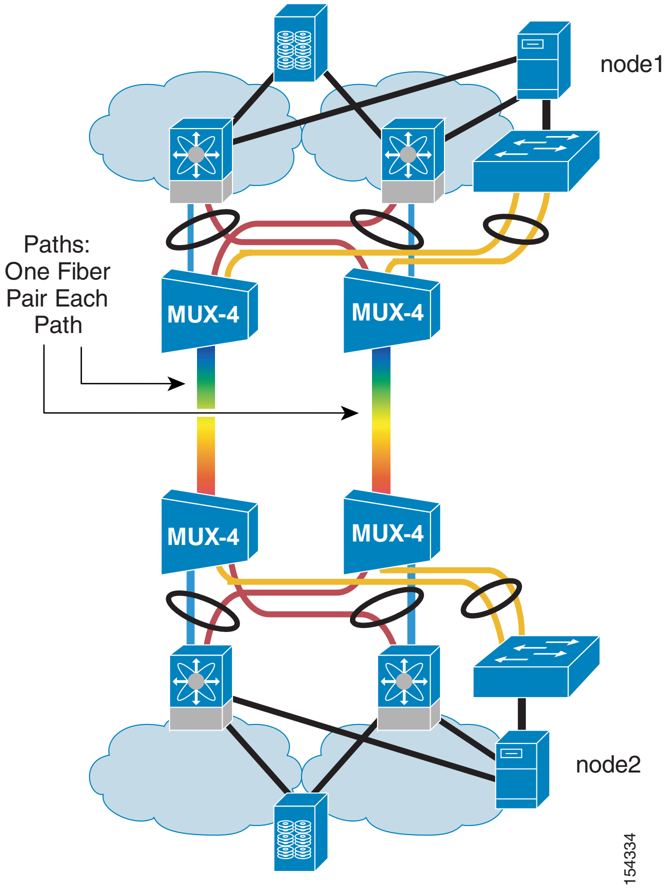
CWDM works on the following SMF fibers:

- ITU-T G.652 (standard SMF)
- ITU-T G.652.C (zero water peak fiber)
- ITU-T G.655 (non-zero dispersion shifted fiber)
- ITU-T G.653 (dispersion shifted fiber)

The CWDM wavelengths are not amplifiable and thus are limited in distance according to the number of joins and drops. A typical CWDM SFP has a 30dB power budget, so it can reach up to ~90 km in a point-to-point topology, or around 40 km in a ring topology with 0.25 db/km fiber loss, and 2x 0.5 db connector loss .

CWDM technology does not intrinsically offer redundancy mechanisms to protect against fiber failures. Redundancy is built with *client protection*. In other words, the device connecting to the CWDM "cloud" must work around fiber failures by leveraging technologies such as EtherChanneling. Figure 2-2 shows an example of a cluster with two nodes, where the SAN and the LAN are extended over ~90 km with CWDM. This topology protects against fiber cuts because port channeling on the Cisco MDS or the Catalyst switch detects the link failure and sends the traffic to the remaining link. When both fibers are available, the traffic can take both paths.

Figure 2-2 Client Protection with CWDM



DWDM

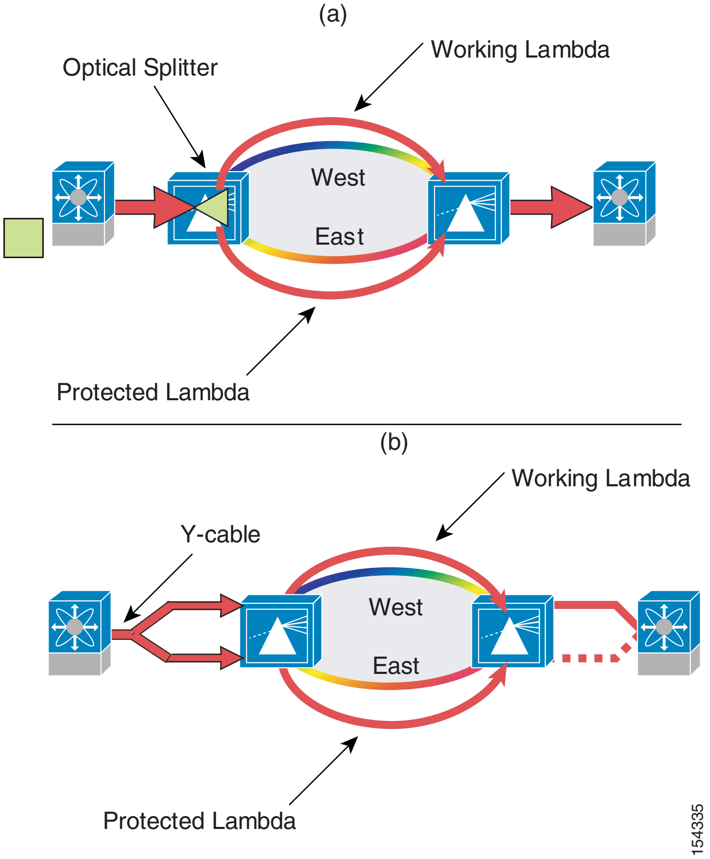
DWDM enables up to 32 channels (lambdas) to share a single fiber pair. Each of the 32 channels can operate at up to 10 Gbps. DWDM networks can be designed either as multiplexing networks similar to CWDM or with a variety of protection schemes to guard against failures in the fiber plant. DWDM is

Erbium-Doped Fiber Amplifier (EDFA)-amplifiable, which allows greater distances. DWDM can transport Gigabit Ethernet, 10 Gigabit Ethernet, Fibre Channel 1 Gbps and 2 Gbps, FICON, ESCON, and IBM GDPS. DWDM runs on SMF ITU-T G.652 and G.655 fibers.

