CCNP SWITCH -Chapter 11 Multilayer Switching - Cisco Express Forwarding

inter-VLAN routing. Catalyst 6500 Multilayer Switch
"router-on-a-stick" aka "one-armed router"
Switch(ed) virtual interface (SVI) - assigning Layer 3 address to a logical interface that represents an entire VLAN

An interface is either in Layer 2 or Layer 3 mode
If the switchport line is shown as enabled, the port is in Layer 2 mode; disabled, is Layer 3 mode:

```
Switch# show interface gigabitethernet 1/0/1 switchport
    Name: Gi1/0/1
    Switchport: Disabled
```

Need Layer 2 functionality? Run the **switchport** command - then configure trunking, access VLANs, etc.
Need Layer 3 interfaces? Run **no switchport** and assign an IP, etc..
For an EtherChannel assign an IP to the port-channel interface- not the links within the channel.

By default is enough TCAM space to perform Layer 3 operation for IPv4. Need IPv6 also?
Reconfigure the SDM template with **sdm prefer dual-ipv4-and-ipv6**

SVI Port Configuration - Making a VLAN itself a L3 interface (SVI)
Enable Layer 3 functionality for an entire VLAN so a network address to be assigned to a logical interface- the VLAN itself. When routing is needed in and out of that VLAN on several ports helps with bandwidth.

```
Switch(config)# interface vlan 10
Switch(config-if)# ip address 10.10.10.1 255.255.255.0
```

The SVI itself has no physical connection - VLAN 10 must extend through a Layer 2 port or trunk to the outside.
Ensure that the new VLAN interface is enabled with **no shutdown**
Creating or configuring the SVI does not create or configure the VLAN; you still must define each independently.

```
Switch(config)# vlan 100
Switch(config-vlan)# name Example_VLAN
Switch(config-vlan)# exit
Switch(config)# interface vlan 100
Switch(config-if)# ip address 192.168.100.1 255.255.255.0
Switch(config-if)# no shutdown
```

An SVI cannot become active until STP has converged with an active link it arrives on.
- SVI **autostate** automatically keeps the SVI down until the VLAN is ready, so no switching or routing can try using the SVI prematurely.
- If you want the SVI to stay up even when no Layer 2 ports are active on the VLAN (such as a port configured for port mirroring to capture traffic) you can run **switchport autostate exclude**

Cisco Express Forwarding (CEF) - Multilayer Switching
- forward packets based on Layer 3 and Layer 4 information

Traditional Multilayer switching (MLS)- NetFlow switching aka route cache switching
- began as a dual effort between a route processor (RP) and a NetFlow-capable switching engine (SE).
- "route once and switch many."
- RP receives the first packet of a new traffic flow between two hosts, routing decision is made, packet forwarded
- SE must know the identity of each RP. The SE then can listen in to the first packet going to the router and also going away from the router.
- If SE can switch the packet in both directions, it can learn a "shortcut path" so that subsequent packets of the same flow can be switched directly to the destination port without passing through the RP.
- legacy Catalyst 6000 Supervisor 1/1a and Multilayer Switch Feature Card (MSFC), Catalyst 5500 with a Route Switch Module (RSM), Route Switch Feature Card (RSFC), or external router.

CEF Overview
Cisco Express Forwarding more efficient form - packet forwarding through the use of dynamic lookup tables. CEF runs by default on the Catalyst switching platforms, specialized hardware.

A CEF-based multilayer switch consists of two basic functional blocks- 1) the Layer 3 engine is involved in building routing information that the 2) Layer 3 forwarding engine can use to switch packets in hardware.

