Data Center Interconnect: Layer 2 Extension Between Remote Data Centers

White Paper

Contents

What You Will Learn **3**
DCI Considerations **3**
DCI Transport Options **4**
Business Reasons for Implementing LAN Extension Between Multiple Data Centers **5**
Business Continuance: High-Availability Clusters 5
Workload Mobility 6
LAN Extension Solution Requirements **7**
End-to-End Loop Prevention 7
Spanning Tree Protocol Isolation to Control Redundant Layer 2 Topology 7
DCI Reference Solutions and Platforms **8**
Transport Option 1: Dark Fiber 10
Virtual Switching System 10
Virtual PortChannel 11
Fiber-Based Layer 2 VPNs with VSS and vPC 12
Transport Option 2: MPLS 17
EoMPLS 18
VPLS 18
A-VPLS 19
EoMPLS, VPLS, and A-VPLS Platform Support and Positioning 20
Transport Option 3: IP 21
EoMPLSoGRE, VPLSoGRE, and A-VPLS 21
EoMPLSoGRE and VPLSoGRE Platform Support and Positioning 23
EoMPLS and VPLS Layer 2 Loop Prevention 24
Cisco IOS EEM Semaphoring in N-PE 25
VPLS and A-VPLS Loop Prevention 29
Conclusion **29**
For More Information **30**

What You Will Learn

This document is intended to help network managers and systems managers understand the various solutions and recommendations that Cisco offers to geographically extend Layer 2 networks over multiple distant data centers. These offerings address the requirements of high performance and fast convergence.

This document also introduces the capabilities of the Cisco® Data Center Interconnect (DCI) solution. Please note that it does not address the requirements for SANs.

Initially, this document describes the types of connectivity (Layer 2, Layer 3, and storage) that can be established between remote data center locations, together with the services that a given enterprise can obtain from a service provider.

The discussion then focuses on LAN extension designs, analyzing some of the most relevant business factors requiring the deployment of a LAN extension, and providing a list of solution requirements.

The document concludes by describing some technical alternatives to provide LAN extension functions, including:

- Point-to-point or point-to-multipoint interconnection, using virtual switching system (VSS), virtual PortChannel (vPC), and optical technologies
- Point-to-point interconnection using Ethernet over Multiprotocol Label Switching (EoMPLS) natively (over an MPLS core) and over a Layer 3 IP core
- Point-to-multipoint interconnections, using virtual private LAN services (VPLS) or advanced VPLS (A-VPLS) natively (over an MPLS core) or over a Layer 3 IP core

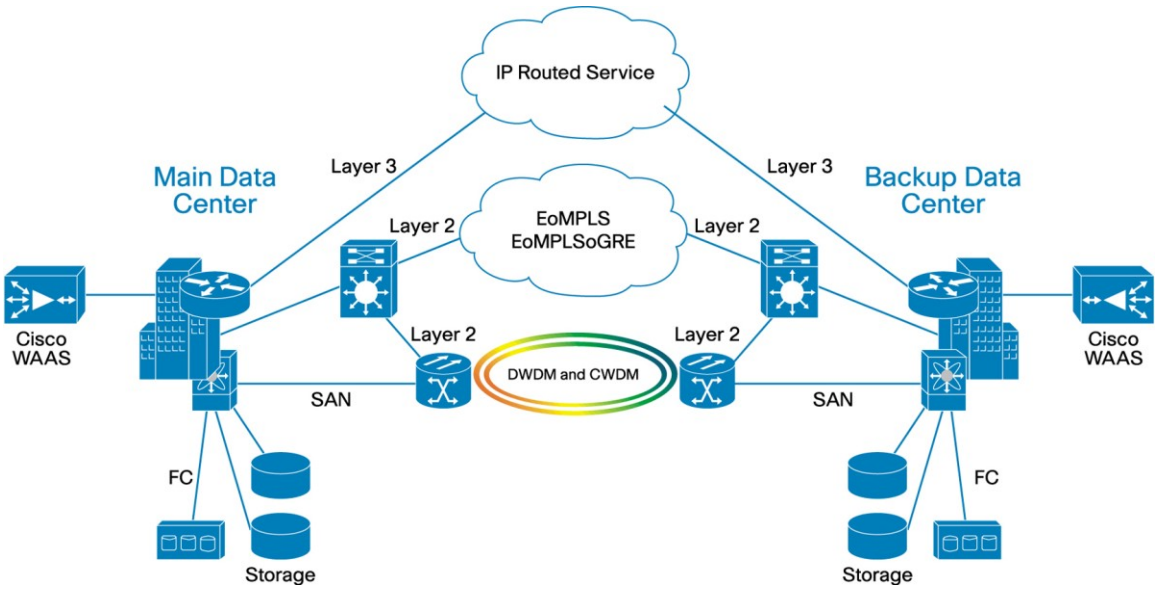
For each technical alternative, specific platform support and positioning information is provided.

DCI Considerations

Figure 1 shows the main considerations when deploying a DCI solution:

- Layer 3 interconnect (typically over an existing enterprise IP core)
- Layer 2 interconnect
- SAN interconnect

Figure 1. DCI Main Considerations



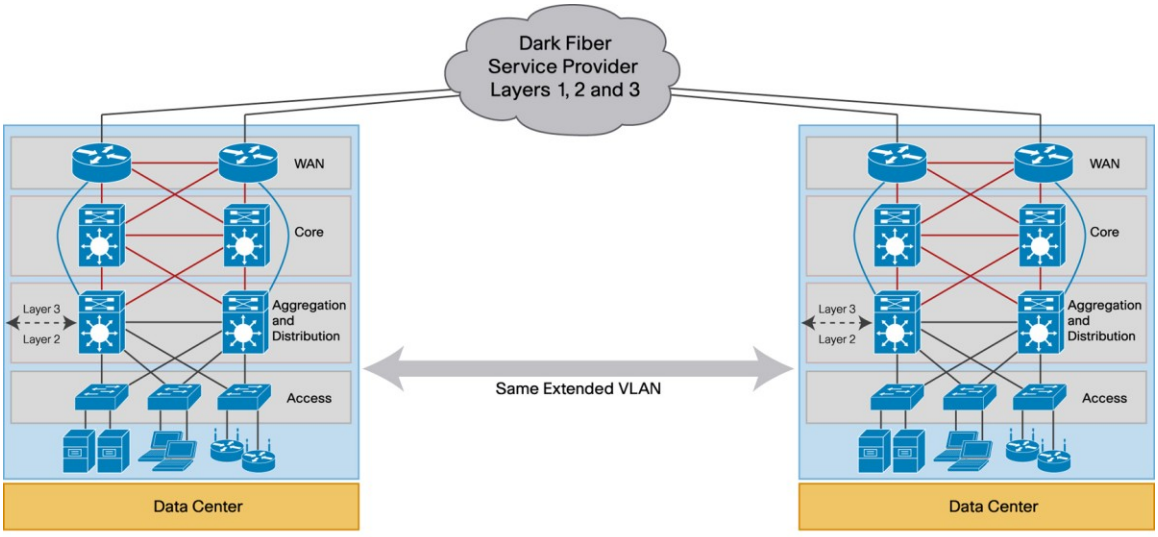
DCI Transport Options

DCI supports two possible transport scenarios, one where the enterprise owns the core infrastructure interconnecting the various data centers, and one where a provider offers the connectivity services.

In addition, the following types of service are possible (Figure 2):

- **Dark Fiber:** This can be considered a Layer 1 type of service. It is popular among many customers today, as it allows the transport of various types of traffic, including SAN traffic. It tends to be expensive, especially as the number of sites increases. Dark fiber offerings are also limited in the distance they can span.
- **Layer 2 Services:** In this case, the LAN extension can be achieved by directly using the provider services. The enterprise sends native Ethernet frames to the provider, which will be delivered to the remote sites. Alternatively, the enterprise can overlay a Layer 2 VPN solution on the service provider service, offering additional operational flexibility.
- **Layer 3 Services:** Service providers deliver these services based on IP or MPLS. In both scenarios, the enterprise must deploy an overlay technology to perform the LAN extension between the various sites. The enterprise's choice of overlay solutions tends to be limited to those based on IP. However, in extremely rare instances, the service provider may be willing to transport and relay MPLS labels on behalf of the enterprise.

Figure 2. DCI LAN Extension Encapsulation Options



WAN Transport	Description	LAN Extension Encapsulation Options
Dark Fiber and Service Provider Layer 1	Customer Owned or Service Provider Leased	Native Ethernet, IP and MPLS
Service Provider Layer 2	Service Provider Layer 2* Service, Ethernet Private Line (EPL), Ethernet Virtual Private Line (EVPL)*	Native Ethernet, IP and MPLS
Service Provider Layer 3	IP Leased-Line Service	IP and MPLS

*HQOS Required for Subrate Layer 2 or EVPL Service.

Figure 2 shows how the available encapsulation options vary with the WAN transport alternatives. Dark Fiber and Layer 2 transport scenarios support native Ethernet, IP, and MPLS encapsulations. For Layer 3 type service, IP encapsulations are mainly used.

The type of service available between the enterprise sites usually dictates the type of LAN extension solution that can be deployed. This will be explained in greater detail in this document.

Business Reasons for Implementing LAN Extension Between Multiple Data Centers

Cisco recommends isolating and reducing Layer 2 networks to their smallest scope, usually limiting them to the access layer. Layer 2 connectivity is required for server-to-server communication, high-availability clusters, networking, and security.

However, in some situations, Layer 2 must be extended beyond the single data center. Specifically, this happens when the framework or scenario developed for a campus has been extended beyond its original geographic area, and is now spread over multiple data centers and across long distances. Such scenarios are becoming more prevalent, as high-speed service provider connectivity becomes more available and cost-effective.

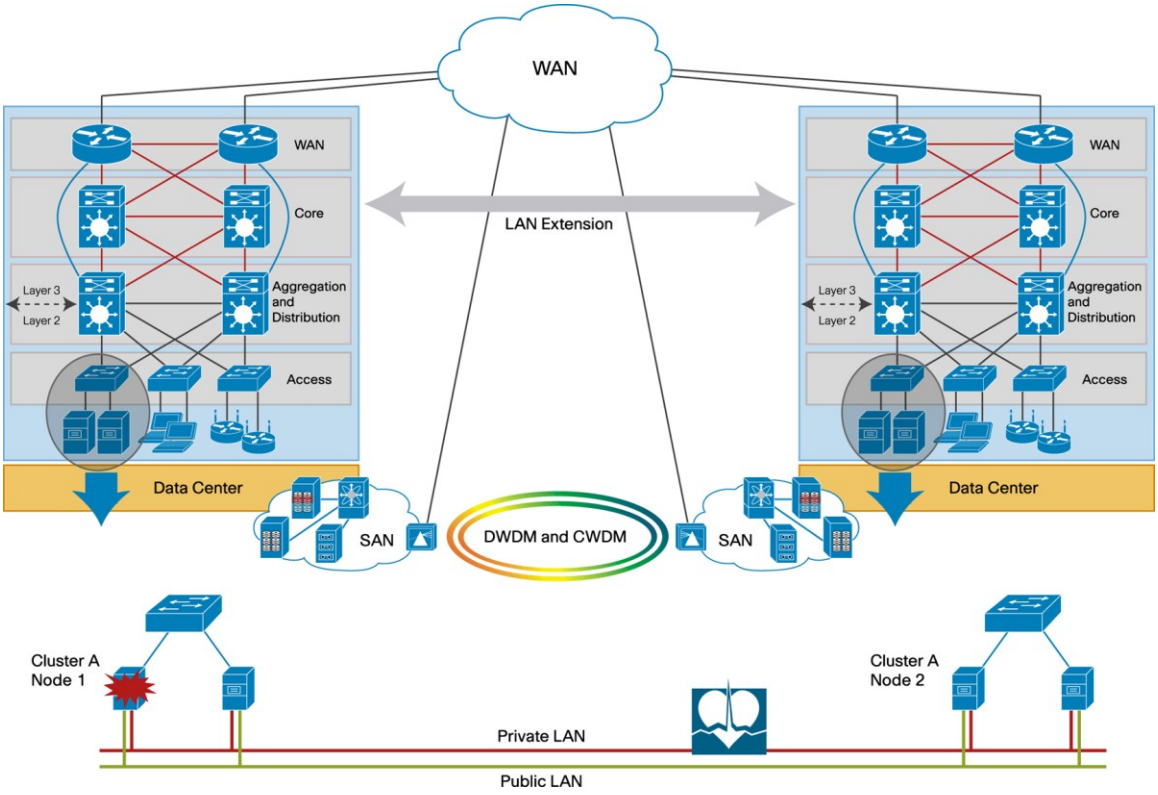
High-availability clusters, server migration, and application mobility are important use cases that require Layer 2 extension.

