Advanced Enterprise Campus High Availability

BRKCAM-3005

Michael Herbert

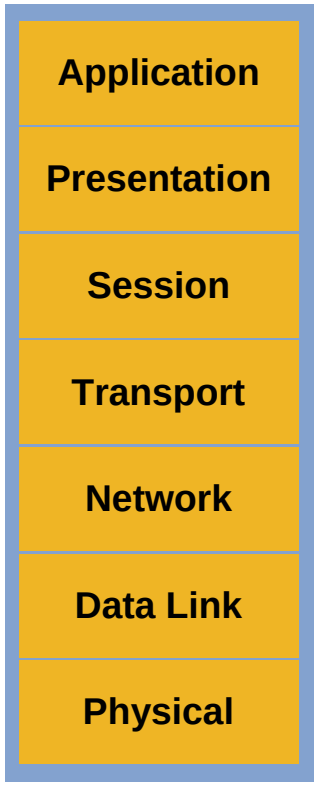
Cisco Networkers 2007

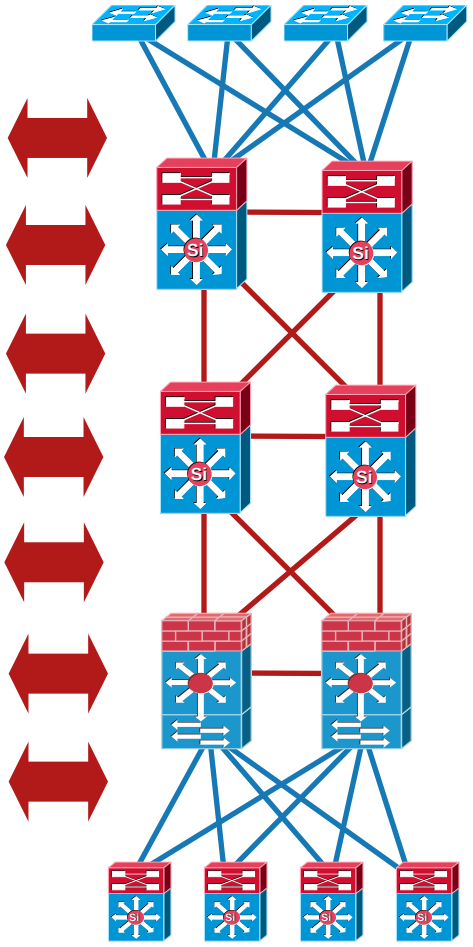
HOUSEKEEPING

- We value your feedback, don't forget to complete your online session evaluations after each session and complete the Overall Conference Evaluation which will be available online from Friday.
- Visit the World of Solutions on Level -01!
- Please remember this is a 'No Smoking' venue!
- Please switch off your mobile phones!
- Please remember to wear your badge at all times including the Party!
- Do you have a question? Feel free to ask them during the Q&A section or write your question on the Question form given to you and hand it to the Room Monitor when you see them holding up the Q&A sign.

High Availability Campus Design
The Resilient Network

- Campus network design is evolving in response to multiple drivers
  - User Expectations: Always ON Access to communications
  - Business Requirements: Globalization means true 7x24x365
  - Technology Requirements: Unified Communications
  - Unexpected Requirements: Worms, Viruses, …
- Designing for availability is no longer just concerned with simple component failures
- Campus design needs to evolve to a ‘resilient’ model

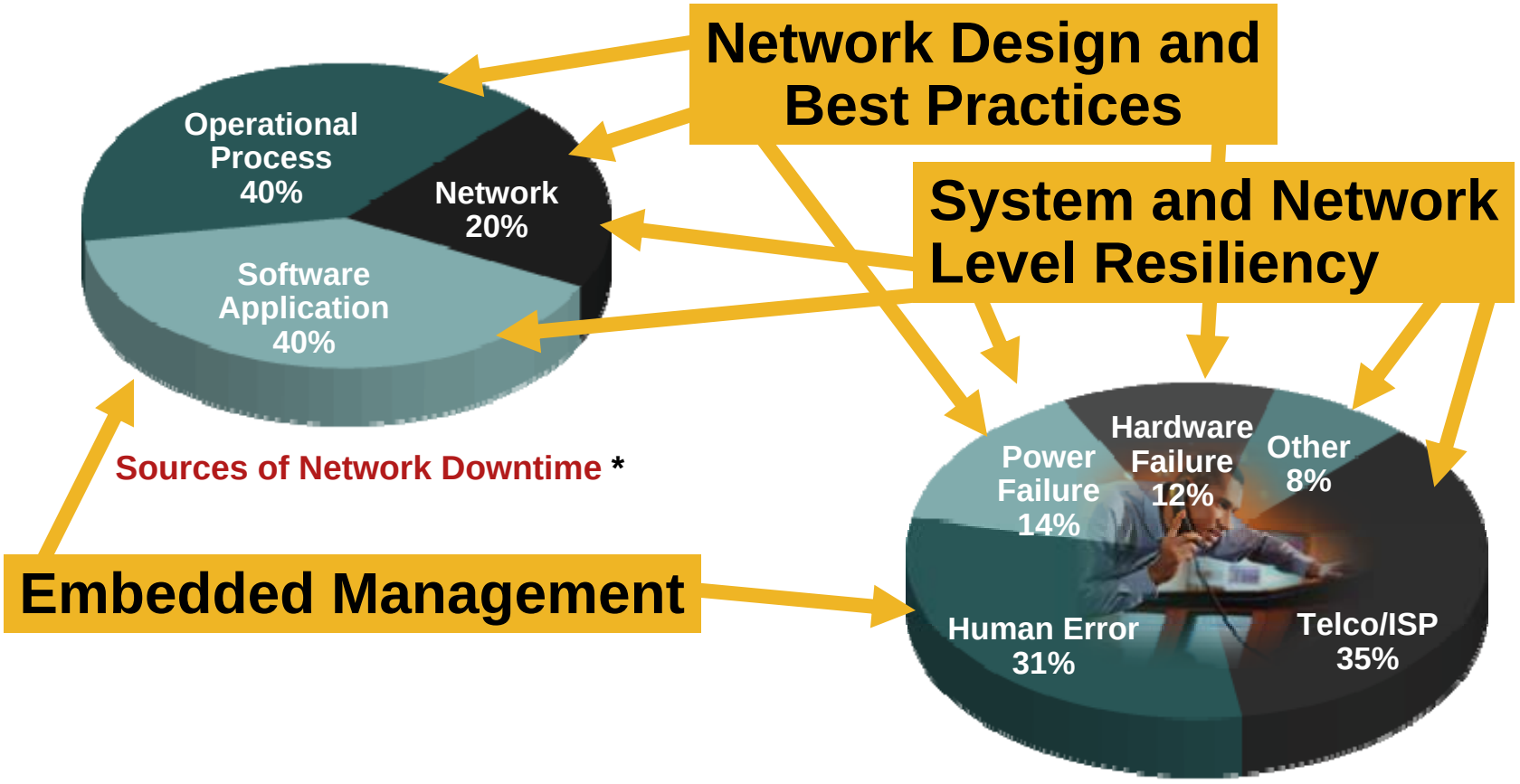




Structured ‘and’ Resilient Design

High Availability Campus Design

Understanding and Addressing all the requirements



*Source: Gartner Group
**Source: Yankee Group The Road to a Five-Nines Network 2/2004

ESE Campus Solution Test Bed

Verified Design Recommendations

Total of 68 Access Switches, 2950, 2970, 3550, 3560, 3750, 4507 SupII+, 4507SupIV, 6500 Sup2, 6500 Sup32, 6500 Sup720 and 40 APs (1200)

