Networkers at Cisco live!

July 22–26, 2007 Anaheim, CA

Networkers at Cisco live!

Cisco Catalyst 4500 Switch Architecture

BRKRST-3445

Session Goal

- To provide attendees with an understanding of the Catalyst 4500 series architecture
- Attendees can understand the characteristics and behavior of features implemented on the 4500.

Related Sessions

- BRKRST-3142: Troubleshooting Cisco Catalyst 4500 Switches

  Presenter: Prashanth Krishnappa , Customer Support Engineer
- BRKRST-3131: Troubleshooting LAN Protocols

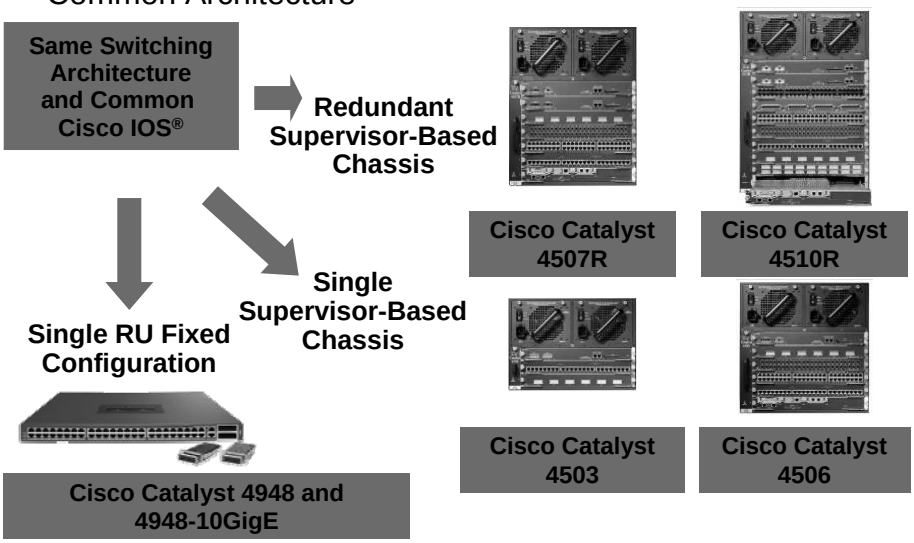
  Presenter: Shridhar Dhodapkar, Customer Support Engineer

Agenda

- 4500 Family Overview
- Supervisor Architecture and Packet Forwarding
- Line Card Architecture
- System High Availability
- QoS, Security, and TCAMs
- System CPU and Control Plane Policing

Cisco Catalyst 4500 Family Overview

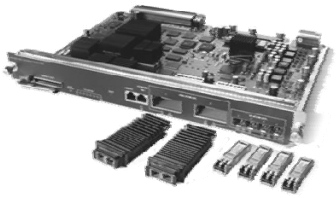
Common Architecture



Cisco Catalyst 4500 Family Overview

Cisco Catalyst 4500 Supervisor Architecture

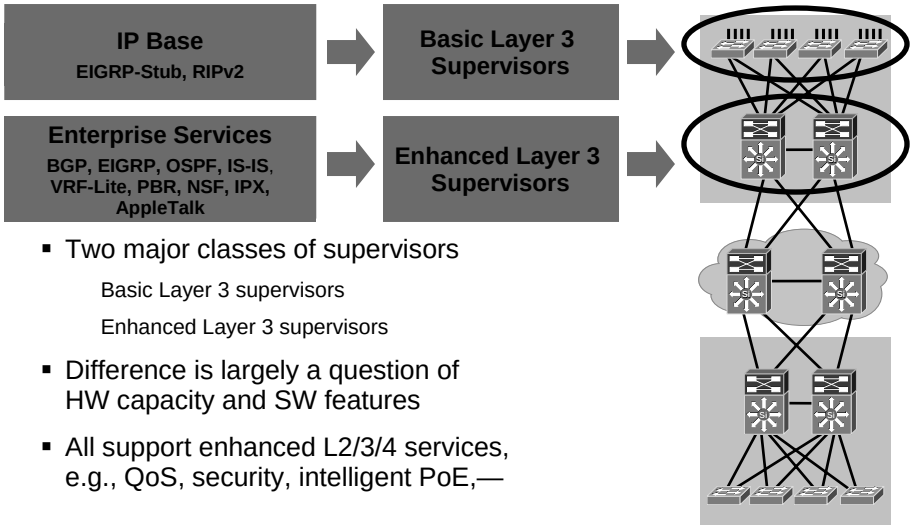
- Common architecture shared between all Cisco Catalyst 4500 supervisors
- Shared memory architecture
- Centralized forwarding ASICs
- Current generation of ASIC is known as K2



- All packets are forwarded via the supervisor
- No distributed line card forwarding or intelligence
- All QoS, ACLs, NetFlow is performed by the supervisor engine

Catalyst 4500 Family Overview

Supervisor Roles



- Two major classes of supervisors
  - Basic Layer 3 supervisors
  - Enhanced Layer 3 supervisors
- Difference is largely a question of HW capacity and SW features
- All support enhanced L2/3/4 services, e.g., QoS, security, intelligent PoE,—

Catalyst 4500 Family Overview

Supervisor Roles

Enhanced Layer 3 Supervisors

- (E)IGRP,OSPF, BGP, ISIS
- 128k IP CEF Entries
- 4096 VLAN's and SVI interfaces
- 28K(L3) 16K (L2) Multicast Routes
- 3000 Spanning Tree instances
- Netflow

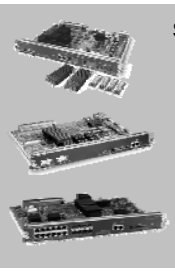
Supervisor V-10GE
- Netflow 2
- 4510R, 4507R, 4506, 4503

Supervisor V
- Netflow 1 (Opt.)
- 4510R, 4507R, 4506, 4503

Supervisor IV
- Netflow 1 (Opt.)
- 4507R, 4506, 4503

Basic Layer 3 Supervisors

- RIP, Static Routes & EIGRP Stub
- 32k IP CEF Entries
- 2048 VLAN's and SVI interfaces
- 12K(L3) 8K (L2) Multicast Routes
- 1500 Spanning Tree instances



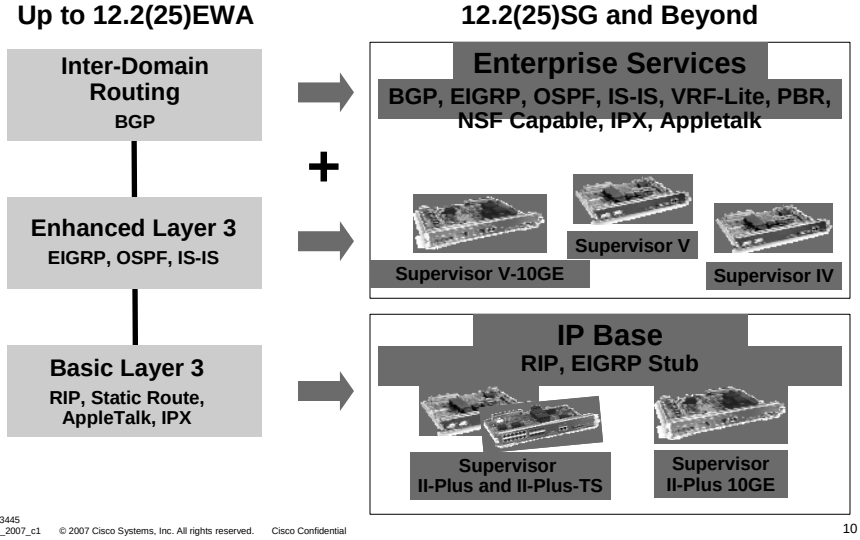
Supervisor II+10GE
- HW BCast & MCast suppression
- 4507, 4506, 4503

Supervisor II+
- 4507, 4506, 4503

Supervisor II+TS
- 4503 only

Cisco Catalyst 4500 Family Overview

Supervisor Roles



Cisco IOS Versions for Cisco IOS-Based

Supervisors

- The GD train 12.1.20E is based on the features in Cisco IOS 12.1.(12c)EW
- 12.1 train is end of life
- 12.2(18)EWx ,12.2(20)EWA,12.2(25)SG, 12.2(3)SG, trains are closed and will not have any more maintenance releases
- 12.2(25)EWAx train contains maintenance releases, but no new features
- 12.2(xx)SG train contains new features

Cisco Catalyst 4500 Supervisor

Supervisor Comparison

Supervisor	II+	II+ 10GigE	IV	V	V-10GigE
Switching Capacity	64 Gbps	108 Gbps	64 Gbps	96 Gbps	136 Gbps
Throughput	48 Mpps	75 Mpps	48 Mpps	72 Mpps	102 Mpps
Multilayer Switching	Basic L2/3/4 Services EIGRP-Stub, RIPv2	Basic L2/3/4 Services EIGRP-Stub, RIPv2	Full L2/3/4 Services EIGRP, OSPF, IS-IS, BGP	Full L2/3/4 Services EIGRP, OSPF, IS-IS, BGP	Full L2/3/4 Services EIGRP, OSPF, IS-IS, BGP
(E)IGRP, OSPF, BGP, ISIS	No	No	Yes	Yes	Yes
RIP, Static Routes, EIGRP Stub	Yes	Yes	Yes	Yes	Yes
Chassis Support	C4006, C4503, C4506, C4507R	C4006, C4503, C4506, C4507R	C4006, C4503, C4506, C4507R	C4006, C4503, C4506, C4507R, C4510R	C4503, C4506, C4507R, C4510R
CPU	266 MHz	666 MHz	333 MHz	400 MHz	833 MHz
IP CEF Entries	32K	32K	128K	128K	128k
SDRAM	256	256/512	512	512	512
Active VLANs	2K	2K	4K	4K	4k
Multicast Entries	12K(L3) 8K (L2)	12K(L3) 8K (L2)	28K(L3) 16K (L2)	28K(L3) 16K (L2)	28K(L3) 16K (L2)

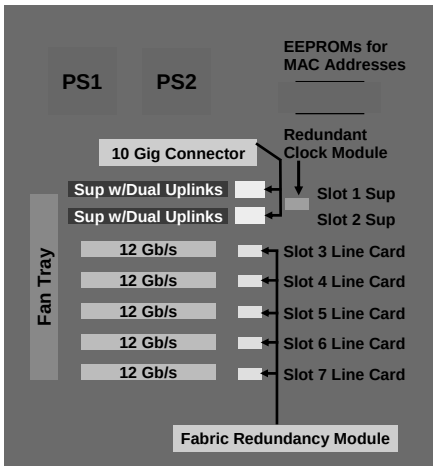
Catalyst 4500 Cisco IOS

Supervisor Comparison

Supervisor	II+	II+ 10GigE	IV	V	V-10GE
STP Instance	1.5K	1.5K	3K	3K	3k
SVI	1K	1K	4K	4K	4k
NVRAM	512 KB	512 KB	512 KB	512 KB	512 KB
NetFlow Support	No	No	Yes (NFL)	Yes (NFL)	Yes (NFL2)
Broadcast Suppression	Software	Hardware	Software	Hardware	Hardware
Multicast Suppression	No	Yes	No	Yes	Yes
QoS Sharing	Nonblocking GE Only	All Ports	Nonblocking GE Only	All Ports	All Ports
QinQ	Pass-Through	In Hardware	Pass-Through	In Hardware	In Hardware
Sup Uplinks	2 GE	2 x 10GE	2 GE	2 GE	2 x 10GE or 4 x GE

Cisco Catalyst 4500 Family Overview

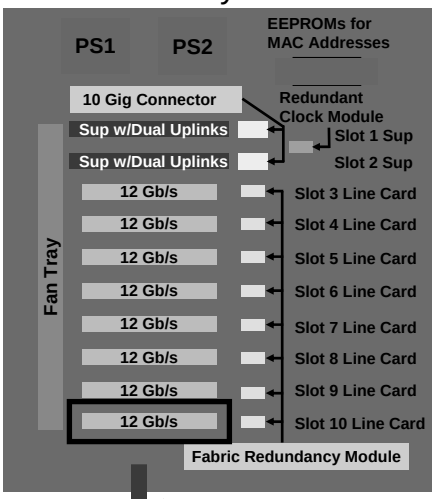
Cisco Catalyst 4507R Chassis Backplane



- Dual data or data + voice PSUs
- Integrated inline power
- Redundant clock
- Chassis EEPROM stores system MAC addresses
- Fan tray for system cooling
- NEBS level three
- Chassis configuration
  - Slot 1 - supervisor engine
  - Slot 2 - redundant supervisor engine (no line cards)
  - Slots 3 to 7 - line card slots
  - Fabric redundancy modules on backplane enable redundancy support with existing line cards

Cisco Catalyst 4500 Family Overview

Cisco Catalyst 4510R Chassis Backplane



Requires Supervisor Engine V or Supervisor Engine V-10GE

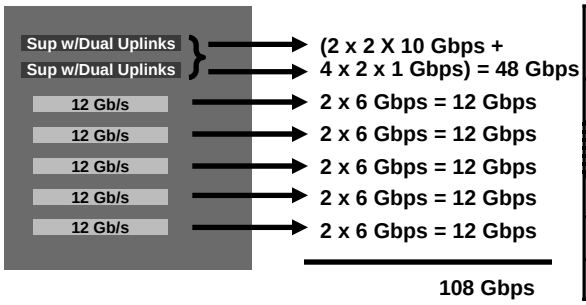
- Supervisor V
  - 96 Gbps of fabric capacity
  - 92 Gbps to line cards + 4 x 1Gbps shared between sups
  - Slot ten operates as a flex-slot and will support 2 x 1GE or WAN
- Supervisor V-10GE
  - 136 Gbps of fabric capacity
  - 96 Gbps to line cards + 8 x 1Gbps shared between sups
  - 'or'
  - 96 Gbps to line cards + 2 x 20 Gbps shared between sups

Full Capacity for Slot Ten Requires a Sup V-10GE

Cisco Catalyst 4500 Family Overview

Supervisor Switching Fabric Capacity

- Switching capacity of the switch is based on the capacity of the Supervisor switching fabric (ASICs and SRAM capacity)
- Total switching fabric capacity is allocated between all line cards and supervisor ports with no oversubscription allowed in the switching fabric itself

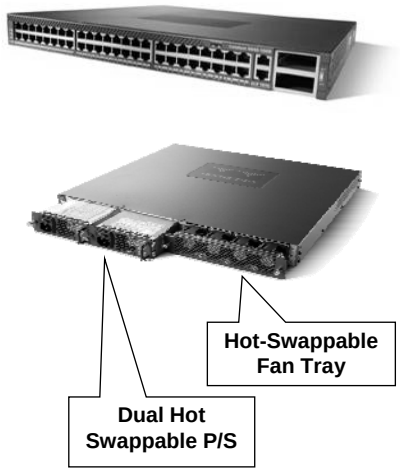


Supervisor	Switching Capacity
Sup II-Plus	64 Gbps
Sup II-Plus 10GE	108 Gbps
Sup IV	64 Gbps
Sup V	96 Gbps
Sup V 10GE	136 Gbps

Cisco Catalyst 4500 Family Overview

Cisco Catalyst 4948 (SupV and SupV-10GE in a Pizza Box)