Business Continuance: High-Availability Clusters

Some network communication between members of high-availability clusters[1] requires some clusters to be Layer 2 (Figure 3):

- Private interprocess communication (such as heartbeat and database replication) used to maintain and control the state of the active node
- Public communication (virtual IP of the cluster)

Figure 3. DCI LAN Extension for High-Availability Clusters



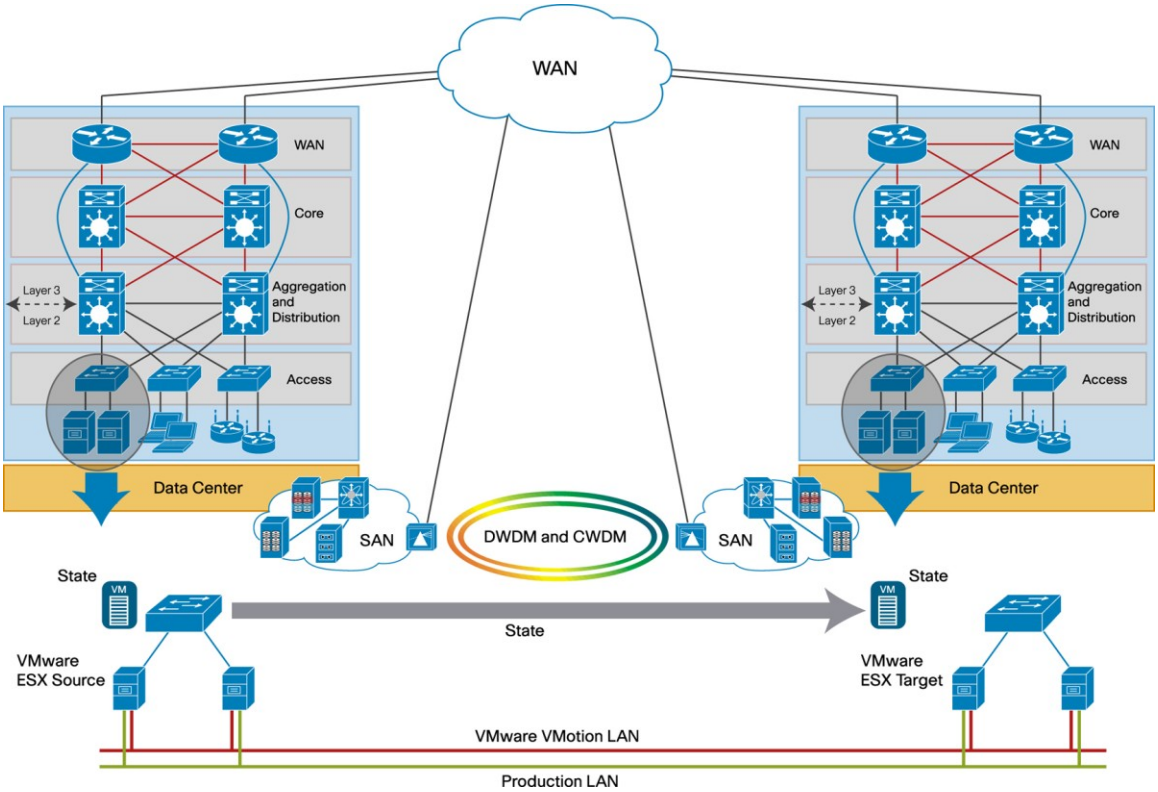
[1] Cluster vendors (Microsoft, Veritas, Sun, etc.) offer geographical high-availability clusters based on Layer 3 interconnection.

Workload Mobility

During the process of migrating physical servers from one data center to a remote site (Figure 4), note the following:

- IP renumbering of servers to be moved is complex and costly. Avoiding IP address renumbering makes physical migration projects easier and reduces cost substantially.
- Some applications may be difficult to readdress at Layer 3 (for example, mainframe applications). In this case, it is easier to extend the Layer 2 VLAN outside the access layer, keeping the original configuration of the systems after the move.
- During phased migration, when only part of the server farm is moving at any given time, Layer 2 adjacency is often required across the whole server farm for business continuity purposes.
- Some applications[2] that offer virtualization of operating systems allow the move of virtual machines between physical servers separated by long distances. To synchronize the software modules of virtual machines during a software move, and keep active sessions running, the same extended VLANs between the physical servers must be maintained.

Figure 4. DCI LAN Extension for VMware VMotion



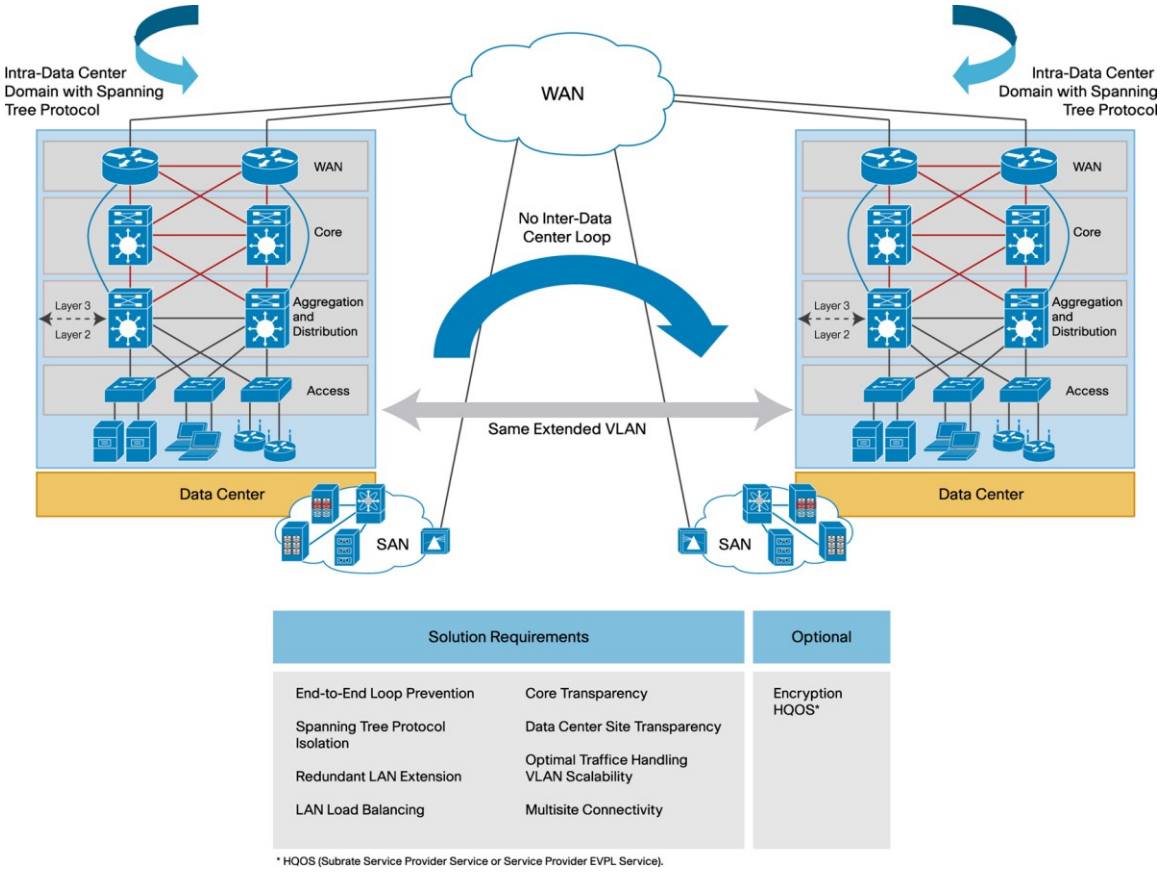
Note: This document covers only extension of the LAN between remote sites. All recommendations described here address sturdiness, optimization of physical links, resiliency, redundancy, performance, and quality of service (QoS). All these software modules are executed in hardware.

[2] The maximum distances depend on the software vendors requirements. Please follow vendor recommendations carefully, as they usually limit the move to a campus area. For high-availability clusters (Veritas, Microsoft Server 2008, or Sun cluster), virtual machines will likely support Layer 3 for the move in the near future.

LAN Extension Solution Requirements

Extension of the Layer 2 network across a WAN (Figure 5) requires special design considerations. These considerations are described in detail in this section.

Figure 5. Inter-Data Center LAN Extension Requirements



End-to-End Loop Prevention

To improve the high availability of the Layer 2 VLAN when it extends between data centers, this interconnection must be duplicated. Therefore, an algorithm must be enabled to control any risk of a Layer 2 loop, and to protect against any type of global disruptions that could be generated by a remote failure.

The first native option to be considered is Spanning Tree Protocol. However, it must be isolated between the remote sites to mitigate the risk of propagating unwanted bad behavior, such as topology change or root bridge movement from one data center to another. These unwanted behaviors could be flooded throughout the Layer 2 network, making all remote data centers and resources unstable, or even inaccessible.

Spanning Tree Protocol Isolation to Control Redundant Layer 2 Topology

Cisco does not recommend extending the Spanning Tree Protocol domain beyond the campus. Spanning TreeIt Protocol is a very conservative, protocol that favors loss of connectivity over temporary looping during its operation. As a result, a Spanning Tree reconvergence generally generates momentary interruption in frame forwarding. Because the different data centers are independent-bridged domains, it is beneficial to isolate their respective Spanning Trees. This way, a change in a particular data center will not cause transient connectivity problems or superfluous flooding in another data center. The segmentation will also make configuration of the topology of the various data centers easier, as it will be computed relative to a local root bridge.

- **Spanning Tree Protocol Enabled as Last Resort:** Spanning Tree Protocol is used in conjunction with Multichassis EtherChannel (MEC). MEC provides redundancy for physical links and switches with all logical PortChannels forwarding (no Layer 2 loop). Spanning Tree Protocol is kept enabled as a last resort. The MEC concept can be deployed using VSS on the Cisco Catalyst® 6500 Series Switches or vPC on the Cisco Nexus® 7000 Series Switches.

Notice that Spanning Tree Protocol becomes useless for controlling redundant Layer 2 topology upward toward the edge switches facing the core. It is kept enabled downward toward the access layer when MEC is enabled at the aggregation layer, as access switches are dual-homed using the same logical PortChannel.

This approach supports an end-to-end, fully redundant Layer 2 network without the need for Spanning Tree Protocol. However, you should keep Spanning Tree Protocol enabled as a last resort.

- **Spanning Tree Protocol Isolation:** Spanning Tree Protocol is fully contained and isolated within each data center with Bridge Protocol Data Units (BPDUs) filtered at the boundary of each edge switch facing the core.
- **Redundant Layer 2 Extension:** One option for achieving redundant access to the core is to use Cisco IOS® Embedded Event Manager (EEM) semaphores to control and activate the WAN forwarding links.
- **WAN Load Balancing:** Typically, WAN links are expensive, so uplinks need to be fully used, with traffic load-balanced across all available uplinks.
- **Core Transparency:** The LAN extension solution needs to be transparent to the existing enterprise core, if available, to reduce any effect on operations.
- **Data Center Site Transparency:** The LAN extension solution should not affect the existing data center network deployment.
- **VLAN Scalability:** The solution must be able to scale to extend up to hundreds or thousands of VLANs.
- **Multisite Scalability:** The LAN extension solution should be able to scale to connect multiple data centers.
- **Hierarchical Quality of Service (HQoS):** HQoS is typically needed at the WAN edge to shape traffic for cases such as, for example, when an enterprise subscribes to a subrate service provider service or a multipoint Ethernet virtual private line (EVPL) service.
- **Encryption:** The requirement for LAN extension cryptography is increasingly prevalent, (for example, to meet federal and regulatory requirements).

DCI Reference Solutions and Platforms

Currently, Cisco recommends three technical approaches that allow Layer 2 VLAN extension between remote sites:

- MEC on high-speed optical Dense Wavelength Division Multiplexing (DWDM) links
- EoMPLS, VPLS, and A-VPLS with MPLS in the core
- EoMPLS over generic routing encapsulation (EoMPLSoGRE), VPLSoGRE, and A-VPLSoGRE with pure Layer 3 IP core

These approaches support wire-rate forwarding, redundancy with less than 5-second failover recovery time at worst, and no Spanning Tree Protocol extended beyond the data center.

MEC is the easiest solution with which to deploy redundant Layer 2 links. It can be implemented with either the Cisco Nexus 7000 Series vPC or the Cisco Catalyst 6500 Series VSS. Cisco specifically recommends MEC for metropolitan area network (MAN) distances between remote sites where the interconnections between physical links are provided using dedicated fibers. This technology provides high availability with all redundant active links as MECs, as well as failover times of less than a second. It can be deployed for point-to-point connectivity between two data centers from the aggregation layer, or it can be deployed for point-to-multipoint interconnections as the number of interconnected

data centers increases. The solution also benefits from the support of IEEE 802.1AE line-rate cryptography on the Cisco Nexus 7000 Series port application-specific integrated circuits (ASICs).

Note: The architecture, configuration, recommendations, and packet flows for VSS and MEC are available and are [described in detail. http://www.cisco.com/en/US/products/ps9335/index.html] The discussion here focuses only on geographical extension of redundant Layer 2 using VSS.

EoMPLS, in conjunction with Cisco IOS EEM, can be used:

- When the links are not dedicated fiber
- When the distances are greater than MAN distances
- When the cost to deploy dedicated fiber is a concern

This technology provides high availability. However, it best applies when the number of data centers to be interconnected is limited to two in a point-to-point fashion. EoMPLS can be established from the network-facing provider-edge (N-PE) switch (edge switches facing the core). The following two deployment alternatives are discussed in this document:

- **Native:** Deployed over Dark Fiber or Layer 2 service provider service
- **EoMPLSoGRE:** Deployed over Layer 3 service (IP or MPLS)

VPLS, in conjunction with Cisco IOS EEM, is the other recommended solution to provide high availability for Layer 2 extension between multiple data centers (more than two data center multipoint-to-multipoint interconnections), without the need to extend Spanning Tree Protocol between the remote sites. VPLS will be established from the node facing the MPLS core, named N-PE. The following two deployment alternatives are discussed in this document:

- **Native:** Deployed over Dark Fiber or Layer 2 service provider service
- **VPLSoGRE:** Deployed over Layer 3 service (IP or MPLS).

