Cisco Catalyst 6500 Supervisor Engine 2T: Layer 2 Scalability Enhancements

White Paper

May 09, 2011

Overview

Network evolution in the recent past and foreseeable future involves more and more virtualization, whether in the campus or data center. Borderless networks and cloud-computing networks that promise service availability anytime, from anywhere, demand hardware that can scale to support such an architecture.

This white paper explains how the Cisco® Catalyst® 6500 Series Supervisor Engine 2T with Policy Feature Card 4 (PFC4) provides Layer 2 scalability enhancements to support these environments. Specifically, the following aspects of Layer 2 scalability will be covered:

- MAC addresses
- Layer 2 interfaces
- EtherChannel hashing
- Bridge domains
- Virtual Private LAN Service (VPLS)

Note: Throughout this paper, the term "Supervisor Engine 2T" will be used. When this term is used in reference to a feature or functionality, the reader should interpret that to mean that the feature or functionality is supported by the Supervisor Engine 2T with PFC4 as well as by any line cards with Distributed Feature Card 4 (DFC4). When the terms "Supervisor Engine 32," Supervisor Engine 720," or "Supervisor Engine 32 and Supervisor Engine 720" are used, the reader should interpret that to mean features or functionality supported by PFC3 and DFC3.
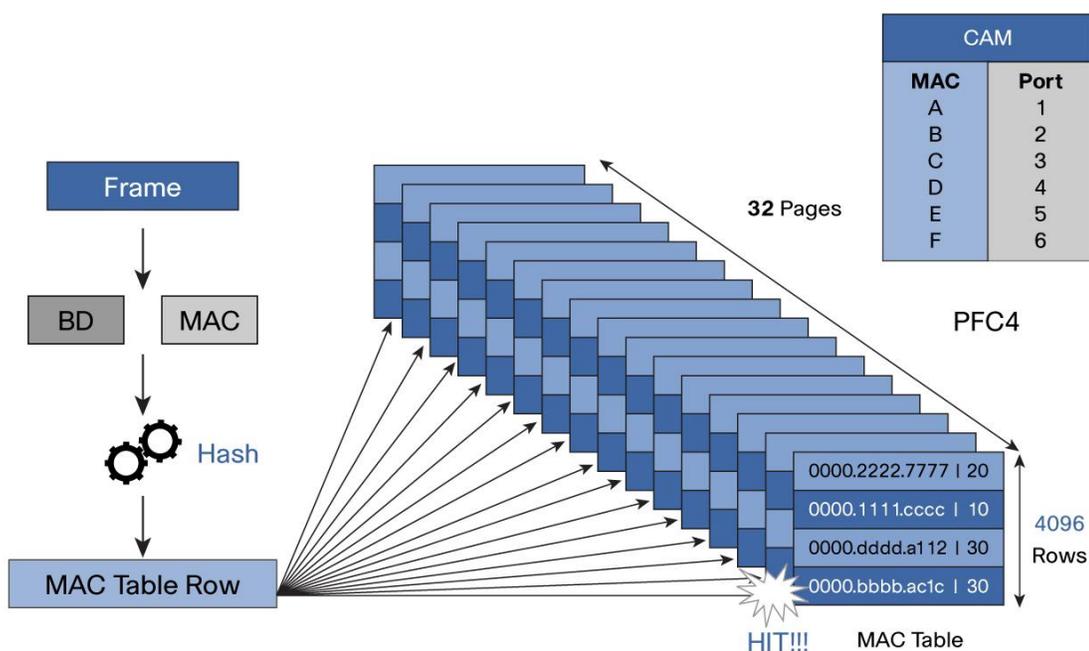
Contents

MAC Address Scalability

As the number of virtual devices increases in network architectures, the number of MAC addresses supported by the network hardware must also increase. The Cisco Catalyst 6500 Series Supervisor Engine 2T provides two major enhancements in the area of MAC address scalability: increased hash efficiency and larger table size.