DWDM offers the following protection mechanisms:

- Client protection—Leveraging EtherChanneling and Fibre Channel Port Channeling, this mechanism protects against fiber or line card failures by using the remaining path, without causing spanning tree or routing protocol recalculations, a new principle selection, or FSPF recalculation. With client protection, you can use both west and east links simultaneously, thus optimizing the bandwidth utilization (be careful if the west and east path have different lengths because this can cause out of order exchanges). For example, you can build a two-port port channel where one port uses a lambda on the west path and the other port uses a lambda on the east path.
- Optical splitter protection—Assume that the DWDM optical devices are connected in a ring topology such as is shown in Figure 2-3 (a). The traffic is split and sent out both a west and east path, where one is the working path and one is the “protected” path. The lambda used on both paths is the same because this operation is performed by a single transponder; also, the power of the signal is 50 percent on each path. The receiving transponder chooses only one of the two signals and sends it out to the client. Traffic is switched from a working path to a protected path in the event of a fiber failure. Switchover times for DWDM are ~50 ms or less and may cause a link up/down. This mechanism does not protect against line card failures.
- Y-cable and redundant transponders—Assume the DWDM optical devices are connected in a ring topology as shown in Figure 2-3 (b). The transceiver connects to two DWDM transponders, which in their turn respectively connect to the west mux and the east mux. The signal is sent on both the west and east path with the same power (because there is one transponder per cable termination). Each side can use a different lambda. Only one of the two receiving transponders transmits to the client.

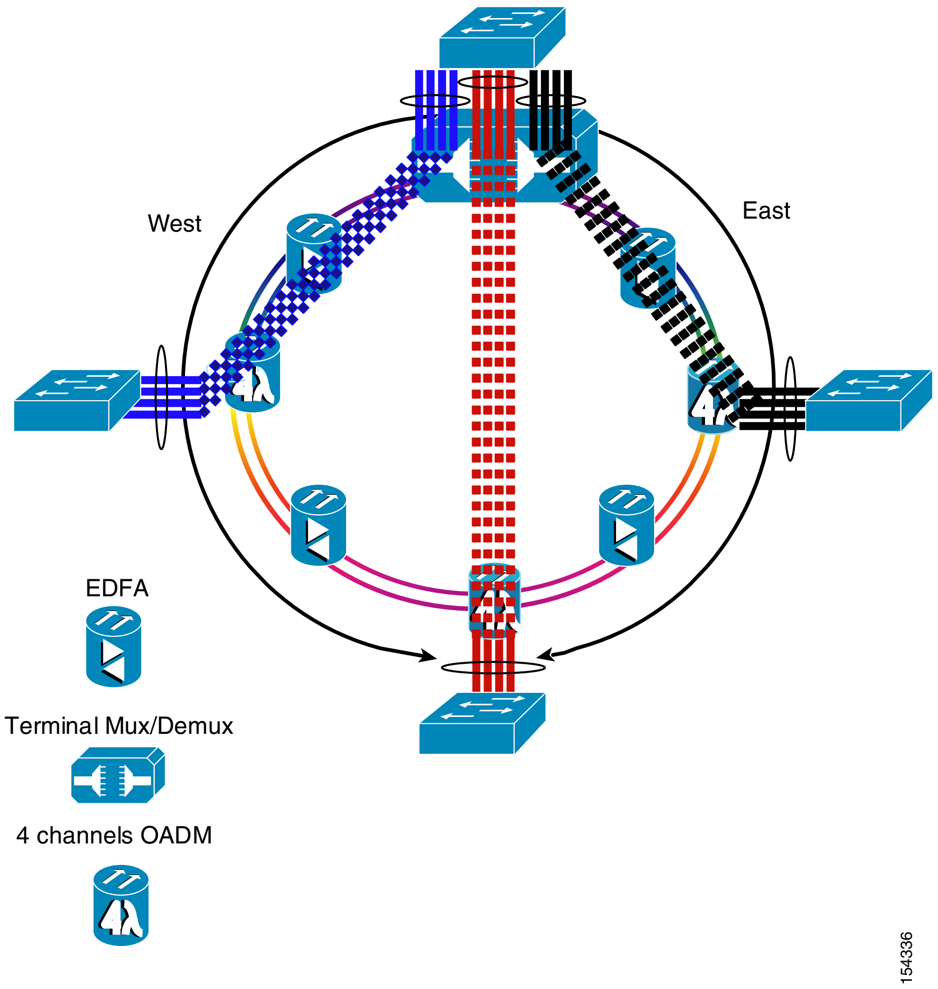
Figure 2-3 DWDM Protection Mechanisms



The basic building blocks for DWDM designs include transponders, optical multiplexers, and demultiplexers, optical add/drop multiplexers (OAMs), optical amplifiers and attenuators, variable optical attenuators (VOA), dispersion compensators (DMC/DCU), and muxponders (which are the devices that multiplex multiple client inputs onto a single channel). With these tools, it is possible to implement several designs.

Figure 2-4 shows a sample DWDM topology connecting four sites.

Figure 2-4 Sample DWDM Hub-and-Spoke Topology



From the primary site, four lambdas are carried to each site and pulled out of the ring by the OAM. In this theoretical design, a Layer 2 switch is present at each site and EtherChanneled to the switch in the primary site. The dotted lines show the logical point-to-point connections to the primary site.

Note A great tool to plan DWDM networks is Metro Planner (MP), which is a GUI-based tool to design Cisco ONS 15454 MSTP DWDM networks. MP calculates power budgets and dispersion, determines the optimized locations for amplifiers and dispersion compensation units, and produces a configuration file that can be exported to the Cisco ONS 15454. Additional details on how to build a DWDM metropolitan area network (MAN) are out of the scope of this document.

A topology such as this one is applicable to an enterprise that manages a MAN. You can configure a physical DWDM ring to provide point-to-point logical topologies, hub-and spoke logical topologies, and fully-meshed logical topologies.

Notice that in Figure 2-4, the Catalyst switches are using protected channels, so they need separate unique channels. You can use a logical ring topology where the same channel is re-used. In this case, you lose the DWDM protection and spanning tree re-routes around link failures.

If the goal of using DWDM is simply to interconnect two sites, it is possible to use a much simpler point-to-point topology with DWDM GBICs and Xenpaks to take full advantage of the available dark fiber via DWDM multiplexers.

For a list of the 32 Cisco DWDM GBICs and DWDM 10 Gigabit Ethernet Xenpaks, see the following URL: http://www.cisco.com/en/US/products/hw/modules/ps5455/products_data_sheets_list.html.

Maximum Distances and BB Credits Considerations

DWDM can provide longer distances than CWDM. Factors that affect the maximum achievable distance include power budget, chromatic dispersion, optical signal-to-noise ratio, and non-linearities.

For example, given the chromatic dispersion of the fiber and the tolerance of the transponder, the maximum achievable distance equals the tolerance of transponder divided by the coefficient of dispersion of the fiber. There are several techniques to compensate the chromatic dispersion with appropriate optical design, but they are out of the scope of this document.