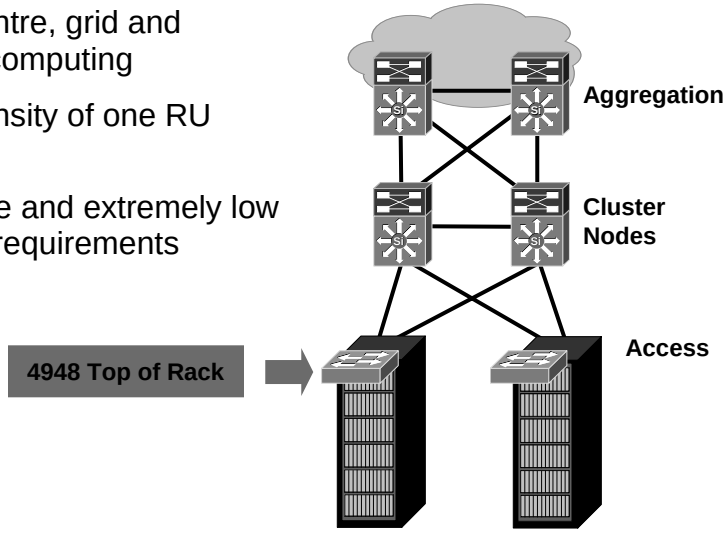
- 4948-10GigE:
  - 136 Gbps capacity and 102 Mpps throughput
  - 48 wire-rate 10/100/1000T GE ports and two wire-rate 10GE uplink ports
- 4948:
  - 96 Gbps capacity and 72 Mpps throughput
  - 48 wire-rate 10/100/1000T GE ports
  - Ports 45–48 alternatively wired for SFP
- One RU form factor
- Dual, hot-swappable, internal power supplies (AC or DC options)
- Hot-swappable fan tray
- Jumbo frames on all ports L2/L3
- Broadcast and multicast suppression in hardware for all ports (L2/3)



Cisco Catalyst 4948

Top of the Rack Data Center Server Switch

- Data centre, grid and cluster computing
- High density of one RU servers
- Wire rate and extremely low latency requirements



Cisco Catalyst 4500 Family Overview

Cisco Catalyst 4948 Comparison

Model	WS-C4948	WS-C4948-10GE
Switching Capacity	96 Gbps	136 Gbps
Throughput	72 Mpps	102 Mpps
Multilayer Switching	Full L2/3/4 Services and Routing	Full L2/3/4 Services and Routing
(E)IGRP,OSPF, BGP, ISIS	Yes	Yes
CPU	266 MHz	666 MHz
IP CEF Entries	32k	32k
SDRAM	256	256
Active VLANs	2K	2k
IGMP Snooping	Yes (8K)	Yes (8k)
STP Instance	1500	1500
SVI	2k	2k
NetFlow Support	No	No
Broadcast Suppression	Hardware	Hardware
Multicast Suppression	Yes	Yes
QoS Sharing	All Ports	All Ports
QinQ	In Hardware	In Hardware
Uplinks	4 SFP	2 10GE (X2)

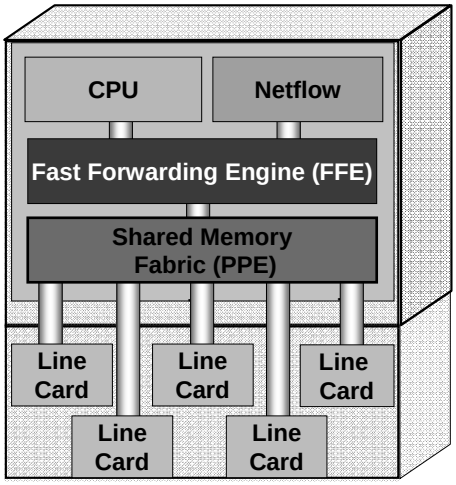
Agenda

- 4500 Family Overview
- Supervisor Architecture and Packet Forwarding
- Line Card Architecture
- System High Availability
- QoS, Security, and TCAMs
- System CPU and Control Plane Policing

Catalyst 4500 Architecture

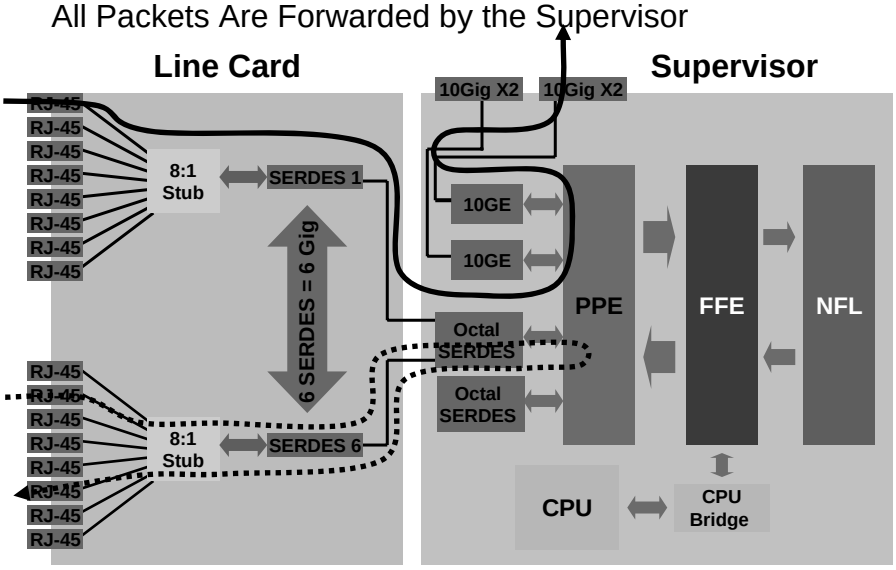
4500 Utilizes a Centralized Architecture

- Catalyst 4500 is a shared memory switch
- All forwarding, queuing, security is implemented on the Supervisor
- The individual line cards are considered to be 'transparent'
  - Contain simple "stub" ASIC's and the PHY's
  - No buffering or local switching
- Each line card has 6 dedicated 1 Gbps (full duplex) connections to the central forwarding engine



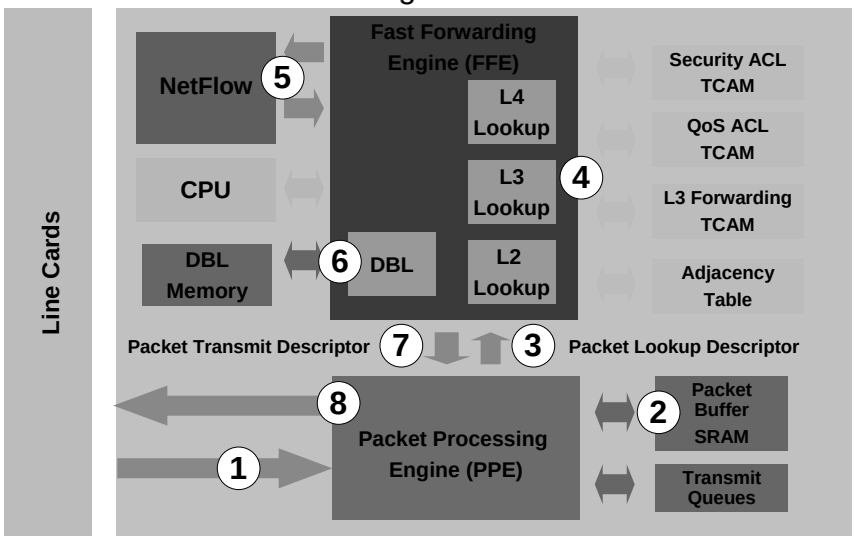
Cisco Catalyst 4500 Architecture

All Packets Are Forwarded by the Supervisor



Supervisor Packet Rorwarding

Unicast Packet Forwarding



Supervisor Packet Forwarding

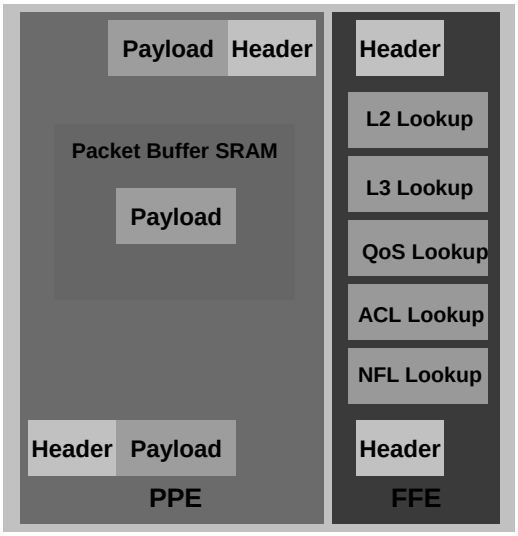
Unicast Packet Forwarding

1. Packet arrives from the line card
2. PPE ASIC buffers the packet data in the packet buffer memory
3. The packet header and flow label is sent to FFE in Packet Lookup Descriptor (PLD)
4. FFE ASIC receives the PLD and performs L2/3/4 forwarding lookups with TCAMs
5. FFE ASIC sends the PLD info to NetFlow 2 ASIC for in-depth QOS, microflow, and packet statistics
6. FFE performs per port, per queue congestion control with DBL by monitoring the amount of buffering per flow
7. FFE ASIC generates the PTD (Packet Transmit Descriptor) and passes it back to the PPE
8. PPE performs QoS scheduling consulting transmit queue memory, rewrites the MAC headers and transmits the packet to egress line card

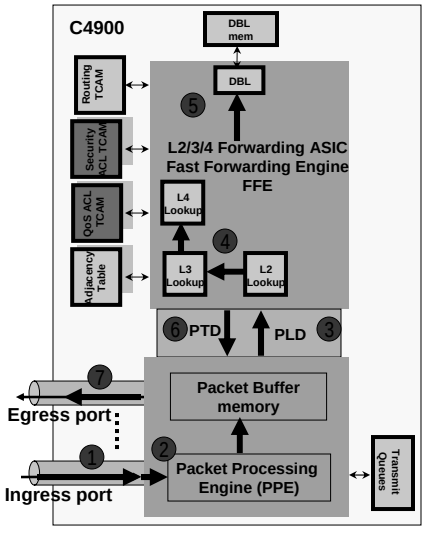
Supervisor Packet Forwarding

Pipeline Architecture

- The FFE utilizes a pipelined architecture
- Packet is processed through ‘all’ stages of the pipeline
- Deterministic latency for ‘all’ traffic in ‘all’ cases



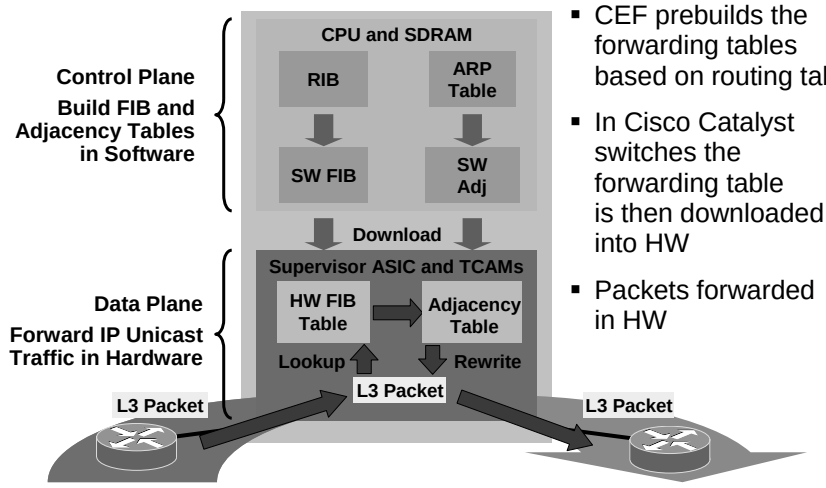
Wire Rate Switching Packet Flow in 4948



1. Packet arrive from any GigE/10Gig port
2. Packets buffered in the shared packet buffer memory
3. The packet header and flow label is sent to Forwarding Engine through the Packet lookup descriptor (PLD)
4. Forwarding Engine performs L2/3/4 forwarding lookups with TCAMs
5. Forwarding Engine performs Per port, per queue congestion control with DBL by monitoring the amount of buffering per flow
6. Forwarding Engine sends the packet transmit descriptor (PTD) to the Packet Processing Engine
7. Packet Processing Engine performs QoS scheduling consulting transmit queue memory, rewrites the MAC headers and transmits the packet to egress port

Supervisor Packet Forwarding

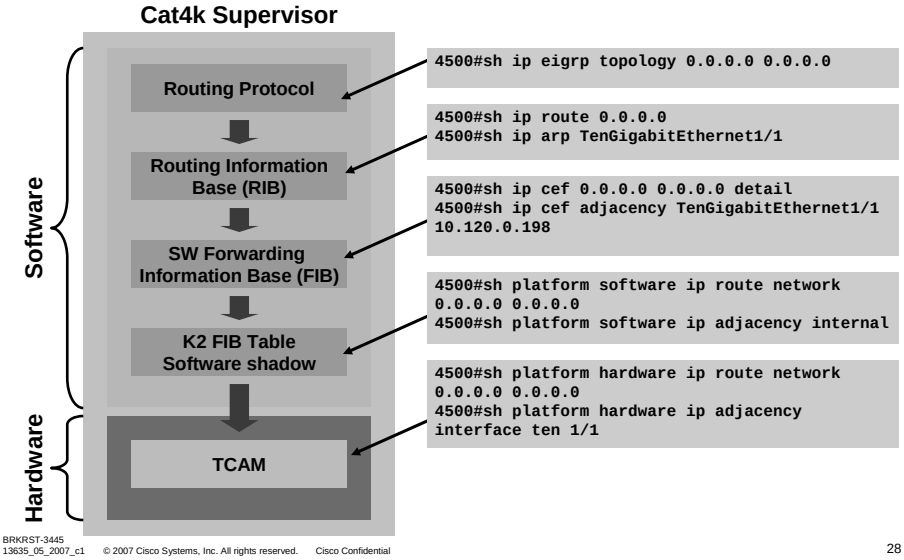
Cisco Express Forwarding (CEF)



- CEF prebuilds the forwarding tables based on routing table
- In Cisco Catalyst switches the forwarding table is then downloaded into HW
- Packets forwarded in HW

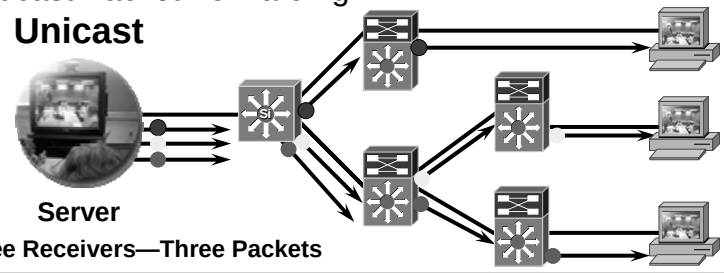
Supervisor Packet Forwarding

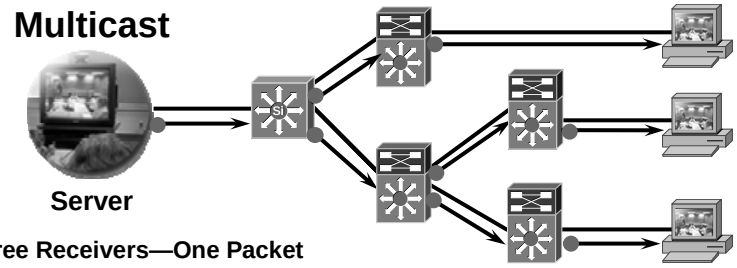
Cisco Catalyst 4500 Implementation of CEF