A-VPLS and A-VPLSoGRE are the recommended solutions to transport Layer 2 Interconnects over an IP or MPLS core. A Layer 2 pseudowire (EoMPLS, VPLS, and A-VPLS) could be deployed natively over an enterprise-managed MPLS core, or in scenarios in which the service offered by a provider is dark fiber or a Layer 2 service.

This document focuses on the following options to extend Layer 2 VLANs between multiple remote data centers:

- Cisco Nexus 7000 Series vPC and Cisco Catalyst 6500 Series VSS for MAN distances
- Virtual Layer 2 links using EoMPLS and VPLS natively (over an MPLS core or over a Layer 1 or 2 type of transport)
- Virtual Layer 2 links using EoMPLS, VPLS, and A-VPLS, but over a Layer 3 core with GRE tunnels (this feature is called VPLSoGRE)

The recommended Cisco platforms for each of these deployments will also be specified.

Transport Option 1: Dark Fiber

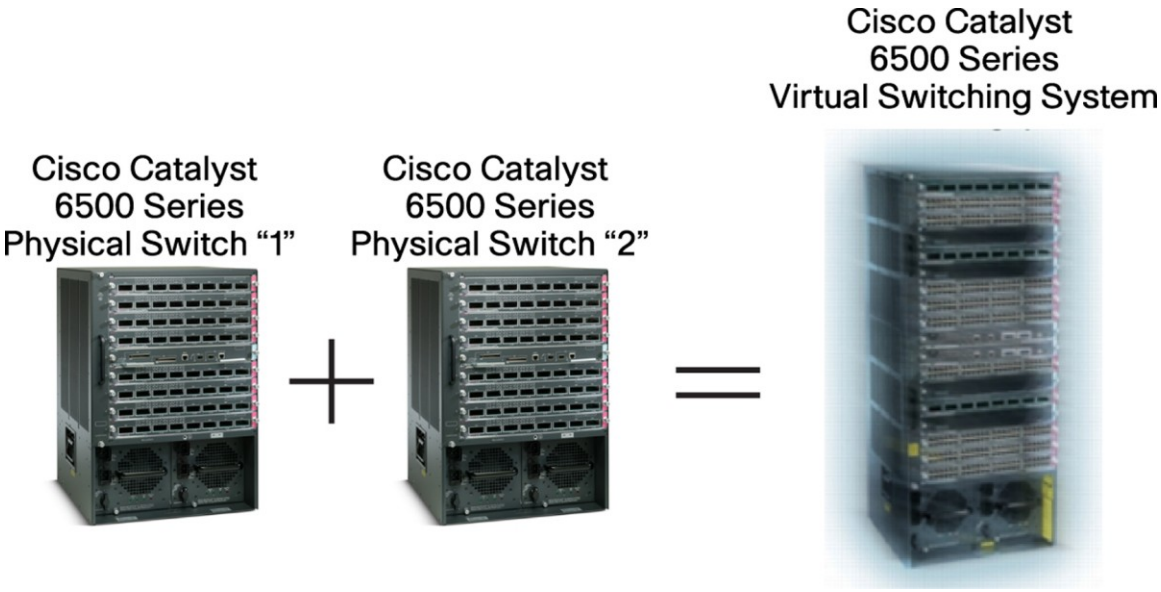
Virtual Switching System

Virtual Switching System (VSS) is a function that can be enabled on the Cisco Catalyst 6500 Virtual Switching Supervisor Engine 720 and Supervisor Engine 2T with 10GE uplinks.

VSS is executed in hardware, with new ASICs developed to support this feature. Therefore, it has no negative effect on performance. Instead, it improves the performance of the data plane and bandwidth used to interconnect all remote sites, using MEC technology.

Although VSS enables a single virtual switch for the management and control plane, it is formed with two physical switches, thereby doubling the performance of the data plane (Figure 6).

Figure 6. Cisco Catalyst 6500 Series VSS



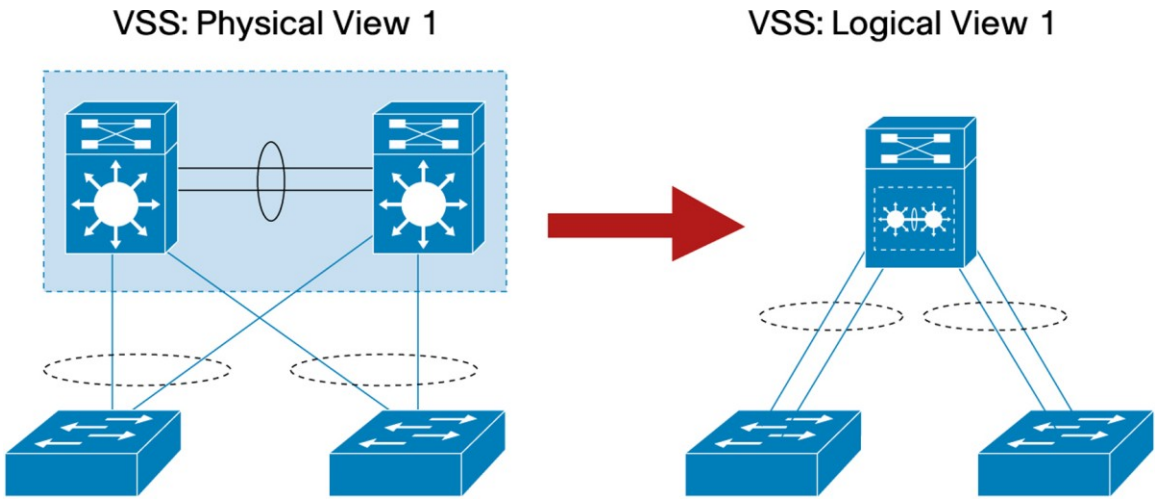
VSS can be configured to execute the control plane and management (configuration and monitoring) in a single system known as the active device. In this case, the data switching (data plane) is executed on both physical fabrics with optimization of the data paths for ingress and egress flows, both upward and downward from the pair of switches.

There is only one single active switch to manage, which encompasses the hardware resources from the active switch, as well as the hardware resources from the standby switch (line cards as well as network services[3]).

As the control plane runs in a single active machine, the Cisco Catalyst 6500 VSS 1440 will authorize the extension of multiple uplinks from any of the two physical switches to build a single logical PortChannel split between both physical switches (Figure 7). This technology is known as Multichassis EtherChannel, or MEC, and it is fully executed in hardware.

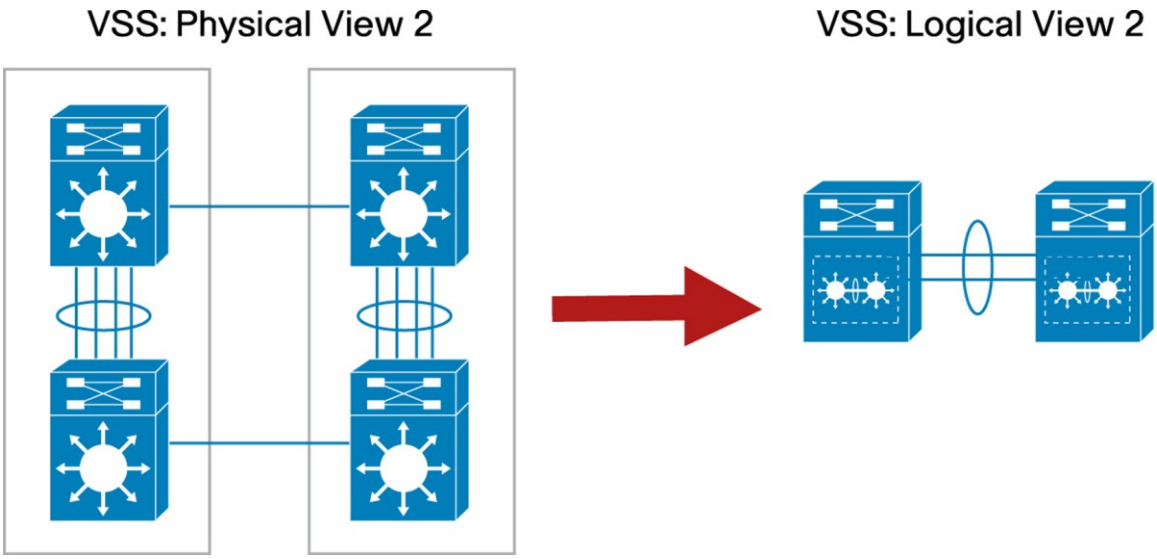
[3] Network and security services such as Cisco Application Control Engine (ACE) and Cisco Catalyst 6500 Series Firewall Services Module (FWSM) are supported with Cisco IOS Software Release 12.2(33)SXI. Currently, only the WS-X6700 series line cards in the Cisco Catalyst 6500 Series, based on centralized forwarding cards (CFCs) or distributed forwarding cards (DFCs), are supported; a mix of CFCs and DFCs is supported as well.

Figure 7. VSS and MEC on Cisco Catalyst 6500 Series



If we use this function between two pairs of VSS devices (two Cisco Catalyst 6500 Series Switches, plus two more Cisco Catalyst 6500 Series Switches), we get two logical switches connected point-to-point (Figure 8).

Figure 8. Back-to-Back Cisco Catalyst 6500 VSS 1440 Systems

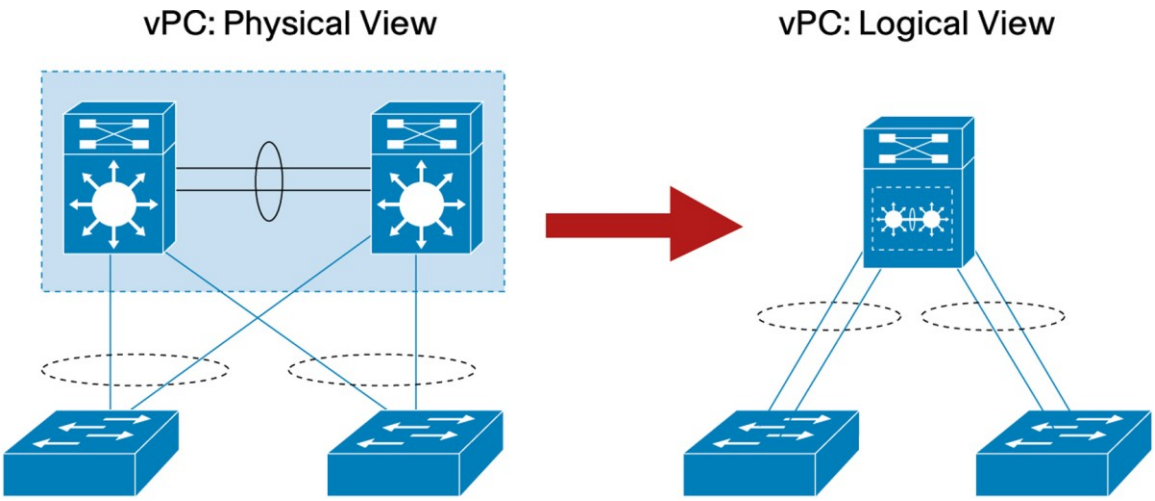


Therefore, we can achieve a fully redundant configuration with total separation of the devices, as well as the interconnection links, with uplink failover recovery time of less than a second (approximately 500 milliseconds [ms]).

Virtual PortChannel

Cisco NX-OS Software provides an MEC technique, called virtual PortChannel (vPC), which allows the network administrator to create a PortChannel with ports that are distributed across different physical devices (Figure 9).

Figure 9. Cisco Nexus 7000 Series vPC Concept



A pair of switches acting as a vPC appear to any PortChannel-attached devices as a single logical entity from the Layer 2 perspective. However, the two device members of the vPC are still two separate devices with independent control planes.

The two independent control planes of the devices participating in a vPC preserve proven network-level redundancy models, such as those provided by IP routing protocols. Each device also independently maintains all its device resiliency attributes, such as element redundancy and stateful restart. Thus, vPCs provide all the resiliency attributes that network architects are familiar with, while improving the operational environment of the bridged network.

The vPC environment combines the benefits of hardware redundancy with the benefits of PortChannel loop management. Most Spanning Tree Protocol dependencies are removed in an all-PortChannel-based loop management model, in which Spanning Tree Protocol is used solely as a fallback mechanism. This approach has very attractive operational implications for the management of bridged environments in the network.

Furthermore, all links are active in a vPC topology, and traffic is distributed over the various links to achieve full utilization of the available cross-sectional bandwidth. Traffic distribution is based on a link aggregation hash algorithm, such as the one defined in the IEEE 802.3ad link aggregation standard. Failover times in a vPC topology are also comparable to those achieved in traditional point-to-point PortChannels, in the less-than-a-second range.