Another limiting factor is the optical signal-to-noise ratio (OSNR), which degrades each time that the signal is amplified. This means that even if from a power budget point of view you can amplify the signal several times, the final OSNR might not be good enough for the receiving transponder. Because of OSNR, it makes sense to place only a maximum number of optical amplifiers in the path, after which you need full O-E-O regeneration.

Also consider that the amplification applies to all the channels, which implies that the maximum achievable length per fiber span depends also on the wavelength speed (for example, 2.5 Gbps is different from 10 Gbps), and on the total number of channels.

For example, with a single span of fiber and one single 10 Gbps channel, you can potentially achieve a maximum of 41dB loss with co-located EDFA amplification, which, on a fiber with 0.25dB/km loss, equals ~164 km. By adding amplifiers in the path, you can design for example a four-span connection with total power loss of 71dB, which equals ~284 km without regeneration.

Note As previously stated, increasing the number of channels and changing the speed of the channel changes the calculations. Many more factors need to be considered, starting with the fiber type, the chromatic dispersion characteristics of the fiber, and so on. A Cisco optical designer can assist the definition of all the required components and the best design.

Distances of thousands of kilometers can be achieved by using O-E-O regeneration with single spans of ~165 km of fiber. When EDFA or regeneration points are required, enterprises may co-locate them in the POP of the service provider from which they have purchased the fiber.

When carrying Fibre Channel on a DWDM network, the limiting factor becomes buffer-to-buffer credits (BB_credits). All data networks employ flow control to prevent data overruns in intermediate and end devices. Fibre Channel networks use BB_credits on a hop-by-hop basis with Class 3 storage traffic. Senders are permitted to send up to the negotiated number of frames (equal to the BB_credit value) to the receiver before waiting for Receiver Readys (R_RDY) to return from the receiver to replenish the BB_credits for the sender. As distance increases, so does latency, so the number of BB_credits required to maintain the flow of data increases. Fibre Channel line cards in many storage arrays have limited BB_credits, so Fibre Channel directors such as the Cisco MDS 9000, which have sufficient BB_credits, are required if extension is required beyond a few kilometers. The MDS 9000 16-port Fibre Channel line card supports up to 255 BB_credits per port, allowing DWDM metro optical fabric extension over 200 km without BB_Credit starvation and resulting performance degradation. The MDS 9000 14/2-port Multiprotocol Services Module offers Extended BB_credits, which allows distances up to 7000 km @ 1 G

FC or 3500 km @ 2 G FC. Up to 2400 BB_credits can be configured on any one port in a four-port quad with remaining ports maintaining 255 BB_credits. Up to 3500 BB_credits can be configured on any one port in a four-port quad when remaining ports shut down.

At the maximum Fibre Channel frame size of 2148 bytes, one BB_Credit is consumed every two kilometers at 1 Gbps and one BB_Credit per kilometer at 2 Gbps. Given an average Fibre Channel frame size for replication traffic between 1600–1900 bytes, a general guide for allocating BB_credits to interfaces is as follows:

- 1.25 BB_credits for every 2 km at 1 Gbps
- 1.25 BB_credits for every 1 km at 2 Gbps

In addition, the 2.5 Gb and 10 Gb datamux cards on the Cisco ONS 15454 provide buffer-to-buffer credit spoofing, allowing for distance extension up to 1600 km for 1Gb/s Fibre Channel and 800 km for 2 Gbps Fibre Channel.

CWDM versus DWDM

CWDM offers a simple solution to carry up to eight channels (1 Gbps or 2 Gbps) on the same fiber. These channels can carry Ethernet or Fibre Channel. CWDM does not offer protected lambdas, but client protection allows re-routing of the traffic on the functioning links when a failure occurs. CWDM lambdas can be added and dropped, thus allowing the creation of hub-and-spoke, ring, and meshed topologies. The maximum achievable distance is ~100 km with a point-to-point physical topology and 40 km with a ring physical topology.

DWDM offers more channels than CWDM (32), more protection mechanisms (splitter protection and Y-cable protection), and the possibility to amplify the channels to reach greater distances. Each channel can operate at up to 10 Gbps.

In addition to these considerations, a transponder-based DWDM solution such as a solution based on ONS-15454 offers better management for troubleshooting purposes than a CWDM or DWDM solution simply based on muxes.

The two main DWDM deployment types are as follows:

- DWDM GBICs and Xenpaks (IPoDWDM) connected to a MUX (for example, Cisco ONS 15216 products)—Enterprises can take advantage of the increased bandwidth of DWDM by connecting DWDM 10 GigE Xenpaks directly to a passive MUX. This approach offers increased bandwidth over CWDM, potentially greater distances (~160 km in a single fiber span with EDFA amplifiers co-located at the data center premises and more), and several hundred kilometers with amplification and multiple fiber spans. The main disadvantage of a DWDM GBCI/MUX-based solution as compared with a DWDM transponder solution or a CWDM solution, is the fact that there is currently no DWDM transceiver for Fibre Channel, which limits the deployment to IP over DWDM.
- Regular Ethernet GBICs and Fibre Channel connected to a transponder (for example, ONS 15454 products)–Enterprises can take advantage of this type of solution to build a MAN, and can use transponders to build a transparent multiservice (including Ethernet and Fibre Channel) transport infrastructure with virtually no distance limits. A transponder-based solution has virtually no distance limitations, and allows building a Layer 2 transport (with O-E-O regeneration) across sites at thousands of kilometers of distance. As previously stated, when buying dark fiber from an SP, an enterprise may be able to co-locate their EDFA or O-E-O gear at the SP POP.

Table 2-1 compares the various solutions.