Three Distribution Blocks
6500 with Redundant Sup720

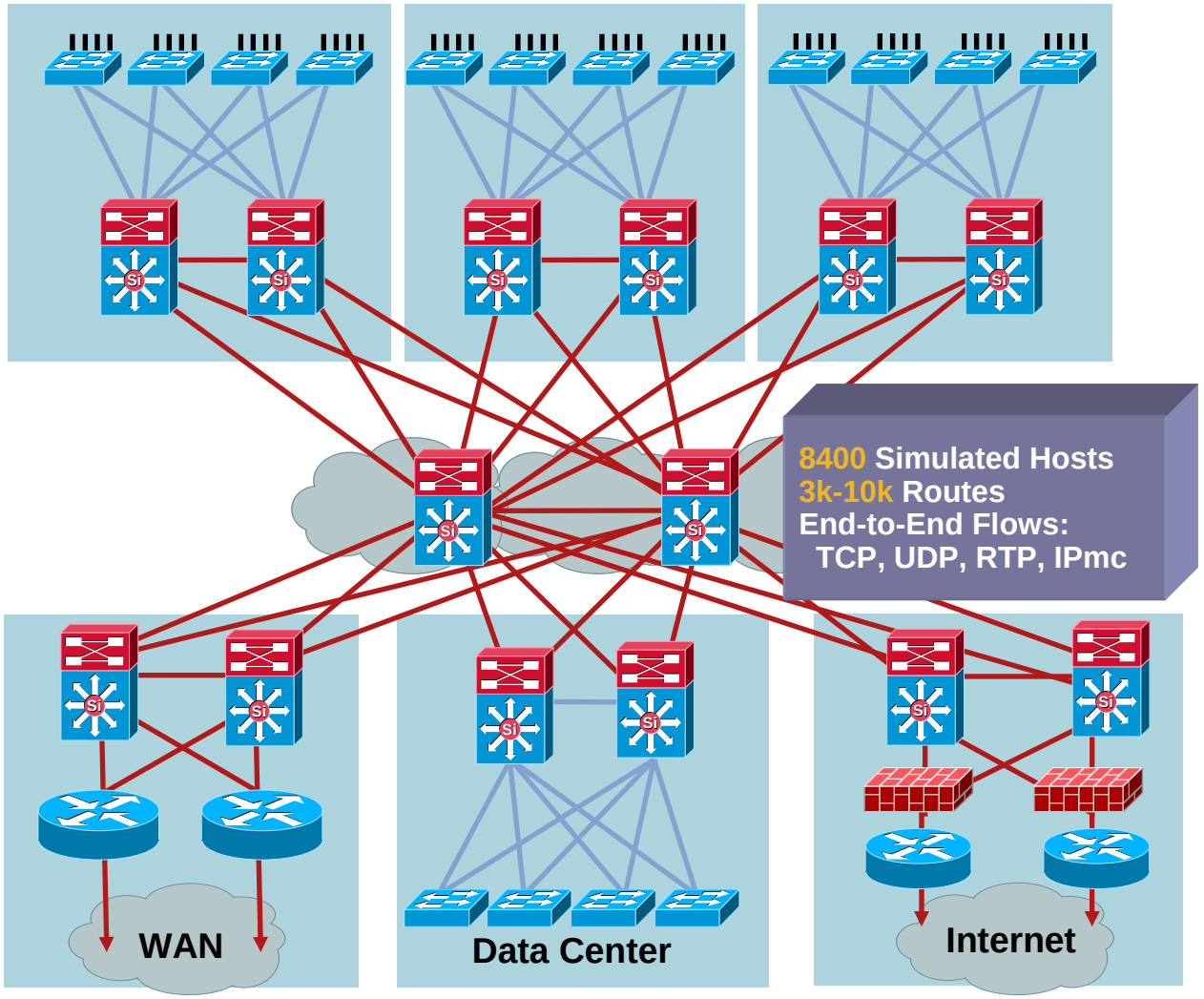
4507 with Redundant SupV

6500 with Redundant Sup720s

Three Distribution Blocks
6500 with Redundant Sup720s

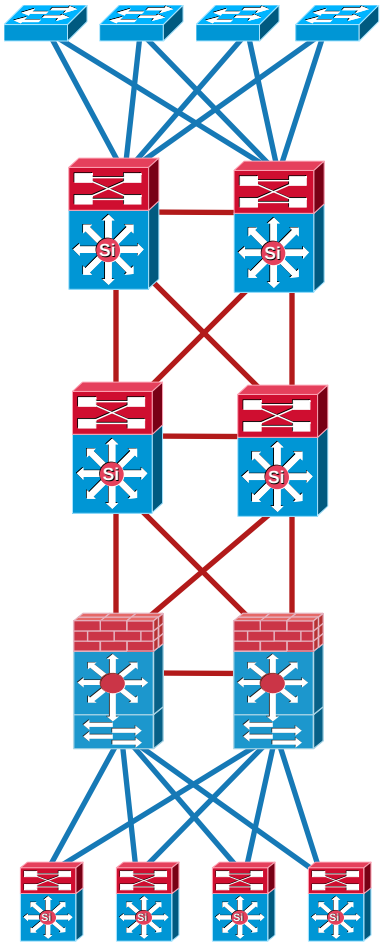
7206VXR NPEG1

4500 SupII+, 6500 Sup720, FWSM, WLSM, IDSM2, MWAM



High Availability Campus Design
Agenda

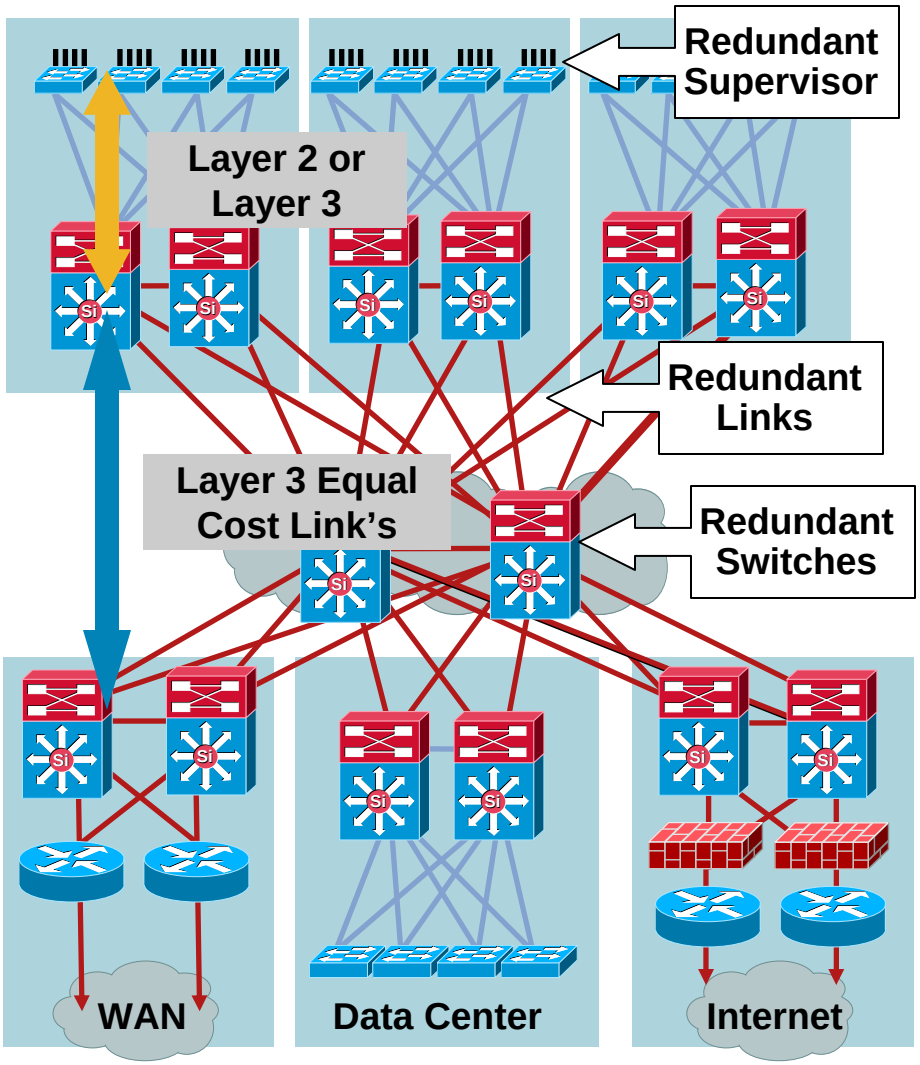
- **Network Level Resiliency**
  - **High Availability Design Principles**
  - **Redundancy in the Distribution Block**
  - **Redundancy and Routing Design**
- **System Level Resiliency**
  - **Integrated Hardware and Software Resiliency**
    - **NSF/SSO**
    - **ISSU & IOS Modularity**
  - **System Management Resiliency**
    - **GOLD & EEM**
- **Hardening the Campus Network Design**



High Availability Campus Design

Structure, Modularity and Hierarchy

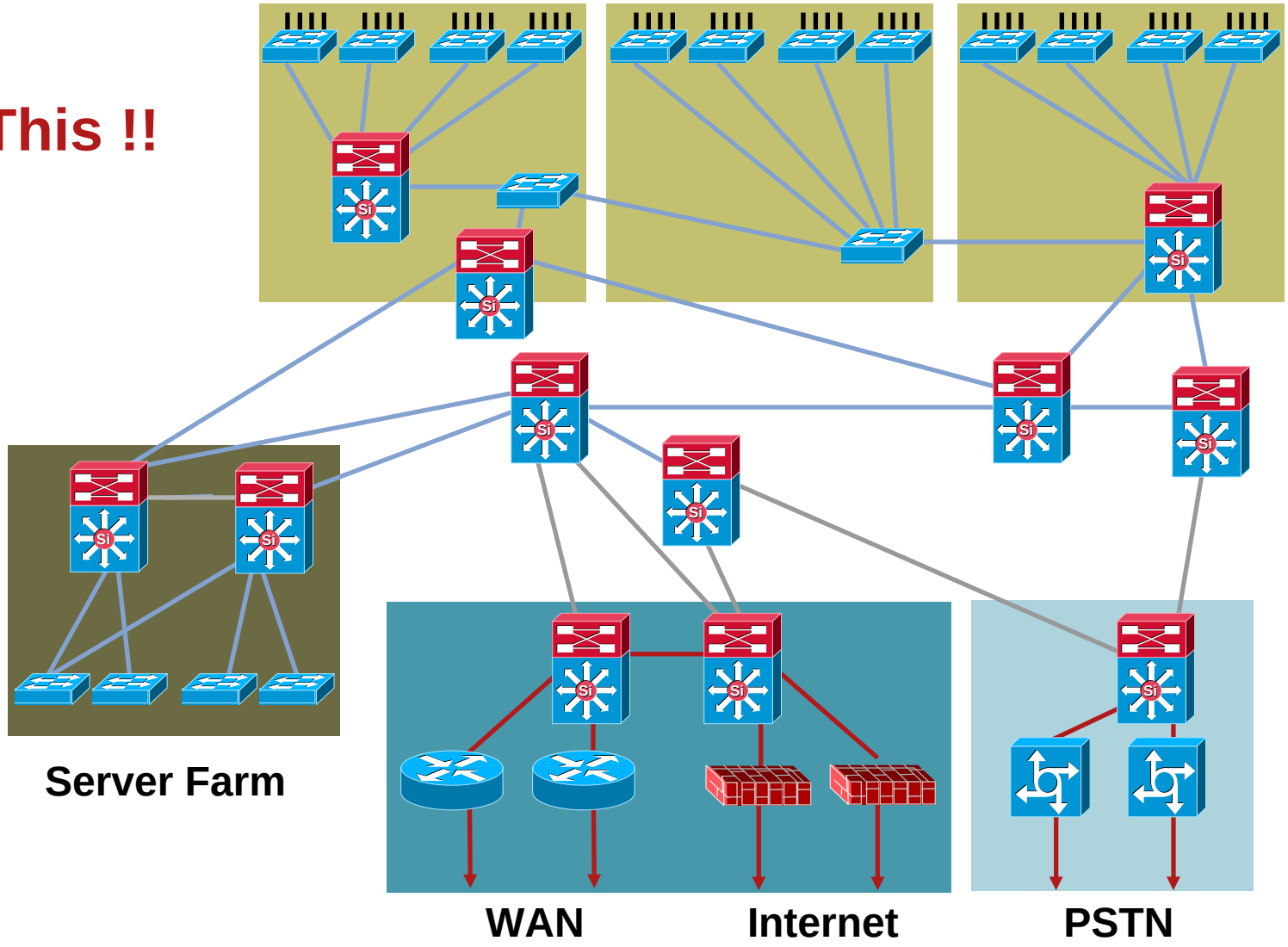
- **Optimize the interaction of the physical redundancy with the network protocols**
  - **Provide the necessary amount of redundancy**
  - **Pick the right protocol for the requirement**
  - **Optimize the tuning of the protocol**
- **The network looks like this so that we can *map the protocols onto the physical topology***
- **We want to build networks that look like this**



Hierarchical Campus Network

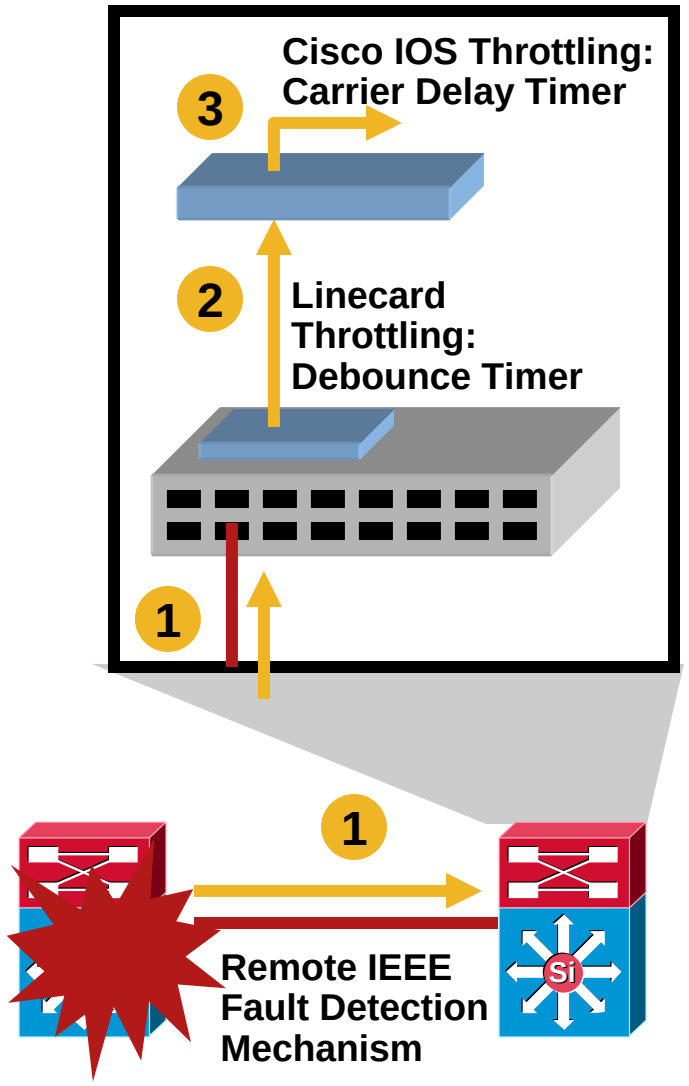
Structure, Modularity and Hierarchy

Not This !!