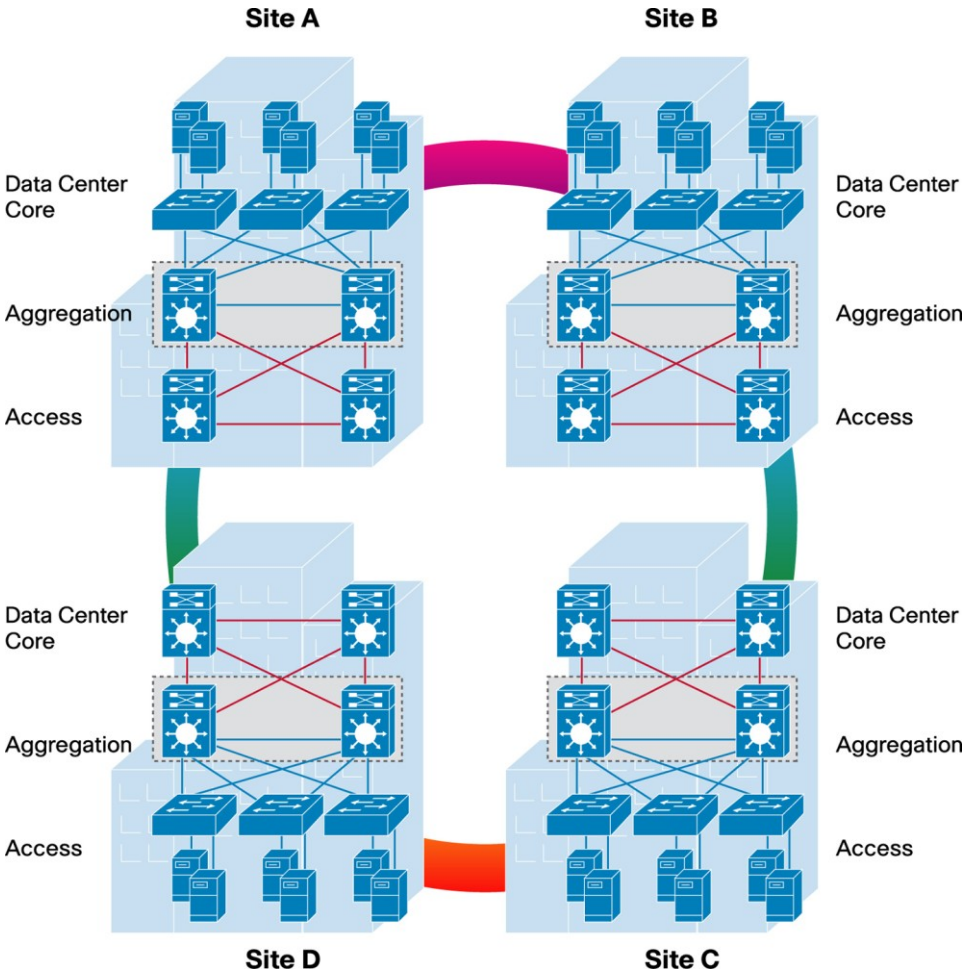
All multichassis PortChannel technologies require a direct link between the two device members of the PortChannel. This link is often much smaller than the aggregate bandwidth of the vPCs connected to the endpoint pair. Cisco technologies, such as vPC and VSS, are specifically designed to optimize unicast and multicast traffic patterns. They limit the use of this interswitch link to the minimum required to switch management traffic and the occasional traffic flow from a failed network port. This approach is crucial to support data center environments in which many terabits of data traffic may be in transit.

Fiber-Based Layer 2 VPNs with VSS and vPC

By extending the VSS and vPC concepts to a geographical topology built with multiple remote sites, a flexible, scalable, end-to-end architecture can be created to support additional data centers.

The example in Figure 10 shows four remote data centers to be interconnected using a DWDM ring. DWDM provides the media layer to initiate the various point-to-point physical layers (built from each available wavelength of the optical link). Cisco recommends not extending this ring beyond 100 kilometers.

Figure 10. VSS and Dedicated Fibers

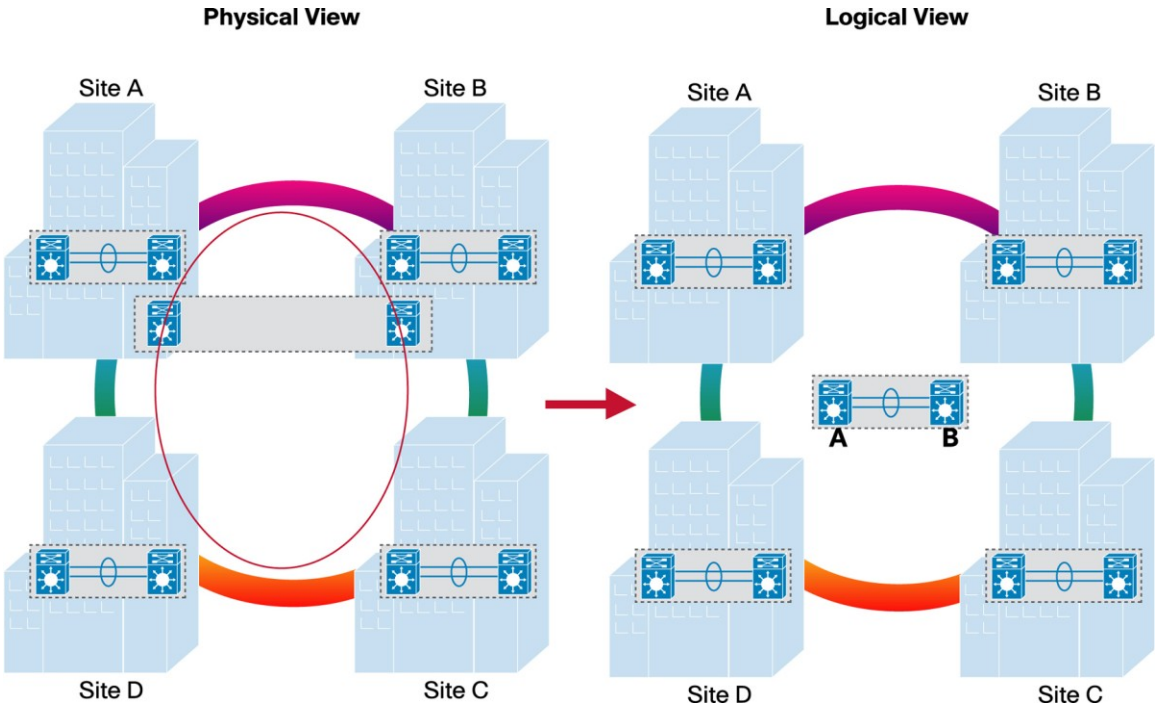


Within each single data center, the aggregation layer can provided by a Cisco Catalyst 6500 VSS 1440, Supervisor 2T, or a Cisco Nexus 7000 Series Switch with vPC. For the purposes of this solution, the MEC capability is what is required; hence, vPC and VSS are equivalent propositions. In Figure 11, a core VSS or vPC is added to the intra–data center aggregation layer. It functions as the distribution point for traffic to all other data centers.

A crucial architecture point is that each physical switch that comprises the core VSS or vPC is spread onto different remote sites, thus offering high availability. Logical point-to-point connections are created using DWDM transport (also called lambda or wavelength) over an optical ring. The logical point-to-point connections are laid out to produce a virtual star topology with the core VSS or vPC as the hub and the aggregation pairs as the spokes. A dedicated wavelength is also used to provide a logical point-to-point link between the members of the VSS or vPC. The interswitch links[4] that provide interconnection between the two VSS or vPC members must be 10-Gbps Ethernet.

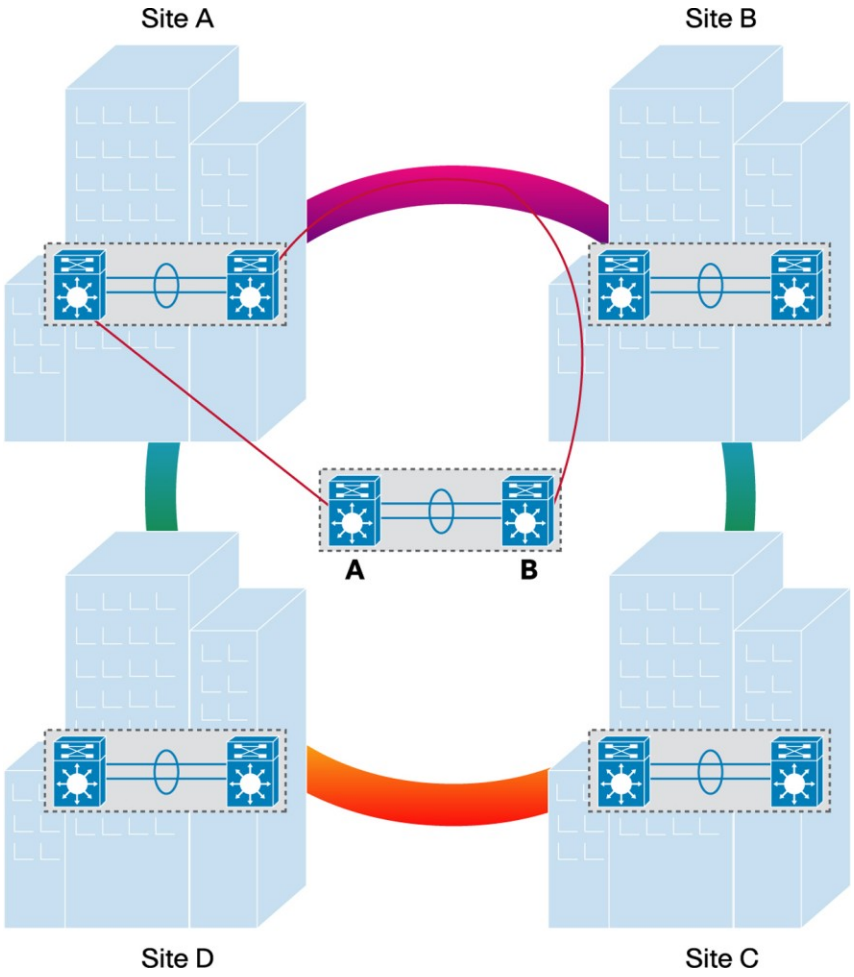
[4] The links that interconnect the two members of a VSS are called virtual switch links (VSLs). A VSL is typically built with the two active 10 Gigabit Ethernet links from the Cisco Catalyst 6500 Series VSS Engine 720 or SUP2T with 10GE uplinks. The links that interconnect the two members of a vPC are called vPC-peer links. vPC-peer links can use any combination of 10 Gigabit Ethernet links on the Cisco Nexus 7000 Series Switches.

Figure 11. Core VSS or vPC



After the core vPC or VSS is configured, the next step is to interconnect each VSS or vPC that comprises the aggregation layer to the core VSS or vPC. In Figure 12, for the interconnection of Site A to the core VSS or vPC, the switch on the left of the core VSS or vPC is physically located in the same data center in Site A. It uses one local fiber (Gigabit Ethernet or 10 Gigabit Ethernet). The physical switch on the right side of the core VSS or vPC, being located in Site B, is connected to Site A through Gigabit Ethernet (or 10 Gigabit Ethernet) built from one of the available wavelengths created between Sites A and B.

Figure 12. Site Interconnection to the Core VSS or vPC (Site A)



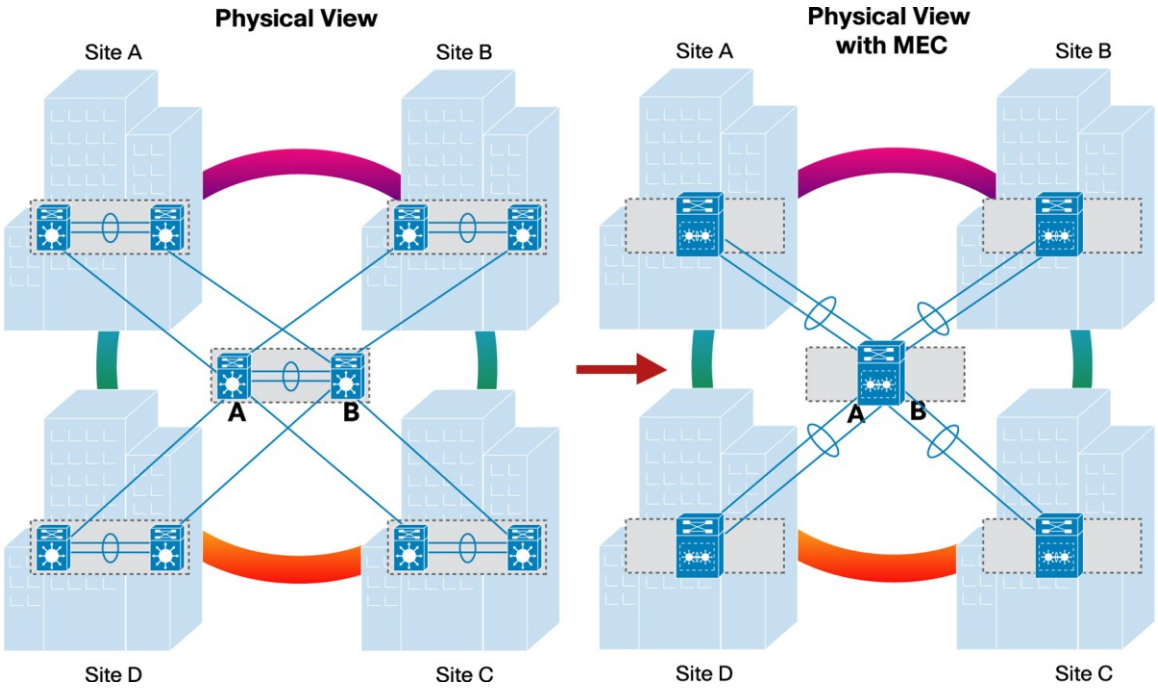
Sites B, C, and D are configured in a similar manner.

Applying this same interconnection methodology to all sites yields a highly flexible design, in which the only limitation is the maximum number of uplinks available from the core VSS or vPC and the maximum number of wavelengths available from the DWDM ring.