Supervisor Packet Forwarding

Multicast Packet Forwarding

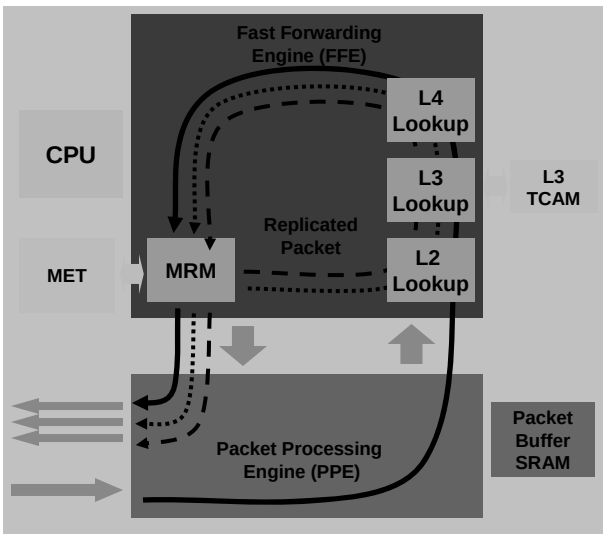




Supervisor Packet Forwarding

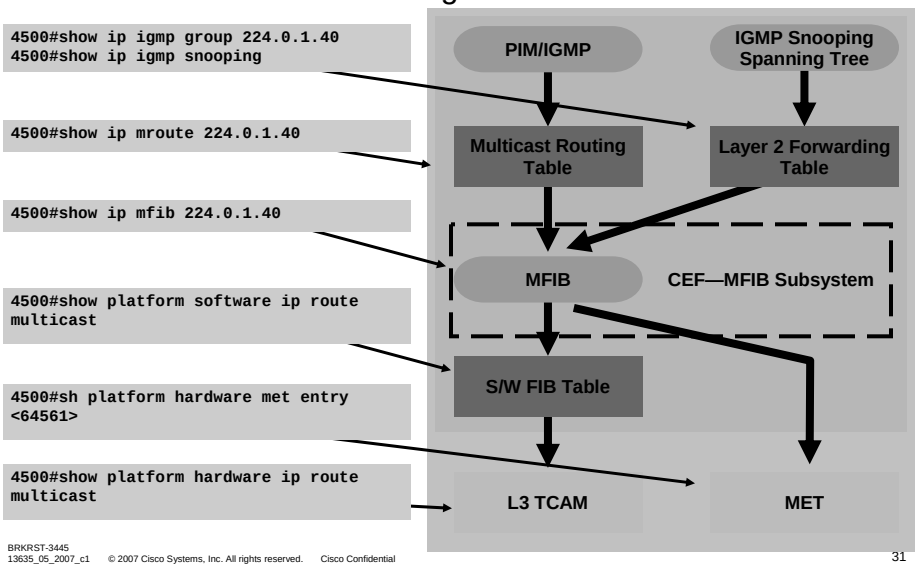
Multicast Packet Forwarding

- First packet is processed through the pipeline
- Multicast Replication Module (MRM) determines this packet requires replication based on Multicast Expansion Table (MET)
- Create a new header that is injected into the pipeline
- All packets pass through pipeline to ensure ACL and QoS policies are enforced



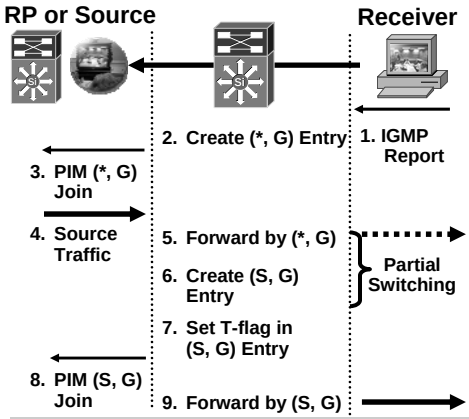
Supervisor Packet Forwarding

Multicast Packet Forwarding

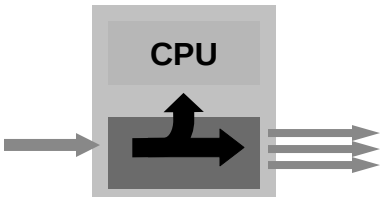


Supervisor Packet Forwarding

Partial Switching




```
cr7-4507-1#sh ip mfib
 . . .
Packets: Fast/Partial/Slow Bytes: Fast/Partial/Slow
(*, 224.0.1.40), flags (IC)
   Packets: 2292/2292/0, Bytes: 512883/512883/0
   TenGigabitEthernet1/1 (A S)
```



- Partial switching packets are those copied to the CPU
- PIM register and IGMP join-group
- Packet is still switched in HW only a copy is sent to the CPU

Supervisor Packet Forwarding

Cisco Catalyst 4500 Multicast Scalability

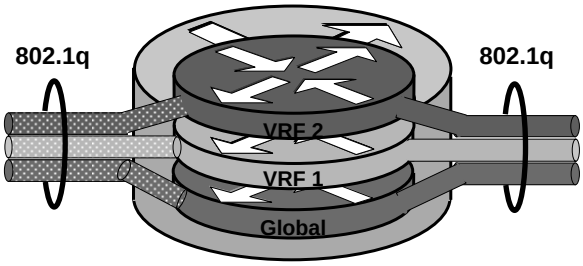
Multicast Routing/Features Implemented in Hardware

Supervisor	Supervisor II-Plus/II-Plus TS	Supervisor IV/V/V—10GE
IP Multicast Routes (PIM Dense Mode)	12,000	28,000
IP Multicast Routes (PIM Sparse Mode)	6000	14,000
IGMP Snooping Group Entries	8000	16,000
QoS Support for IP Multicast Packets	Full, Including Four Queues per Port	Full, Including Four Queues per Port
TCAM Entries	32,000 (Shared by IP Unicast, Mcast Entries)	128,000 (Shared by IP Unicast, Mcast Entries)

- Supervisor engines program both (S,G) and (*,G) for sparse mode, this halves the hardware supported mroutes
- Multicast switching features—IGMP snooping (V1, 2, 3) and CGMP server

Supervisor Packet Forwarding

VRF (Virtual Routing and Forwarding)



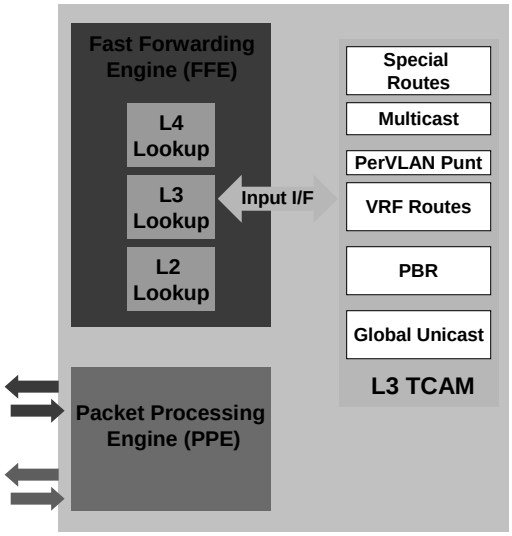
Traffic Is Routed from 802.1q VLAN to 802.1q VLAN

- VRF allows for the creation of multiple logical forwarding tables
  - Distinct Routing Information Base (RIB)
  - Distinct Forwarding Information Base (FIB)
- Each VRF has a unique RIB and FIB, in addition the switch has the common global RIB and FIB
- Cisco Catalyst 4500 does not currently support tag switching (MPLS)

Supervisor packet forwarding

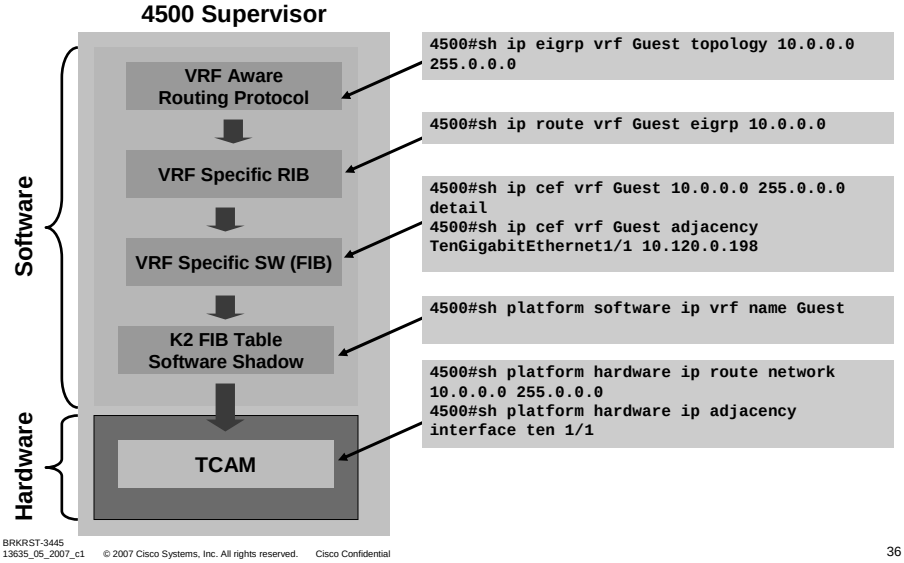
VRF Packet Forwarding (Unicast)

- Routes are sorted in the TCAM
- Provides for special processing, e.g., VRF and PBR
- TCAM lookup uses ‘input interface’ as part of the n-tuple lookup criteria
- By selecting the table entry based on interface VLAN it is possible to select a unique VRF route based on 802.1q input tag



Supervisor Packet Forwarding

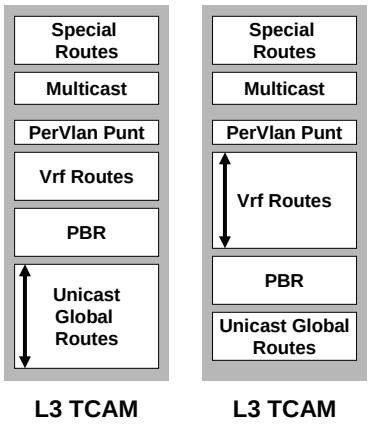
VRF Packet Forwarding (Unicast)



Supervisor Packet Forwarding

VRFs and TCAMs

- The unicast global FIB and all of the VRF-specific FIB tables share a common TCAM (Ternary Content Addressable Memory)
- The TCAM allows for dynamic growth of any specific region
- VRF's FIB size increases based on (# routes) x (# VLANs in the VRF)
- VRFs that can not have their entire FIB loaded in HW will have 'all' traffic for that VRF forwarded in SW



Supervisor Packet Forwarding

VRFs and TCAMs

- Watch for and manage TCAM exhaustion when implementing VRFs
- `C4K_L3HWFORWARDING-4-FWDCAMOUTOFSPACEFORVRF ROUTINGTABLE: Insufficient TCAM resources to load VRF Guest routing table. Switching to software forwarding for this VRF.`

```
cr7-4507-1(config)#ip vrf Guest
cr7-4507-1(config-vrf)#maximum routes 1000 ?
  <1-100>       Threshold value (%) at which to generate a warning msg
  warning-only  Only give a warning message if is limit exceeded
```

```
cr7-4507-1#sh platform hardware ip route summary
TCAM running in 144 bit mode. (16 routes per block)
63 blocks used out of 8192 (0.76%)
620 K2Fib TCAM entries used out of 131072 (0.47%)
(448 entries are fixed overhead)
152 K2FibAdjs used out of 32768 (0.46%)
167 IrmFibEntries used out of 262144 (0.06%)
11 IrmMfibEntries used out of 65536 (0.01%)
154 IrmFibAdjs used out of 32768 (0.46%)
. . .
```

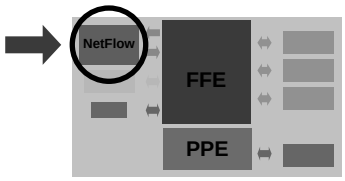
% of FIB Used

% of Adjacency Table Used

Cisco Catalyst 4500 Netflow

NetFlow Support

- NetFlow feature card utilized to track traffic flows
  - Collect statistics on traffic flows forwarded in hardware
  - Provides capability to support flow-based policiers (User Based Rate Limiting—UBRL)
- Statistics includes switched (L2) and routed (L3) traffic as well as software switched packets
- NetFlow engine I and NetFlow engine II
- NDE versions 1, 5, and 8 are supported

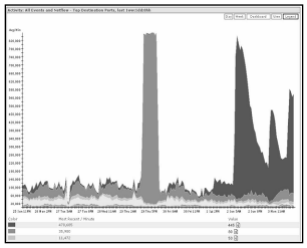


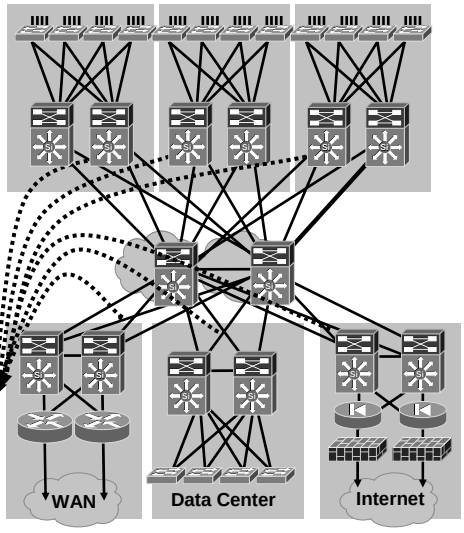
	NFL1	NFL2
Supervisors	Sup IV and V	Sup V-10GE
Optional	Daughter Card	Included
Flow Policing	No	85 Flows, 511 Unique Flow Rates
Total Entries	128K	128K
Effective Entries	64k	85k

Cisco Catalyst 4500 Netflow

Scalable Monitoring for Network Worms

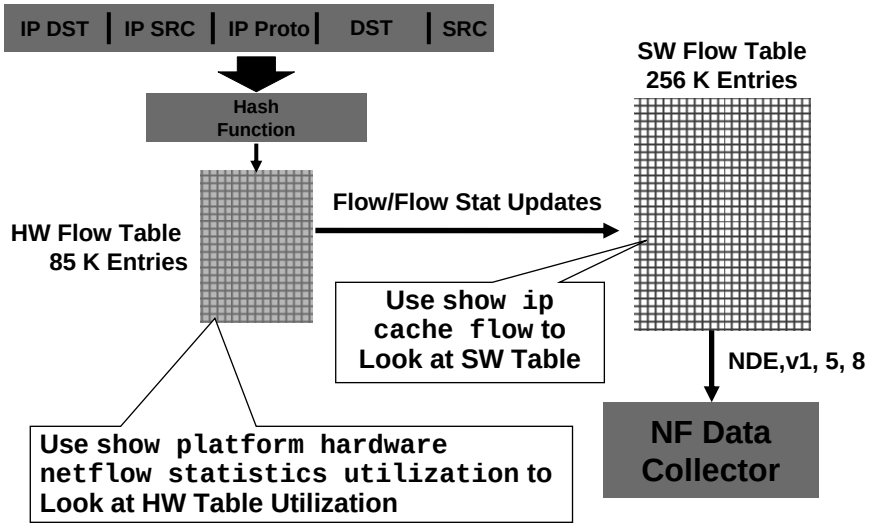
- NetFlow's anomaly detection using statistical profiling, pinpoints day zero attacks like worm outbreaks
- CS-MARS detects anomalies comparing the previous data against current data; upon anomaly is detected, CS-MARS starts to save that data and creates an incident





Cisco Catalyst 4500 Netflow

NetFlow HW-SW Cache Interaction



Agenda

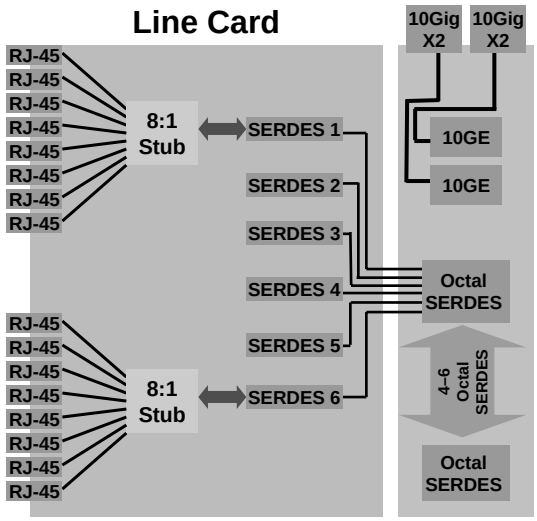
- 4500 Family Overview
- Supervisor Architecture and Packet Forwarding
- Line Card Architecture
- System High Availability
- QoS, Security, and TCAMs
- System CPU and Control Plane Policing