Table 2-1 Solution Comparison

	CWDM	DWDM GBIC/Xenpak	DWDM Transponder
Number of channels	8	32	32
Available speeds	2 Gbps	10 Gbps	10 Gbps
Protection	Client	Client, splitter, Y-cable	Client, splitter, Y-cable
Ethernet support	Yes	Yes	Yes
Fibre Channel support	Yes	No	Yes
Amplifiable	No	Yes	Yes
Buffer-to-buffer options	255 BB_credits from the MDS	Up to 3500 BB_credits on the MDS 14+2 cards for distances up to 7000 km	Extended BB_credits on MDS 14+2 for distances up to 7000 km or BB spoofing on the ONS 15454 for distances up to ~1600 km
Distance	~100 km	~160 km in a single span, virtually unlimited with regeneration stations	Virtually unlimited with regeneration stations
Management	None	Some management	Rich management and troubleshooting tools

Fiber Choice

When deploying an optical solution, it is important to identify the existing fiber type, or to choose the fiber type that needs to be deployed. Before using the fiber, verify the fiber conditions by using the optical time domain reflectometer (OTDR). It is also important to test the fiber for polarization mode dispersion (PMD). For example, standard fiber at 1550 nm has a dispersion of 17ps/nm/km, which may be inadequate, depending on the desired distance. For example, if the dispersion tolerance is 1800 ps/nm for the path between a transmitter and receiver and the fiber dispersion is 18 ps/nm-km, the maximum distance is 100 km between end nodes (1800/18). The dispersion tolerance or limitation is inversely proportional to the data rate, and is typically not an issue at speeds below OC-192 (10 Gbps).

For SMF, you typically have to choose between the following types:

- ITU-T G.652 (standard SMF, also known as SMF28)—Optimized for 1310 nm (SONET). Works with CWDM and DWDM. There is an attenuation peak at 1383 nm (0.50dB/km).
- ITU-T G.652.C (zero water peak fiber)—Optimized for CWDM. Most systems support CWDM in the 1470–1610 range. From a chromatic dispersion point of view, this is just like a G.652. This is also referred to as extended band, because it eliminates the water peak. Also works with DWDM.
- ITU-T G.655 (non-zero dispersion shifted fiber)—Best for DWDM. There is a little dispersion at 1550 nm, and 4ps/nm/km in the 1530–1570 nm range. This type of fiber addresses non-linearity in DWDM, and more specifically four-wave mixing (FWM). Works with CWDM, and for TDM at 1310 nm and TDM at 1550 nm.
- ITU-T G.653 (dispersion-shifted fiber [DSF])—Changes the chromatic and waveguide dispersion to cancel at 1550 nm. Works with CWDM, good for TDM at 1550 nm. Not good for DWDM.

For more information on various fibers, see the following URLs:

- Lucent/OFS fibers—http://www.ofsoptics.com/product_info/ofs-fitel.shtml
- Corning fibers— http://www.corning.com/opticalfiber/
- Alcatel fibers—http://www.alcatel.com/opticalfiber/index.htm
- Pirelli fibers—
http://www.pirelli.com/en_42/cables_systems/telecom/product_solutions/optical_fibres.jhtml

SONET/SDH

Although DWDM allows Layer 2 connectivity at continental distances, it is more typical for enterprises to connect sites at continental distances via SONET offerings from service providers (SPs). SONET can transport several interface types, such as T1, DS3, N x STS-1, and so on. The Cisco ONS15454 platform offers SONET/SDH client connection options, in addition to Gigabit Ethernet and Fibre Channel (FC).

SONET/SDH-based architectures can support both sub-rate (less than 1 Gbps) and line rate (1 Gbps or 2 Gbps) services. SPs have already installed large SONET/SDH rings and can leverage this existing architecture to provide storage, Ethernet, and data center connectivity. Ethernet support includes 10/100/1000 Mbps interfaces. SAN interfaces include Fibre Channel, FICON, and ESCON.

The following line cards are deployed to support Ethernet and storage support on SONET/SDH networks:

- E-Series Ethernet card
- G-Series gigabit Ethernet card (G1000-4/G1K-4)
- ML-Series Ethernet cards (Ethernet /FCIP)
- SL-Series FC card

From the standpoint of the Fibre Channel switch, the connection looks like any other optical link. However, it differs from other optical solutions in that it *spoofs* R_RDY frames to extend the distance capabilities of the Fibre Channel transport. All Fibre Channel over optical links use BB_credits to control the flow of data between switches. R_RDY frames control the BB_credits. As the distance increases, so must the number of BB_credits. The spoofing capability of the SL line card extends this distance capability to 2800 km at 1 Gbps.

The Cisco ONS 15454 offers a number of SONET/SDH protection options, in addition to client-level protection through Fibre Channel switches. Port channels in conjunction with VSAN trunking are recommended where multiple links are used.

Just as with other optical solutions, FC over SONET/SDH is suitable for synchronous replication deployments subject to application performance constraints. Latency through the FC over SONET/SDH network is only negligibly higher than other optical networks of the same distance because each frame is serialized in and out of the FC over SONET/SDH network. The latency is 10µs per maximum-sized FC frame at 2 Gbps.

SONET/SDH Basics

The details of SONET and its history are out of the scope of this document, but it is important to consider some of the key principles behind this technology. SONET offers the following hierarchy of transport rates:

- STS-1 (51.84 Mbps)
- STS-3 (155.52 Mbps)

- STS-12 (622.08 Mbps)
- STS-24 (1244.16 Mbps)
- STS-48 (2488.32 Mbps)
- STS-192 (9.953.28 Mbps)

Note This guide refers to STS-1 and OC-1 interchangeably, and similarly for STS-3 and OC-3, and so on.

STS-3 is made of 3 STS-1s. One STS-1 can be transporting Fibre Channel, another can be transporting voice, and so on, or these three channels can be “bundled”; that is, *concatenated* in an OC-3c.

SONET/SDH has been designed to facilitate multiplexing and demultiplexing of multiple low-rate traffic. It has also been designed to carry legacy traffic such as DS1 (often referred to as T1), DS3, ATM, and so forth. This is done by subdividing an STS-1 into multiple virtual tributary groups (VTG). VTGs can transport VT1.5 (to carry DS1), VT2 (to carry E1 frames), and so on.

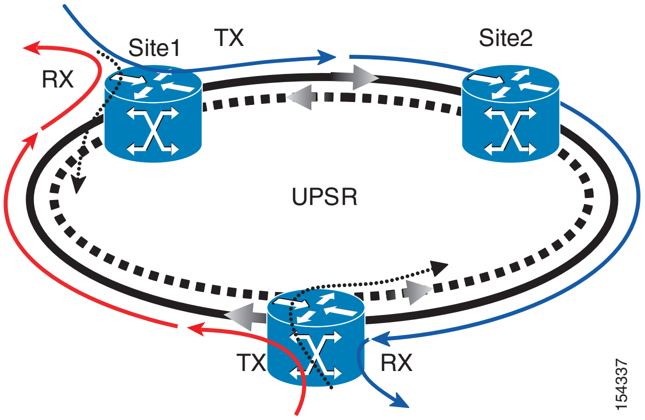
Several components comprise a SONET network, and for the purpose of this document, it is enough to focus simply on the use of add-drop-multiplexers; these are the device that insert them or remove DSs and STSs from an OC-N network.