Figure 13, which shows four remote data centers, uses up to eight dedicated fibers (or eight lambdas) as follows:

- Two 10 Gigabit Ethernet for the interswitch link of the core VSS or vPC
- Six Gigabit Ethernet for all VSSs or vPCs that build the four aggregation layers: site A (one lambda), site B (one lambda), site C (two lambdas), and site D (two lambdas); the others use a local fiber (site A and B connections to the core VSS or vPC)

Figure 13. Layer 2 Extension over Dark Fiber DWDM WAN Network



The aggregation layer of each site is provided by a pair of Cisco Catalyst 6500 Series Switches running VSS or a pair of Cisco Nexus 7000 Series Switches running vPC. Each access switch (responsible for direct attachment to the servers) can therefore be dual-homed to the aggregation layer with both uplinks active (MEC). If the access switch itself supports MEC (Cisco Catalyst 6500 Series and Cisco Nexus 7000 and 5000 Series), the dual-homed attachment of the servers can be substantially improved by enabling Link Aggregation Control Protocol (LACP) on the server connection and making multiple network interface cards (NICs) active at the same time for transmit and receive traffic I/O but distributed to two different switches (Figure 14).

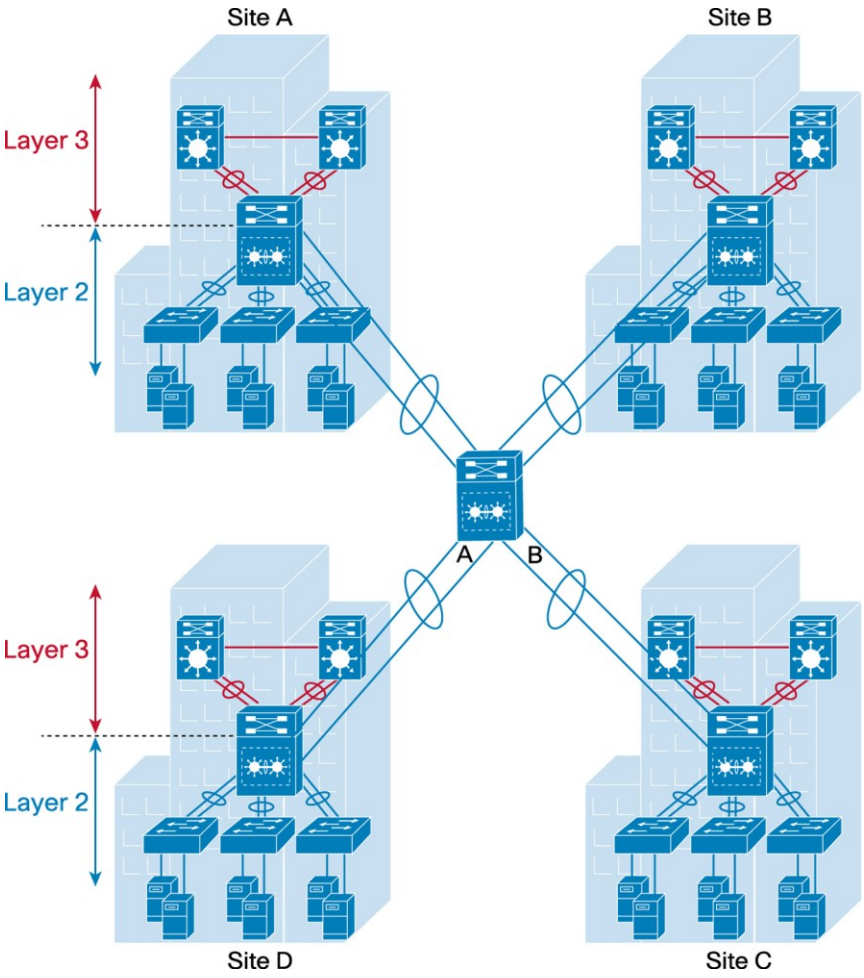
The redundant Layer 2 extends from the access layer to the aggregation layer. It is connected to the core VSS or vPC, building a full end-to-end redundant and active Layer 2 extension. There is no need to enable Spanning Tree Protocol to control Layer 2 looping, as all uplinks are active with MEC. However, keeping Spanning Tree Protocol as the last resort in case of software or fiber or cable misconfiguration is preferred.

While not mandatory with this VSS or vPC solution, Cisco strongly recommends enabling Spanning Tree Protocol locally within each data center access layer as a failsafe mechanism. This will protect against physical errors such as mistakes in cabling, patching, and configuration.

When Spanning Tree Protocol is not used to control the Layer 2 loop in this scenario, but instead enabled to provide additional security, it may disturb the entire extended network. For example, topology change notification (TCN) could run beyond a local data center. This scenario can be improved by creating a Multiple Spanning Tree (MST) region for each site.

With MST, you can isolate the topology of the various MST instances so that, for example, a blocked port in Region 2 will not move as a result of a topology change in Region 1. In addition, with MEC, the logical link to interconnect each data center to the core VSS will use both physical dedicated fibers. Alternatively, since the star topology in the core VSS or vPC is loop-free, BPDUs can be filtered on the interfaces facing the core VSS or vPC. BPDU filtering on these interfaces allows isolation of the Spanning Tree domains to each site and prevents the formation of a single Spanning Tree across all sites.

Figure 14. MEC Across a Global Data Center



Transport Option 2: MPLS

The MPLS transport option is relevant in two scenarios:

- Enterprise owns the MPLS enabled core network
- Enterprise acquires a Layer 1 or Layer 2 type of service from a provider, and MPLS is run between the enterprise devices at the edge of the provider's cloud

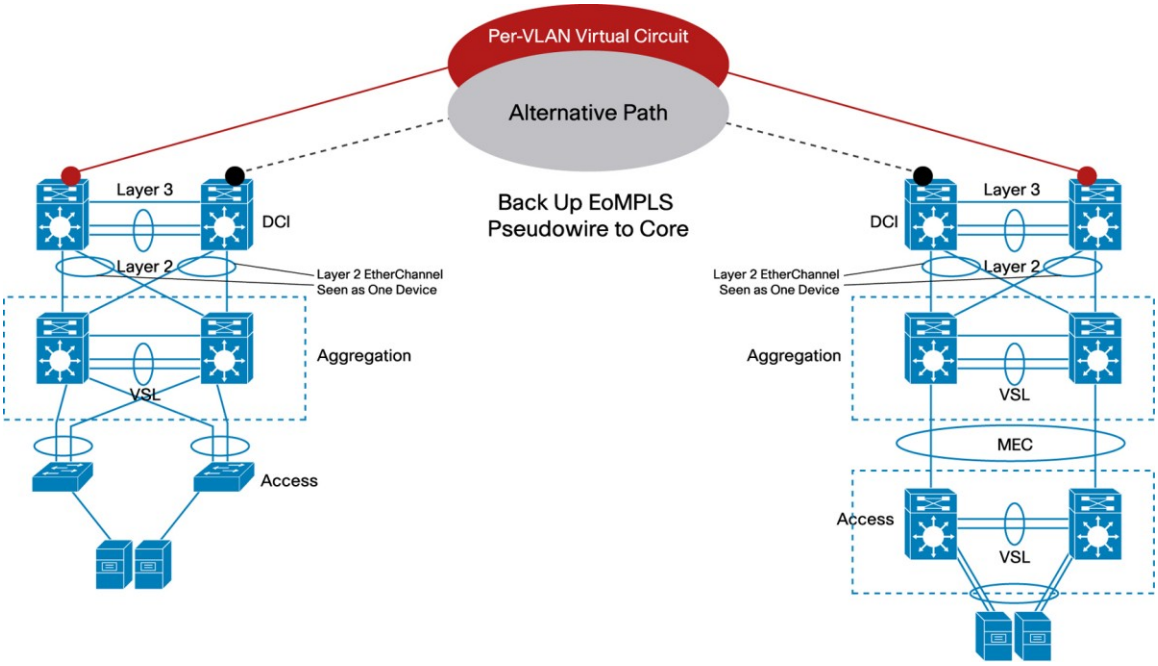
Regardless of the specific scenario, the idea is that the Layer 2 extension technology is initiated from a pair of DCI devices owned by the enterprise and deployed in each site that needs to be connected. From a technology perspective, two approaches are possible:

- **EoMPLS:** Usually positioned for point-to-point scenarios in which the enterprise needs to connect two data center sites
- **VPLS:** Used for point-to-multipoint deployments in which LAN extension needs to be provided between multiple sites

EoMPLS

EoMPLS supports cross-connect access to ports in a one-to-one fashion. Any ingress traffic will be transported and delivered to the remote port as is, whether it is data or a control packet. EoMPLS performs this action by using dual-label tagging: one tag for the core path pointing to the edge switch, and one tag for the virtual circuit pointing to the edge port. This dual-label path, called a pseudowire, is the virtual wire or fiber extension of the cross-connections between the two remote interfaces used for private interconnects (Figure 15).

Figure 15. Layer 2 Extension Across Data Centers Using EoMPLS



EoMPLS delivers the following benefits:

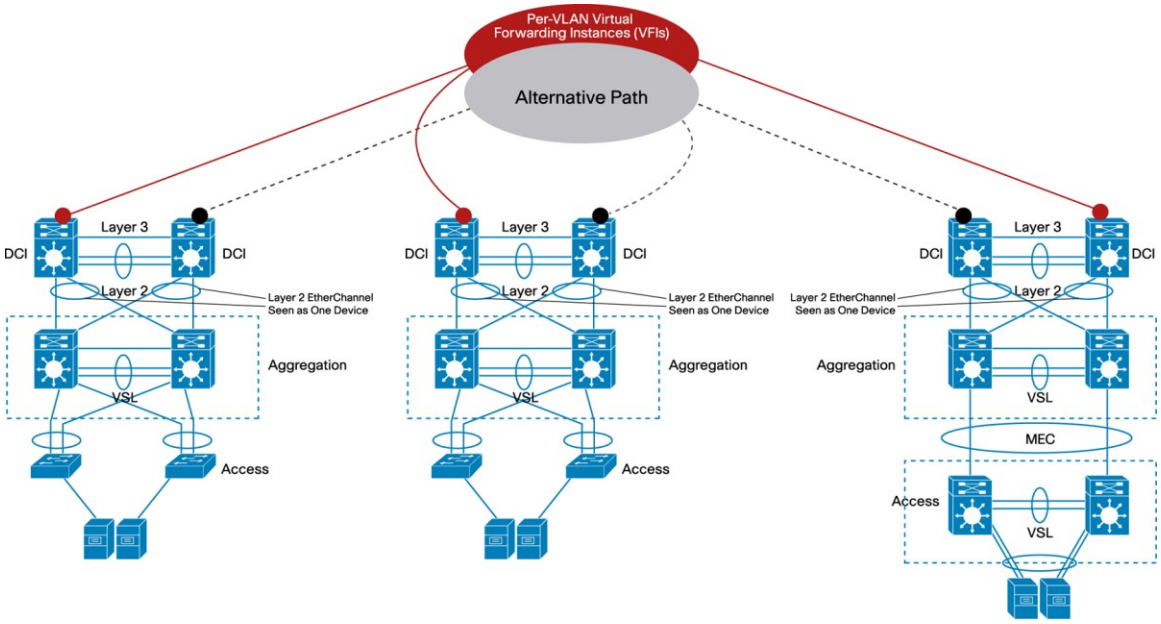
- Private network transport that complies with cluster vendor recommendations to avoid the use of Spanning Tree Protocol
- Redundancy addressed without the need to enable Spanning Tree Protocol in the core, as the failover on the physical layer is controlled by the Layer 3 network
- Layer 3 technology responds upon failure, allowing very fast convergence and maintaining stability
- The overlay of the Layer 2 connection on the Layer 3 transport hides any physical convergence, thereby increasing Layer 2 stability overall

VPLS

VPLS is a class of VPN that supports the connection of multiple sites in a single bridged domain over a managed IP or MPLS network. VPLS presents an Ethernet interface to customers, simplifying the LAN or WAN boundary for enterprise customers, and supporting rapid and flexible service provisioning, since the service bandwidth is not tied to the physical interface. All services in a VPLS network appear to be on the same LAN, regardless of location (Figure 16).

VPLS uses edge routers that can learn, bridge, and replicate on a per-VPN basis. These routers are connected by a full mesh of tunnels, enabling any-to-any connectivity. VPLS operation emulates an IEEE Ethernet bridge.

Figure 16. DCI LAN Extension with VPLS



A-VPLS

A-VPLS is a enhancement to the existing VPLS solution. It is a flow-aware pseudowire used to extend VLANs across data centers over an IP or MPLS core network, while providing a switch-port-based provisioning model for enterprise customers.

A-VPLS introduces a new virtual interface–based command-line interface (CLI) concept for VPLS and switch port configuration called virtual Ethernet. The virtual Ethernet Interface is used to configure the trunking and pruning of VLANs and the A-VPLS neighbor list and its associated attributes. It greatly simplifies existing VPLS configuration and is enterprise friendly and scalable (Figures 17 and 18). The virtual Ethernet interface supports multipoint by nature using a single interface.

A-VPLS also uses the Cisco Catalyst 6500 Series VSS feature to provide native dual-homing, eliminating the need to use Cisco IOS EEM scripts or Spanning Tree Protocol to provide redundancy and load balancing. It also provides flow-based load balancing over equal-cost multipath (ECMP) paths. The load balancing occurs at the local provider-edge uplinks, as well as at the IP routers.