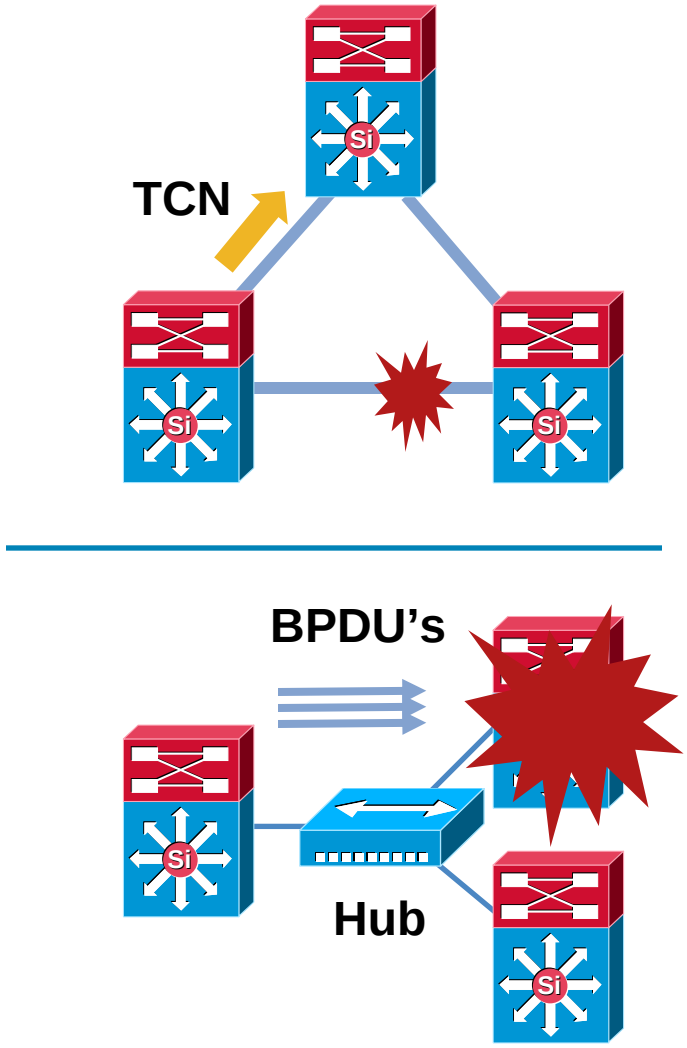
Redundancy and Protocol Interaction
Link Redundancy and Failure Detection

- Direct point to point fiber provides for fast failure detection
- IEEE 802.3z and 802.3ae link negotiation define the use of Remote Fault Indicator & Link Fault Signaling mechanisms
- Bit D13 in the Fast Link Pulse (FLP) can be set to indicate a physical fault to the remote side
- Do not disable auto-negotiation on GigE and 10GigE interfaces
- Carrier-Delay
  - 3560, 3750 & 4500 - 0 msec
  - 6500 – leave it at default 50 msec
- The default debounce timer on GigE and 10GigE fiber linecards is 10 msec.
- The minimum debounce for copper is 300 msec



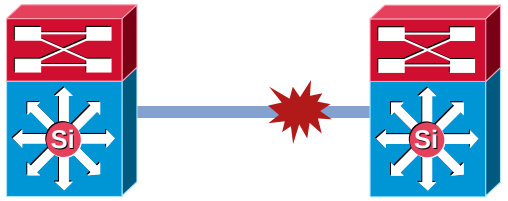
Redundancy and Protocol Interaction
Link Neighbour Failure Detection

- Indirect link failures are harder to detect
- With no direct HW notification of link loss or topology change convergence times are dependent on SW notification
- In certain topologies the need for TCN updates or dummy multicast flooding (uplink fast) is necessary for convergence
- Indirect failure events in a bridged environment are detected by Spanning Tree Hello's
- You should not be using hubs in an high availability design

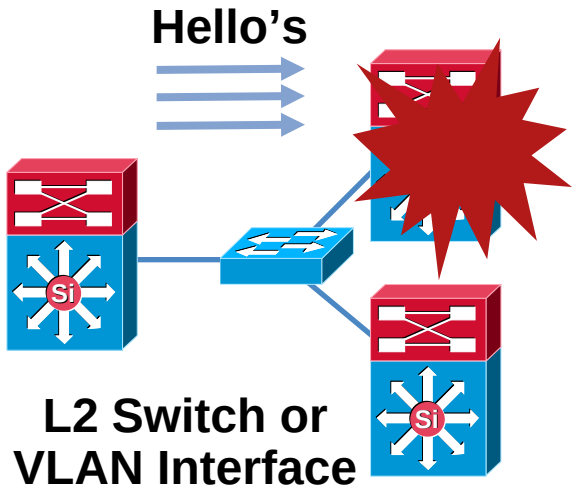


Redundancy and Protocol Interaction
Link Neighbour Failure Detection

- When using routed interfaces in the event of a physical interface state change the routing processes are notified directly
- In event of a logical L3 interface (e.g. SVI) physical events trigger L2 spanning tree changes first which then trigger RP notification
- Indirect failures require a SW process to detect the failure
- To improve failure detection
  - Use routed interfaces between L3 switches
  - Decrease interface carrier-delay to 0s
  - Decrease IGP hello timers



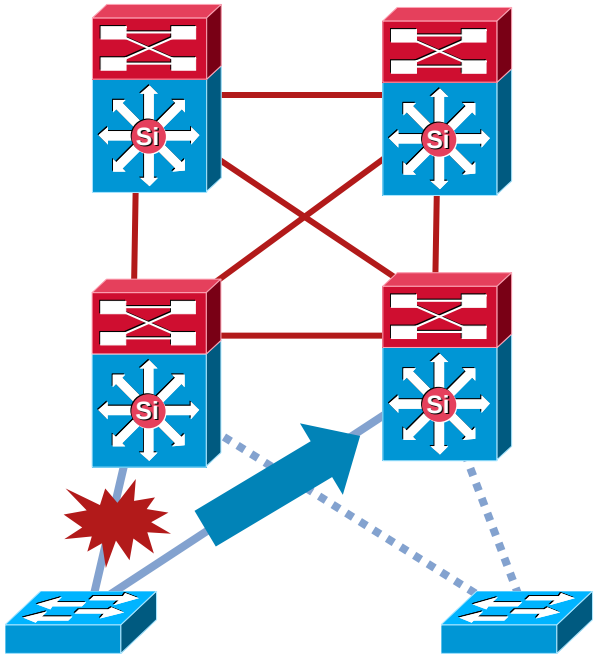
SVI Interface—
L2 Link Down Then L3 Interface Down



Redundancy and Protocol Interaction

Keep All Paths Open

- In the recommended distribution block design recovery of access to distribution link failures is accomplished based on L2 CAM updates not spanning tree
- Time to restore traffic flows is based on
  - Time to detect link failure
  - Update the HW CAM
- No dependence on external events (no need to wait for spanning tree convergence)
- Behavior is deterministic

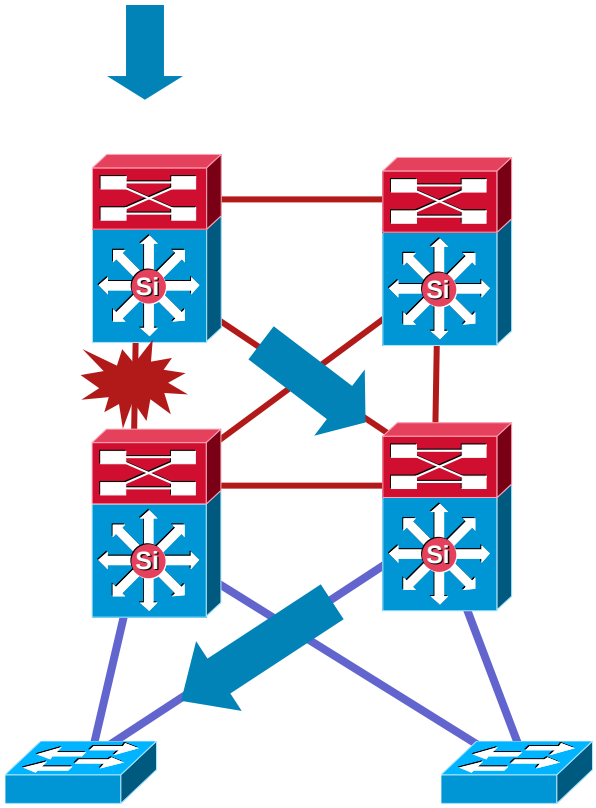


All Links Forwarding: In an Environment with All Links Active Traffic Is Restored Based on HW Recovery

Redundancy and Protocol Interaction

CEF Equal Cost Path Recovery

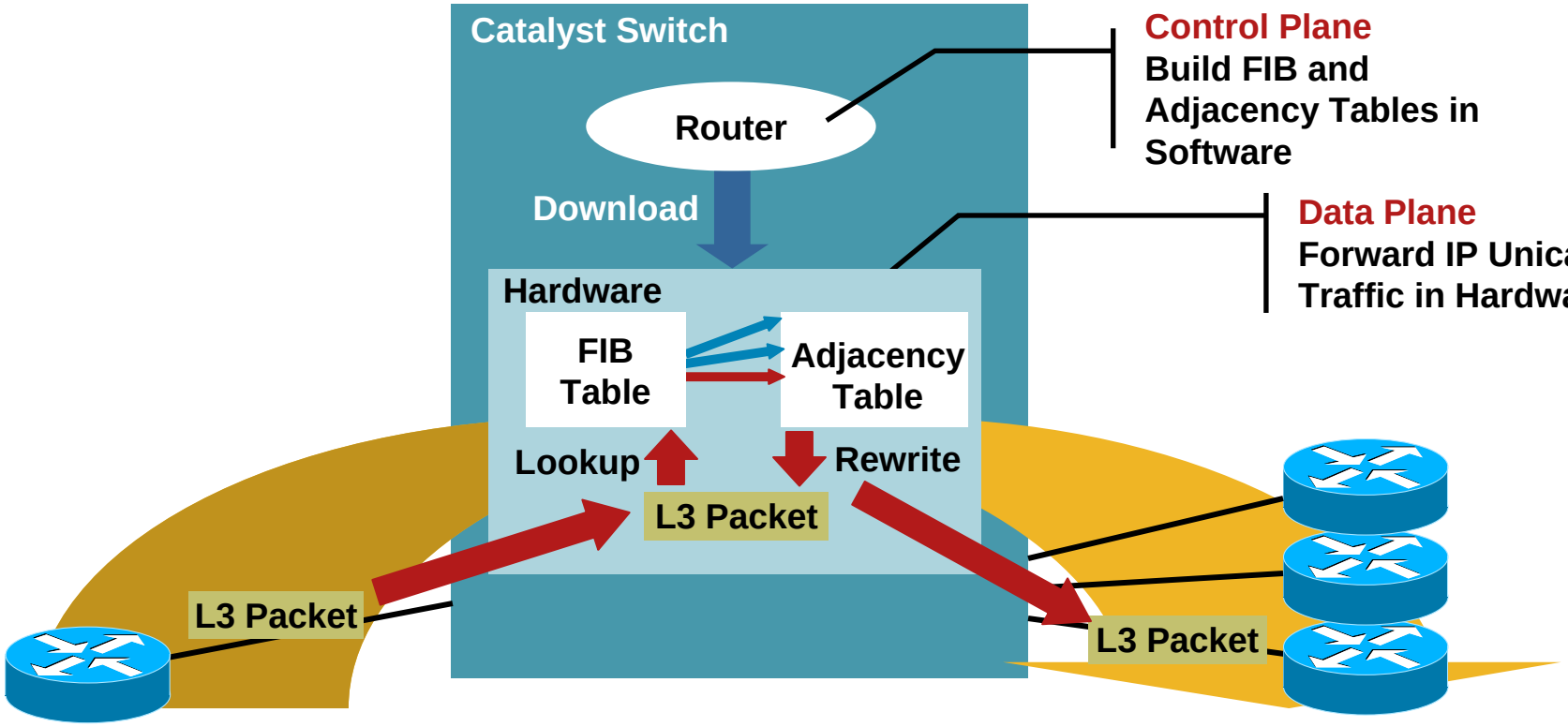
- In the recommended design the recovery from most component failures is based on L3 CEF equal cost path recovery
- Time to restore traffic flows is based on
  - Time to detect link failure
  - Process the removal of the lost routes from the SW FIB
  - Update the HW FIB
- No dependence on external events (no routing protocol convergence required)
- Behavior is deterministic