SONET UPSR and BLSR

The two most fundamentals topologies used in SONET are the unidirectional path-switched ring (UPSR) and the bidirectional line-switched ring (BLSR).

With UPSR, the traffic between two nodes travels on two different paths. For example, in Figure 2-5, the traffic on the outer fiber travels clockwise, and the protection path (inner ring) travels counter-clockwise. As a result, Site1-to-Site3 traffic takes a different path than Site3-to-Site1 traffic.

Figure 2-5 SONET UPSR Topology



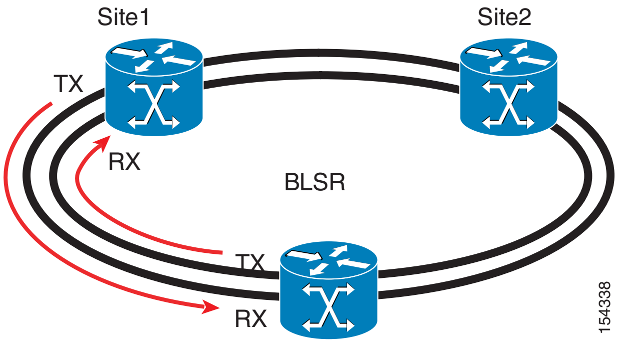
Note that Site1 TX sends traffic on both the outer and the inner ring. Site3 selects only the traffic coming from the outer ring. In case a failure occurs, the receiving end selects the traffic from the protected path. This topology works well for short distances, because the TX-RX paths can yield very different latency values, which can affect flow control.

From a bandwidth utilization perspective, the communication between two nodes always involves the full ring. For example, if Site1-to-Site2 is using one STS-1 and the ring is an OC-3 ring, there are only 2 STS-1s left for the other nodes to use.

From a logical topology point of view, USPR rings are more suited for hub-and-spoke topologies.

Figure 2-6 shows the BLSR topology.

Figure 2-6 SONET BLSR Topology



In this case, the communication between Site1 and Site3 does not involve the full ring. This means that if Site1-to-Site3 is using 6 STS-1s on an OC-12 ring, Site1-to-Site2 can still use 6 STS-1s and Site2-to-Site3 can also use 6 STS-1s.

Note that only half of the available channels can be used for protection reasons. In other words, if one link between Site1-and-Site3 fails, Site1 and Site3 need to be able to communicate over the path that goes through Site2. Six STS-1s need to be available along the alternate path.

BLSR offer several advantages: TX and RX between two sites travel on the same path, and bandwidth utilization is more optimized.

Ethernet Over SONET

Several options are available to transport Ethernet Over SONET. With the ONS products, the following three families of line cards are often used for this purpose:

- *E-Series* cards include the E100T-12/E100T-G and the E1000-2/E1000-2. An E-Series card operates in one of three modes: multi-card EtherSwitch group, single-card EtherSwitch, or port-mapped. E-Series cards in multicard EtherSwitch group or single-card EtherSwitch mode support Layer 2 features, including virtual local area networks (VLANs), IEEE 802.1Q, STP, and IEEE 802.1D. Port-mapped mode configures the E-Series to operate as a straight mapper card and does not support these Layer 2 features. Within a node containing multiple E-Series cards, each E-Series card can operate in any of the three separate modes.
- *G-Series* cards on the Cisco ONS 15454 and ONS 15454 SDH map up to four Gigabit Ethernet ports onto a SONET/SDH transport network and provide scalable and provisionable transport bandwidth at signal levels up to STS-48c/VC4-16 per card. The G-Series cards provide line rate forwarding for all Ethernet frames (unicast, multicast, and broadcast) and can be configured to support Jumbo frames (defined as a maximum of 10,000 bytes). The card maps a single Ethernet port to a single STS circuit. You can independently map the four ports on a G-Series card to any combination of STS-1, STS-3c, STS-6c, STS-9c, STS-12c, STS-24c, and STS-48c circuit sizes, provided that the sum of the circuit sizes that terminate on a card do not exceed STS-48c.

- *ML-Series* cards are independent Gigabit Ethernet (ML1000-2) or Fast Ethernet (ML100T-12 and ML100X-8) Layer 3 switches that process up to 5.7 Mpps. The cards are integrated into the ONS 15454 SONET or the ONS 15454 SDH. The ML-Series card uses Cisco IOS Release 12.2(28) SV, and the Cisco IOS command-line interface (CLI) is the primary user interface for the ML-Series card. The ML100T-12 features twelve RJ-45 interfaces, and the ML100X-8 and ML1000-2 feature two SFP slots to support short wavelength (SX) and long wavelength (LX) optical modules. All three cards use the same hardware and software base and offer similar feature sets. The ML-Series card features two virtual packet-over-SONET/SDH (POS) ports, which function in a manner similar to OC-N card ports.

  The ML-Series cards support the following:

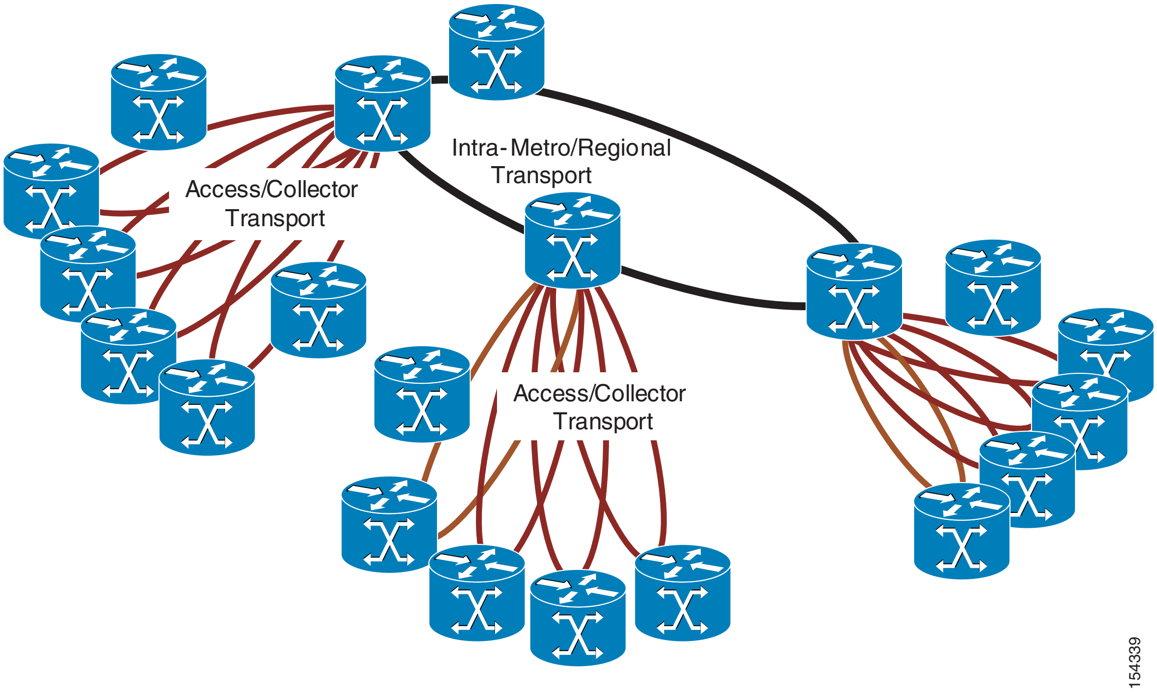
  - Transparent bridging
  - MAC address learning
  - Aging and switching by hardware
  - Multiple Spanning-Tree (MST) protocol tunneling
  - 255 bridge groups maximum
  - IEEE 802.1q VLAN tunneling
  - IEEE 802.1d and IEEE 802.1w Rapid Spanning-Tree Protocol (RSTP)
  - Resilient packet ring (RPR)
  - Ethernet over Multiprotocol Label Switching (EoMPLS)
  - EtherChanneling
  - Layer 3 unicast and multicast forwarding
  - Access control lists
  - Equal-cost multipath (ECMP)
  - VPN Routing and Forwarding (VRF)-lite
  - EIGRP, OSPF, IS-IS, PIM, BGP, QoS and more

Service Provider Topologies and Enterprise Connectivity

Most service providers deploy SONET rings where UPSRs are used at the edge of the network, while the backbone may be provided by BLSR rings. From an enterprise point of view, the end-to-end connectivity between sites over an SP SONET offering can consist of several rings of different types.

Figure 2-7 shows a typical SP topology with multiple SONET rings.

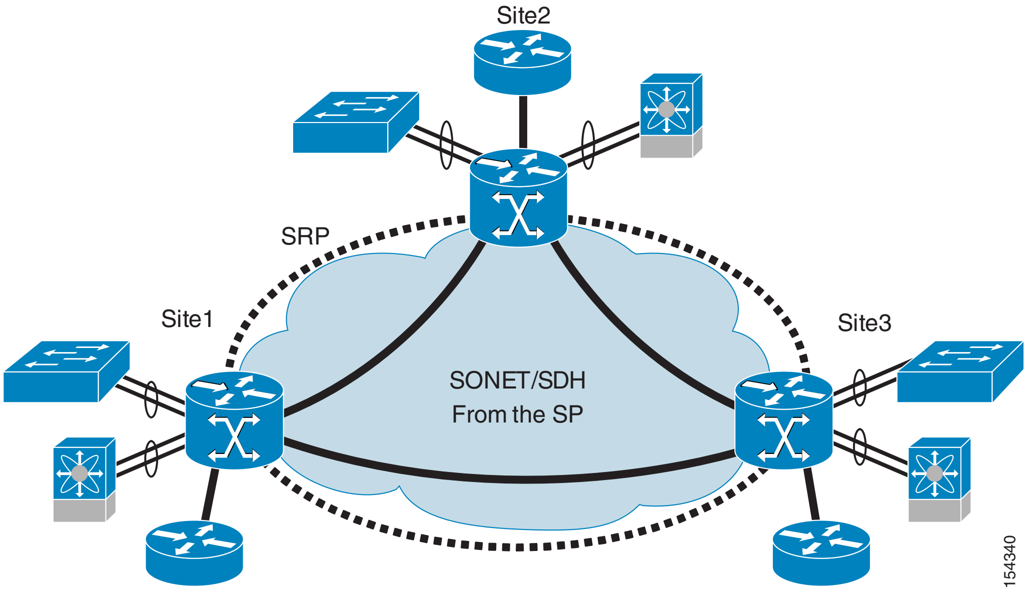
Figure 2-7 Typical SP Topology with Multiple SONET Rings



An enterprise with SONET connectivity between sites and with the need to extend an Ethernet segment across sites as well as the Fibre Channel network can consider using Ethernet over SONET and Fibre Channel over SONET by means of the ML-series, G-series, and SL-series cards connected to an ONS-15454 device.

Figure 2-8 shows the use of SONET to bridge Ethernet and Fibre Channel.

Figure 2-8 Use of SONET to Bridge Ethernet and Fibre Channel



Resilient Packet Ring/Dynamic Packet Transport

The Resilient Packet Ring (RPR)/Dynamic Packet Transport (DPT) protocol overcomes the limitations of SONET/SDH and spanning tree in packet-based networks. RPR/DPT convergence times are comparable to SONET/SDH and faster than 802.1w. RPR/DPT uses two counter-rotating rings where fibers are concurrently used to transport both data and control traffic. DPT rings run on a variety of transport technology including SONET/SDH, wavelength division multiplexing (WDM), and dark fiber. IEEE 802.17 is the working group that defines the standard for RPRs. The first major technical proposal for an RPR protocol was submitted by Cisco based on the DPT protocol.