Increased Hash Efficiency

Figure 1 describes the Layer 2 forwarding operation of the Policy Feature Card 4 (PFC4) on the Supervisor Engine 2T. This operation also applies to any Distributed Forwarding Card 4 (DFC4) equipped line cards.

Figure 1.   Layer 2 Forwarding Operation of PFC4 on Cisco Catalyst 6500 Series Supervisor Engine 2T

1. Hash result identifies the starting Page and Row in MAC table.
2. Lookup key (BD+MAC) compared to contents of indexed line on each page, sequentially.
3. Destination lookup: Match returns destination interface(s), Miss results in Flood.
4. Source lookup: Match updates age of matching entry, Miss installs new entry in table.

Step 1 of the Layer 2 forwarding operation performed by a PFC4 or DFC4 is to use the bridge domain (BD) and MAC address of the incoming frame as inputs into a hash function, which identifies the starting page and row in the MAC address table. A hash function is any well-defined procedure or mathematical function that converts a large, variable-sized amount of data into a small piece of data that serves as an index. Hash functions are mostly used to speed up table lookups, such as finding items in a database. The downside of hash functions is that they may map two or more sets of inputs to the same index, resulting in a collision. In this case, a collision means that a MAC address cannot be learned by the PFC4 or DFC4 performing the Layer 2 forwarding operation.

The more efficient a hash function is, the more MAC address table entries it can fill, and the fewer collisions it produces. The hash function used by the PFC4 and DFC4 is 99 percent efficient, meaning that the system can populate 99 percent of the MAC address table, at a minimum (the system can populate 100 percent of the MAC address table if there are no collisions). If a collision does occur and a MAC address cannot be learned, the system floods in hardware (see Step 3 in Figure 1). All Layer 2 operations are hardware based, so no CPU impact is realized in this case.

While previous PFCs and DFCs for the Cisco Catalyst 6500 Series Switch support Layer 2 forwarding operations in hardware just as the PFC4 and DFC4 do, the hash functions used by the PFC4 and DFC4 have increased in efficiency to 99 percent. Table 1 shows the changes in the hash efficiency for the different generations of PFCs and DFCs for the Cisco Catalyst 6500 Series Switch.

Table 1.   Hash Efficiency of PFCs and DFCs for Cisco Catalyst 6500 Series Switch

PFC/DFC Generation	Maximum MAC Table Size	Hash Efficiency	Guaranteed MAC Table Population
PFC2/DFC	128K	25%	32K
PFC3A/3B/3BXL DFC3A/3B/3BXL	64K	50%	32K
PFC3C/3CXL DFC3C/3CXL	96K	90%	86K
PFC4/4XL DFC4/4XL	**128K**	**99%**	**127K**

NOTE: 1 K = 1024

By supporting a more efficient hash function, the Supervisor Engine 2T with PFC4 provides a more efficient infrastructure that will prevent unwanted traffic on the network. This is critical for virtualized infrastructures that require more efficient link utilization because the deployment of server virtualization technologies increases the number of services supported on a single link.

Larger Table Size

Table 1 showed the evolution in MAC address table size for the various generations of PFCs and DFCs. The biggest change occurred from PFC2 to PFC3, when the MAC address table was moved from a standalone chip on the PFC2 into the Layer 2 forwarding chip on the PFC3. Over time, the size of the MAC address table has gradually increased, and just as importantly the hash efficiency has increased as well.