Equal Cost Links: Link/Box Failure Does Not Require Multi-Box Interaction

Redundancy and Protocol Interaction

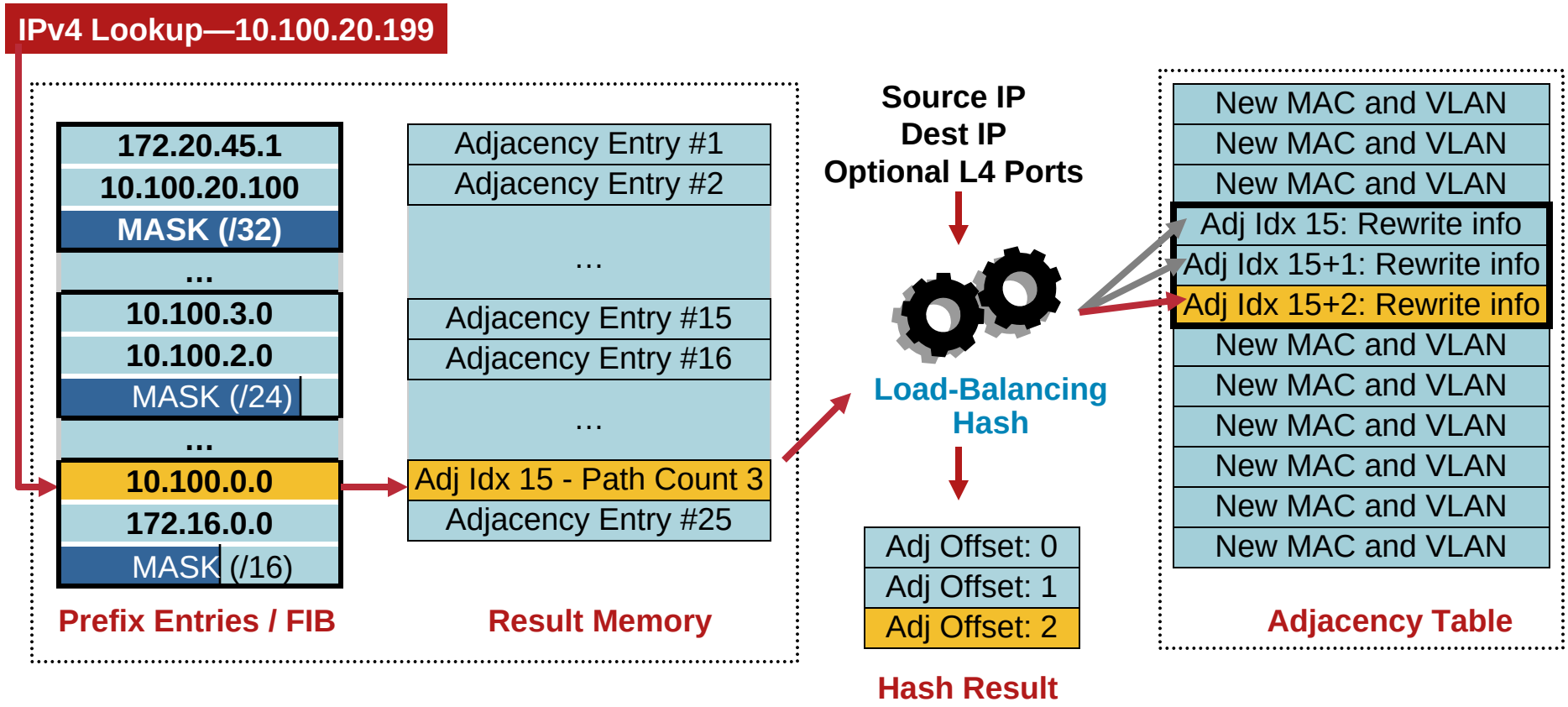
CEF Equal Cost Path Recovery



Networks using Layer 3 redundant equal cost links support fast convergence due to the behaviour of HW Cisco Express Forwarding (CEF)

Redundancy and Protocol Interaction

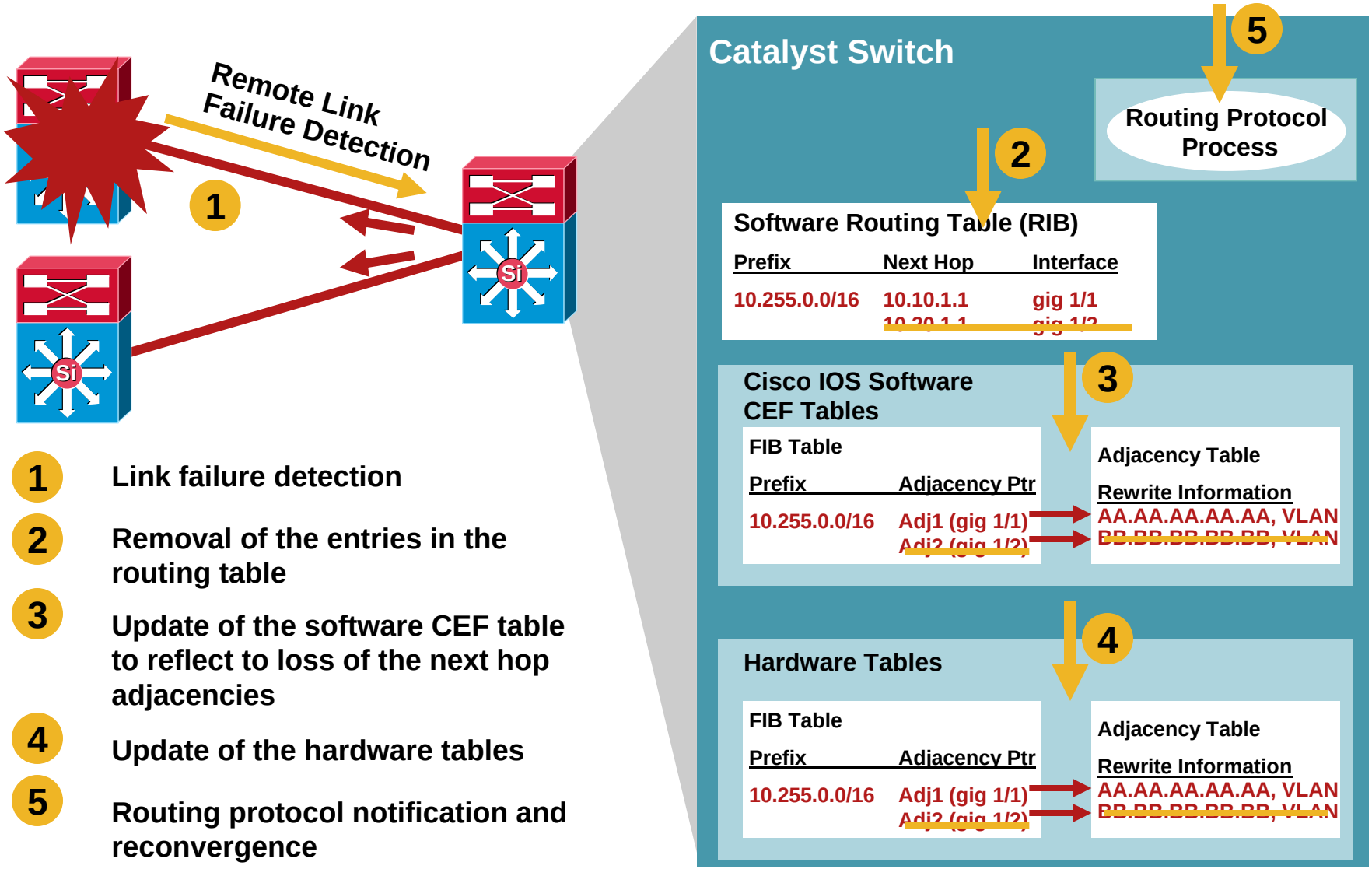
CEF Equal Cost Path Recovery




```
Switch#show mls cef exact-route 10.77.17.8 10.100.20.199
Interface: Gi1/1, Next Hop: 10.10.1.2, Vlan: 1019, Destination Mac: 0030.f272.31fe
Switch#show mls cef exact-route 10.44.91.111 10.100.20.199
Interface: Gi2/2, Next Hop: 10.40.1.2, Vlan: 1018, Destination Mac: 000d.6550.a8ea
```

Redundancy and Protocol Interaction

Time to Recovery CEF paths

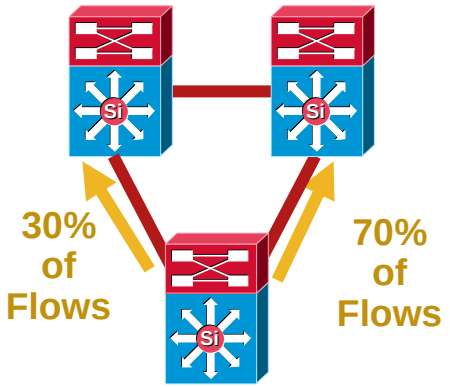


1. Link failure detection
2. Removal of the entries in the routing table
3. Update of the software CEF table to reflect to loss of the next hop adjacencies
4. Update of the hardware tables
5. Routing protocol notification and reconvergence

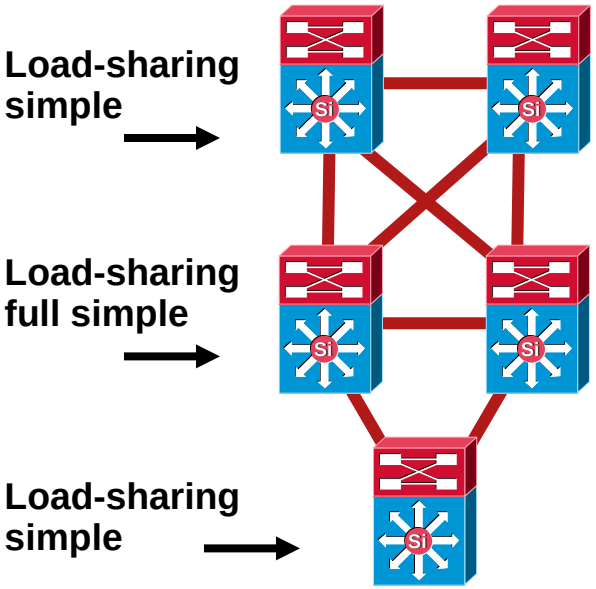
Equal Cost Multi-Path

Optimizing CEF Load-Sharing

- **Up to eight equal cost CEF paths are supported in HW today**
- **Depending on the traffic flow patterns, one algorithm may provide better load-sharing results than another**

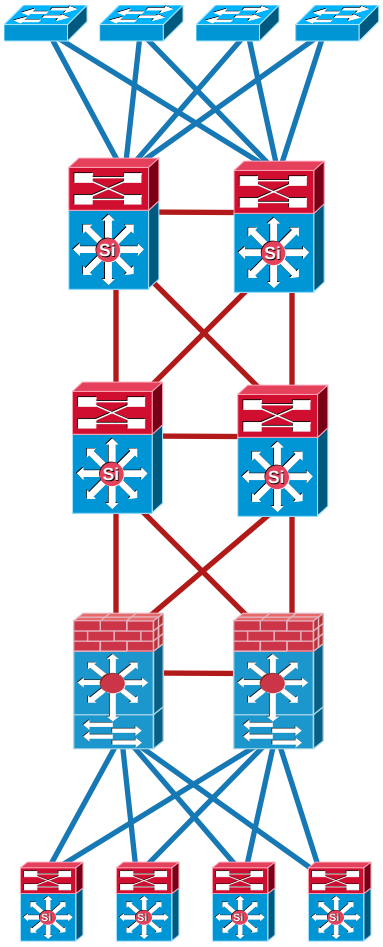


Catalyst 4500 Load-Balancing Options	
Original	Src IP + Dst IP
Universal	Src IP + Dst IP + Unique ID
Include Port	Src IP + Dst IP + (Src 'or' Dst Port) + Unique ID
Catalyst 6500 PFC3* Load-Balancing Options	
Default	Src IP + Dst IP + Unique ID
Full	Src IP + Dst IP + Src Port + Dst Port + opt.
Full Exclude Port	Src IP + Dst IP + (Src 'or' Dst Port)
Simple	Src IP + Dst IP
Full Simple	Src IP + Dst IP + Src Port + Dst Port



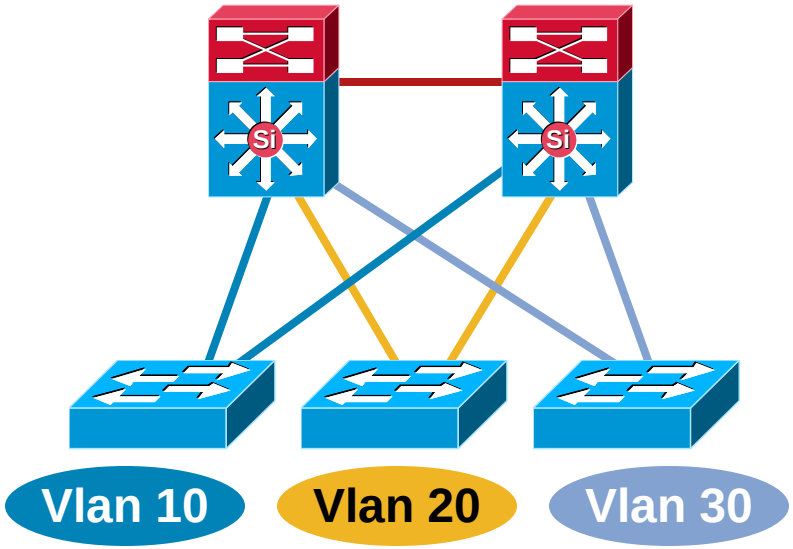
High Availability Campus Design
Agenda

- **Network Level Resiliency**
  - High Availability Design Principles
  - Redundancy in the Distribution Block
  - Redundancy and Routing Design
- **System Level Resiliency**
  - Integrated Hardware and Software Resiliency
    - NSF/SSO
    - ISSU & IOS Modularity
  - System Management Resiliency
    - GOLD & EEM
- **Hardening the Campus Network Design**