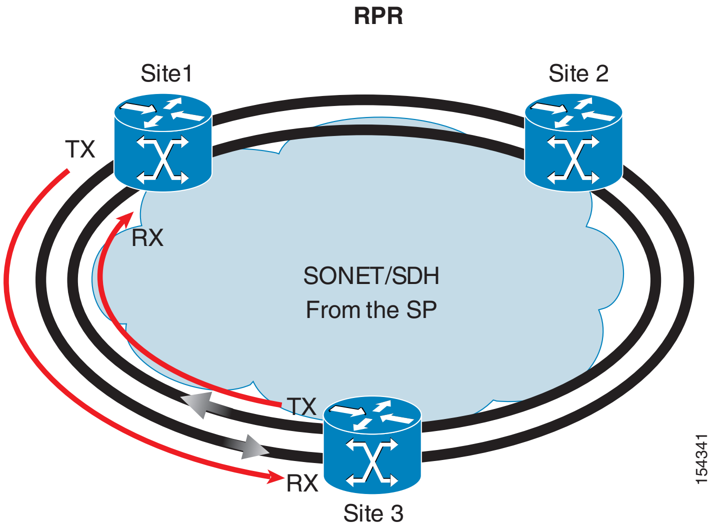
RPR/DPT provides sophisticated protection switching for self-healing via the intelligent protection switching (IPS) algorithm. IPS enables sub-50 ms protection for rapid IP service restoration. RPR/DPT rings use automatic procedures for address assignment and resolution, ring topology and status discovery, and control message propagation, which optimizes ring traffic routing and management procedures.

Spatial Reuse Protocol

Spatial Reuse Protocol (SRP) was developed by Cisco for ring-based media, and is the underlying media-independent protocol used in the DPT products. SRP uses two counter-rotating rings (inner and outer ring), and performs topology discovery, protection switching (IPS), and bandwidth control. Spatial reuse is among the key characteristics of SRP.

Figure 2-9 shows the communication between Site1and Site3.

Figure 2-9 Use of SONET to Build an RPR Ring



This communication in traditional ring technologies involves the full ring. With SRP, the bandwidth utilization is more efficient, because the destination strips off the frame from the ring (only multicast frames are stripped from the source). By using this mechanism, DPT rings provide packet-by-packet spatial reuse wherein multiple segments can concurrently exchange traffic at full ring bandwidth without interference.

Another important aspect of the RPR operation is how the ring is selected. Site1 sends out an Address Resolution Protocol (ARP) request to a ring that is chosen based on a hash. Site3 responds to the ARP by looking at the topology and choosing the ring with the shortest path. Site1 then uses the opposite ring to communicate with Site3. This ensures that the communication path is the shortest.

The SRP Fairness Algorithm (SRP-fa) ensures that both global fairness and local bandwidth optimization are delivered on all segments of the ring.

RPR and Ethernet Bridging with ML-series Cards on a SONET Network

RPR/DPT operates at the Layer 2 level and operates on top of protected or unprotected SONET/SDH. It is well-suited for transporting Ethernet over a SONET/SDH ring topology and enables multiple ML-Series cards to become one functional network segment or shared packet ring (SPR). Although the IEEE 802.17 draft was used as reference for the Cisco ML-Series RPR implementation, the current ML-Series card RPR protocol does not comply with all clauses of IEEE 802.17 because it supports enhancements for Ethernet bridging on an RPR ring.

The ML-Series cards in an SPR must connect directly or indirectly through point-to-point STS/STM circuits. The point-to-point STS/STM circuits are configured on the ONS node and are transported over the SONET/SDH topology of the ONS node with either protected or unprotected circuits. On circuits unprotected by the SONET/SDH mechanism, RPR provides resiliency without using the capacity of the redundant protection path that a SONET/SDH-protected circuit requires. This frees this capacity for additional traffic. RPR also utilizes the bandwidth of the entire ring and does not block segments as does spanning tree.

Differently from IEEE 802.17, the ML-Series cards perform destination stripping both for routed and bridged traffic. IEEE 802.17 performs destination stripping only for routed traffic; bridged frames are flooded on the ring. The Cisco DPT implementation has a local MAC address table on each node; if the traffic matches a MAC address that is local to the line card, it is not sent out on the ring, thus preserving the ring bandwidth. The size of the MAC address table is optimized because RPR transit traffic is not learned by the Layer 2 forwarding table. The Layer 2 forwarding table is a CAM table for wire rate Layer 2 forwarding.

Metro Offerings

Customers who need to connect data centers can choose from several service provider offerings, some of which have been already described in this guide: dark fiber (which can in turn be used in conjunction with CWDM or DWDM) and SONET/SDH point-to-point connectivity.

In addition to these offerings, enterprises can also buy Metro Ethernet connectivity. Ethernet is attractive because it allows for rate limiting in increments not provided by time division multiplexing (TDM) service providers. For example, enterprise customers can purchase a 10 Mbps service (or less) instead of committing to a DS3 (44.736 Mbps) connection. With this solution, enterprise customers looking for a transparent LAN service can obtain connections starting at 0.5–1000 Mbps. Metro Ethernet supports both SAN and LAN traffic. Typical applications include Ethernet connectivity (server-to-server and client-server communications) and asynchronous/synchronous storage applications. Metro Ethernet can be used to transport FC over IP (FCIP) along with the traditional server/client communication.

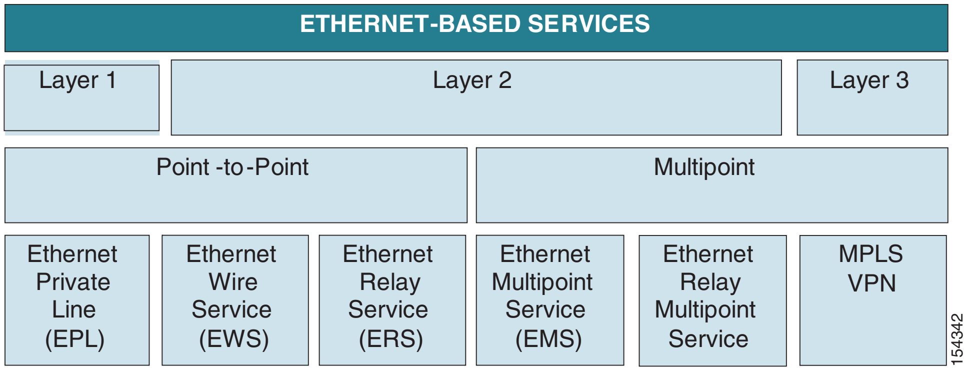
These are typically available in either of the following formats:

- Ethernet Relay Service (ERS) or Ethernet Virtual Circuit Service (EVCS)—Provides a point-to-point Ethernet circuit between customer premises equipment (CPEs) over the metro network. Multiple ERSes can be mapped from a single CPE over the user-to-network interface (UNI) of the SP. Each circuit is associated with and mapped to a single SP VLAN. In this way, ERS emulates the traditional Frame Relay service in which the VLAN is analogous to the data-link connection identifier (DLCI). This type of transport does not carry Bridge Protocol Data Units (BPDUs), Cisco Discovery Protocol (CDP), VLAN Trunk Protocol (VTP), and so on.