Cisco Catalyst 4500 Architecture

Transparent Line Cards

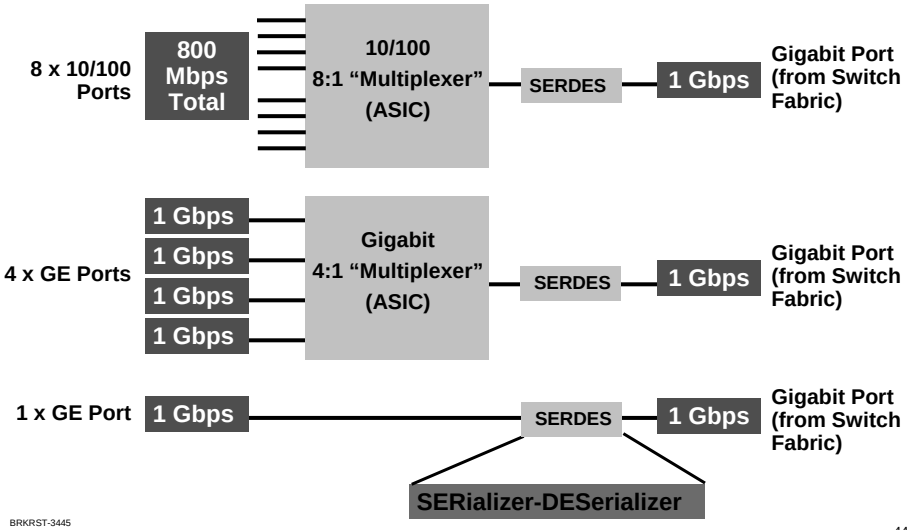
- Dedicated six full-duplex gigabit connections to the switch fabric per line card
- Transparent

  No local forwarding, all packets go to supervisor
- Gigabit connections from switch fabric straight to front-panel port or connect to stubs



Switching Module Architecture

Line Cards Stub ASICs



Switching Module Architecture

Oversubscription on GigE Ports

- A port that does not oversubscribe access to the switching fabric is a nonblocking GE port
- A port that oversubscribes access to the switching fabric is a blocking GE port
- On over-subscribed GE ports, each of the front-panel port can burst up to one Gigabit
- On transmit and receive side of Stub ASIC, each of the front panel port is serviced round-robin on a per-packet basis
- Guaranteed rate per-port, if all of them are bursting at same packet size:

  8:1—125 Mbps

  4:1—250 Mbps

Switching Module Architecture

Blocking and Nonblocking GE Ports

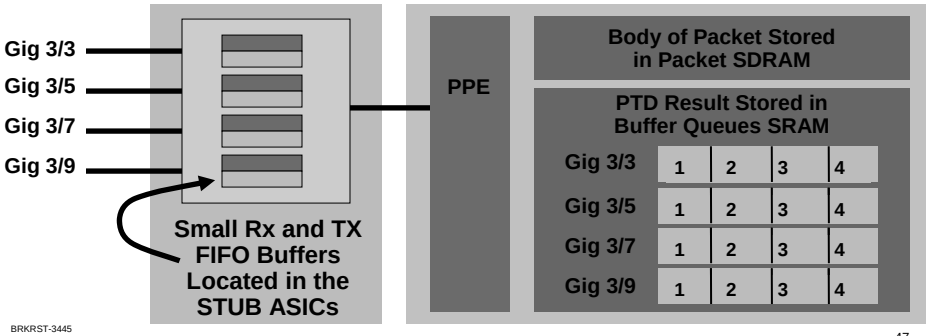
Nonblocking GE Line Cards/Chassis	Blocking GE Line Card	Oversubscription Ratio for Blocking Line Cards
Supervisor Uplink Ports	All Ports on the Following WS-X4424-GB-RJ45 WS-X4524-GB-RJ45V	4:1
All Ports on the Following WS-X4302-GB WS-X4306-GB WS-X4506-GB-T WS-X4013+TS WS-C4948 WS-C4948-10GE	All Ports on the Following WS-X4448-GB-RJ45 WS-X4448-GB-SPF WS-X4448-GB-LX WS-X4548-GB-RJ45 WS-X4548-GB-RJ45V	8:1
Two 1000 Base-X Ports on the WS-X4232-GB-RJ	1000 Base-T Ports on the WS-X4412-2GB-TX	4:1
First Two Ports on WS-X4418-GB	Last 16 Ports on the WS-X4418-GB	4:1

Jumbo Frames (Up to 9216 Byte) Are Supported Only on Nonblocking Ports

Switching Module Architecture

Ingress (Rx) and Egress (Tx) Queuing

- Due to the nonblocking architecture of the supervisor there are no ingress (Rx) queues on the supervisor
- Egress (Tx) queues are allocated on a per port basis in the queue SRAM
- On return from FFE the PTD result is stored in the correct Tx queue
- The packet itself is stored in packet SRAM



Switching Module Architecture

Egress (Tx) Queue Sizes (# of Packets per Queue)

```
cr39-4507-1#sh mod

Mod Ports Card Type                              Model             Serial No.
---+-----+--------------------------------------+-----------------+-----------
 1     2  Supervisor IV 1000BaseX (GBIC)         WS-X4515          JAB0627065V
 2     2  Supervisor IV 1000BaseX (GBIC)         WS-X4515          JAB064907TY
 3    18  1000BaseX (GBIC)                       WS-X4418          JAB030800Q0
 4     6  1000BaseX (GBIC)                       WS-X4306          JAE044709VR
 5    18  1000BaseX (GBIC)                       WS-X4418-GB       JAE0646014E
 6    24  10/100/1000BaseT (RJ45)                WS-X4424-GB-RJ45  JAB052406EF

cr39-4507-1#show platform software interface gigabitEthernet 6/24 tx-queue
Switch Phyport Gi6/24 Tx-Queue Software State
Phyport TxQ BaseAddr Size     Shape      Share
                          Mant.  Exp. Mant. Exp.
-------------------------------------------------
Gi6/24  0   0x15540  240    0     0     0    0
Gi6/24  1   0x15630  240    0     0     0    0
Gi6/24  2   0x15720  240    0     0     0    0
Gi6/24  3   0x15810  240    0     0     0    0

cr39-4507-1#show platform software interface gigabitEthernet 4/6 tx-queue
Switch Phyport Gi4/6 Tx-Queue Software State
Phyport TxQ BaseAddr Size     Shape      Share
                          Mant.  Exp. Mant. Exp.
-------------------------------------------------
Gi4/6   0   0x00000  1920   0     0     43   14
Gi4/6   1   0x00780  1920   0     0     43   14
Gi4/6   2   0x00F00  1920   0     0     43   14
Gi4/6   3   0x01680  1920   0     0     43   14
```

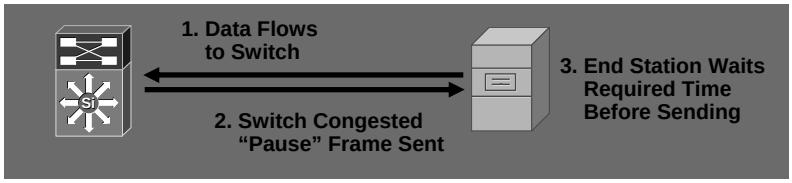
On Blocking Ports Queue Memory Is Divided Amongst Port Tx Queues

1920 Packets per Queue on Nonblocking Ports

Switching Module Architecture

IEEE 802.3 Flow Control (Blocking GE Ports)

- 802.3x is an IEEE standards-based mechanism used to control data flow
- 802.3x utilizes pause frames (DA MAC 01-80-C2-00-00-00-0F) to signal flow control between end station and switch
- Flow control operation steps
  - Data flows to switch
  - Switch congested so “pause” frame sent
  - End station waits required time before sending
- Cisco Catalyst 4500 supervisors support both Tx and Rx pause frames

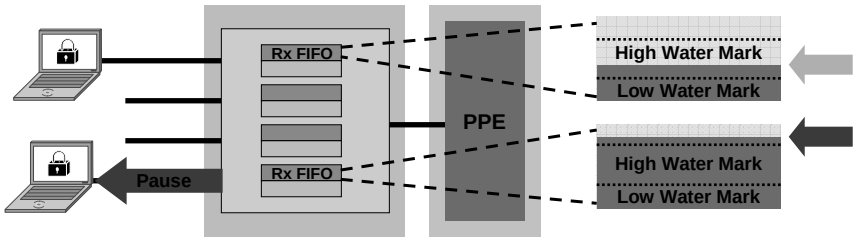



```
cr7-4507-1# show interface GigabitEtherent 4/7 counters detail
<output skipped>
Port          Rx-No-Pkt-Buff     RxPauseFrames     TxPauseFrames     PauseFramesDrop
Gi4/7                      0             35648              4854                   0
```


Switching Module Architecture

IEEE 802.3 Flow Control

- Each stub ASIC FIFO Rx and Tx buffer has a high and low water mark
- When the buffer fills past the high water mark it will issue a pause frame to the end station (wait 33 microseconds)
- This allows the end station to queue traffic while waiting for Rx buffer to drain
- In a similar manner stub ASIC applies back pressure to the supervisor to buffer traffic in the Tx queues if end station signals the switch to pause transmission



Switching Module Architecture

IEEE 802.3 Flow Control Configuration

Type of Interface	Send	Receive
Stub GE (Blocking)	On	Desired
Nonstub GE (Nonblocking)	Off (Not Needed)	Desired
Ten GE (Nonblocking)	Off (Not Needed)	On

Recommended Configuration Is the Default

```
cr7-4507-1(config-if)#flowcontrol receive ?
  desired  Allow but do not require flow-control packets on port
  off      Disable flow-control packets on port
  on       Enable flow-control packets on port
!
!
cr7-4507-1#sh interfaces gigabitEthernet 3/48 flowcontrol
Port       Send FlowControl  Receive FlowControl  RxPause TxPause
           admin    oper     admin    oper
---------  -------- -------- -------- --------    ------- -------
Gi3/48     on       disagree desired  off         0       0
```


Switching Module Architecture

Checking Capabilities of an Interface

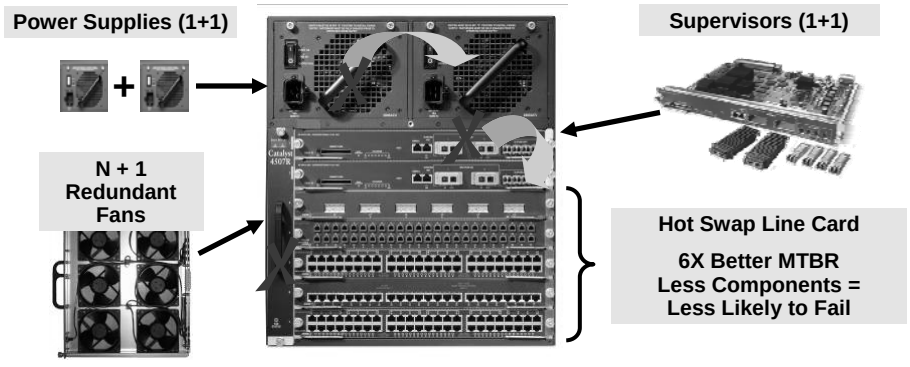
```
cr7-4507-1# sh interfaces gigabitEthernet 3/48 capabilities
GigabitEthernet3/48
  Model:                 WS-X4448-GB-RJ45-RJ-45
  Type:                  10/100/1000-TX
  Speed:                 10,100,1000,auto
  Duplex:                half,full,auto
  Trunk encap. type:     802.1Q,ISL
  Trunk mode:            on,off,desirable,nonegotiate
  Channel:               yes
  Broadcast suppression: percentage(0-100), hw
  Flowcontrol:           rx-(off,on,desired),tx-(off,on,desired)
  VLAN Membership:       static, dynamic
  Fast Start:            yes
  Queuing:               rx-(N/A), tx-(1p3q1t, Sharing/Shaping)
  CoS rewrite:           yes
  ToS rewrite:           yes
  Inline power:          no
  SPAN:                  source/destination
  UDLD:                  yes
  Link Debounce:         no
  Link Debounce Time:    no
  Port Security:         yes
  Dot1x:                 yes
  Maximum MTU:           1552 bytes (Baby Giants)
  Multiple Media Types:  no
  Diagnostic Monitoring: N/A
```


Agenda

- 4500 Family Overview
- Supervisor Architecture and Packet Forwarding
- Line Card Architecture
- System High Availability
- QoS, Security, and TCAMs
- System CPU and Control Plane Policing

System High Availability

Integrated Hardware and Software Redundancy

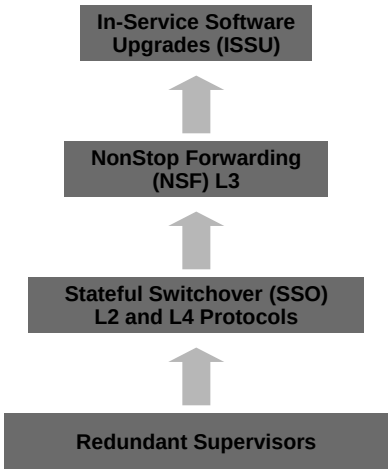


- Chassis hardware redundancy matched to the appropriate software redundancy mechanisms
- SSO, NSF, ISSU, and fault detection

System High Availability

Redundant Supervisors—SSO, NSF, and ISSU

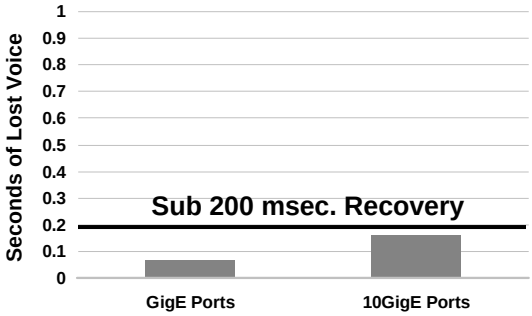
- Cisco Catalyst 4507R and 4510R support the use of redundant supervisors
- Chassis slots one and two are configured for redundant supervisors (not usable for line cards)
- Supervisor hardware redundancy will leverage three software mechanisms to improve network resiliency and provide for enhanced operational change processes
  - SSO—Stateful Switchover
  - NSF—NonStop Forwarding
  - ISSU—In Service Software Upgrade

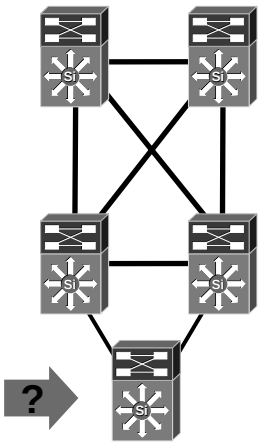


Redundant Supervisors

Where Do They Make Sense?