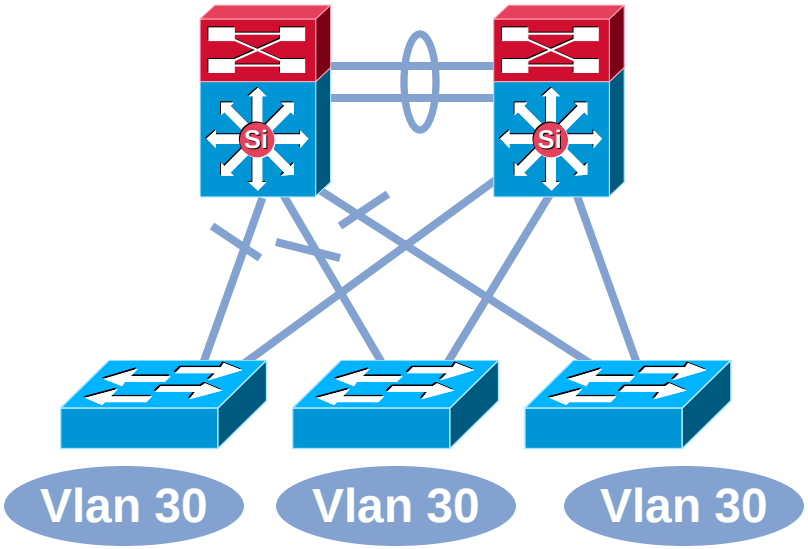


Multilayer Network Design

Layer 2 Access with Layer 3 Distribution



- Each access switch has unique VLAN's
- No layer 2 loops
- Layer 3 link between distribution
- No blocked links

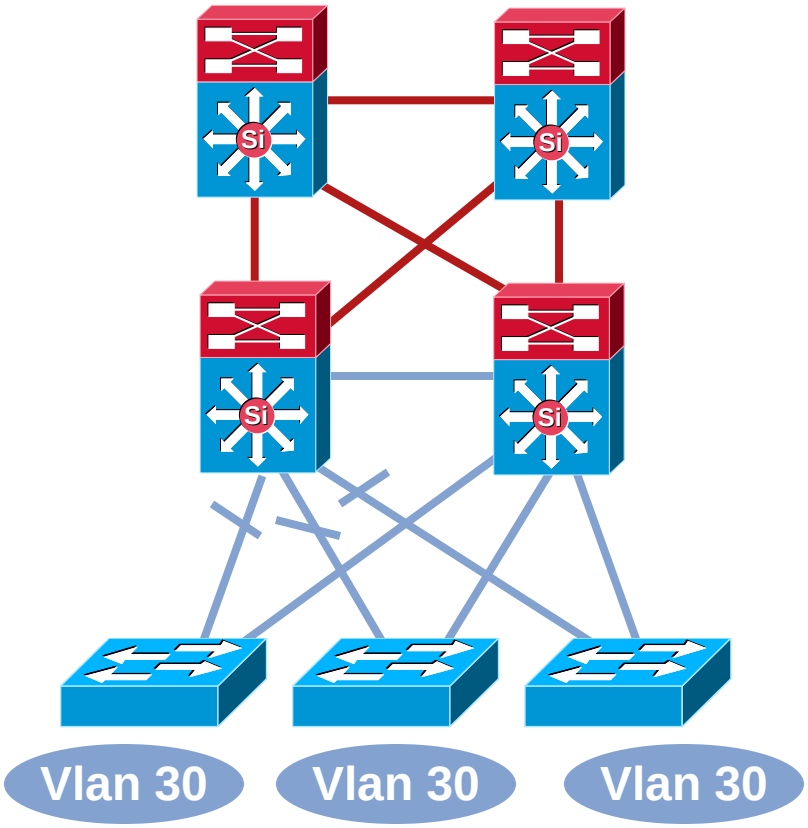


- At least some VLAN's span multiple access switches
- Layer 2 loops
- Layer 2 and 3 running over link between distribution
- Blocked links

Layer 2 Access with Layer 3 Distribution

Layer 2 Loops and Spanning Tree

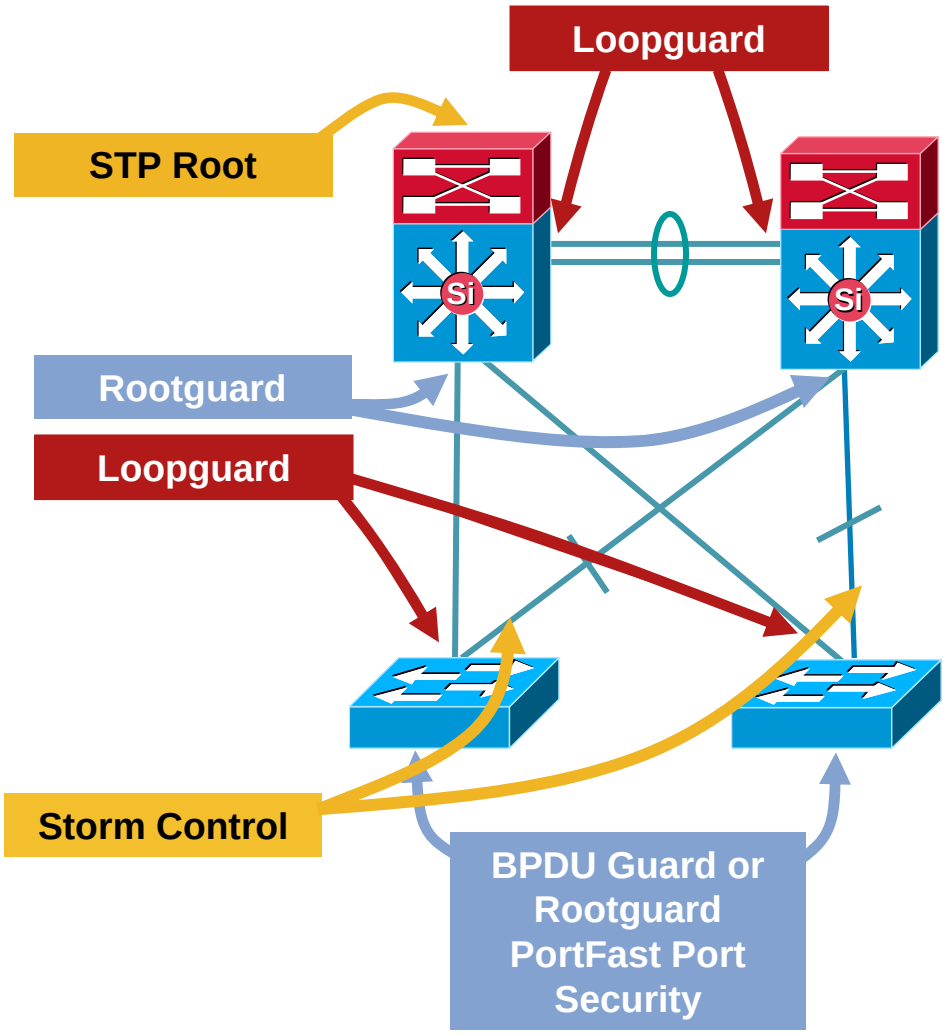
- Implement physical L2 loops only when you have to
- Spanning tree protocol is very, very rarely the problem
- L2 has no native mechanism to dampen down a problem
- When you have a physically looped topology use all the tools you have to provide that mechanism
- Utilize Rapid PVST+ for best convergence
- Take advantage of the Spanning Tree Toolkit to help prevent a problem
- Leverage storm control to help dampen the problem



Layer 2 Loops and Spanning Tree

Spanning Tree Should Behave the Way You Expect

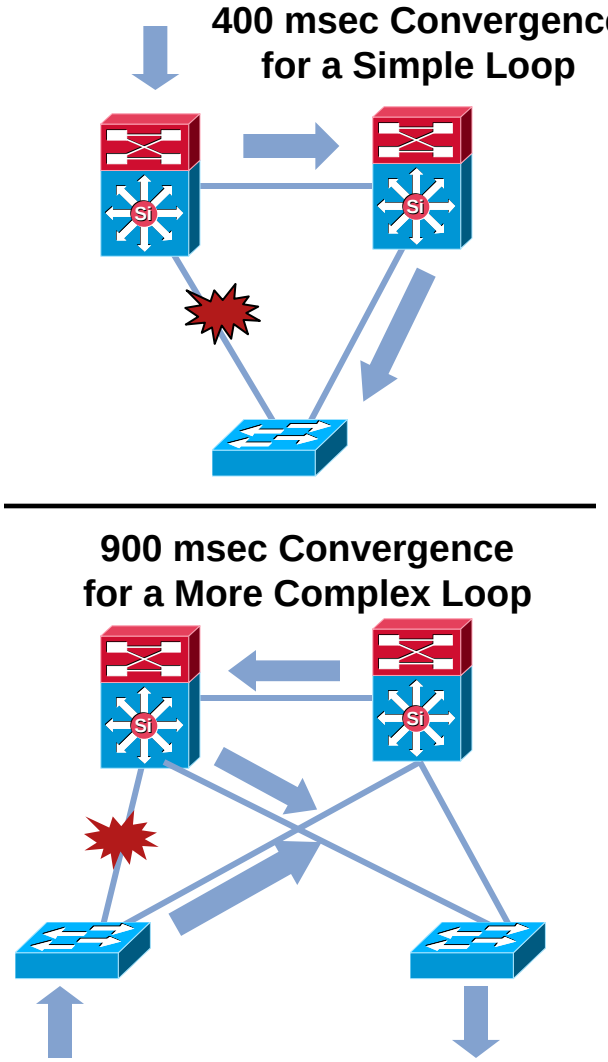
- **The root bridge should stay where you put it**
  - **Loopguard and rootguard**
  - **UDLD**
- **Only end station traffic should be seen on an edge port**
  - **BPDU guard**
  - **Port-Security**
- **There is a reasonable limit to B-Cast and M-Cast traffic volumes**
  - **Configure storm control on backup links to aggressively rate limit B-Cast and M-Cast**
  - **Utilize Sup720 rate limiters or SupIV/V with HW queuing structure**



Optimizing L2 Convergence

Complex Topologies Take Longer to Converge

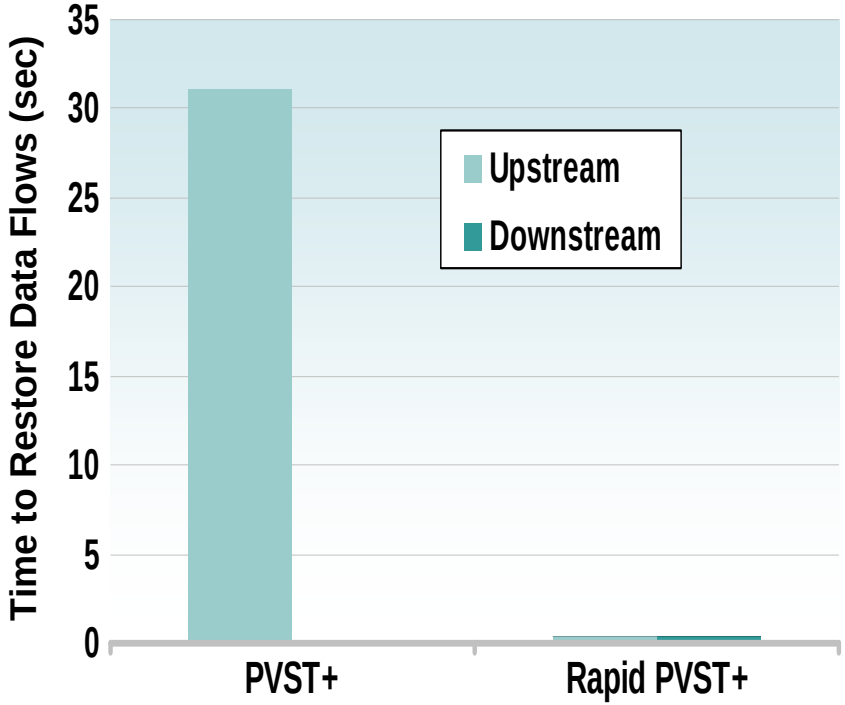
- **Time to converge is dependent on the protocol implemented 802.1d, 802.1s or 802.1w (all now a part of IEEE 802.1d 2004 spec)**
- **It is also dependent on:**
  - **Size and shape of the L2 topology (how deep is the tree)**
  - **Number of VLAN's being trunked across each link**
  - **Number of ports in the VLAN on each switch**
- **Complex Topologies Take Longer to Converge**
- **Restricting the topology is necessary to reduce convergence times**