The question that often arises from a discussion of this topic is: Why are such large MAC address tables needed in enterprise networks? The answer, as Figure 2 illustrates, is that more and more enterprise networks are using virtualization to a greater extent.
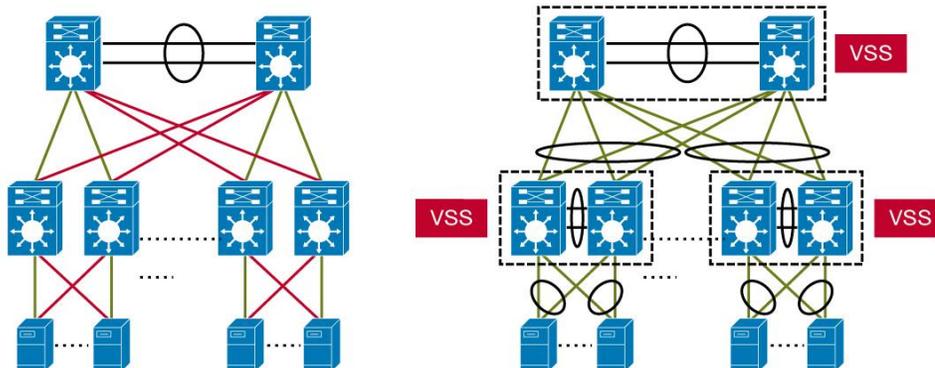
Figure 2.   MAC Address Scalability Requirements

Legacy Compute Environment
1 Active Uplink per Switch
1 Active NIC per Server
0 VMs

Virtualized Compute Environment
2 Active Uplinks per Switch
2 Active NICs per Server
Up to 42,666 VMs with 3 MACs each

In the legacy compute environment, one physical server runs one application and has a single active connection to the access layer. At the aggregation layer, each switch has only half of its interfaces active (most likely 10 Gigabit Ethernet in a data center architecture) as a result of Hot Standby Router Protocol (HSRP) and Spanning Tree Protocol (STP) operations. Since there are no virtual machines (VMs) running in this environment, the maximum number of MAC addresses required at the distribution layer is roughly:

> Maximum 10G density = 130 ports (6509-E with 8 x WS-X6716-10G-3C/3CXL)
>
> Maximum 10G Density density with HSRP/STP operations: 65
>
> Maximum 1G density at access layer = 384 ports (8 x WS-X6748-GE-TX)
>
> Number of MAC addresses per 1G host = 1
>
> Maximum number of MAC addresses at distribution = (65 * 384 * 1) = 24,576

In the virtualized compute environment, each physical server runs multiple VMs and has multiple active connections to the access layer because of the Cisco Catalyst 6500 Series Virtual Switching System (VSS). At the distribution layer, each switch has all of its interfaces active as a result of VSS. The maximum number of MAC addresses dictates the total number of VMs that can be supported in a pod connecting to a VSS. With a table size of 128 K, the PFC4 can support up to 42,666 VMs since each VM has three MACs each - one for each network interface card (NIC) and one for server migration (for example, through VMware vMotion). This is an increase of over 10,000 VMs compared to what the PFC3 could support (32,000).

The ability to support a higher number of VMs also means that a Supervisor Engine 2T-based network does require as many network elements or servers. This leads to lower power consumption, a smaller footprint, and lower cooling costs, allowing an organization to more easily meet their energy efficiency and green policies.

Layer 2 Interfaces

The Supervisor Engine 2T has an integrated 2-Tbps switch fabric on the baseboard of the supervisor. This switch fabric is comprised of 26 fabric channels that are distributed among the slots of the various Cisco Catalyst 6500-E Series chassis. Since there are 26 fabric channels, every slot in every chassis will receive two fabric channels that can be used by the line cards placed in those slots. The previous switch fabrics - whether on dedicated line cards as was the case with the Supervisor 2 or on the baseboard, as is the case with Supervisor Engine 720 (all versions) - contain 18 fabric channels. This allowed all chassis except the 6513 to receive dual fabric channels in all slots. Figure 3 shows the fabric channel distribution in the Cisco Catalyst 6513 Switch chassis with an 18-channel switch fabric.