Figure 17. Two-Site Setup with A-VPLS

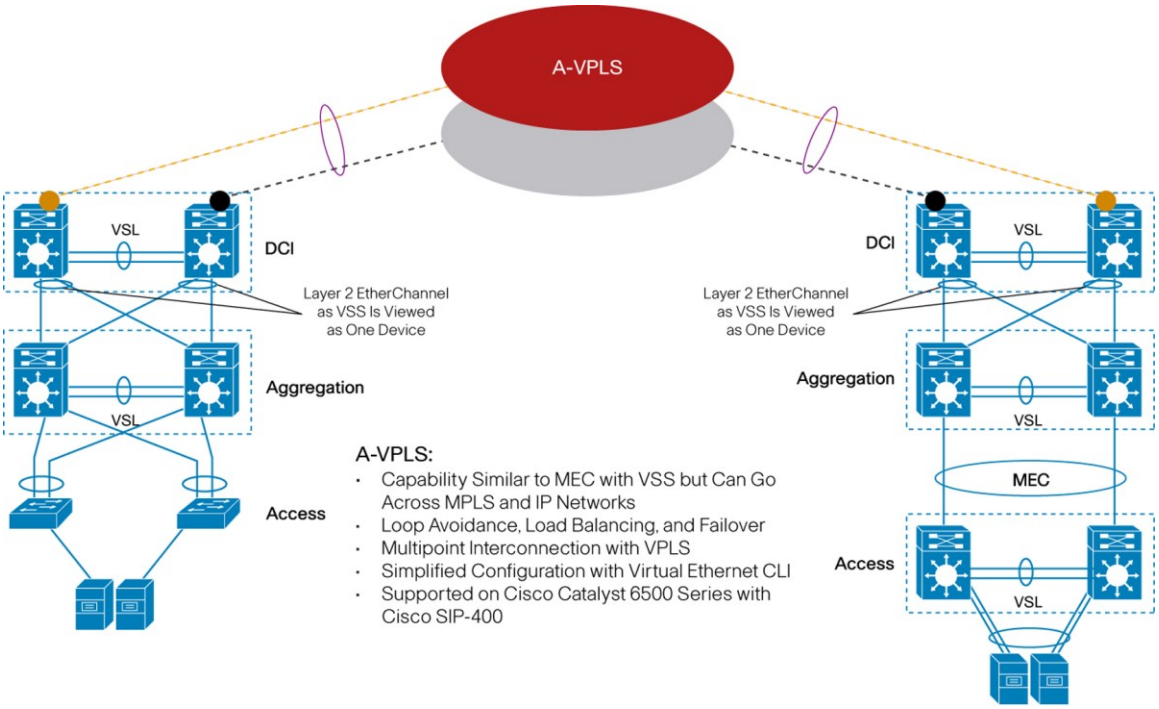
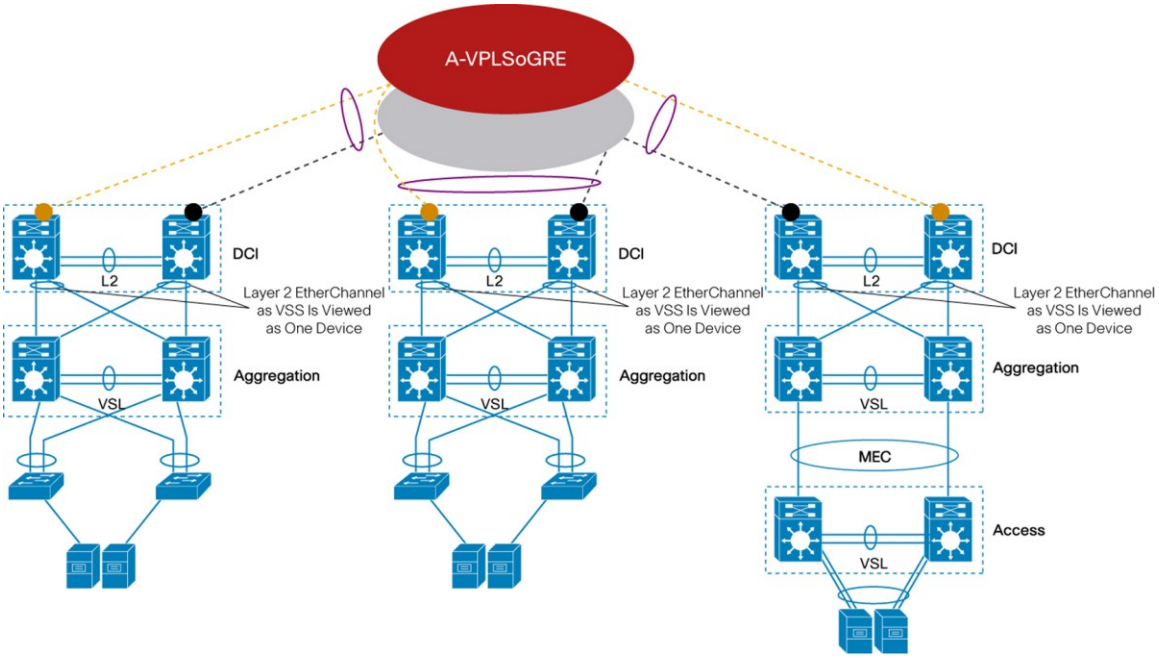


Figure 18. Three-Site Setup with A-VPLS



EoMPLS, VPLS, and A-VPLS Platform Support and Positioning

Table 1 shows the Cisco platforms and solutions for transporting Layer 2 packets over an MPLS-based WAN network.

Table 1. EoMPLS, VPLS, and A-VPLS Platform Support and Positioning

	DCI Solution	
Requirement	**Solution**	**Platform**
Layer 2 over MPLS	EoMPLS	• Cisco Catalyst 6500 Series with Cisco Shared-Port Adapter (SPA) Interface Processor-400 (SIP-400) • Cisco Catalyst 6500 Series with Cisco SIP-600 (Special Bundle)* • Cisco Catalyst 6500 Series with Cisco ES+ Line Card • Cisco Catalyst 6500 Series with Cisco Sup 2T** • Cisco ASR 1000 Series Aggregation Services Routers
	VPLS A-VPLS	• Cisco Catalyst 6500 Series with Cisco SIP-400 • Cisco Catalyst 6500 Series with Cisco SIP-600 (Special Bundle)* • Cisco Catalyst 6500 Series with Cisco ES+ Line Card • Cisco Catalyst 6500 Series with Cisco Sup 2T**
Encryption	IEEE 802.1ae	• Cisco Nexus 7000 Series
	IP Security (IPsec)	• Cisco Catalyst 6500 Series with Cisco SPA Services Card 600 (SSC-600) or VPN SPA (VSPA) • Cisco ASR 1000 Series
Multilevel QoS	HQoS	• Cisco Catalyst 6500 Series with Cisco SIP-400 • Cisco Catalyst 6500 Series with Cisco SIP-600 • Cisco Catalyst 6500 Series with Cisco ES+ Line Card • Cisco ASR 1000 Series

* New reduced-pricing product IDs: VPLS-2x10GE-LAN, VPLS-2x10GE-XFP, A-VPLS, and A-VPLSoGRE are supported on Cisco SIP-400 and ES+ Line Cards only.
** SUP2T will support VPLS, H-VPLS and A-VPLS natively, and will not require any special WAN cards like SIP-400, SIP-600, or ES+ (please check feature navigator for exact feature support)

Transport Option 3: IP

The IP transport option applies when the enterprise-edge device is peering at Layer 3 with the first provider device. In this case, you cannot use EoMPLS, VPLS, or A-VPLS natively, and an overlay needs to be created to logically interconnect the enterprise devices in the different data centers.

The typical way of achieving this is with GRE tunnels. EoMPLS, VPLS, and A-VPLS can then be run on top of the logical overlay created by the mesh of GRE tunnels, as discussed in the following sections.

EoMPLSoGRE, VPLSoGRE, and A-VPLS

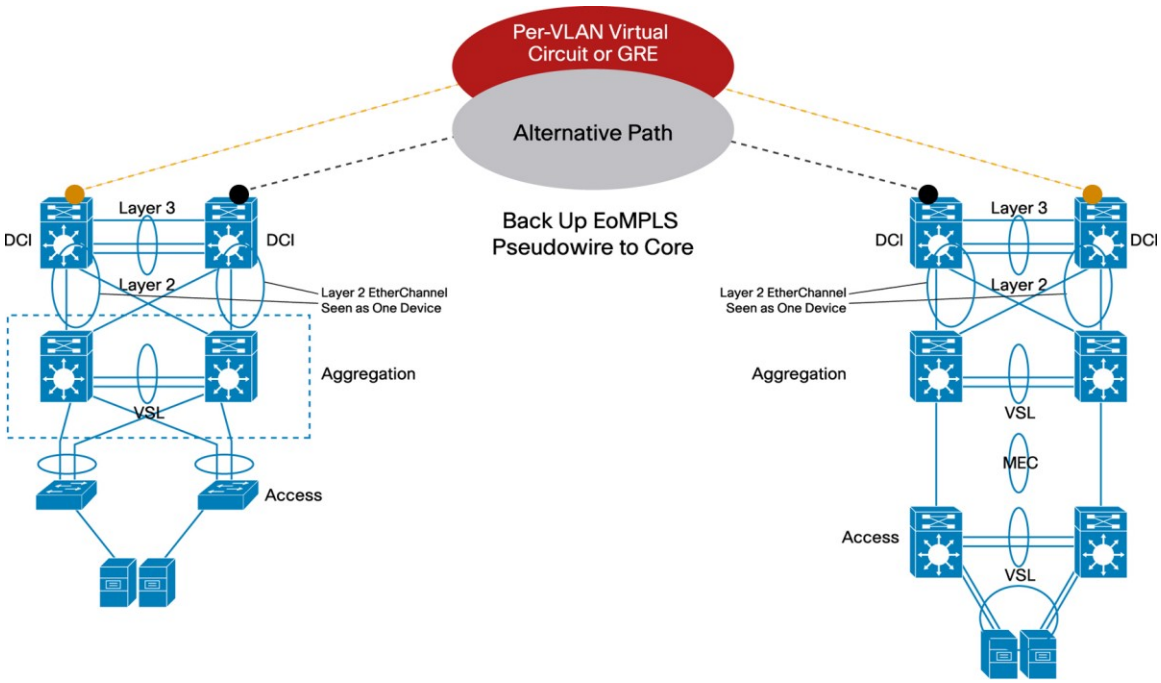
To facilitate the adoption of Layer 2 extension, Cisco offers solutions to encapsulate the A-VPLS, VPLS, and EoMPLS traffic on a GRE tunnel. Encapsulation helps enable the transport of all Layer 2 flows over the existing IP core, eliminating the need for a complex migration process.

This solution is called Any Transport over MPLS (AToM) over GRE (AToMoGRE) or Layer 2 VPN over GRE (L2VPNoGRE). It creates a GRE tunnel, hardware-switched and with high performance, that encapsulates AToM frames, such as EoMPLS or VPLS, on its top.

The L2VPN-over-IP design is identical to the deployment over MPLS. EoMPLS port cross-connect is the default option for point-to-point connection, and VPLS or A-VPLS is effectively the option for multisite interconnection. However, in the future, if interconnection of additional data centers is likely, you should enable A-VPLS or VPLS for point-to-point interconnection, for smooth interconnection of data centers.

In the EoMPLSoGRE design in Figure 19, the GRE connection is established between the two data center core switches. Then, the MPLS link-state packet (LSP) is tunneled over. From this point, any AToM session is established over this MPLSoGRE connection.

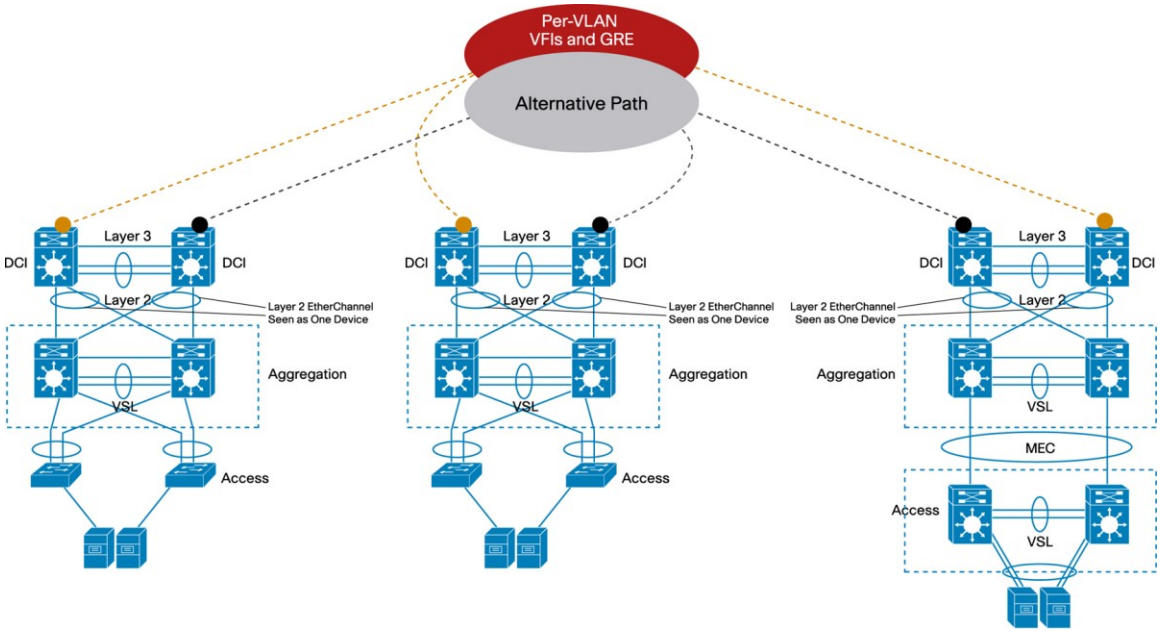
Figure 19. DCI LAN Extension with EoMPLSoGRE



This approach allows the enterprise to build EoMPLS point-to-point across connections between two sites, while these connections are being transported over the existing IP core. MPLS does not need to be deployed in the core network.