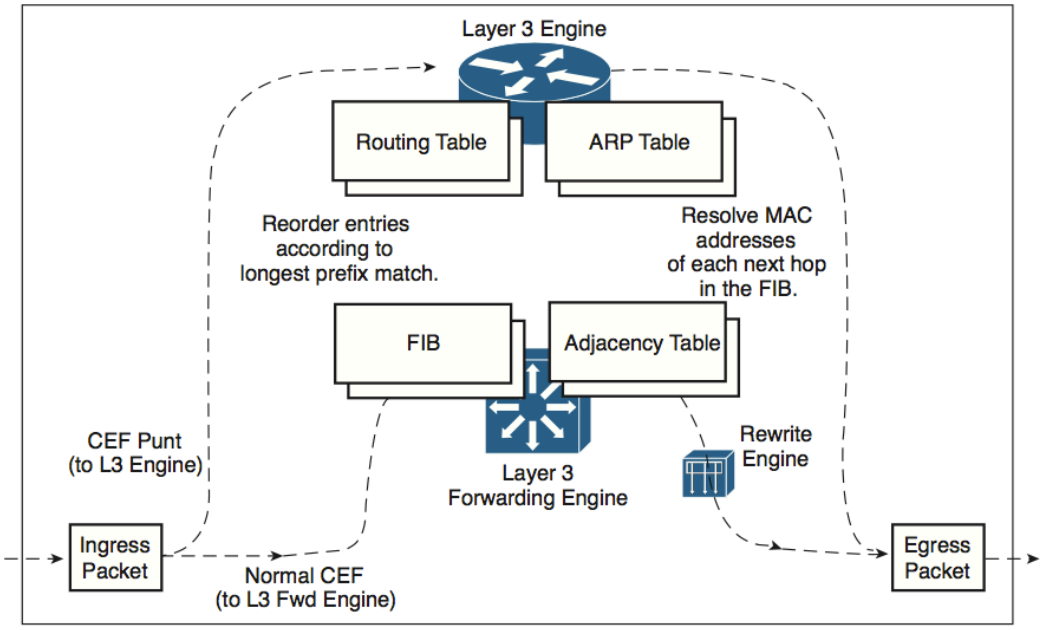

Figure 11-3 *Packet Flow Through a CEF-Based Multilayer Switch*

Forwarding Information Base

The Layer 3 engine (basically a router) maintains routing info, (static routes or dynamic routing protocols).
- Routing table is reformatted into an ordered list (most specific route first), for each IP destination subnet
- This is the Forwarding Information Base (FIB) - routing or forwarding info the network prefix can reference.

A route to 10.1.0.0/16 might be in the FIB along with routes to 10.1.1.0/24 and 10.1.1.128/25
- Notice that these examples are increasingly more specific subnets, as designated by longer subnet masks.
- In the FIB, these would be ordered with the most specific match first, followed by less specific subnets.
- When the switch receives a packet, it can examine the destination and find the longest-match destination

Each entry also has the next-hop address in the FIB so it is found simultaneously.
FIB also contains host route (subnet mask 255.255.255.255) entries without them being advertised or manually configured- maintained for efficient routing lookup to directly connected or adjacent hosts.

FIB is dynamic -when the Layer 3 engine sees any change in the routing topology, it updates the FIB
- If the routing table gets a change to a route prefix or next-hop address, the FIB updates
- If a next-hop address is changed or aged out of the ARP table, the FIB must reflect the same change.

You can display FIB table entries related to a specific interface or VLAN with **show ip cef**
Switch# **show ip cef** [*type member/module/number* | **vlan** *vlan-id*] [**detail**]

Example 11-1 *Displaying FIB Table Entries for a Specified VLAN*

```
Switch# show ip cef vlan 101
Prefix              Next Hop            Interface
10.1.1.0/24         attached            Vlan101
10.1.1.2/32         10.1.1.2            Vlan101
10.1.1.3/32         10.1.1.3            Vlan101
Switch#
```

You also can view FIB entries by specifying an IP prefix address and mask
Switch# **show ip cef** [*prefix-ip prefix-mask*] [**longer-prefixes**] [**detail**]
- Normally, only an exact match of the IP prefix and mask will be displayed
- To see other longer match entries, you can add the **longer-prefixes** keyword.

- This displays any subnet within 10.1.0.0/16 that is known, regardless of the prefix or mask length

Example 11-2 *Displaying FIB Table Entries for a Specified IP Prefix Address/Mask*

```
Switch# show ip cef 10.1.0.0 255.255.0.0 longer-prefixes
Prefix              Next Hop            Interface
10.1.1.0/24         attached            Vlan101
10.1.1.2/32         10.1.1.2            Vlan101
10.1.1.3/32         10.1.1.3            Vlan101
10.1.2.0/24         attached            Vlan102
10.1.3.0/26         192.168.1.2         Vlan99
                    192.168.1.3         Vlan99
10.1.3.64/26        192.168.1.2         Vlan99
                    192.168.1.3         Vlan99
10.1.3.128/26       192.168.1.4         Vlan99
                    192.168.1.3         Vlan99
```