Optimizing L2 Convergence
PVST+, Rapid PVST+ or MST

- **Rapid-PVST+ greatly improves the restoration times for any VLAN that requires a topology convergence due to link UP**
- **Rapid-PVST+ also greatly improves convergence time over Backbone fast for any indirect link failures**

- **PVST+ (802.1d)**
  - **Traditional Spanning Tree Implementation**
- **Rapid PVST+ (802.1w)**
  - **Scales to large size (~10,000 logical ports)**
  - **Easy to implement, proven, scales**
- **MST (802.1s)**
  - **Permits very large scale STP implementations (~30,000 logical ports)**
  - **Not as flexible as Rapid PVST+**



Optimizing L2 Convergence
Spanning Tree Protocol Scaling (Catalyst 6500)

- **Spanning Tree scalability is limited by the number of logical interfaces the switch needs to track & number of BPDU's to process**
- **Clear unnecessary VLANs off trunk configs (VTP Pruning does not remove vlan port instances)**
- **Distribute Trunks across line cards to space out virtual ports**
- **In recommended Campus Designs RPVST+ provides more than sufficient capacity and provides for more design flexibility**

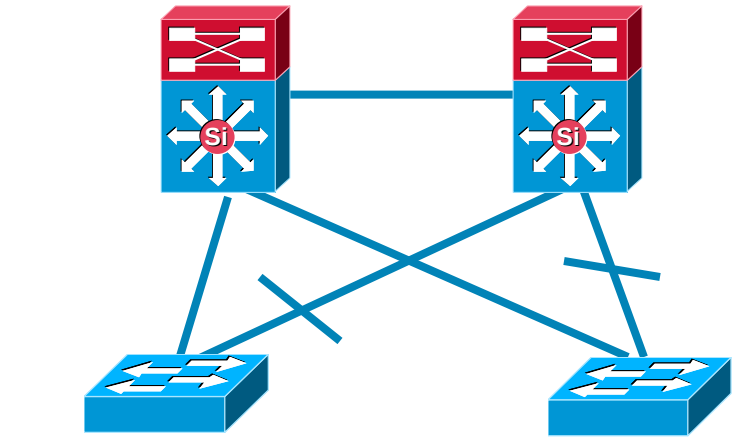
	MST	RPVST+	PVST+
Total Active STP Logical Interfaces	50,000 total 30,000 total with Release 12.2(17b)SXA	10,000 total	13,000 total
Total Virtual Ports per LineCard	6,000[1] per switching module	1,800 [1] per switching module	1,800 [1] per switching module

1 10 Mbps, 10/100 Mbps, and 100 Mbps switching modules support a maximum of 1,200 logical interfaces per module

Flex Link

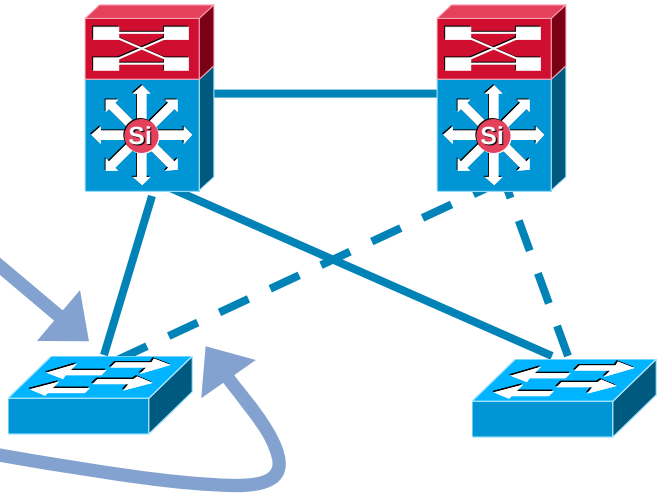
Link Redundancy (Back-Up Link)

- Flex Link provides a backup interface for an access switch uplink
- On failure of the prime link the backup link will start forwarding
- Link failure detection is processed locally on the switch
- Supported on 2970, 3550, 3560, 3750 and 6500



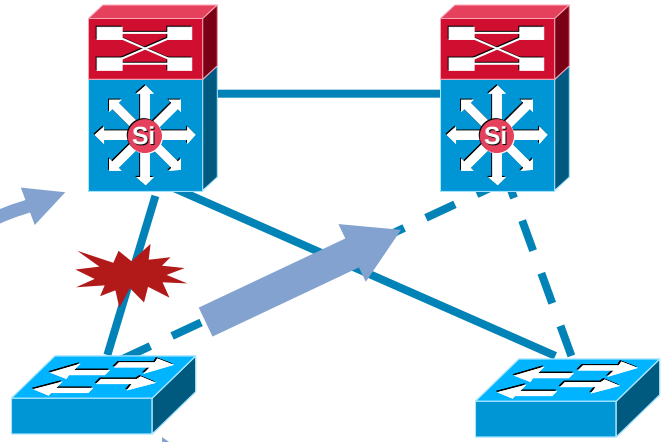
```
interface gigabitethernet1/0/1
 switchport trunk encapsulation dot1q
 switchport mode trunk
 switchport backup interface gigabitethernet1/0/2
```

```
interface gigabitethernet1/0/2
 switchport trunk encapsulation dot1q
 switchport mode trunk
```



Flex Link
Design Considerations

- **Distribution switch does not participate directly in the link recovery**
  - **Default behaviour is to send dummy multicast packets upstream**
  - **Use of move multicast update to speed up downstream convergence (currently Cisco 3750 only)**
- **Possible to configure preemption to force link back to 'primary'**



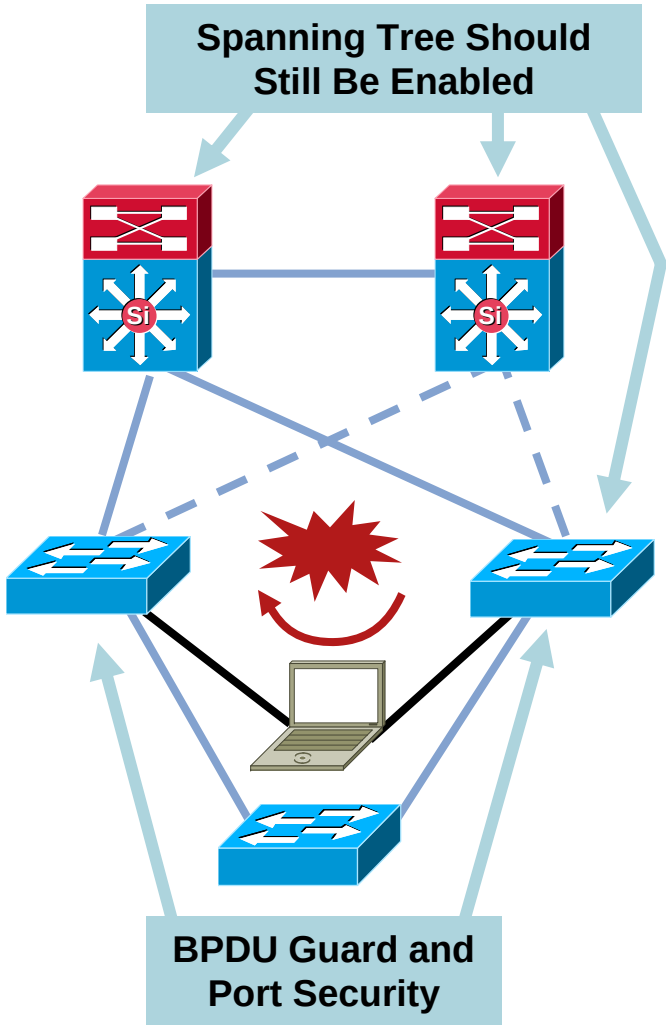
Flood Multicast Packets Upstream to Force Re-Learning

```
! Enable Receipt for MMU on the distribution switch
3750-Dist(conf)# mac address-table move update receive
```

```
! Enable Transmission of MMU (MAC Move Update) on the access switch
3750(conf)# mac address-table move update transmit
3750(conf)# interface gigabitethernet1/0/1
3750(conf-if)#switchport backup interface gigabitethernet1/0/2 preemption mode forced
3750(conf-if)#switchport backup interface gigabitethernet1/0/2 preemption delay 50
3750(conf-if)# switchport backup interface gigabitethernet0/2 mmu primary vlan 2
3750(conf-if)# exit
```

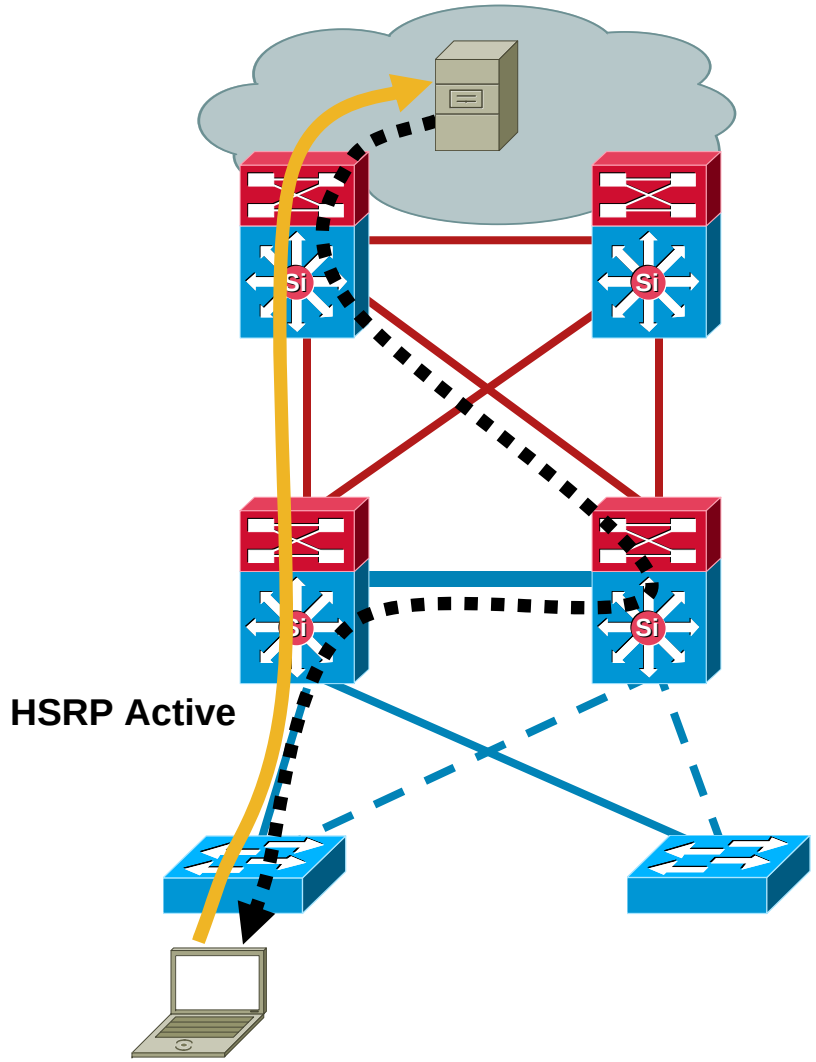
Flex Link
Design Considerations

- Spanning tree is not involved in link recovery however the network is ‘not’ L2 loop free
- Access switch blocks BPDU’s on both the active and the backup Flex Link ports
- Spanning tree should still be configured on access and distribution switches
- Follow best practice spanning tree configuration on all ports not configured as Flex Links
- Flex Link reduces size of the spanning tree topology but does not make the network loop free



Flex Link
Design Considerations

- Need to consider overall traffic flows
- 50% of return path traffic will pass across the dist-dist link in an ECMP design
- Upstream traffic for all VLAN's will pass through the same distribution switch
  - Use HSRP not GLBP
  - HSRP Active for voice and data need to be on the same switch
- Flexlink backs up the physical link not the VLAN