If the enterprise requirement wants to interconnect multiple sites in a multipoint fashion, VPLS is the recommended technology. Like EoMPLS, it can be transported over IP using a GRE tunnel. Again, MPLS does not need to be deployed in the core network (Figure 20).

Figure 20. DCI LAN Extension with VPLSoGRE



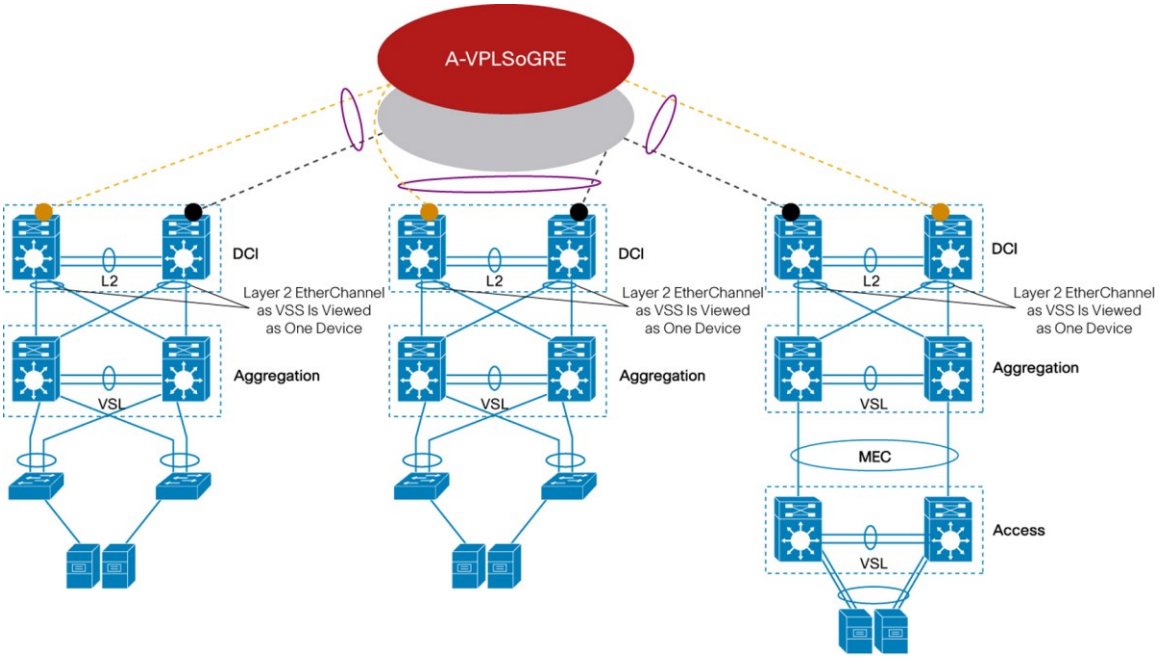
The power of VPLS lies in the capability to create virtual switching instances, (VSIs, also called virtual forwarding instances, or VFIs) fully meshed across sites without any risk of creating a Layer 2 loop in the core. This autoprotection loop breaker, called split-horizon protection, prevents retransmission of any packet received from the core to a VFI over any other core connection.

The multipoint characteristic of VPLS allows easy evolution of the global architecture, with flexibility to incorporate new sites without service disruption.

The hierarchical VPLS (H-VPLS) mode supports highly scalable bridging domains. It also offers VLAN overlapping at the edge, a critical feature in multiple-tenant data centers.

The added advantages of A-VPLS can be applied over an IP network as well by deploying A-VPLSoGRE, as shown in Figure 21.

Figure 21. DCI LAN Extension with A-VPLSoGRE



EoMPLSoGRE and VPLSoGRE Platform Support and Positioning

Table 2 shows the Cisco platforms and solutions supporting Layer 2 transport over an IP-based WAN network.

Table 2. EoMPLSoGRE and VPLSoGRE Platform Support and Positioning

	DCI Solution	
Requirement	**Solution**	**Platform**
Layer 2 over IP	EoMPLSoGRE	• Cisco Catalyst 6500 Series with Cisco SIP-400 • Cisco Catalyst 6500 Series with Cisco ES+ Line Card • Cisco Catalyst 6500 Series with Cisco Sup 2T** • Cisco ASR 1000 Series
	VPLSoGRE A-VPLSoGRE	• Cisco Catalyst 6500 Series with Cisco SIP-400 • Cisco Catalyst 6500 Series with Cisco ES+ Line Card • Cisco Catalyst 6500 Series with Cisco Sup 2T**
Encryption	IEEE 802.1ae	• Cisco Nexus 7000 Series
	IPsec	• Cisco Catalyst 6500 Series with Cisco SSC-600 or VSPA • Cisco ASR 1000 Series
Multilevel QoS	HQoS	• Cisco Catalyst 6500 Series with Cisco SIP-400 • Cisco Catalyst 6500 Series with Cisco ES+ Line Card • Cisco ASR 1000 Series

** SUP2T will support VPLSoGRE, H-VPLSoGRE and A-VPLSoGRE natively and will not require any special WAN cards such as SIP-400, SIP-600, or ES+ (please check feature navigator for exact feature support)

EoMPLS and VPLS Layer 2 Loop Prevention

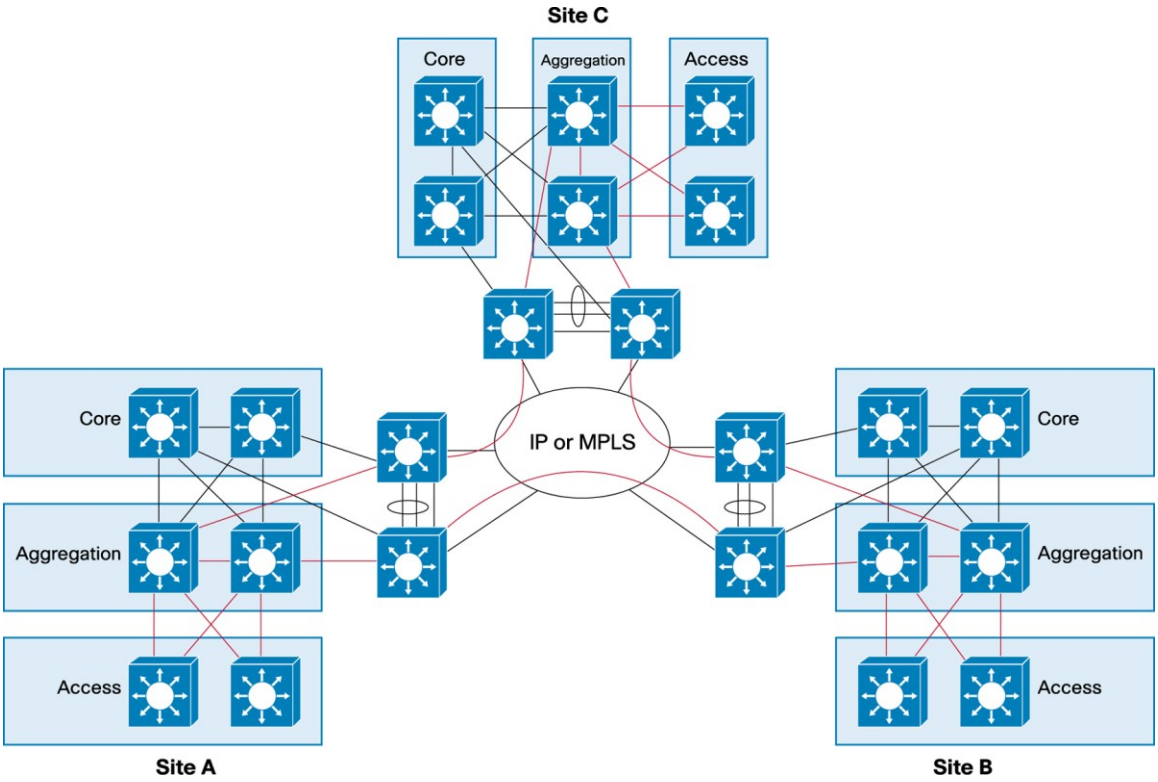
Layer 2 bridging technology was designed to work on a traditional campus, over a stable dark fiber network, and on a limited scale. The requirement for Layer 2 extension across multiple sites is pushing bridging beyond its intended scalability and capability to accommodate medium-quality links. However, Layer 3 technology offers proven support for long-distance links and scalability.

Layer 3 is natively very stable and uses fast convergence technologies (Fast Interior Gateway Protocol [IGP], Bidirectional Forwarding Detection [BFD], or even Fast Reroute [FRR]), so it provides transport of Layer 2 frames over very stable pseudowires. Any problems occurring in the physical network are completely transparent to the tunneled Layer 2 traffic.

With split-horizon protection, VPLS offers loopless multipoint bridging without the need to activate Spanning Tree Protocol in the core. Nevertheless, because data centers have to be dual-connected to the VPLS core (as shown in Figure 16), understanding how the redundant devices manage loops and redundancy is essential. Similar considerations can be applied to the EoMPLS scenario depicted in Figure 15.

Every bridging domain (a domain that resides on one side in the data centers sites, and on the other end in the Layer 2 pseudowires core) uses its own loop-breaker system. However, the topology must be considered end-to-end. Figure 22 shows how a loop can be created by associating independent loop-free domains.

Figure 22. Global End-to-End Layer 2 Loop with Dual Connections to Loop-Free Domain



In this example, every core pseudowire (in red) is guaranteed to be forwarding and not producing any loop inside the core (IP or MPLS) However, because each data center is dual-connected, a global loop is still created.

To break this global loop, several technologies can be used:

- Extension of Spanning Tree Protocol end-to-end
- Cisco IOS EEM semaphores
- VSS Support for Layer 2 extension technologies

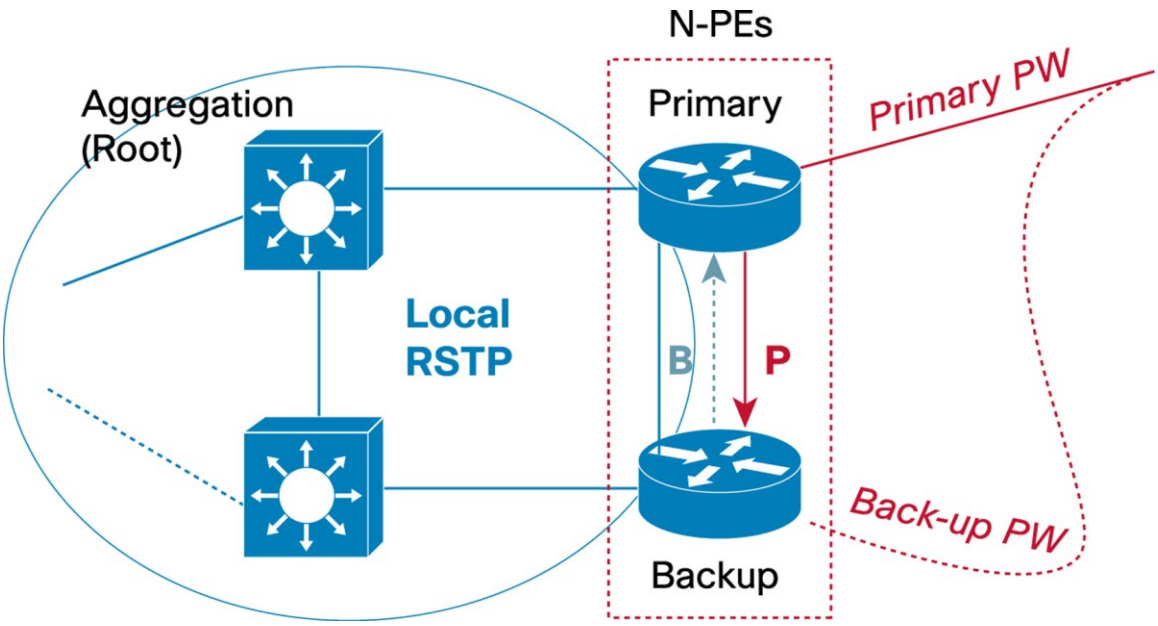
Because of high-availability requirements, end-to-end Spanning Tree Protocol across all sites is not recommended. Therefore, the Cisco IOS EEM approach is discussed here.

Note: The discussion in the following section applies to both EoMPLS and VPLS deployments independent of whether these technologies are deployed over MPLS - or IP-based transport.

Cisco IOS EEM Semaphoring in N-PE

The Cisco IOS EEM and N-PE solution is the most complete and flexible approach available. It can be adapted to any kind of data center topology to help ensure redundancy and scalability. The solution's insertion into an existing data center is smooth, without any need to change the root bridge placement. It is also fully compatible with the use of VSS in the distribution layer (Figure 23).