Add the **detail** keyword to see more information

Example 11-3 *Displaying Detailed CEF Entry Information*

```
Switch# show ip cef 10.1.3.0 255.255.255.192 detail
10.1.3.0/26, version 270, epoch 0, per-destination sharing
0 packets, 0 bytes
  via 192.168.1.2, Vlan99, 0 dependencies
    traffic share 1
    next hop 192.168.1.2, Vlan99
    valid adjacency
  via 192.168.1.3, Vlan99, 0 dependencies
    traffic share 1
    next hop 192.168.1.3, Vlan99
    valid adjacency
  0 packets, 0 bytes switched through the prefix
  tmstats: external 0 packets, 0 bytes
           internal 0 packets, 0 bytes
Switch#
```

- Version number represents the number of times the entry has been updated since the table was generated
- Epoch number is the number of times the whole CEF table has been flushed and regenerated
- The 10.1.3.0/26 subnet has two next-hop addresses, so the local switch is using per-destination load sharing between the two routers.

[The previous diagram of packet flow through the switch is referenced here]. After the FIB is built, packets can be forwarded along the bottom dashed path. This hardware switching process where time- consuming operations are avoided by design. Sometimes the FIB can't be used (see below), so packets are marked as “CEF punt” and sent to the Layer 3 engine for further processing, as shown in the top dashed path This happens when:

- No match found in the FIB, or the FIB table is full.
- Ether the IP TTL has expired or the MTU is exceeded (so the packet must be fragmented)
- An ICMP redirect is involved.
- The encapsulation type is not supported.
- Packets are tunneled, requiring a compression or encryption operation.
- An access list with the **log** option is triggered or NAT operations must be performed.

CEF also can be optimized through the use of specialized forwarding hardware, using the following techniques:

- Accelerated CEF (aCEF): CEF is distributed across multiple Layer 3 forwarding engines, typically located on individual line cards in chassis-based Catalyst switches. These engines do not have the capability to store and use the entire FIB, so only a portion of the FIB is downloaded to them at any time. This functions as an FIB

“cache,” containing entries that are likely to be used again. If FIB entries are not found in the cache, requests are sent to the Layer 3 engine for more FIB information. The net result is that CEF is accelerated on the line cards, but not necessarily at a sustained wire-speed rate.

Distributed CEF (dCEF): CEF can be distributed completely among multiple Layer 3 forwarding engines for even greater performance. Because the FIB is self-contained for complete Layer 3 forwarding, it can be replicated across any number of independent Layer 3 forwarding engines. For example, the Catalyst 6500 has line cards that support dCEF, each with its own FIB table and forwarding engine. A central Layer 3 engine maintains the routing table and generates the FIB, which is then dynamically downloaded in full to each of the line cards.

Adjacency Table (holds MAC addresses living "next door")

A router normally maintains a routing table and an ARP table.

The FIB keeps the L3 next-hop address for each entry, but also corresponding L2 information for every next-hop entry. This part of the FIB is called the adjacency table, holding MAC addresses a single L2 hop away.

Switch# **show adjacency** [*type member/module/number* | **vlan** *vlan-id*] [**summary** | **detail**]

Total number of adjacencies known can be displayed with **show adjacency summary** command:

Example 11-4 *Displaying the Total Number of Known Adjacencies*

```
Switch# show adjacency summary
Adjacency Table has 106 adjacencies
  Table epoch: 0 (106 entries at this epoch)
  Interface                 Adjacency Count
  Vlan99                    21
  Vlan101                   3
  Vlan102                   1
  Vlan103                   47
  Vlan104                   7
  Vlan105                   27
Switch#
```

Example 11-5 *Displaying Detailed Information About Adjacencies*

```
Switch# show adjacency vlan 99 detail
Protocol Interface                 Address
IP       Vlan99                    192.168.1.2(5)
                                   0 packets, 0 bytes
                                   epoch 0
                                   sourced in sev-epoch 0
                                   Encap length 14
                                   000A5E45B145000E387D51000800
                                   L2 destination address byte offset 0
                                   L2 destination address byte length 6
                                   Link-type after encap: ip
                                   ARP
IP       Vlan99                    192.168.1.3(5)
                                   1 packets, 104 bytes
                                   L2 destination address byte offset 0
                                   L2 destination address byte length 6
                                   Link-type after encap: ip
                                   ARP
                                   000CF1C909A0000E387D51000800
                                   L2 destination address byte offset 0
                                   L2 destination address byte length 6
                                   Link-type after encap: ip
                                   ARP
```

The MAC address could be shown as the first six octets of the long string of hex or on a line by itself. The remainder of the string of hex digits contains the MAC address of the Layer 3 engine's interface (six octets, corresponding to the Vlan99 interface in the example) and the EtherType value (two octets, where 0800 denotes IP).
The adjacency table information is built from the ARP table. Example 11-5 shows adja- cency with the age of its ARP entry. As a next-hop address receives a valid ARP entry, the adjacency table is updated. If an ARP entry does not exist, the FIB entry is marked as "CEF glean." This means that the Layer 3 forwarding engine cannot forward the packet
in hardware because of the missing Layer 2 next-hop address. The packet is sent to the Layer 3 engine so that it can generate an ARP request and receive an ARP reply. This is known as the CEF glean state, in which the Layer 3 engine must glean the next-hop des- tination's MAC address.
The glean state can be demonstrated in several ways, as demonstrated in Example 11-6.