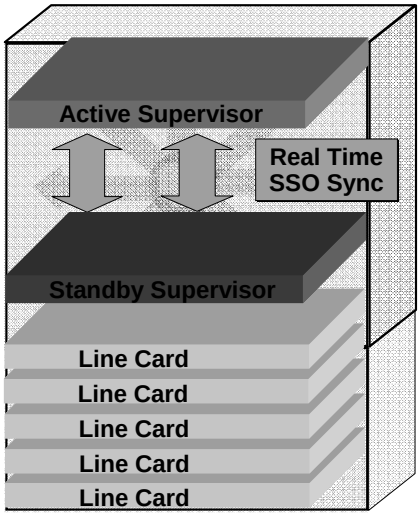
- Access switch is the single point of failure in best practices HA campus design
- Supervisor failure is most common cause of access switch service outages
- Layer 2 SSO or Layer 3 NSF/SSO makes sense in the access





System High Availability

SSO—Supervisor Redundancy



- Stateful Switchover (SSO)
- Utilizes IOS High Availability Framework
- Synchronizes HW states in real time between Supervisors
- SW processes may be clients of HA Framework

SSO Synchronizes:	
Port Security	802.1x
IGMP Snooping	ARP/DHCP
VLANs/Trunks/Ports	STP/VTP/DTP
PAgP/LACP	802.1Q
ACL/QoS	Voice VLAN with PoE

System High Availability

SSO Redundancy

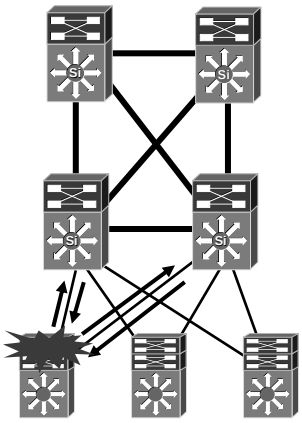
```
cr39-4507-1(config)#redundancy
cr39-4507-1(config-red)#mode ?
  rpr  Route Processor Redundancy
  sso  Stateful Switchover
```

```
cr39-4507-1#sh mod
Chassis Type : WS-C4507R

Power consumed by backplane : 40 Watts

Mod Ports Card Type                              Model             Serial No.
---+-----+--------------------------------------+-----------------+-----------
 1     2  Supervisor IV 1000BaseX (GBIC)         WS-X4515          JAB0627065V
 2     2  Supervisor IV 1000BaseX (GBIC)         WS-X4515          JAB064907TY
 3    18  1000BaseX (GBIC)                       WS-X4418          JAB030800Q0
 4     6  1000BaseX (GBIC)                       WS-X4306          JAE044709VR
 5    18  1000BaseX (GBIC)                       WS-X4418-GB       JAE0646014E
 6    24  10/100/1000BaseT (RJ45)                WS-X4424-GB-RJ45  JAB052406EF

 <snip>

Mod  Redundancy role     Operating mode      Redundancy status
----+------------------+-------------------+------------------
 1   Active Supervisor   SSO                 Active
 2   Standby Supervisor  SSO                 Standby hot
```


System High Availability

NSF Recovery (Routing Protocol Recovery)

- Non-Stop Forwarding (NSF) provides the capability for the routing protocols to gracefully restart after an SSO fail-over
- The newly active redundant supervisor continues forwarding traffic using the synchronized HW forwarding tables
- The NSF capable Routing Protocol requests a graceful neighbor start
- Routing neighbors reform with no loss of traffic



No Route Flaps During Recovery

System High Availability

Enabling NSF in the Routing Protocol

```
cr39-4507-1(config)#router eigrp 100
cr39-4507-1(config-router)#nsf

cr39-4507-1(config-router)#timers nsf ?
  converge    EIGRP time limit for convergence after switchover
  route-hold  EIGRP hold time for routes learned from nsf peer
  signal      EIGRP time limit for signaling NSF restart
```

```
cr39-4507-1(config)#router ospf 100
cr39-4507-1(config-router)#nsf

cr39-4507-1(config-router)#nsf ?
  enforce  Cancel NSF restart when non-NSF-aware neighbors detected
```

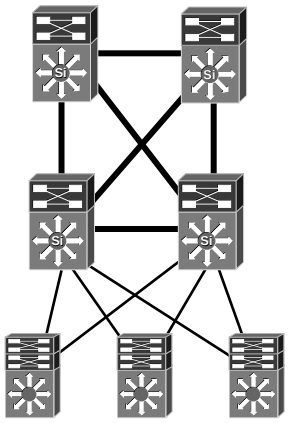
```
cr39-4507-1(config)#router isis level2
cr39-4507-1(config-router)#nsf cisco
        'or'
cr39-4507-1(config)#router isis level2
cr39-4507-1(config-router)#nsf ietf
```

```
cr39-4507-1(config-router)#bgp graceful-restart ?
  restart-time    Set the max time needed to restart and come back up
  stalepath-time  Set the max time to hold onto restarting peer's stale paths
  <cr>

cr39-4507-1(config-router)#bgp graceful-restart
```


System Resiliency

NSF OSPF Example

```
Switch#*Aug 11 15:37:49: %OSPF-5-ADJCHG: Process 100, Nbr
   100.1.1.1 on Vlan608 from LOADING to FULL, Loading Done
Switch#show ip ospf
 <snip>
 Non-Stop Forwarding enabled, last NSF restart 00:00:23
   ago (took 31 secs)
 <snip>
 Switch#show ip ospf neighbor detail
 Neighbor 100.1.1.1, interface address 172.26.197.67
  <snip>
  LLS Options is 0x1 (LR), last OOB-Resync 00:00:41 ago
  Dead timer due in 00:00:33
  <snip>
```

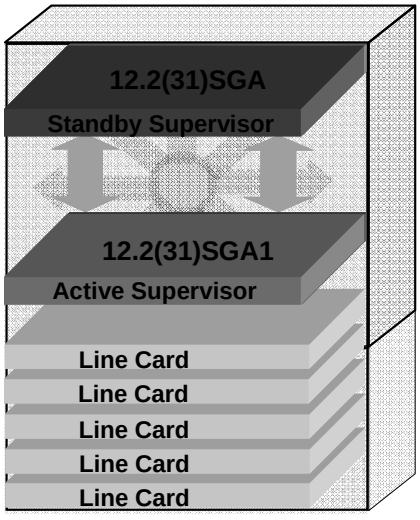


- OSPF-ADJCHG messages appear on the switches after a switchover even though no routes flaps occur during an NSF switchover

No Route Flaps During Recovery

Supervisor Processor Redundancy

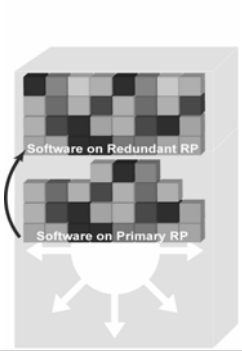
In Service Software Upgrade (ISSU)



- ISSU provides a mechanism to perform software upgrades and downgrades without taking the switch out of service
- Leverages the capabilities of NSF and SSO to allow the switch to forward traffic during supervisor IOS upgrade (or downgrade)
- ISSU provides for full image upgrade which allows for both
  - The addition of new features
  - Upgrades to address any defects (PSIRT's and patches)

System High Availability

In Service Software Upgrade (ISSU)

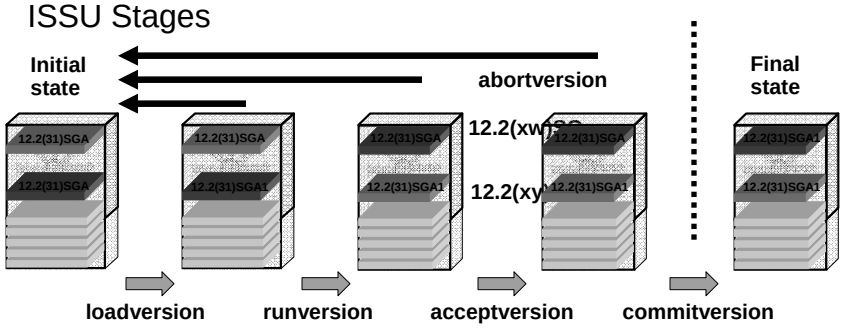


- Catalyst 4500 is utilizing full image upgrade ISSU
- The use of full image upgrade allows for both
  - The addition of new features
  - Upgrades to address any defects (PSIRTS and patches)

•Full image upgrade
• New features and patches

Catalyst 4500 ISSU

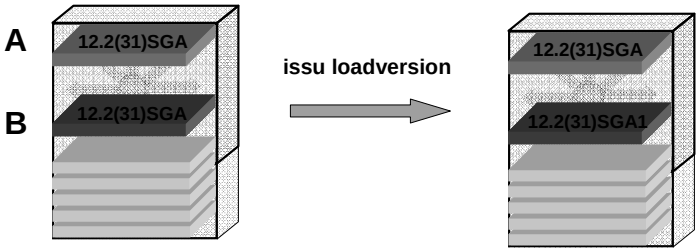
ISSU Stages



- ISSU upgrade is a 4 step process
- Possible to rollback (abort) up until you complete the 4th step (commit to final state)
- Leverages NSF/SSO to implement supervisor transition
- Requires that the two images are compatible for upgrade/downgrade processing

ISSU Upgrade Process

Step 1—loadversion



- Initiate the process by loading the new image into flash on both supervisors
- Issue the 'issu loadversion' command to reboot the standby supervisor (B) using the new image
- If an incompatible image is detected and SSO mode not achievable the switch automatically aborts the ISSU process and reboots the standby (B) with the previous version
- An abort issued now causes the standby (B) to reset and load original image

12.2(xw)SG

12.2(xy)SG

ISSU Upgrade Process

Step 1—loadversion

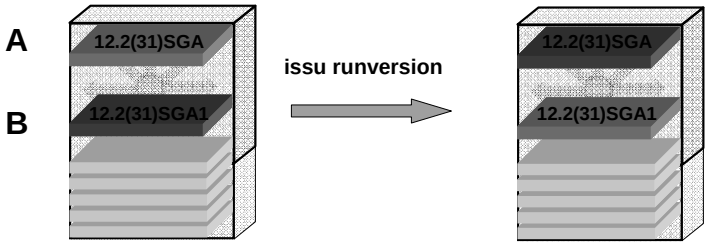
```
4507R#sh issu state
                        Slot = 1
                   RP State = Active
                  ISSU State = Init
                 Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12;

                        Slot = 2
                   RP State = Standby
                  ISSU State = Init
                 Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12;


4507R#issu loadversion 1 bootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin 2
        slavebootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin
```

Standby Supervisor Reboots with new image

```
4507R#sh issu state
                        Slot = 1
                   RP State = Active
                  ISSU State = Load Version
                 Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12

                        Slot = 2
                   RP State = Standby
                  ISSU State = Load Version
                 Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12
```


ISSU Upgrade Process

Step 2—runversion



- Issue the 'issu runversion' command to initiate an SSO failover to the standby supervisor (B) running the new image **12.2(xw)SG**
- Old active supervisor (A) reboots with the old image into standby mode **12.2(xy)SG**
- System is still in SSO mode and rollback timer is started
- An abort issued now causes the newly active supervisor (B) to failover to the standby supervisor (A) running the old image and will also cause the rebooting supervisor (B) to load the original image

ISSU Upgrade Process

Step 2—runversion

```
4507R#issu runversion 2 slavebootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin
This command will reload the Active unit.  Proceed ? [confirm]
```

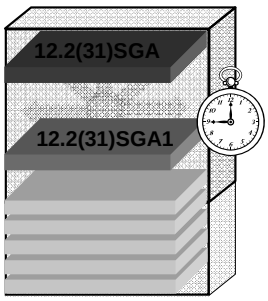
SSO Failover to Redundant Supervisor running new image

```
4507R#sh issu state
                    Slot = 2
               RP State = Active
            ISSU State = Run Version
          Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12

                    Slot = 1
               RP State = Standby
            ISSU State = Run Version
          Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12


4507R#sh version
 . . .

Uptime for this control processor is 8 minutes
System returned to ROM by Stateful Switchover
System image file is "bootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin"

cisco WS-C4507R (MPC8540) processor (revision 10) with 524288K bytes of memory.
```


Catalyst 4500 ISSU

Rollback Timer

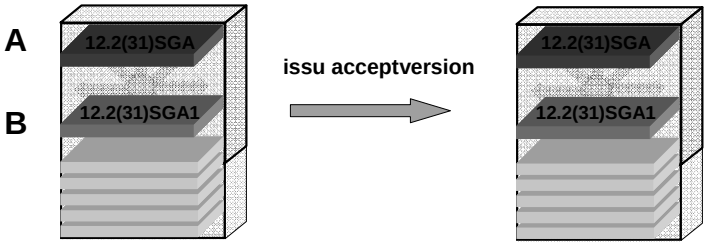
- On issuing the ‘runversion’ command the system activates the rollback timer
- Provides a recovery mechanism to trigger an abort to return the switch to it’s original state if you lose connectivity during upgrade
- You can disable the rollback by setting the timer to ‘0’




```
cr39-4507-1#sh issu rollback-timer
       Rollback Process State = Not in progress
     Configured Rollback Time = 45:00


cr39-4507-1(config)#issu set rollback-timer ?
  <0-7200>  Rollback timer value
```


ISSU Upgrade Process

Step 3—acceptversion



- Prior to issuing the ‘issu acceptversion’ command the system will be counting down the rollback timer **12.2(xw)SG**
- If ‘issu acceptversion’ is not completed before rollback timer expires and automatic abort will occur **12.2(xy)SG**
- No features available on the new version will work yet
- An abort issued now causes the newly active supervisor (B) to failover to the standby supervisor (A) running the old image and will also cause the rebooting supervisor (B) to load the original image

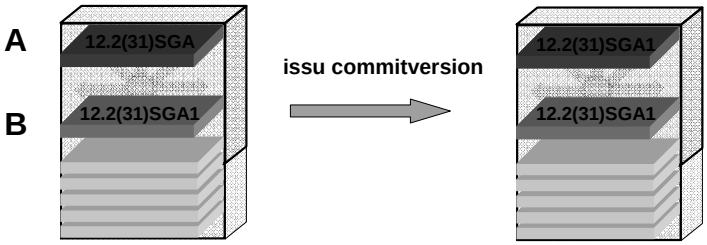
ISSU Upgrade Process

Step 3—acceptversion

```
4507R#show issu rollback-timer
        Rollback Process State = In progress
      Configured Rollback Time = 45:00
       Automatic Rollback Time = 39:19

4507R#issu acceptversion 2
% Rollback timer stopped. Please issue the commitversion command.

4507R#show issu rollback-timer
        Rollback Process State = Not in progress
      Configured Rollback Time = 45:00


4507R#show bootvar
BOOT variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12
CONFIG_FILE variable does not exist
BOOTLDR variable does not exist
Configuration register is 0x2102

Standby BOOT variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12
Standby CONFIG_FILE variable does not exist
Standby BOOTLDR variable does not exist
Standby Configuration register is 0x2102
```


ISSU Upgrade Process

Step 4—commitversion



- The acceptversion state is not meant for long term network operation
- Once network is confirmed stable and change evaluation criteria are met issue the ‘issu commitversion’ command
- On commitversion the standby supervisor (A) reboots and loads the new image coming up in standby mode
- New IOS features are enabled at this point
- If required to back out change will need to restart the 4 step ISSU process implementing a software downgrade

ISSU Upgrade Process

Step 4—commitversion

```
4507R#$issu commitversion 1 slavebootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin
```