**Figure 3.**   Cisco Catalyst 6513 Switch Fabric Channel Distribution
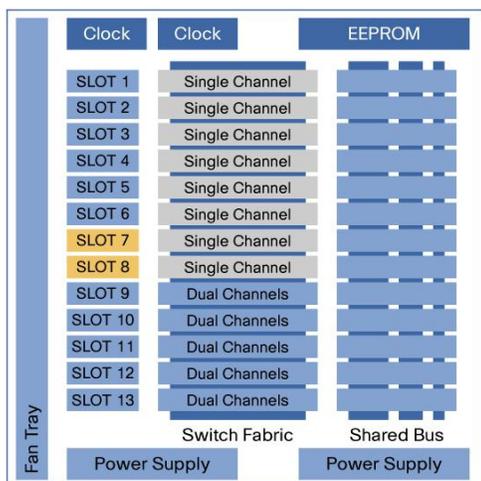
Notice that slots 1-8 support single fabric channels while slots 9-13 support dual-fabric channels. This means that any line card that requires dual fabric channels, such as all 6700 line cards (except 6724), are limited to slots 9-13. This restriction reduces the density of high-performance interfaces in the 6513 chassis, making the 6509-E chassis a better option when the highest density of high-performance interfaces is required (see Table 2).

Table 2.   High-Performance Interface Density in the Cisco Catalyst 6509-E Switch Chassis Versus the 6513 Switch Chassis

Interface Type	6509-E Density	6513 Density
48-port 10/100/1000 WS-X6748-GE-TX	384	240
48-port GE SFP WS-X6748-SFP	384	240
4-port 10G Fiber WS-X6704-10GE	32	20
8-port 10G Fiber WS-X6708-10G-3C	64	40
16-port 10G Fiber WS-X6716-10G-3C	128	80
16-port 10G Copper WS-X6716-10G-3C	128	80

In June 2010, the Cisco Catalyst 6513-E Switch chassis was introduced. This chassis increases the density of Layer 2 interfaces when using a Supervisor Engine 2T by providing dual-fabric channels for all 13 slots of the chassis. Figure 4 shows the fabric channel distribution of the 6513-E chassis.

Figure 4.   Cisco Catalyst 6513-E Switch Fabric Channel Distribution
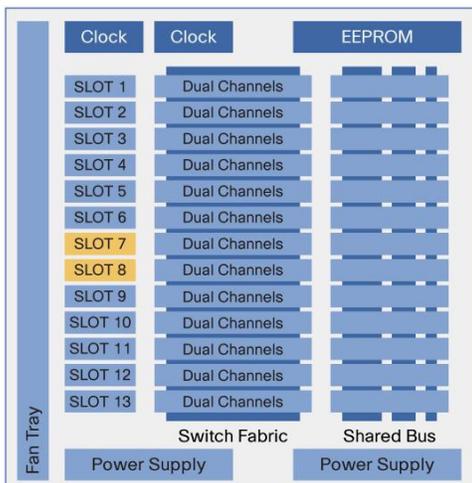
In order to take advantage of the dual fabric channels in slots 1-8 of the 6513-E chassis, a Supervisor Engine 2T must be used since it has the 8 extra fabric channels for these slots. If a Supervisor Engine 720 (any version) is used, the fabric channel pattern in the 6513-E chassis will be identical to that of the 6513 chassis as seen in Figure 3.

Since the Supervisor Engine 2T can support high-speed interfaces in all line card slots of the 6513-E chassis, the maximum density that can be achieved for these types of interfaces in a single chassis increases by up to 120%. Table 3 shows the difference between densities with the Supervisor Engine 720 with the 6509-E and 6513 chassis and the Supervisor Engine 2T with 6513-E chassis.

Table 3.   High-Performance Interface Density: Cisco Catalyst 6509-E and 6513 with Switch Supervisor 720 Versus Cisco Catalyst 6513-E with Switch Supervisor 2T

Interface Type	Supervisor 720 6509-E Density	Supervisor 720 6513 Density	Supervisor 2T 6513-E Density	% Increase (6509-E/6513)
48-port 10/100/1000 WS-X6748-GE-TX	384	240	528	**+37.5%/+120%**
48-port GE SFP WS-X6748-SFP	384	240	528	**+37.5%/+120%**
4-port 10G Fiber WS-X6704-10GE	32	20	44	**+37.5%/+120%**
16-port 10G Fiber WS-X6716-10G-3C	128	80	176	**+37.5%/+120%**
16-port 10G Copper WS-X6716-10G-3C	128	80	176	**+37.5%/+120%**
8-port 10G Fiber (1:1) WS-X6908-10G	N/A	N/A	88	N/A
4-port 40G Fiber WS-X6904-40G-CFP	N/A	N/A	44	N/A