Example 11-6 *Displaying Adjacencies in the CEF Glean State*

```
Switch# show ip cef adjacency glean
Prefix              Next Hop            Interface
10.1.1.2/32         attached            Vlan101
127.0.0.0/8         attached            EOBC0/0
[output omitted]
Switch# show ip arp 10.1.1.2
Switch# show ip cef 10.1.1.2 255.255.255.255 detail
10.1.1.2/32, version 688, epoch 0, attached, connected
0 packets, 0 bytes
  via Vlan101, 0 dependencies
    valid glean adjacency
Switch#
```

Notice that the FIB entry for directly connected host 10.1.1.2/32 is present but listed in the glean state. The **show ip arp** command shows that there is no valid ARP entry for the IP address.
During the time that an FIB entry is in the CEF glean state waiting for the ARP resolu- tion, subsequent packets to that host are immediately dropped so that the input queues do not fill and the Layer 3 engine does not become too busy worrying about the need for duplicate ARP requests. This is called ARP throttling or throttling adjacency. If an ARP reply is not received in 2 seconds, the throttling is released so that another ARP request can be triggered. Otherwise, after an ARP reply is received, the throttling is released, the FIB entry can be completed, and packets can be forwarded completely in hardware.
The adjacency table also can contain other types of entries so that packets can be han- dled efficiently. For example, you might see the following adjacency types listed:
■ **Null adjacency:** Used to switch packets destined for the null interface. The null interface always is defined on a router or switch; it represents a logical interface that silently absorbs packets without actually forwarding them.

■ **Drop adjacency:** Used to switch packets that cannot be forwarded normally. In effect, these packets are dropped without being forwarded. Packets can be dropped because of an encapsulation failure, an unresolved address, an unsupported proto- col, no valid route present, no valid adjacency, or a checksum error. You can gauge drop adjacency activity with the following command:

```
Switch# show cef drop
CEF Drop Statistics
Slot  Encap_fail  Unresolved Unsupported    No_route     No_adj  ChkSum_Err
RP      8799327           1       45827     5089667         32           0
Switch#
```

- **Discard adjacency:** Used when packets must be discarded because of an access list or other policy action.
  Punt adjacency: Used when packets must be sent to the Layer 3 engine for further processing. You can gauge the CEF punt activity by looking at the various punt adja- cency reasons listed by the **show cef not-cef-switched** command:

```
Switch# show cef not-cef-switched
CEF Packets passed on to next switching layer
Slot  No_adj  No_encap  Unsupp'ted  Redirect  Receive  Options  Access  Frag
RP   3579706        0           0         0  41258564        0       0     0
Switch#
```

- The reasons shown are as follows:
- **No_adj:** An incomplete adjacency
- **No_encap:** An incomplete ARP resolution
- **Unsupp'ted:** Unsupported packet features
- **Redirect:** ICMP redirect
- **Receive:** Layer 3 engine interfaces; includes packets destined for IP addresses that are assigned to interfaces on the Layer 3 engine, IP network addresses, and IP broad- cast addresses
- **Options:** IP options present
- **Access:** Access list evaluation failure
- **Frag:** Fragmentation failure **Packet Rewrite** When a multilayer switch finds valid entries in the FIB and adjacency tables, a packet is almost ready to be forwarded. One step remains: The packet header information must be rewritten. Keep in mind that multilayer switching occurs as quick table lookups to find the next-hop address and the outbound switch port. The packet is untouched and still has the original destination MAC address of the switch itself. The IP header also must be adjusted, as if a traditional router had done the forwarding.

Packet Rewrite When a multilayer switch finds valid entries in the FIB and adjacency tables, a packet is almost ready to be forwarded. One step remains: The packet header information must be rewritten. Keep in mind that multilayer switching occurs as quick table lookups to find the next-hop address and the outbound switch port. The packet is untouched and still has the original destination MAC address of the switch itself. The IP header also must be adjusted, as if a traditional router had done the forwarding.