- Ethernet Wire Service (EWS)—Emulates a point-to-point virtual wire connection, and appears to the customer as a "clear channel" pipe. This is sometimes referred to as an Ethernet private line. A customer using EWS does not see the SP network, and the connection appears as if it were a local Ethernet segment. All data passes transparently over the connection and the following Layer 2 (L2) control protocols STP, CDP, and VTP. Data transparency means that the data is transported intact with the VLAN ID untouched.
- Ethernet Multipoint Service (EMS), also known as Transparent LAN Service (TLS)—Multipoint-to-multipoint virtual wire connection. EMS is the multipoint extension of the EWS, and has the same service characteristics, such as data transparency and L2 control protocol tunneling. EMS is analogous to a multipoint Ethernet private line service.

Figure 2-10 shows the relation between Metro Ethernet services, their network layer, and point-to-point versus point-to-multipoint classification.

Figure 2-10 Metro Ethernet Services


ETHERNET-BASED SERVICES
Layer 1
Layer 2
Layer 3
Point -to-Point
Multipoint
Ethernet Private Line (EPL)
Ethernet Wire Service (EWS)
Ethernet Relay Service (ERS)
Ethernet Multipoint Service (EMS)
Ethernet Relay Multipoint Service
MPLS VPN
154342

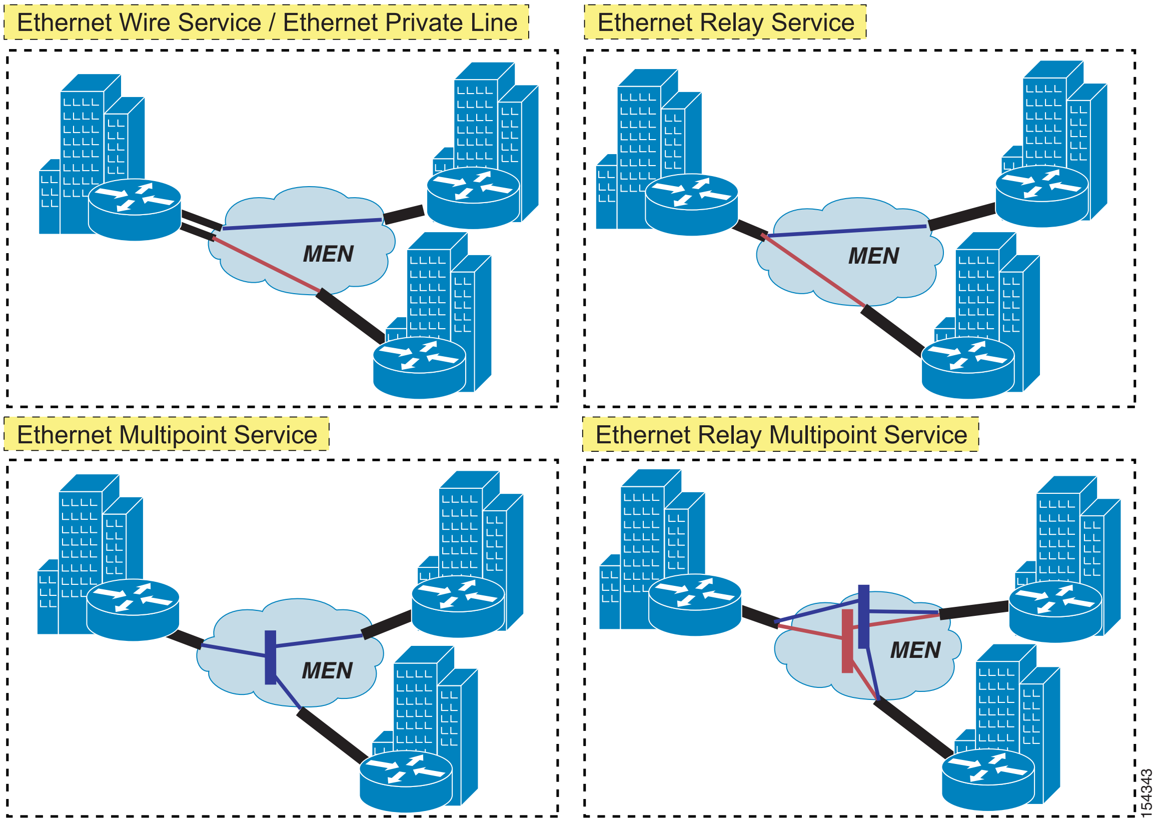
MEF identifies the following Ethernet services:

- Ethernet Line Service Type (E-Line)—Point-to-point Ethernet service; that is, a single point-to-point Ethernet circuit provisioned between two UNIs.
- Ethernet LAN Service Type (E-LAN)—Multipoint-to-multipoint Ethernet service; that is, a single multipoint-to-multipoint Ethernet circuit provisioned between two or more UNIs.
- Ethernet Private Line (EPL)—Port-based point-to-point E-Line service that maps Layer 2 traffic directly onto a TDM circuit.
- Ethernet Wire Service (EWS)—Point-to-point port-based E-Line service that is used primarily to connect geographically remote LANs over an SP network.
- Ethernet Relay Service (ERS)—Point-to-point VLAN-based E-Line service that is used primarily for establishing a point-to-point connection between customer routers.
- Ethernet Multipoint Service (EMS)—Multipoint-to-multipoint port-based E-LAN service that is used for transparent LAN applications.
- Ethernet Relay Multipoint Service (ERMS)—Multipoint-to-multipoint VLAN-based E-LAN service that is used primarily for establishing a multipoint-to-multipoint connection between customer routers.

- ERS Access to MPLS VPN—Mapping of an Ethernet connection directly onto an MPLS VPN. It provides Layer 2 access using an ERS UNI, but is a Layer 3 service because it traverses the MPLS VPN.
- ERS Access to ATM Service Interworking (SIW)—Point-to-point VLAN-based E-Line service that is used for Ethernet-to-ATM interworking applications.

Figure 2-11 shows a variety of metro Ethernet services.

Figure 2-11 Metro Ethernet Services


Ethernet Wire Service / Ethernet Private Line
MEN
Ethernet Relay Service
MEN
Ethernet Multipoint Service
MEN
Ethernet Relay Multipoint Service
MEN
154343