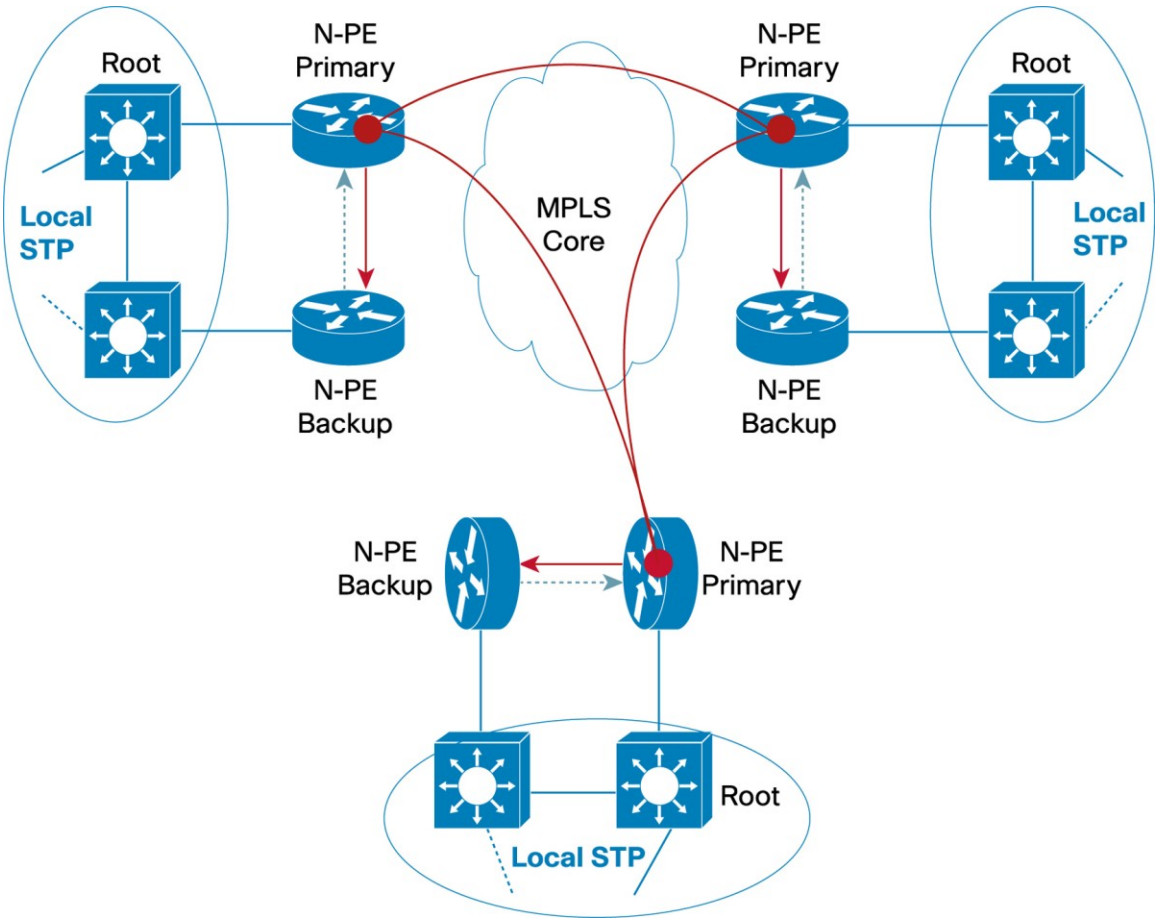
Figure 23. Cisco IOS EEM and N-PE



The Cisco IOS EEM semaphore approach is relatively simple. One of the N-PE devices (EoMPLS or VPLS edge device) is designated as primary, and the other is designated as backup. The backup pseudowire stays in standby mode while the primary one is active, and is activated upon failure. Because only one pseudowire is active at a time, no loop can exist, even at the global topology.

From the primary N-PE, a signal (called a semaphore) indicates to the backup N-PE that it is still active. As shown in Figure 24, as long as the backup node receives this signal, no action is taken.

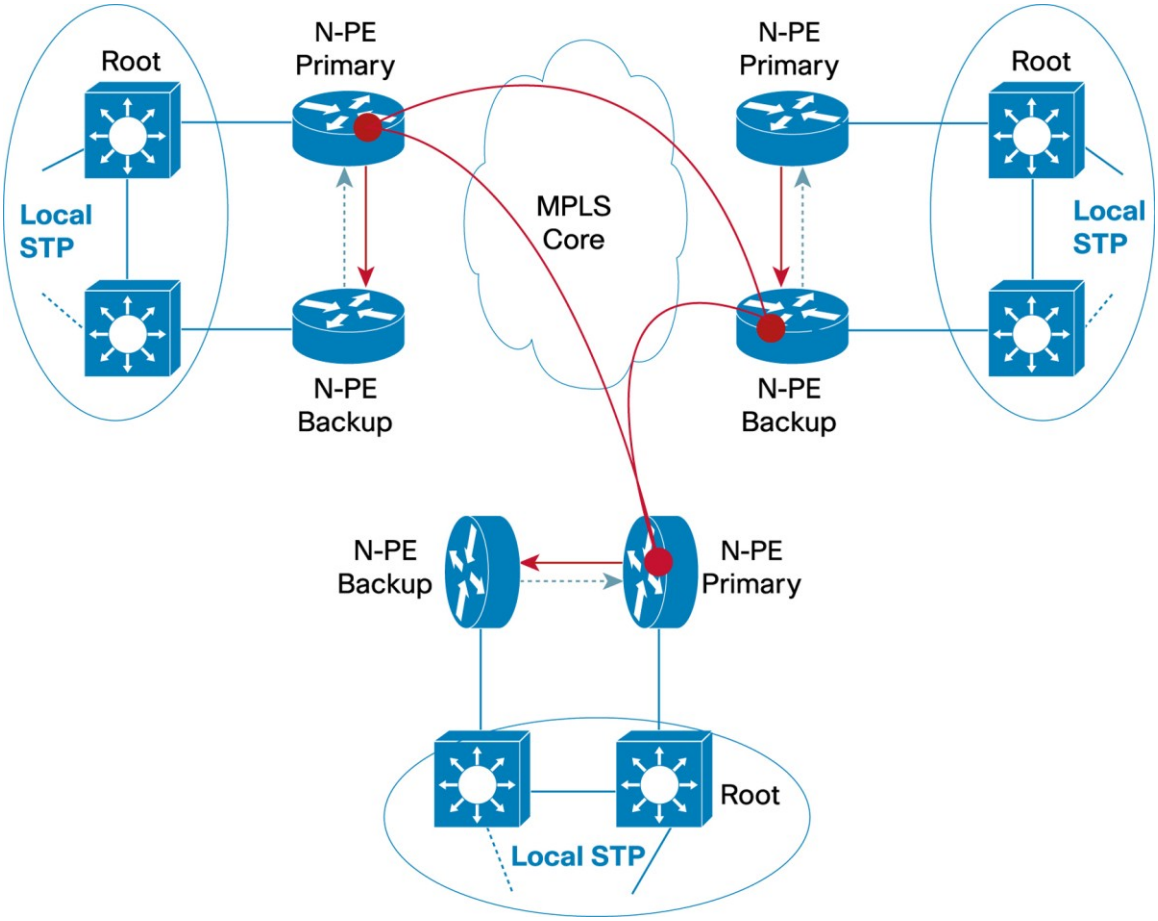
Figure 24. Primary N-PE Is Active



If the red P (primary) semaphore is up, this forces the backup node into standby status, which is acknowledged by a signal showing that the B (backup) semaphore is down. With the B semaphore down, the primary node can be active.

If the P signal disappears, the backup N-PE immediately relays the site connection through setup of the backup pseudowire. Note that the backup pseudowire is connected to the primary VFI of the other sites, which is the same VFI that owns the primary pseudowire, providing split-horizon protection (Figure 25).

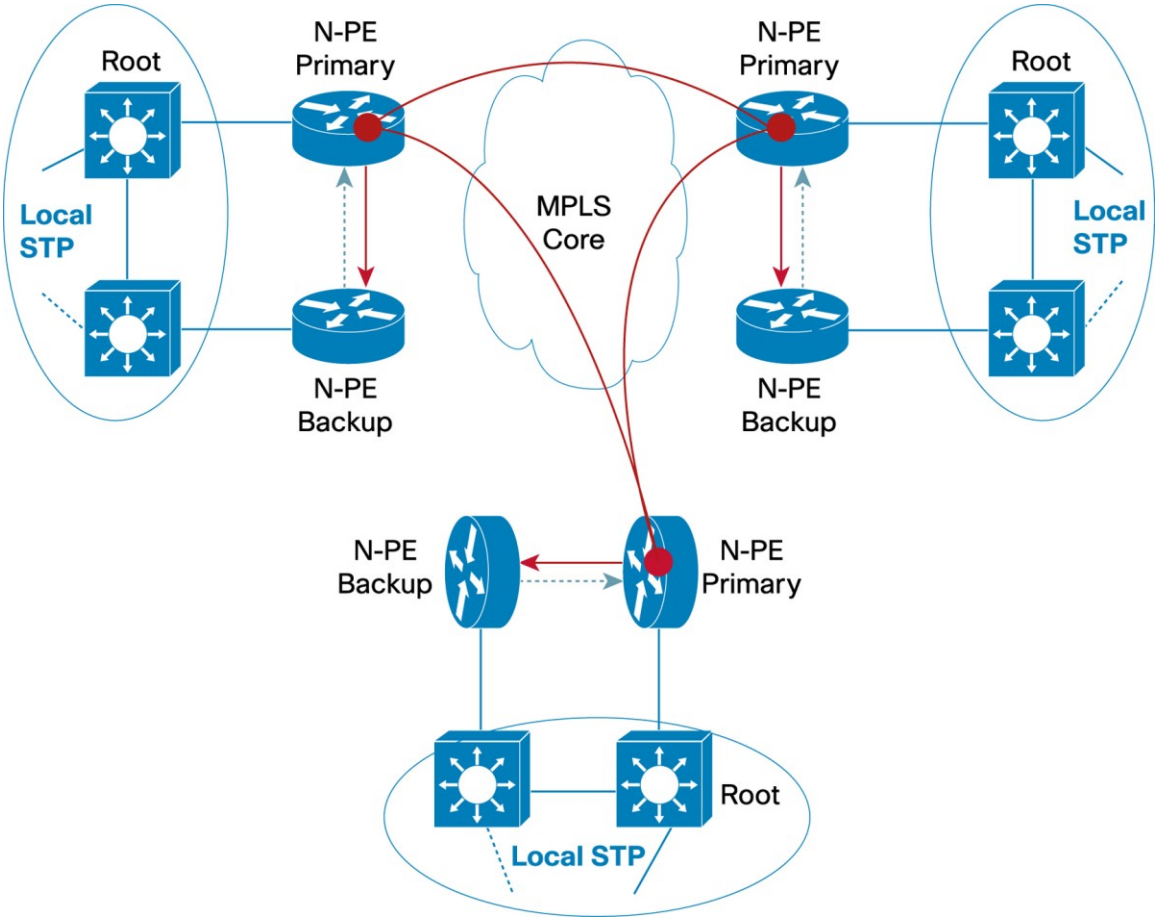
Figure 25. Primary N-PE Is Down



If the red primary P semaphore goes down, the backup N-PE becomes active immediately, and inserts the green B semaphore to prevent the primary N-PE from running in an active-active state.

In Figure 26, when the primary N-PE state comes back ready, it is still receiving the green B semaphore, and therefore, it stays in standby mode. In the meantime, it raises the red P signal destined for the backup node as a request to be ready to become active. A tunable probing timer is then started, to verify primary node stability. When this delay has expired, the backup node runs in standby mode and shuts down its B semaphores, allowing the primary node to become fully active.

Figure 26. Primary N-PE Is Active Again After a Probing Delay



The secondary node is now ready for the next backup.

VPLS and A-VPLS Loop Prevention

A-VPLS provides native loop prevention using the following two features:

VPLS and A-VPLS over VSS: The two VSS devices appear as one single device to the directed Label Distribution Protocol (LDP) peer and use dual-homing support similar to MEC. This eliminates the need for Spanning Tree Protocol and Cisco IOS EEM mechanisms to track availability or redundant devices as described earlier.

Split horizon: VPLS provides split horizon to help ensure that packets received from remote provider-edge devices are not being forwarded back through the same learning interface.

Conclusion

Cisco recommends retaining Layer 2 domains within each data center. However, to meet new application framework requirements or for migration purposes, an enterprise may have to extend Layer 2 beyond the data center. If there is no Layer 3 alternative, Cisco recommends the three approaches listed here. These approaches allow the safeguarding of sturdy extended Layer 2 networks, while maintaining multipath redundancy with fast convergence and high performance. All these solutions are executed in hardware, and enterprises can deploy different classes of service (CoS) by application.

These recommendations address the concerns brought by the use of Spanning Tree Protocol beyond the data center such as a partial or total disruption of the Layer 2 forwarding links on any extended part of the bridging domain or very poor recovery time in case of failover.

The Cisco enterprise Layer 2 extension solutions are:

- Cisco Catalyst 6500 VSS 1440, Cisco Catalyst 6500 SUP2T and Cisco Nexus 7000 Series vPC over MAN distances using dedicated optical fiber links with very high throughput
- EoMPLS on Cisco Catalyst 6500 Series and Cisco ASR 1000 Series or VPLS and A-VPLS on Cisco Catalyst 6500 Series with an MPLS core over long distances
- EoMPLSoGRE on Cisco Catalyst 6500 Series and Cisco ASR 1000 Series and VPLSoGRE and A-VPLSoGRE on Cisco Catalyst 6500 Series with an IP core over long distances

A-VPLS and A-VPLSoGRE on the Catalyst 6500 are new enhancements that provide native dual-homing, and a fully redundant and highly scalable solution to deploy Layer 2 Interconnects over an MPLS or IP core. The new virtual Ethernet CLI makes deployment of A-VPLS enterprise-friendly.

For more Information on VPLS on SUP2T please refer to the SUP2T VPLS White Paper.

To isolate Spanning Tree Protocol between remote data centers, Cisco recommends:

- BPDU filtering on specific interfaces.
- Cisco IOS EEM in the N-PE (edge device facing the core) with semaphores to control and validate the states of all components of the Layer 2 architecture to improve the flexibility and scalability of the end-to-end design.
- Using VPLS or A-VPLS, which blocks STP from going across, and uses Split Horizon to prevent loops

For More Information

Visit the Cisco DCI page.http://www.cisco.com/go/dci

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-493718-05 07/11