The switch has an additional functional block that performs a packet rewrite in real time. The packet rewrite engine (shown in Figure 11-3) makes the following changes to the packet just before forwarding:
■ **Layer 2 destination address:** Changed to the next-hop device's MAC address ■ **Layer 2 source address:** Changed to the outbound Layer 3 switch interface's MAC
address
■ **Layer 3 IP TTL:** Decremented by one because one router hop has just occurred
■ **Layer 3 IP checksum:** Recalculated to include changes to the IP header
■ **Layer 2 frame checksum:** Recalculated to include changes to the Layer 2 and Layer 3 headers
A traditional router normally would make the same changes to each packet. The multi- layer switch must act as if a traditional router were being used, making identical changes. However, the multilayer switch can do this very efficiently with dedicated packet-rewrite hardware and address information obtained from table lookups.

Configuring CEF
CEF-capable switches have it on by default, and some it can't be conventionally disabled.
You can disable CEF on a per-interface basis with **no ip route-cache cef** and **no ip cef**

Verifying Multilayer Switching
verify how a switch is forwarding packets

Verifying Inter-VLAN Routing
To verify the configuration of a Layer 2 port, you can use the following EXEC command:
Switch# **show interface** *type member/module/number* **switchport**

The output from this command displays the access VLAN or the trunking mode and native VLAN. The administrative modes reflect what has been configured for the port, whereas the operational modes show the port's active status.

You can use this same command to verify the configuration of a Layer 3 or routed port. In this case, you should see the switchport (Layer 2) mode disabled, as in Example 11-7.

Example 11-7 *Verifying Configuration of a Layer 3 Switch Port*

```
Switch# show interface gigabitethernet 1/0/1 switchport
Name: Gi1/0/1
Switchport: Disabled
Switch#
```

To verify the configuration of an SVI, you can use the following EXEC command:
Switch# **show interface vlan** *vlan-id*
The VLAN interface should be up, with the line protocol also up. If this is not true, either the interface is disabled with the **shutdown** command, the VLAN itself has not been defined on the switch, or there are no active Layer 2 switch interfaces configured to use the VLAN. Use the **show vlan** command to see a list of configured VLANs. Example 11-8 shows the output produced from the **show vlan** command. Notice that each defined VLAN is shown, along with the switch ports that are assigned to it.

Example 11-8 *Displaying a List of Configured VLANs*

```
Switch# show vlan

VLAN Name                             Status    Ports
---- -------------------------------- --------- -------------------------------
1    default                          active    Gi1/0/1, Gi1/0/2, Gi1/0/3
                                                Gi1/0/4, Gi1/0/5, Gi1/0/6
                                                Gi1/0/7, Gi1/0/8, Gi1/0/9
                                                Gi1/0/10, Gi1/0/11, Gi1/0/12
                                                Gi1/0/13, Gi1/0/14, Gi1/0/15
                                                Gi1/0/16, Gi1/0/17, Gi1/0/18
                                                Gi1/0/19, Gi1/0/20, Gi1/0/21
                                                Gi1/0/25, Gi1/0/26, Te1/0/1
                                                Te1/0/2
2    VLAN0002                         active    Gi1/0/22
5    VLAN0005                         active
10   VLAN0010                         active
11   VLAN0011                         active    Gi1/0/23
12   VLAN0012                         active
99   VLAN0099                         active    Gi1/0/24
Switch#
```

You also can display the IP-related information about a switch interface with the **show ip interface** command, as demonstrated in Example 11-9.

Displaying IP-Related Information About a Switch Interface
Switch# **show ip interface vlan 101** Vlan101 is up, line protocol is up
  Internet address is 10.1.1.1/24
  Broadcast address is 255.255.255.255
  Address determined by setup command
  MTU is 1500 bytes
  Helper address is not set
  Directed broadcast forwarding is disabled
  Outgoing access list is not set
  Inbound access list is not set
  Proxy ARP is enabled
  Local Proxy ARP is disabled
  Security level is default
  Split horizon is enabled
  ICMP redirects are always sent
  ICMP unreachables are always sent
  ICMP mask replies are never sent
  IP fast switching is enabled
  IP fast switching on the same interface is disabled

```
  IP Flow switching is disabled
  IP CEF switching is enabled
  IP Feature Fast switching turbo vector
  IP Feature CEF switching turbo vector
  IP multicast fast switching is enabled
  IP multicast distributed fast switching is disabled
  IP route-cache flags are Fast, Distributed, CEF
  Router Discovery is disabled
  IP output packet accounting is disabled
  IP access violation accounting is disabled
  TCP/IP header compression is disabled
  RTP/IP header compression is disabled
  Probe proxy name replies are disabled
  Policy routing is disabled
  Network address translation is disabled
  WCCP Redirect outbound is disabled
  WCCP Redirect inbound is disabled
  WCCP Redirect exclude is disabled
  BGP Policy Mapping is disabled
  Sampled Netflow is disabled
  IP multicast multilayer switching is disabled
Switch#
```