NOTE: When using Supervisor Engine 2T with the 6513-E chassis, only the Supervisor Engine 2T can placed in the supervisor slots (7 and 8). This is why all of the Supervisor Engine 2T density numbers are multiples of 11. This is the only supervisor and chassis combination that has this restriction.

The Supervisor Engine 2T supports Virtual Switching System (VSS) mode, which doubles the number of interfaces, as shown in Table 3.

EtherChannel Hashing

The Cisco Catalyst 6500 Series Switch supports the ability to form EtherChannel bundles, which are single logical links that provide the aggregate bandwidth of up to eight physical links. For example, four 10 Gigabit Ethernet physical links can be bundled together to form a single 40 Gigabit Ethernet logical link, as shown in Figure 5.

Figure 5.   EtherChannel Concept
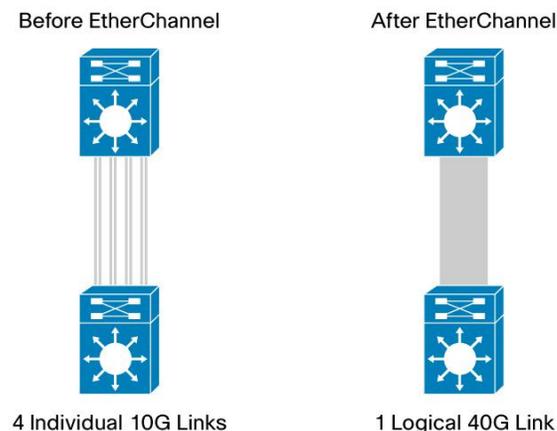
Even though the system sees a single logical link after creating the EtherChannel, the logical link is still made up of individual physical links. When making a forwarding decision, the system uses a hash function to determine which link in the EtherChannel a flow will take. (The Cisco Catalyst 6500 Series Switch does not support per-packet load balancing.) The inputs to this hash function are determined by the user, with the default being the source IP and destination IP of the flow (for a complete list see the **port-channel load-balance** command). The output from the hash function is a hex string that matches a single bit out of an 8-bit string.

The Supervisor Engine 2T enhancement to EtherChannel involves the number of results that are possible from the hash algorithm calculation. In supervisors prior to the Supervisor Engine 2T, the hash function gives a 3-bit result, meaning that one of eight results is possible with each calculation. With the Supervisor Engine 2T, the hash function supports an 8-bit result, meaning that one of 256 results is possible with each calculation. This means that with the Supervisor Engine 2T, flows can be more evenly distributed among the links in an EtherChannel since there is a larger pool of results when doing the hash. Tables 4 and 5 illustrate how increasing the number of results can achieve more effective link utilization.

Table 4. EtherChannel Hash Results per Link prior to Supervisor Engine 2T

Number of Links in the EtherChannel	Number of Results per link
1	8
2	4
3	2 links have 3 results 1 link has 2 results
4	2
5	3 links have 2 results 2 links have 1 result
6	2 links have 2 results 4 links have 1 result
7	1 link has 2 results 6 links have 1 result
8	1

Table 5. EtherChannel Hash Results per Link with Supervisor Engine 2T

Number of Links in the EtherChannel	Number of Results per link
1	256
2	128
3	1 link has 86 results 2 links have 85 results
4	64
5	1 link has 52 results 4 links have 51 results
6	4 links have 43 results 2 links have 42 results
7	4 links have 37 results 3 links have 36 results
8	32