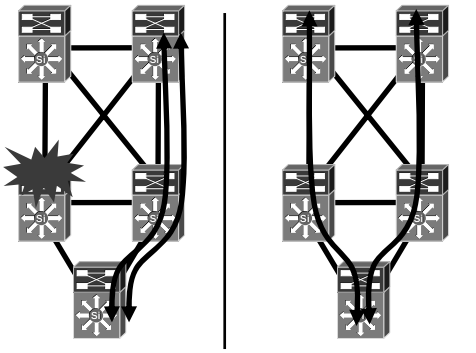
Standby Supervisor Reboots with new image

```
4507R#show issu state
                        Slot = 2
                   RP State = Active
                 ISSU State = Init
                Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin,12;
                        bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12;

                        Slot = 1
                   RP State = Standby
                 ISSU State = Init
                Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin,12;
                        bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12;

4507R#show bootvar
BOOT variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin,12;
          bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12;
CONFIG_FILE variable does not exist
BOOTLDR variable does not exist
Configuration register is 0x2102

Standby BOOT variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin,12;
          bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12;
Standby CONFIG_FILE variable does not exist
Standby BOOTLDR variable does not exist
Standby Configuration register is 0x2102
```


Redundant Supervisors

In Service Software Upgrade (ISSU)

- In redundant topology standard maintenance practice is to shut down devices during upgrade and let the network converge
- ISSU provides the ability to upgrade software in place without having to shut down
- In the access layer or any other single point of failure this can be a significant improvement in operational practices

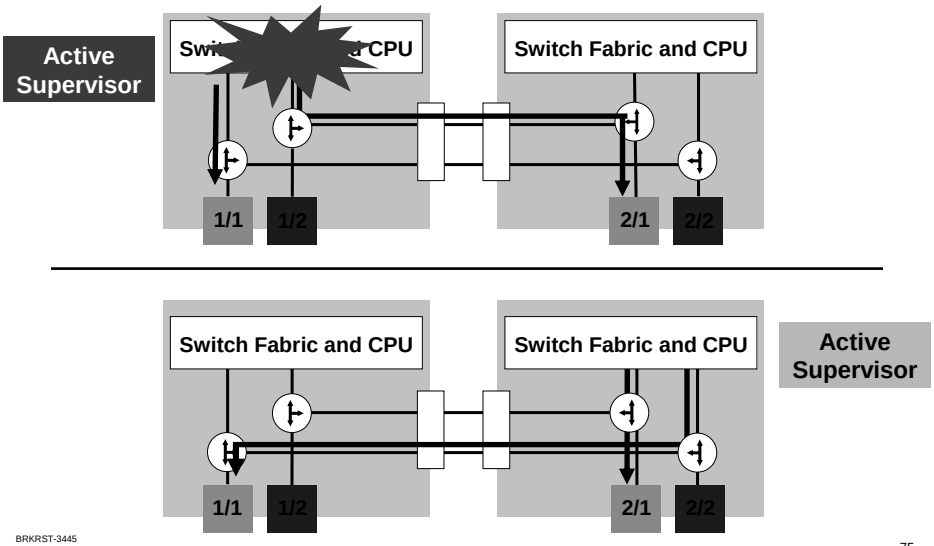


Scheduled Maintenance—Half Capacity

ISSU—All Paths and Switches Active During Upgrade

Redundant Supervisor

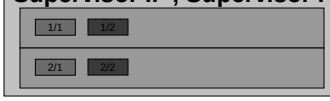
Uplink Behavior with Supervisor Redundancy



Catalyst 4507R Supervisor Redundancy

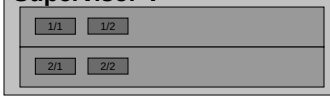
Uplink Redundancy as of 12.2(25)SG

Supervisor II+, Supervisor IV



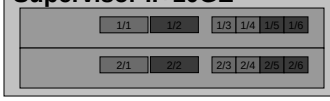
- **2 x GigE ports are active**

Supervisor V



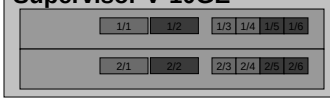
- **4 x GigE ports are active concurrently**

Supervisor II+10GE



- **2 x 10GE and 4 x GigE ports are active**

Supervisor V-10GE

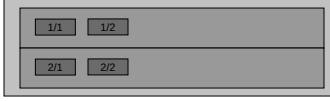


- **2 x 10GE and 4 x GigE ports are active**

Catalyst 4510R Supervisor Redundancy

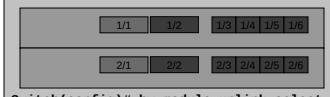
Uplink Redundancy as of 12.2(25)SG

Supervisor V



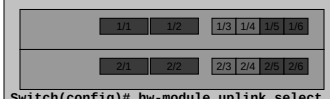
- **4 x GigE ports are active concurrently (10th Slot FlexSlot – only usable with WS-X4302-GB)**

Supervisor V-10GE (3 selectable modes – "hw-module uplink select ..")



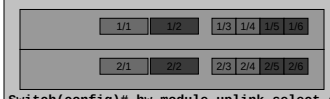
- **2 x 10GE ports are active (10th Slot available for line card) - Default**

```
Switch(config)# hw-module uplink select tengigabitethernet
```



- **4 x GigE ports are active (10th Slot available for line card)**

```
Switch(config)# hw-module uplink select gigabitethernet
```



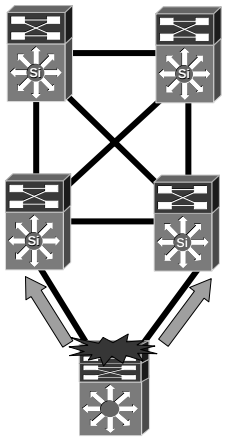
- **2 x 10GE ports and 4 x GigE ports are active concurrently (10th Slot FlexSlot – only usable with WS-X4302-GB)**

```
Switch(config)# hw-module uplink select all
```


Redundant Supervisors

All Uplinks Stay Active

- Standby uplink port is active and forwarding traffic as long as standby Supervisor is fully inserted
- HSRP does not flap & PIM DR does not move
- You maintain Network Topology and Capacity
- Note: If one of the uplinks on either the active or standby supervisor goes down (e.g. fiber cut) the other inactive uplink ports will 'not' become active



System High Availability

System Diagnostics

Power on Diagnostics

- Supervisor
  - Backplane connections
  - PPE ASIC
  - FFE ASIC
  - Memory
  - Ports
- On-line diagnostics
  - Power Supply
  - Fan Tray
  - Modules

On-Going Health Checks

- Module ASICs
- Supervisor memory
- Supervisor redundancy
- Software-hardware state consistency (e.g., L3 table consistency)
- Temperature
- Power-supply
- Fan tray

User Notified via Console/Syslog Error Messages and Crash Information for Development Analysis

System High Availability

Checking Diagnostics

- Detailed diagnostics results including on-going memory tests on supervisor

```
show diagnostic result module <slot_id> detail
```

- Diagnostics test result saved in bootflash device if the diagnostics has failed on bootup
- Reset active or standby supervisor or other modules using the following command

```
hw-module module <slot_id> reset
```

- Resetting modules force the switch to perform online diagnostics

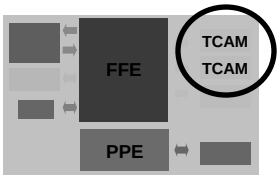
Agenda

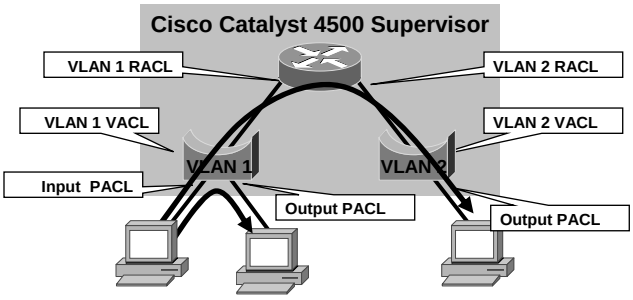
- 4500 Family Overview
- Supervisor Architecture and Packet Forwarding
- Line Card Architecture
- System High Availability
- QoS, Security, and TCAMs
- System CPU and Control Plane Policing

PACLs, RACLs, And VACLs

Security Features Utilize TCAM Resources

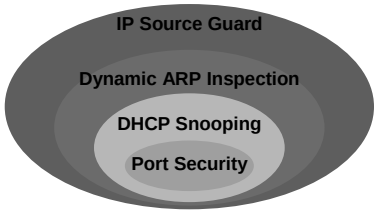
- Each packet being forwarded by the switch is processed through multiple logical ACLs
- All of these ACLs are implemented in security (feature) TCAMs in the supervisor





Catalyst Integrated Security Features

Security Features Utilize TCAM resources



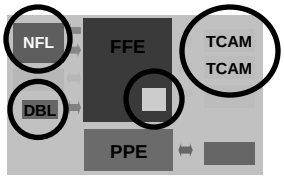
- Port security prevents MAC flooding attacks
- DHCP snooping prevents client attack on the switch and server
- Dynamic ARP Inspection adds security to ARP using DHCP snooping table
- IP source guard adds security to IP source address using DHCP snooping table

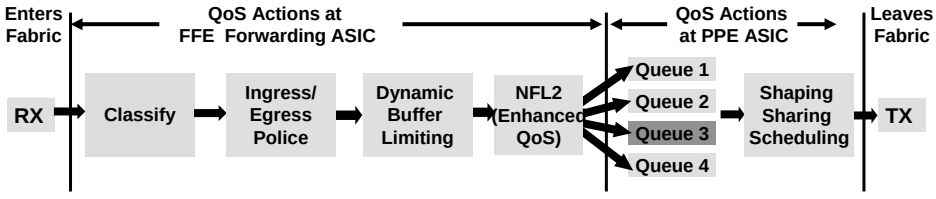
- A wide variety of common security vulnerabilities are mitigated through the use of the CISF's features
- Each of these features requires packets to pass multiple security checks
- These security checks are implemented in hardware utilizing a series of ACL's stored in the same TCAM's as the regular ACL's

Cisco Catalyst 4500 QoS Features

Security Features Utilize TCAM Resources

- Catalyst 4500 implements a sophisticated suite of QoS features
- These QoS features are implemented with three major components
  - TCAMs (Policers)
  - Netflow Feature (UBRL on SupV-10GE)
  - Dynamic Buffer Limiting (DBL)





Dynamic Buffer Limiting

- Congestion avoidance technique
- Flow-Based and maintains flow table per queue
- Flow is identified by source/destination/protocol fields in IP header
- Optionally can include Layer4 ports and VLAN
- Tracks buffer usage and credits available of each flow on tracked interface
- Limits the amount of buffer used by per-flow on a per queue per interface basis
- Packets exceeding limit can either be dropped or marked Explicit Congestion Notification (ECN) bit in the ToS byte of IP header
- DBL is implemented in hardware (no performance penalty)

DBL: Flow Classification

- DBL classifies the flows in to two categories

  Adaptive flows—respond to congestion notification (dropped packet, or ECN) by the switch by reducing the rate of transmission at the source

  Aggressive flows—Do not take any such corrective action in response to a congestion notification

DBL Allows Adaptive Flows a Fair Use of the Transmit Queue in Presence of Aggressive Flows

```
qos
qos dbl
class-map match-all vlan30
  match any
!
policy-map vlan30_policy
  class vlan30
    dbl
!
interface Vlan30
service-policy input vlan30_policy
```


Security, QoS And TCAMs

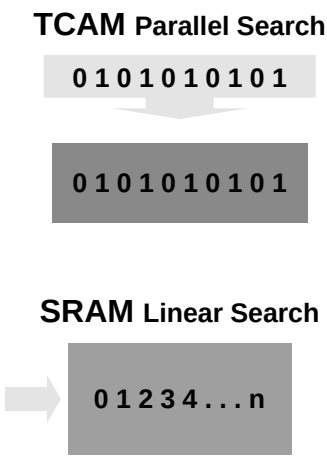
This Is not a Good Sign

- %C4K_HWACLMAN-4-ACLHWPROGERRREASON: (suppressed one time) input(null,12/normal) Security: 140—insufficient hardware TCAM masks
- %C4K_HWACLMAN-4-ACLHWPROGERR: (suppressed four times) input security: 140— hardware TCAM limit, some packet processing will be software switched

Security, QoS And TCAMs

Okay what Is a TCAM

- Ternary Content Addressable Memory (TCAM)
- Unique form of high-speed memory that allows lookups using a variety of inputs
- Provides maskable lookups which allows for searches on a sophisticated set of criteria
- Memory lookup is accomplished in parallel rather than in a serial search fashion resulting in a deterministic lookup time for every query



Security, QoS And TCAMs

TCAMs on GE Supervisors (Sup IV, Sup II-Plus, …)

- One mask per eight entries
- Example

```
access-list 101 permit
ip host 8.1.1.1 any

access-list 101 deny ip
8.1.1.0 255.255.255.0
any
```

Number of Masks Used = 2
Number of Masks Available = 0
Number of Entries Used = 2
Number of Entries Available = 14
(for the Two Masks Defined)

Masks	Patterns
Mask One Match: All 32 Bits of Source IP Address	Src IP = 8.1.1.1
	Empty 2
	Empty 3
	Empty 4
	Empty 5
Don't Care: All Remaining Bits	Empty 6
	Empty 7
	Empty 8
Mask Two Match: Most Significant 24 Bits of Source IP Addr	Src IP = 8.1.1.1
	Empty 2
	Empty 3
	Empty 4
	Empty 5
Don't Care: All Remaining Bits	Empty 6
	Empty 7
	Empty 8

Security, QoS And TCAMs

TCAMs on 10GE Supervisors (Sup V 10GE, …)

- One mask per one entry
- Example

```
access-list 101 permit ip
host 8.1.1.1 any

access-list 101 deny ip
8.1.1.0 255.255.255.0 any
```

Number of Masks Used = 2
Number of Masks Available = 14
Number of Entries Used = 2
Number of Entries Available = 14

Additional TCAM Masks Left = 87.5 %
Optimized for Security/QoS
Features Such as IPSG/pvQoS

Masks	Patterns
Mask 32 Bits for IP1	Src IP = 8.1.1.1
Mask 24 Bits for IP2	Src IP = 8.1.1.0
Empty Mask 3	Empty 3
Empty Mask 4	Empty 4
Empty Mask 5	Empty 5
Empty Mask 6	Empty 6
Empty Mask 7	Empty 7
Empty Mask 8	Empty 8
Empty Mask 9	Empty 9
Empty Mask 10	Empty 10
Empty Mask 11	Empty 11
Empty Mask 12	Empty 12
Empty Mask 13	Empty 13
Empty Mask 14	Empty 14
Empty Mask 15	Empty 15
Empty Mask 16	Empty 16

TCAM Scaling

GE vs. 10GE Supervisors

```
SupIV#sh platform hardware acl statistics utilization brief
                          Entries/Total(%)  Masks/Total(%)
                          ----------------  --------------
Input  Acl(PortAndVlan)     0 / 8112 (  0)     0 / 1014 (  0)
Input  Acl(PortOrVlan)      0 / 8112 (  0)     0 / 1014 (  0)
Input  Qos(PortAndVlan)     0 / 8128 (  0)     0 / 1016 (  0)
Input  Qos(PortOrVlan)      0 / 8128 (  0)     0 / 1016 (  0)
Output Acl(PortAndVlan)     0 / 8112 (  0)     0 / 1014 (  0)
Output Acl(PortOrVlan)      0 / 8112 (  0)     0 / 1014 (  0)
Output Qos(PortAndVlan)     0 / 8128 (  0)     0 / 1016 (  0)
Output Qos(PortOrVlan)      0 / 8128 (  0)     0 / 1016 (  0)


SupV-10GE#sh platform hardware acl statistics utilization brief
                          Entries/Total(%)  Masks/Total(%)
                          ----------------  --------------
Input  Acl(PortAndVlan)     0 / 8112 (  0)     0 / 8112 (  0)
Input  Acl(PortOrVlan)      0 / 8112 (  0)     0 / 8112 (  0)
Input  Qos(PortAndVlan)     0 / 8128 (  0)     0 / 8128 (  0)
Input  Qos(PortOrVlan)      0 / 8128 (  0)     0 / 8128 (  0)
Output Acl(PortAndVlan)     0 / 8112 (  0)     0 / 8112 (  0)
Output Acl(PortOrVlan)      0 / 8112 (  0)     0 / 8112 (  0)
Output Qos(PortAndVlan)     0 / 8128 (  0)     0 / 8128 (  0)
Output Qos(PortOrVlan)      1 / 8128 (  0)     1 / 8128 (  0)

L4Ops: used 2 out of 128
```