First Hop Redundancy

Sub-second Timers

VRRP Config

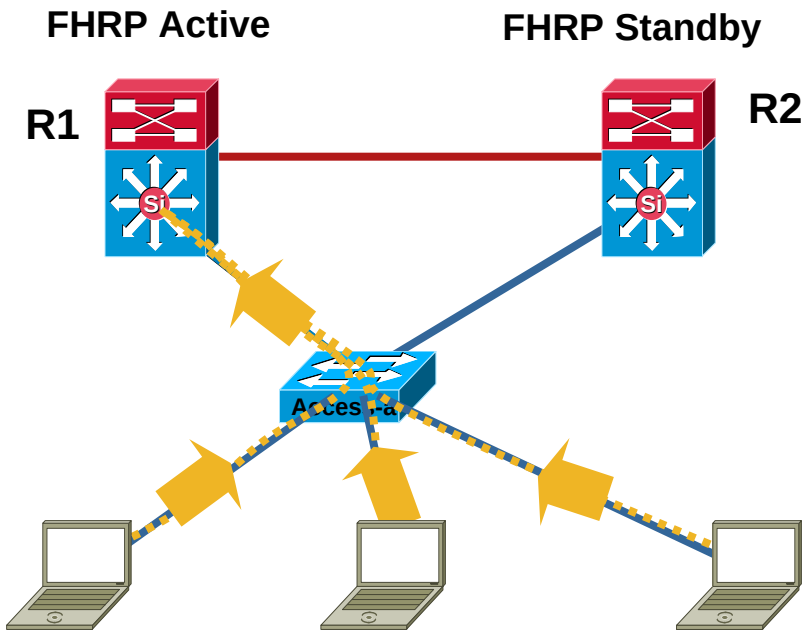
```
interface Vlan4
 ip address 10.120.4.1 255.255.255.0
 ip helper-address 10.121.0.5
 no ip redirects
 vrrp 1 description Master VRRP
 vrrp 1 ip 10.120.4.1
 vrrp 1 timers advertise msec 250
 vrrp 1 preempt delay minimum 180
```

HSRP Config

```
interface Vlan4
 ip address 10.120.4.2 255.255.255.0
 standby 1 ip 10.120.4.1
 standby 1 timers msec 250 msec 750
 standby 1 priority 150
 standby 1 preempt
 standby 1 preempt delay minimum 180
```

GLBP Config

```
interface Vlan4
 ip address 10.120.4.2 255.255.255.0
 glbp 1 ip 10.120.4.1
 glbp 1 timers msec 250 msec 750
 glbp 1 priority 150
 glbp 1 preempt
 glbp 1 preempt delay minimum 180
```

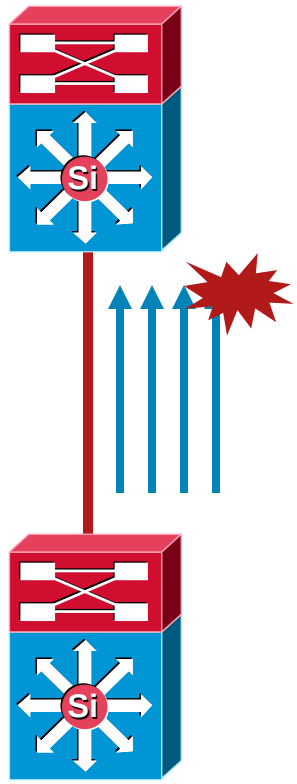


- **Sub-second Hello timer enables < 1 Sec traffic recovery upstream**
- **Preempt delay avoids black holing traffic when ACTIVE gateway recovers and preempt the backup, as upstream routing and link may not be active**

Sub-second Timer Considerations

HSRP, GLBP, OSPF, PIM

- **Evaluate your network before implementing any sub-second timers**
- **Certain events can impact the ability of the switch to process sub-second timers**
  - **Application of Large ACL's**
  - **OIR of line cards in 6500**
- **The volume of control plane traffic can also impact the ability to process**
  - **250/750 msec GLBP & HSRP timers are only valid in designs with less than 150 VLAN instances (Catalyst 6500 in the distribution)**
  - **Spanning Tree size discussed above**
- **Check size of input queue**




```
interface GigabitEthernet3/2
 description Downlink to Access
 hold-queue 2000 in
 hold-queue 2000 out
```

Sub-second Timer Considerations

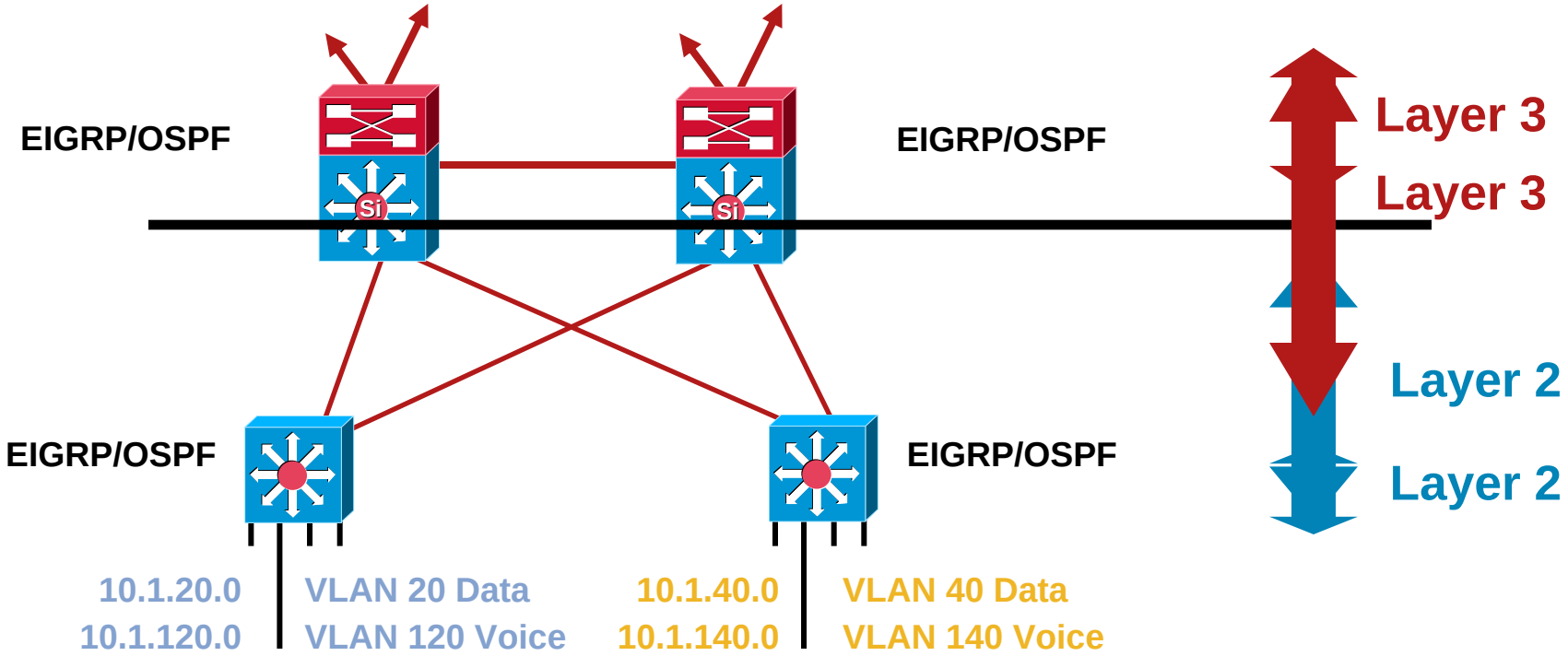
OIR Recommendations – Catalyst 6500

- During Online Insertion and Removal (OIR) of classic line cards in a Catalyst 6500 a stall signal is generated on the backplane bus to prevent backplane data corruption
- Bus stall prevents packets from being transmitted to the backplane: this results in traffic interruption for the duration of the stall
- Switches with only DFC enabled line cards are not impacted by Bus stall as forwarding decision is made locally and traffic flows over the switch fabric (WS-X6748, WS-6704, WS-6708, …)
- OIR of new linecards (WS-X6148A-GE-TX, WS-X6148A-RJ-45, WS-X6148-FE-SFP, Enhanced FlexWAN, SIP) will not stall the BUS resulting in *zero* packet loss
- Continuing improvements please continue to refer to release notes

Interruption	OIR (Insertion and Removal)	Power sequence (Down and Up)
Redundant WS-SUP32-8GE	140ms	N/A
WS-X6408A-GBIC	500ms	300ms
WS-X6148A-GE-TX	0s	0s

Routing to the Edge

Layer 3 Distribution with Layer 3 Access



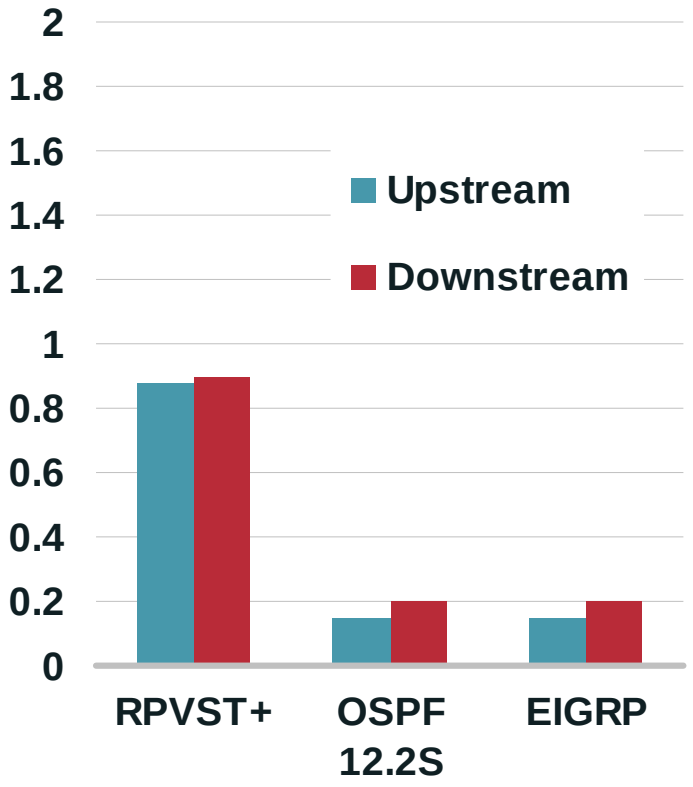
- **Move the Layer 2/3 demarcation to the network edge**
- **Upstream convergence times triggered by hardware detection of light lost from upstream neighbor**
- **Beneficial for the right environment**

Routing to the Edge

Advantages, Yes in the Right Environment

- **Ease of implementation, less to get right**
  - **No matching of STP/HSRP/ GLBP priority**
  - **No L2/L3 multicast topology inconsistencies**
- **Single control plane and well known tool set**
  - **traceroute, show ip route, show ip eigrp neighbor, etc.**
- **Most Cisco Catalysts support L3 switching today**
- **EIGRP converges in <200 msec**
- **OSPF with sub-second tuning converges in <200 msec**
- **RPVST+ convergence times dependent on GLBP/ HSRP tuning**

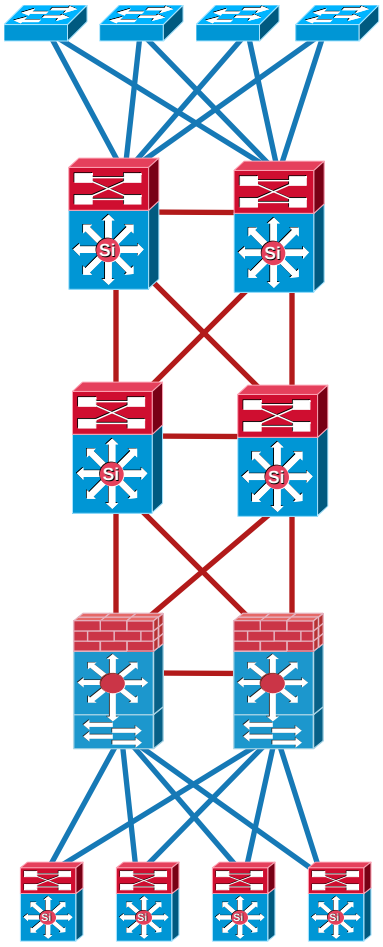
Both L2 and L3 Can Provide Sub-Second Convergence