By supporting a larger number of results per link of an EtherChannel, the Supervisor Engine 2T provides a more efficient utilization of the links in an EtherChannel, especially for EtherChannels whose number of links is not a power of 2. In a virtualized environment where the numbers of IPs, MACs, and Layer 4 ports is likely to be highly variable, the ability to utilize a larger number of hash results increases the likelihood of flows being more evenly distributed among the links of an EtherChannel.
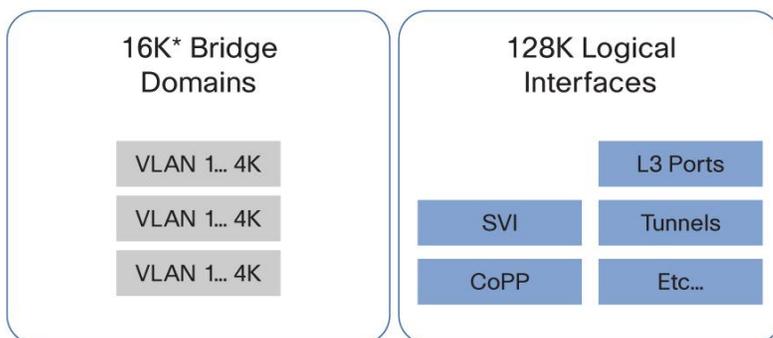
Bridge Domains

Two new concepts are being introduced on the Cisco Catalyst 6500 Series Switch with Supervisor Engine 2T: bridge domains (BDs) and logical interfaces (LIFs). Bridge domains are used to represent Layer 2 VLANs while logical interfaces are used to represent Layer 3 interfaces such as switched virtual interfaces (SVIs), tunnel interfaces, router ports, and others.

Prior to the introduction of Supervisor Engine 2T, VLANs were used internally by the system to represent not only Layer 2 VLANs but also Layer 3 interfaces. Figures 6 and 7 show the differences between Supervisor Engine 2T, Supervisor Engine 32, and Supervisor Engine 720.

Figure 6.　Supervisor Engine 32 and Supervisor Engine 720 VLAN Usage

Figure 7.　Supervisor Engine 2T VLAN Usage

[*] 4-K bridge domains will be supported in the first release of software.

By separating Layer 2 and Layer 3 resources, the Supervisor Engine 2T can provide improved scalability for both types of interfaces. Whereas the total pool for Layer 2 and Layer 3 interfaces was 4096 with Supervisor Engine 32 and Supervisor Engine 720, the pools for Layer 2 and Layer 3 for the Supervisor Engine 2T are 16,384 (4096 supported in first software release) and 131,072, respectively.

In addition to increasing the numbers of available interfaces, the Supervisor Engine 2T bridge domain feature allows for VLAN reuse on different interfaces. For example, VLAN 10 can be configured on two different physical interfaces and belong to two separate bridge domains. This is possible because with Supervisor Engine 2T, a VLAN is significant at the port level, whereas with Supervisor Engine 32 and Supervisor Engine 720 a VLAN was globally significant.

A single port can support up to 4096 VLANs, and the total number of port - VLAN combinations is bounded only by the available BD pool. For example, a system can be configured with 40 ports each and have the same 102 VLANs, but each will belong to different bridge domain.

Virtual Private LAN Service (VPLS)

The increase in adoption of server virtualization technologies is changing the way data center architectures are being built. In the past, it was not recommended that VLANs be spread across data centers, but that recommendation is changing with the adoption of server migration capabilities that require Layer 2 connectivity between multipoint, geographically separated locations.

To support these requirements, the Cisco Catalyst 6500 Series Switch supports virtual private LAN services (VPLS). VPLS is a technology that provides Ethernet-based, multipoint-to-multipoint connectivity over an IP or Multiprotocol Label Switching (MPLS) backbone.

Figure 8 shows an example of a VPLS architecture. The customer edge (CE) device connects to a provider edge (PE) device that then establishes pseudowires across the backbone to other PE devices. The pseudowires allow the geographically dispersed data centers to be part of the same Ethernet broadcast domain even if the backbone of the architecture is a Layer 3 backbone. This supports the ability for servers to move between locations if needed since the servers see Layer 2 connectivity between all sites.