Eight ACEs to One Mask TCAM2

1000 IPSG Addresses

One ACEs to One Mask TCAM3

5000 IPSG Addresses

Security, QoS, And TCAMs

TCAM Regions

Input Acl (PortAndVlan) (8k Entries)
Input Acl (PortorVlan) (8k Entries)
Input QoS (PortAndVlan) (8k Entries)
Input QoS (PortOrVlan) (8k Entries)

Output Acl (PortAndVlan) (8k Entries)
Output Acl (PortorVlan) (8k Entries)
Output QoS (PortAndVlan) (8k Entries)
Output QoS (PortOrVlan) (8k Entries)

```
cr7-4507-1#sh platform hardware acl statistics utilization brief
                          Entries/Total(%)  Masks/Total(%)
                          ----------------  --------------
Input  Acl(PortAndVlan)     0 / 8112 (  0)     0 / 8112 (  0)
Input  Acl(PortOrVlan)      0 / 8112 (  0)     0 / 8112 (  0)
Input  Qos(PortAndVlan)     0 / 8128 (  0)     0 / 8128 (  0)
Input  Qos(PortOrVlan)      0 / 8128 (  0)     0 / 8128 (  0)
Output Acl(PortAndVlan)     0 / 8112 (  0)     0 / 8112 (  0)
Output Acl(PortOrVlan)      0 / 8112 (  0)     0 / 8112 (  0)
Output Qos(PortAndVlan)     0 / 8128 (  0)     0 / 8128 (  0)
Output Qos(PortOrVlan)      1 / 8128 (  0)     1 / 8128 (  0)

L4Ops: used 2 out of 128
```


Security, QoS, And TCAMs

TCAM Utilization (Examples)

The ACL and/or QoS Rules Are Programmed into a Specific TCAM Region Depending on How the ACL Needs to Be Applied

```
cr7-4507-1#sh platform hardware acl statistics utilization brief
                          Entries/Total(%)  Masks/Total(%)
                          ----------------  --------------
Input  Acl(PortAndVlan)                                    IP Source Guard
Input  Acl(PortOrVlan)                                     RACLs, VACLs, PACL, MAC ACL in (L2-4)
Input  Qos(PortAndVlan)                                    Port and VLAN Features Merge*
Input  Qos(PortOrVlan)                                     Inbound QoS ACLs (Service-Policy Input)
Output Acl(PortAndVlan)                                    Port and VLAN Features Merge*
Output Acl(PortOrVlan)                                     RACLs, VACLs, PACL, MAC ACL out (L2-4)
Output Qos(PortAndVlan)                                    Port and VLAN Features Merge*
Output Qos(PortOrVlan)                                     Outbound QoS ACLs (Service-Policy Input)
L4Ops: used 2 out of 128                                   Layer 4 Operators within ACL
```

***Used Anytime You Have Port and VLAN Features (ACL/Qos) E.G., Ingress PACL and VACL**

TCAM Optimization

Packed vs. Scattered Supported as of 12.2(20)EW

- Multiple ACLs with a regular pattern may not be optimally stored in the TCAMs
- As an example when IP source guard is configured multiple ACLs with a common format and common mask need to be programmed
- If using GE supervisors (Sup II-Plus, Sup IV, Sup V) changing the method used to program ACLs into the TCAM to ‘scattered’ may improve utilization
- Caution: confirm the result in a lab or during a change window prior to implementing in production

```
cr39-4507-1(config)#access-list hardware entries ?
  packed     Program entries packed
  scattered  Program entries scattered

cr39-4507-1(config)#access-list hardware entries scattered
```


TCAM Optimization

TCAM Resizing Supported as of 12.1(31)SG

- By default, each of the PortOrVLAN and PortAndVLAN regions in TCAM are allocated 50% of the total region
- TCAM resizing allows resizing based on percentage allocation between the two regions components
- On resizing, all the ACLs will be unloaded and reloaded back in the hardware
- During the time ACLs are unloaded and reloaded, no ACLs will be applied to the traffic
- The TCAM resizing cannot be done between regions belonging to different features/areas

```
cr39-4507-1(config)#access-list hardware region ?
  feature  Configure regions in the Feature TCAM
  qos      Configure regions in the QoS TCAM
```

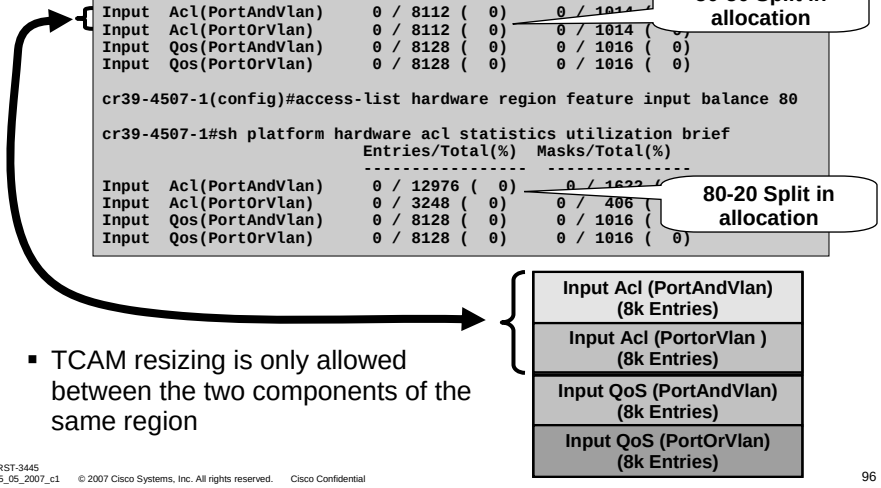

TCAM Optimization

TCAM Resizing Supported as of 12.2(31)SG

```
cr39-4507-1#sh platform hardware acl statistics utilization brief
                          Entries/Total(%)  Masks/Total(%)
                          ----------------  --------------
Input  Acl(PortAndVlan)     0 / 8112 (  0)     0 / 1014 (  0)
Input  Acl(PortOrVlan)      0 / 8112 (  0)     0 / 1014 (  0)
Input  Qos(PortAndVlan)     0 / 8128 (  0)     0 / 1016 (  0)
Input  Qos(PortOrVlan)      0 / 8128 (  0)     0 / 1016 (  0)

cr39-4507-1(config)#access-list hardware region feature input balance 80

cr39-4507-1#sh platform hardware acl statistics utilization brief
                          Entries/Total(%)  Masks/Total(%)
                          ----------------  --------------
Input  Acl(PortAndVlan)     0 / 12976 (  0)     0 / 1622 (  0)
Input  Acl(PortOrVlan)      0 / 3248 (  0)     0 /  406 (  0)
Input  Qos(PortAndVlan)     0 / 8128 (  0)     0 / 1016 (  0)
Input  Qos(PortOrVlan)      0 / 8128 (  0)     0 / 1016 (  0)
```

50-50 Split in allocation

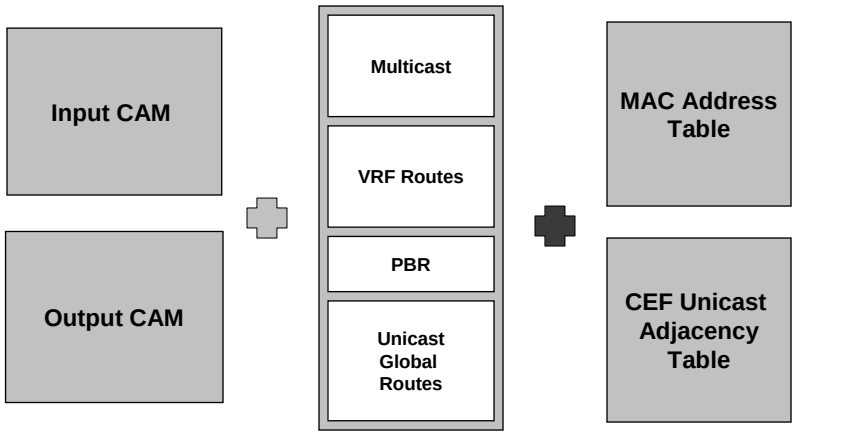
80-20 Split in allocation



- TCAM resizing is only allowed between the two components of the same region

TCAM Optimization

TCAM Capacity



(32k + 32k) **128k** **(55,000 + 32,000)**

TCAM Optimization

Supervisor ACL Resources Comparison

Product	Feature TCAM (Per Direction)	QoS TCAM (Per Direction)	L4 Operators (GT, LT, NEQ, Range)
Supervisor II+ Supetrvisor II+ TS	8K Entries 1K Masks	8K Entries 1K Masks	64 (6 per ACL)
Supervisor IV/V WS-C4948	16K Entries 2K Masks	16K Entries 2K Masks	64 (6 per ACL)
Supervisor II+10GE	8K Entries 8K Masks	8K Entries 8K Masks	128 (8 per ACL)
Supervisor V-10GE WS-C4948-10GE	16K Entries 16K Masks	16K Entries 16K Masks	128 (8 per ACL)

When Using Security Features That Require Large TCAM Resources It Is Recommended to Run 12.2(20)EW or Later Due to TCAM Programming Optimizations

Agenda

- 4500 Family Overview
- Supervisor Architecture and Packet Forwarding
- Line Card Architecture
- System High Availability
- QoS, Security, and TCAMs
- System CPU and Control Plane Policing

Software Switching and CPU Traffic

Hardware Switches Still Use the CPU

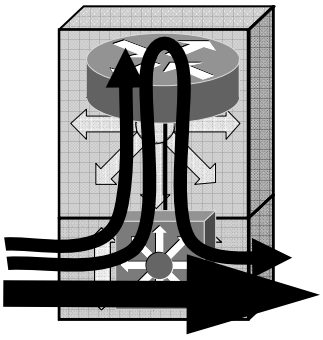
- Cisco Catalyst switches operate by forwarding traffic in hardware
- Some traffic is sent to the CPU for processing

  ESMP, BPDU, 802.1X, DCHP, ARP, IGMP, CDP, EIGRP, OSPF, —

  Telnet, SSH, SNMP, —
- Some traffic is forwarded in HW but also copied to the CPU for processing

  ACL logging, SA learning, —
- Some traffic is forwarded by the CPU

  IPX, Appletalk, GRE, —



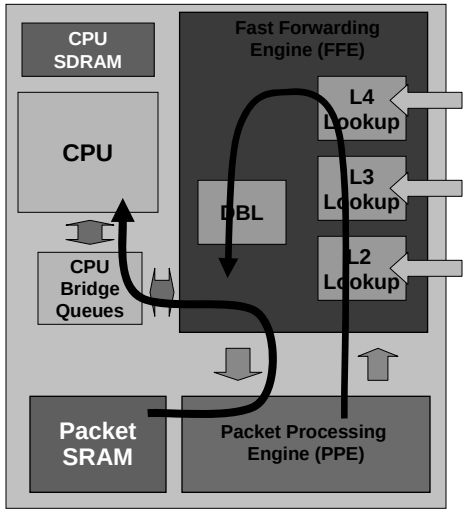
Software Switching and CPU Traffic

CPU Bound Traffic

- Packets are identified at multiple stages in the FFE pipeline as candidates for punting or copying to the CPU
- L2 lookups based on well know MAC addresses

  BPDU (1:80:c2:00:00 - 0f)

  CDP/PAgP (01:00:0c:cc:cc:cc)
- Packets with DMAC = Switch L3 interface
- Input ACL's also defined to identify and classify specific traffic types needing CPU processing



Software Switching and CPU Traffic

L2 Control Packets Sent to the CPU

```
cr39-4507-1#show platform hardware l2-lookup-cam
Bucket Protocols
------ ---------
0      Ipv4 Esmp ArpIpv4 RarpIpv4
1      Ipx
2      AppleTalk DecNet
3      Ipv6 Other Vines Xns Arp Rarp

Flags are:
----------
R - to router
D - drop

Index  Mac Address     Vlan   Pkt Flags Type       Destination Port(s)
-----  --------------  -----  --- ----- ---------- ----------------------------
0      0180.C200.0000  0      0         agg 519    Cpu aggport
masks  FFFF.FFFF.FFF0  0x000  0
1      0100.0CCC.CCCC  0      0         agg 519    Cpu aggport
masks  FFFF.FFFF.FFF0  0x000  0
2      0180.C200.0020  0      0         agg 519    Cpu aggport
masks  FFFF.FFFF.FFF0  0x000  0
3      0010.7BAB.9932  0      0         agg 519    Cpu aggport
masks  FFFF.FFFF.FFFF  0x000  0
4      0010.7BAB.996F  0      0         agg 519    Cpu aggport
masks  FFFF.FFFF.FFFF  0x000  0
5      0010.7BAB.9933  0      0   D     agg 520    Drop aggport
masks  FFFF.FFFF.FFFF  0xFFF  3
6      0010.7BAB.9933  0      0   D     agg 520    Drop aggport
masks  FFFF.FFFF.FFFF  0xFFF  3
7      0010.7BAB.9933  0      0   D     agg 520    Drop aggport
masks  FFFF.FFFF.FFFF  0xFFF  3
```

Packets Matching the Lookup Criteria Are Marked to Be Forwarded to the CPU Gigaport (Virtual Interface for the CPU)

Software Switching and CPU Traffic

Static Input ACLs

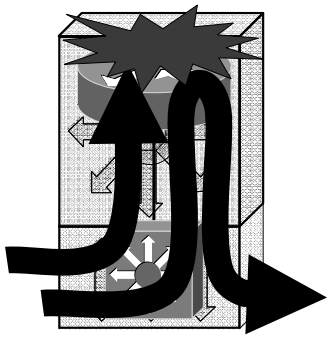
```
cr39-4507-1#show platform hardware acl input entries sw-allocated

        Input Acl Cam Table

Idx: 17 / 3
    VTag: 0 / 0 PTag: 0 / 0 FlType: 0 / 1
    L4PortOpCode: 0 / 0
    Mac DA : 01:80:C2:00:00:03 / FF:FF:FF:FF:FF:FF
    The protocol family is completely masked out.
RawFlowLabel: 00006030 800000C0 00000000 00000000 /
003FFFFF FFFFFFC0 00000000 00000000

    ActIdx: 23 CTCOnce: false

. . .

Idx: 64 / 8
    VTag: 0 / 0 PTag: 0 / 0 FlType: 1 / 1
    L4PortOpCode: 0 / 0
    ipSrcAddr : 0.0.0.0 / 0.0.0.0
    ipDstAddr : 224.0.0.0 / 255.255.255.0
    ipTos : 0 / 0
    ipProtocol : ospf / 255
    ipMoreFragment : false / false
    ipFragment : false / false
    l4Data : 0 / 0
    rsvd : 0 / 0

    ActIdx: 17 CTCOnce: false
```

Input ACLs Provide More Specific Filtering for L2 and Provide L3/4 Selection Criteria for CPU Traffic