You can use the **show ip interface brief** command to see a summary listing of the Layer 3 interfaces involved in routing IP traffic, as demonstrated in Example 11-10.

Example 11-10 *Displaying a Summary Listing of Interfaces Routing IP Traffic*

```
Switch# show ip interface brief
Interface                  IP-Address      OK? Method Status                Protocol
Vlan1                      unassigned      YES NVRAM  administratively down down
Vlan54                     10.3.1.6        YES manual up                    up
Vlan101                    10.1.1.1        YES manual up                    up
GigabitEthernet1/0/10      10.1.5.1        YES manual up                    up
[output omitted]
Switch#
```

Verifying CEF

CEF operation depends on the correct routing information being generated and down- loaded to the Layer 3 forwarding engine hardware. This information is contained in the FIB and is maintained dynamically. To view the entire FIB, use the following EXEC command:

Switch# **show ip cef**

Example 11-11 shows sample output from this command.

Example 11-11 *Displaying the FIB Contents for a Switch*

```
Switch# show ip cef
Prefix              Next Hop            Interface
0.0.0.0/32          receive
192.168.199.0/24    attached            Vlan1
192.168.199.0/32    receive
192.168.199.1/32    receive
192.168.199.2/32    192.168.199.2       Vlan1
192.168.199.255/32  receive
Switch#
```

On this switch, only VLAN 1 has been configured with the IP address 192.168.199.1 255.255.255.0. Notice several things about the FIB for such a small configuration:

- **0.0.0.0/32:** An FIB entry has been reserved for the default route. No next hop is defined, so the entry is marked "receive" so that packets will be sent to the Layer 3 engine for further processing.
- **192.168.199.0/24:** The subnet assigned to the VLAN 1 interface is given its own entry. This is marked "attached" because it is connected directly to an SVI, VLAN 1.
- **192.168.199.0/32:** An FIB entry has been reserved for the exact network address. This is used to contain an adjacency for packets sent to the network address, if the network is not directly connected. In this case, there is no adjacency, and the entry is marked "receive."
- **192.168.199.1/32:** An entry has been reserved for the VLAN 1 SVI's IP address. Notice that this is a host route (/32). Packets destined for the VLAN 1 interface must be dealt with internally, so the entry is marked "receive."
- **192.168.199.2/32:** This is an entry for a neighboring multilayer switch, found on the VLAN 1 interface. The Next Hop field has been filled in with the same IP address, denoting that an adjacency is available.
- **192.168.199.255/32:** An FIB entry has been reserved for the 192.168.199.0 subnet's broadcast address. The route processor (Layer 3 engine) handles all directed broad- casts, so the entry is marked "receive."
- To see complete FIB table information for a specific interface, use the following EXEC command: Switch# **show ip cef** *type member/module/number* [**detail**]

Table 11-3 Inter-VLAN Routing Configuration Commands

Put a port into Layer 2 mode.	Switch(config-if)# **switchport**
Put a port into Layer 3 mode.	Switch(config-if)# **no switchport**
Define an SVI.	Switch(config)# **interface vlan** vlan-id

Table 11-4 Multilayer Switching Verification Commands

Show a Layer 2 port status.	**Switch# show interface type member/ module/number switchport**
Show a Layer 3 port status.	**Switch# show interface type member/ module/number**
Show an SVI status.	**Switch# show interface vlan vlan-id**
View the FIB contents.	**Switch# show ip cef**
View FIB information for an interface.	**Switch# show ip cef [type member/ module/number | vlan vlan-id] [detail]**
View FIB information for an IP prefix.	**Switch# show ip cef [prefix-ip prefix- mask] [longer-prefixes] [detail]**
View FIB adjacency information.	**Switch# show adjacency [type member/ module/number | vlan vlan-id] [summary | detail]**

View counters for packets not switched by CEF. Switch# show cef not-cef-switched