Figure 8. VPLS Architecture
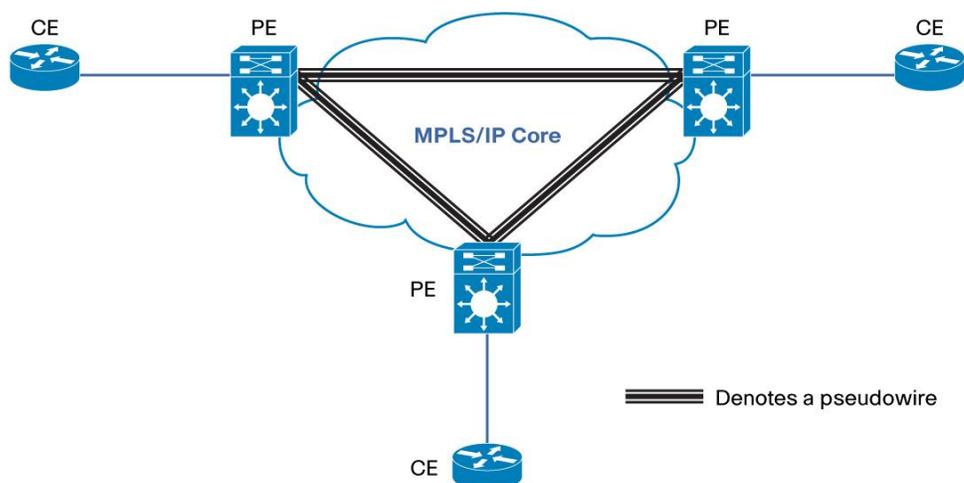
The Supervisor Engine 2T will greatly increase the number of pseudowires that a Cisco Catalyst 6500 Series system can support when compared to prior Supervisors. In Cisco IOS® Software Release 12.2(50)SY, the first version of code to be supported with Supervisor Engine 2T, up to 4096 VPLS instances are supported compared to 2048 instances for previous supervisors. Table 6 outlines other enhancements introduced with Supervisor Engine 2T.

Table 6. VPLS Enhancements with Supervisor Engine 2T PFC4 and DFC4

Feature	Supervisor 32/720 (PFC3/DFC3)	Supervisor 2T (PFC4/DFC4)
Number of VPLS Instances	2048	4096*
VPLS Core Facing Interfaces	Requires SIP/SPA	Any Ethernet Interface
H-VPLS Core Facing Interfaces	Requires SIP/SPA	Any Ethernet Interface

* 16384 with future code enhancements

In addition to increasing the number of VPLS instances, the Supervisor Engine 2T allows any Ethernet interface in the system to be used as a core facing interface for a VPLS, or hierarchical VPLS (H-VPLS), network. With previous Supervisor Engine 32- and Supervisor Engine 720-based systems, a shared port adapter (SPA) interface processor (SIP) was required to perform this function. By removing this hurdle, VPLS networks can be more easily deployed and at a lower cost since specialized modules are no longer required.

For additional information about VPLS and H-VPLS with the Cisco Catalyst 6500 Series Switch and Supervisor Engine 2T, see *Catalyst 6500 Supervisor Engine 2T VPLS Guide* on Cisco.com.

Summary

This document shows how the Supervisor Engine 2T with Policy Feature Card 4 (PFC4) and line cards with the Distributed Feature Card 4 (DFC4) support L2 Scalability Enhancements to support virtualized Layer 2 environments. Increases in MAC Address table size, Layer 2 Interface density, and Virtual Private LAN Service instances along with more efficient EtherChannel Hashing and the introduction of Bridge Domains make the Cisco Catalyst 6500 Series Switch with Supervisor Engine 2T an ideal choice for deployment in networks where virtualization is required.

Printed in USA   C11-663636-01   09/11