Software Switching and CPU Traffic

CPU Overload Protection Mechanisms

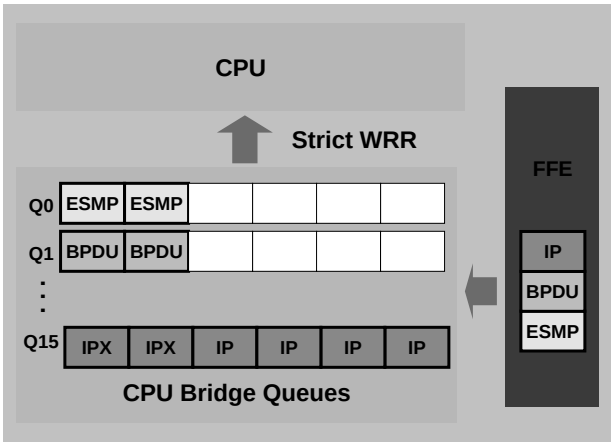
- When the volume of CPU traffic being sent to the CPU exceeds the processor capacity the switch can become unstable
- A number of mechanisms have been built into the switch to prevent this
  - CPU input queues
  - DBL policing of CPU bound traffic
  - IOS based SW throttling mechanisms, e.g.
    - Selective Packet Discard
    - ARP, DHCP, DAI rate limiting
- Control Plane Policing (CoPP)



Software Switching And CPU Traffic

CPU Queues

- 16 Queues implemented in the CPU bridge
- Each traffic type is assigned to a unique queue
- CPU drains each queue using a strict weighted round-robin algorithm
- Guarantees control plane packets receive priority



Software Switching and CPU Traffic

CPU Queues

	Queue Name	Packets Using the Queued
0	ESMP	ESMP Packets (Internal Management Packets) for the Line Card ASICs
1	Control	L2 Control Plane Packets, Such as STP, CDP, PAgP, LACP, or UDLD
2	Host Learning	Packets Copied to the CPU to Build the L2 Forwarding Table
3, 4, 5	L3 Forwarded Traffic	Packets That Must Be L3 Forwarded in Software • GRE-tunneled packets • ARP requests
6, 7, 8	L2 Forwarded Traffic	Packets That Are Bridged to the CPU MAC Address • IPX, Appletalk • ARP request and response • IP header options, expired TTL, non-ARPA encapsulation
9, 10	L3 Rx High, L3 Rx Low	L3 Control Plane Traffic, Routing Protocols, Telnet, SNMP, SSH, —
11	RPF Failure	Multicast Packets That Failed the RPF Check
12	ACL Fwd (Snooping)	DHCP Snooping, Dynamic ARP Inspection, or IGMP Snooping Features
13	ACL Log, IP Unreachable	ACL with the Log Keyword or ICMP Unreachable Messages
14	ACL SW Processing	Punted to the CPU Due to a Lack of Additional ACL Hardware Resources
15	MTU Fail/Invalid	Packets That Need to Be Fragmented

Software Switching and CPU Traffic

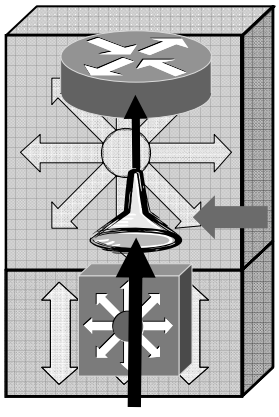
CPU Queues

```
cr39-4507-1#sh platform cpu packet driver
Hardware Accelerated Packet Engine Rev 2
Tx packets 1353409
Tx descriptors used 0
Tx dropped 0
Restarted 0
   Queue              rxTail   received  all guar allJ gurJ   rxDrops  rxDelays
 0 Esmp               1C6F53D4    1313450   98  100    0    5         0         0
 1 L2/L3Control       1C6F6AF8      11431 2497 2500    0    5         0         0
 2 Host Learning      1C6F7C34         34  498  500    0    5         0         0
 3 L3 Fwd High        1C6F837C          0  300  300    0    5         0         0
 4 L3 Fwd Medium      1C6F882C          0  500  500    0    5         0         0
 5 L3 Fwd Low         1C6F8FFC          0  900  900    0    5         0         0
 6 L2 Fwd High        1C6F9E0C          0  300  300    0    5         0         0
 7 L2 Fwd Medium      1C6FA2BC          0  500  500    0    5         0         0
 8 L2 Fwd Low         1C6FAFA0      15625  899  900    0    5         0         0
 9 L3 Rx High         1C6FB990         61  299  300    0    5         0         0
10 L3 Rx Low          1C6FC174      31466  298  300    0    5         0         0
11 RPF Failure        1C6FC1FC          0  200  200    0    5         0         0
12 ACL fwd(snooping)  1C6FC51C          0  100  100    0    5         0         0
13 ACL log, unreach   1C6FC6AC          0  200  200    0    5         0         0
14 ACL sw processing  1C6FC9CC          0  100  100    0    5         0         0
15 MTU Fail/Invalid   1C6FCB5C          0  102  102    0    5         0         0
```


Software Switching and CPU Traffic

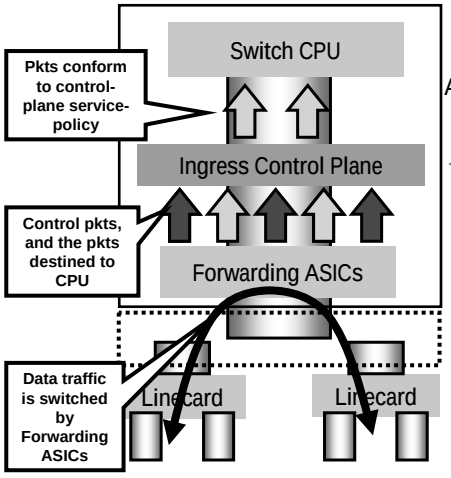
Control Plane Policing (CoPP)

- While CPU queues provide a HW-based mechanism to prioritize traffic to the CPU they do not prevent high CPU conditions nor control which Telnet traffic to prioritize
- Utilizing the HW-based QoS features of the switch you can add an additional layer of CPU control and protection
- Control plane policing uses 4500 QoS policiers applied to the virtual CPU interface ‘control-plane’ to rate limit traffic



Apply Inbound QoS Policy via CoPP

H/W CoPP Overview



Applying Policy

- Pre-configured System Traffic Types
- User Configurable Traffic Types

- Hardware-based mechanisms
- Rate limit CPU bound traffic
- Protect from DoS attacks
- User defined police actions

Control Plane Policing

CoPP Configuration

```
cr39-4507-1(config)#qos
cr39-4507-1(config)#macro global apply system-cpp
```

1. **Enable QoS globally**
2. **Apply the predefined system-cpp macro**

```
cr39-4507-2#sh run
!
class-map match-all system-cpp-cdp
  match access-group name system-cpp-cdp
class-map match-all system-cpp-pim
  match access-group name system-cpp-pim
     . . .
class-map match-all system-cpp-cgmp
  match access-group name system-cpp-cgmp
!
policy-map system-cpp-policy
  class system-cpp-dot1x
  class system-cpp-bpdu-range
      . . .
  class system-cpp-dhcp-ss
!
control-plane
 service-policy input system-cpp-policy
```

Defines class-map Statements for All the Control Plane Traffic Types

Defines the CoPP Policy Map (‘*no*’ Policing Actions Defined by Default)

Applies the Policy Map to the CPU Interface

Control Plane Policing

Predefined ACLs

Predefined Named ACL	Description
system-cpp-dot1x	MacDA = 0180.C200.0003
system-cpp-bpdu-range	MacDA = 0180.C200.0000–0180.C200.000F
system-cpp-cdp	MacDA = 0100.0CCC.CCCC (UDLD/DTP/VTP/Pagp)
system-cpp-garp-range	MacDA = 0180.C200.0020–0180.C200.002F
system-cpp-sstp	MacDA = 0100.0CCC.CCCD
system-cpp-cgmp	Mac DA = 01-00-0C-DD-DD-DD
system-cpp-ospf	IP Protocol = OSPF, IPDA Matches 224.0.0.0/24
system-cpp-igmp	IP Protocol = IGMP, IPDA Matches 224.0.0.0/3
system-cpp-pim	IP Protocol = PIM, IPDA Matches 224.0.0.0/24
system-cpp-all-systems-on-subnet	IPDA = 224.0.0.1
system-cpp-all-routers-on-subnet	IPDA = 224.0.0.2
system-cpp-ripv2	IPDA = 224.0.0.9
system-cpp-ip-mcast-linklocal	IP DA = 224.0.0.0/24
system-cpp-dhcp-cs	IP Protocol = UDP, L4SrcPort = 68, L4DstPort = 67
system-cpp-dhcp-sc	IP Protocol = UDP, L4SrcPort = 67, L4DstPort = 68
system-cpp-dhcp-ss	IP Protocol = UDP, L4SrcPort = 67, L4DstPort = 67

Control Plane Policing

Predefined ACLs (Not Visible in show run)

```
cr39-4507-1#show access-lists

Extended IP access list system-cpp-all-routers-on-subnet
    10 permit ip any host 224.0.0.2
Extended IP access list system-cpp-all-systems-on-subnet
    10 permit ip any host 224.0.0.1
Extended IP access list system-cpp-dhcp-cs
    10 permit udp any eq bootpc any eq bootps (304 matches)
Extended IP access list system-cpp-dhcp-sc
    10 permit udp any eq bootps any eq bootpc
Extended IP access list system-cpp-dhcp-ss
Extended MAC access list system-cpp-bpdu-range
    permit any 0180.c200.0000 0000.0000.000f (1948 matches)
Extended MAC access list system-cpp-cdp
    permit any host 0100.0ccc.cccc (492 matches)
Extended MAC access list system-cpp-cgmp
    permit any host 0100.0cdd.dddd
Extended MAC access list system-cpp-dot1x
    permit any host 0180.c200.0003
Extended MAC access list system-cpp-garp-range
    permit any 0180.c200.0020 0000.0000.000f
Extended MAC access list system-cpp-sstp
    permit any host 0100.0ccc.cccd
```


Control Plane Policing

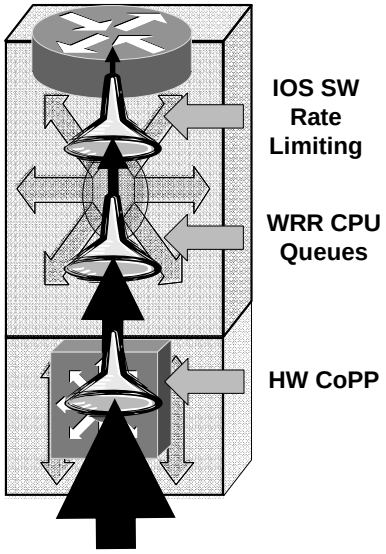
CoPP Configuration

```
cr39-4507-2(config)#policy-map system-cpp-policy
cr39-4507-2(config-pmap)#class system-cpp-dhcp-cs
cr39-4507-2(config-pmap-c)#police 32000 1000 conform-action transmit exceed-action drop

cr39-4507-2(config)#access-list 140 deny tcp 10.120.200.0 0.0.0.255 any eq telnet
cr39-4507-2(config)#access-list 140 permit tcp any any eq telnet

cr39-4507-2(config)#class-map Network-Operations
cr39-4507-2(config-cmap)#match access-group 140
cr39-4507-2(config-cmap)#exit

cr39-4507-2(config)#policy-map system-cpp-policy
cr39-4507-2(config-pmap)#class Network-Operations
cr39-4507-2(config-pmap-c)#police 80000 1000 conform-action transmit exceed-action drop

cr39-4507-2#sh policy-map system-cpp-policy
  Policy Map system-cpp-policy
    Class system-cpp-dot1x
 . . .

    Class system-cpp-dhcp-cs
      police 32000 bps 1000 byte conform-action transmit exceed-action drop
    Class system-cpp-dhcp-sc
    Class system-cpp-dhcp-ss
    Class Network-Operations
      police 80000 bps 1000 byte conform-action transmit exceed-action drop
```


Software Switching And CPU Traffic

Multiple Tiers of Protection

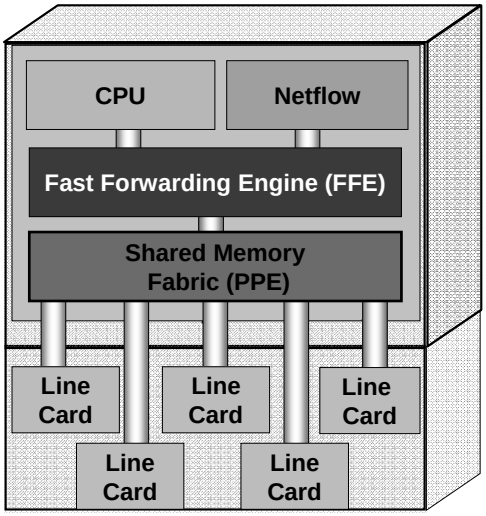
- When both HW and SW based throttling features are enabled both are enforced
- HW policing is performed first
- Then whatever traffic makes it to the CPU is SW throttled
- IOS has a number of inherent throttles
  - ARP and Selective Packet Discard
- Catalyst switches can also configure selective SW throttles
  - DHCP Snooping
  - IGMP Snooping
  - DAI



Catalyst 4500 Architecture

It All Happens in the Sup

- Catalyst 4500 Supervisor is where all functions on the switch are performed
- Wire rate Security and QoS services
- Redundant Supervisors provide for overall enhanced system resiliency
- Understand and manage you TCAM utilization
- You have the security tools use them



More Information

- Catalyst 4500 Power over Ethernet Capabilities White Paper

  http://www.cisco.com/en/US/partner/products/hw/switches/ps4324/products_white_paper09186a00801f44be.shtml

- Catalyst 4500 Security Features Best Practices for Supervisors

  http://www.cisco.com/en/US/partner/products/hw/switches/ps4324/products_white_paper09186a00801faa79.shtml

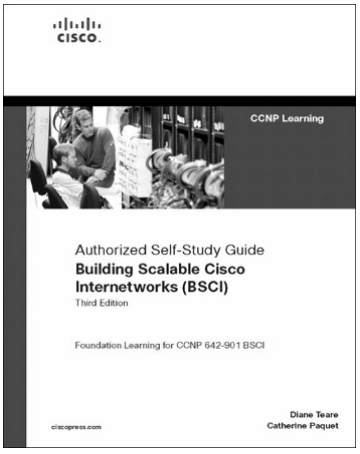
- Catalyst 4500 ISSU Deployment Guide

  http://www.cisco.com/en/US/products/hw/switches/ps4324/products_white_paper0900aecd805e6a95.shtml

Q & A

Recommended Reading

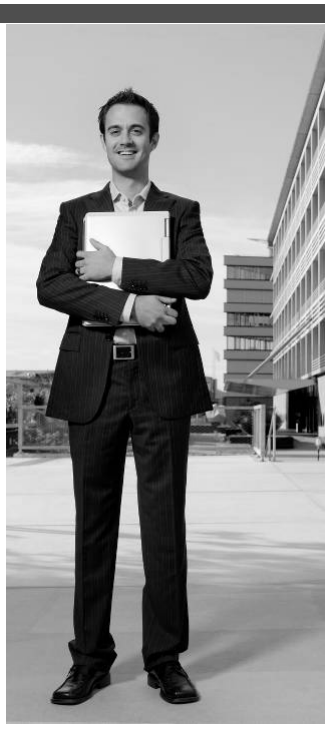
- Continue your Networkers at Cisco Live learning experience with further reading from Cisco Press
- Check the Recommended Reading flyer for suggested books



Available Onsite at the Cisco Company Store

Complete Your Online Session Evaluation

- Win fabulous prizes; give us your feedback
- Receive ten Passport Points for each session evaluation you complete
- Go to the Internet stations located throughout the Convention Center to complete your session evaluation
- Winners will be announced daily at the Internet stations



CISCO

BRKRST-3445
13635_05_2007_c1 © 2007 Cisco Systems, Inc. All rights reserved. Cisco Confidential