BRKCAM-3004 - Deploying a Fully Routed Campus Network

High Availability Campus Design
Agenda

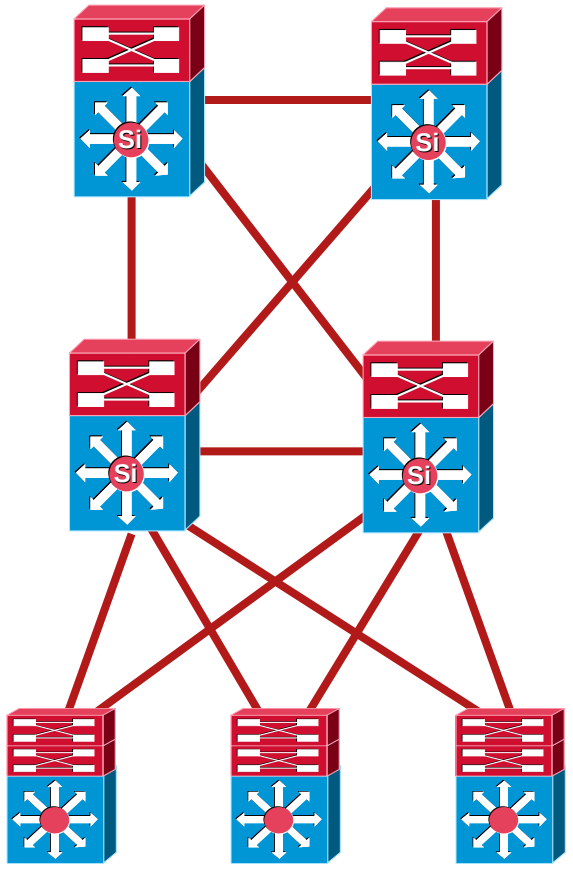
- **Network Level Resiliency**
  - High Availability Design Principles
  - Redundancy in the Distribution Block
  - Redundancy and Routing Design
- **System Level Resiliency**
  - Integrated Hardware and Software Resiliency
    - NSF/SSO
    - ISSU & IOS Modularity
  - System Management Resiliency
    - GOLD & EEM
- **Hardening the Campus Network Design**



Multilayer Network Design

Core and Distribution Routing Design

- **Good routing design forms the foundation of the HA campus design**
- **Needed to quickly re-route around failed node/links while providing load balancing over redundant paths**
- **Build full meshed equal cost path designs for deterministic convergence**
- **Only peer on links that you intend to use as transit**
- **Insure redundant L3 paths to avoid black holes**
- ***Map the protocol design to the physical design***



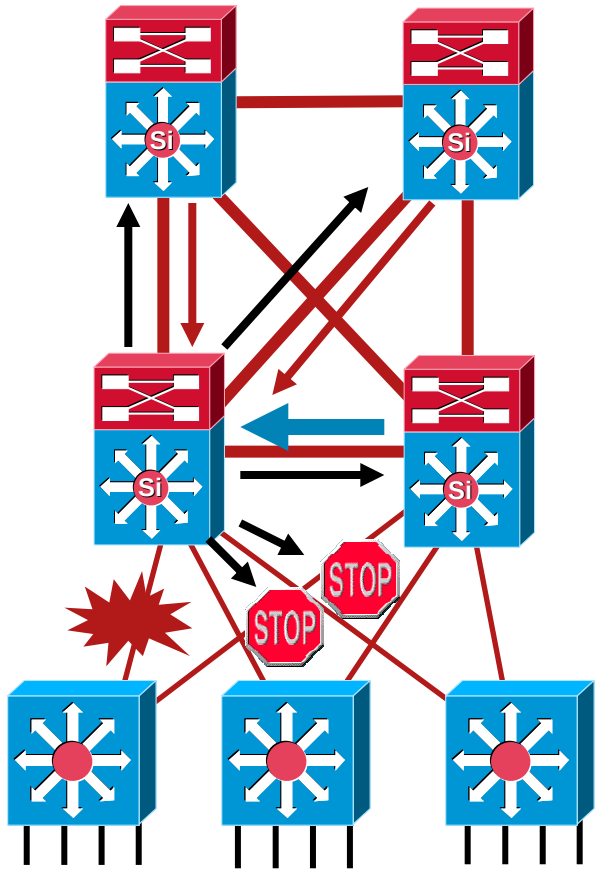
EIGRP Design Rules for HA Campus

High-Speed Campus Convergence

- EIGRP convergence is largely dependent on query response times
- Minimize the number and time for query response to speed up convergence
- Summarize distribution block routes upstream to the core
- Configure all access switches as EIGRP stub routers
- Filter routes sent down to access switches

```
interface TenGigabitEthernet 4/1
 ip summary-address eigrp 100 10.120.0.0 255.255.0.0 5

router eigrp 100
 network 10.0.0.0
 distribute-list Default out <mod/port>

ip access-list standard Default
 permit 0.0.0.0
```

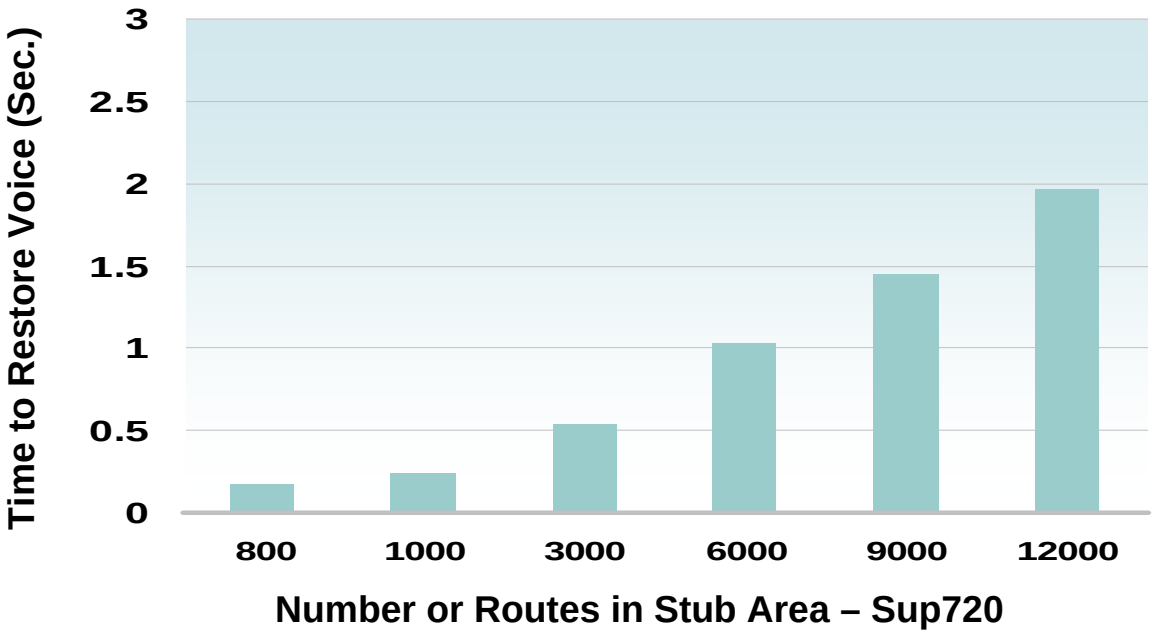
```
router eigrp 100
 network 10.0.0.0
 eigrp stub connected
```

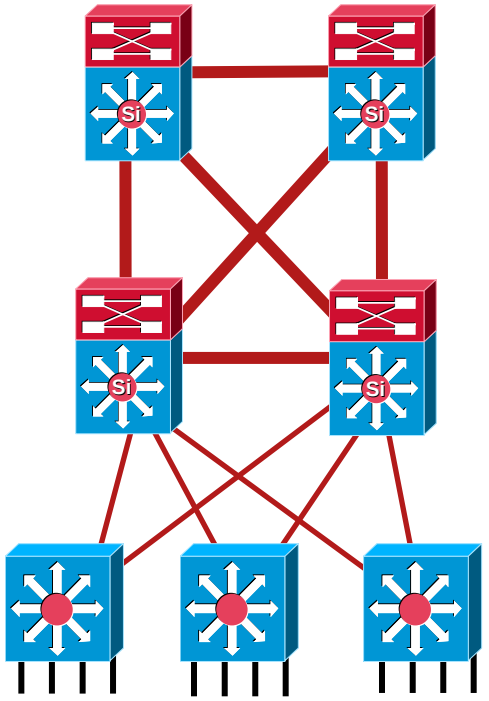


OSPF Design Rules for HA Campus

Manage the Area Boundaries

- **Managing the number of routes in the network is important**
- **Both EIGRP and OSPF need summarization**
- ***Map the protocol to the topology***

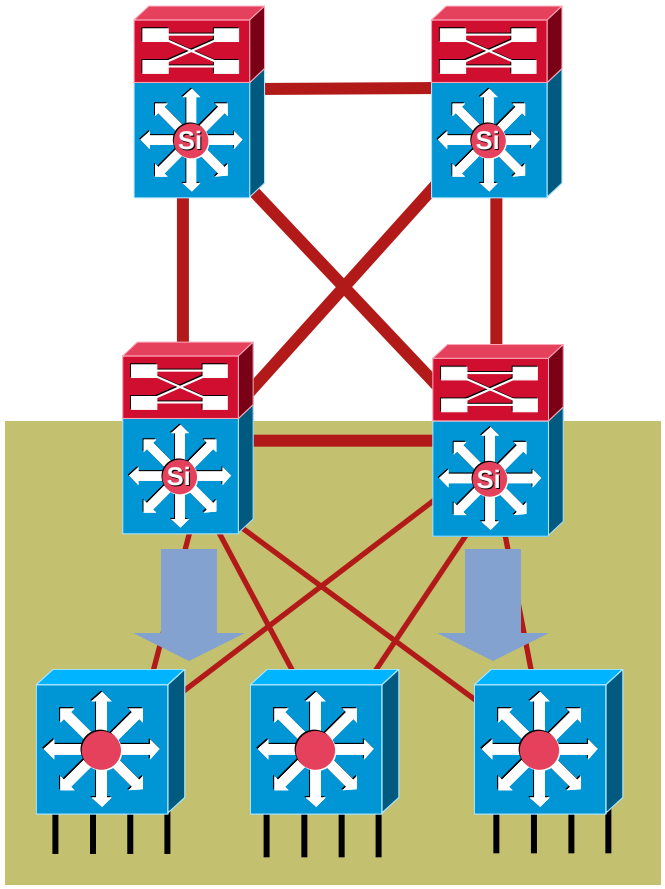




Area Types

Reduce SPF and LSA Load in Distribution Area

- **ABR for a regular area forwards**
  - **Summary LSAs (Type 3)**
  - **ASBR summary (Type 4)**
  - **Specific externals (Type 5)**
- **Stub area ABR forwards**
  - **Summary LSAs (Type 3)**
  - **Summary default (0.0.0.0)**
- **A totally stubby area ABR forwards**
  - **Summary default (0.0.0.0)**

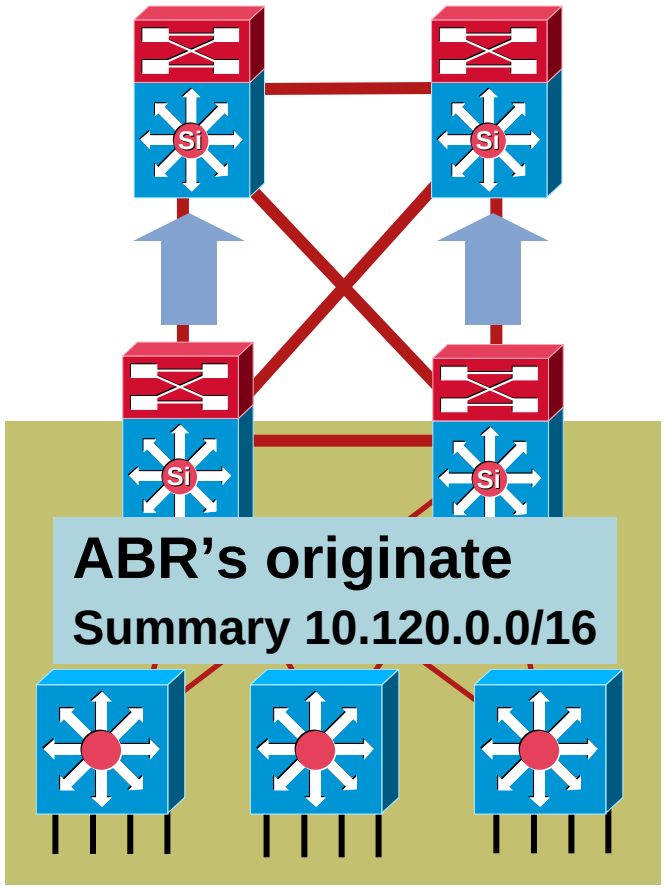



```
router ospf 100
 area 120 stub no-summary
 area 120 range 10.120.0.0 255.255.0.0 cost 10
 network 10.120.0.0 0.0.255.255 area 120
 network 10.122.0.0 0.0.255.255 area 0
```

Summarization Distribution to Core

Reduce SPF and LSA Load in Area 0

- Summarize routes from the distribution block upstream into the core
- Minimize the number of LSA's and routes in the core
- Reduce the need for SPF calculations due to internal distribution block changes
- Incremental SPF (iSPF) is a mechanism to reduce the computational load of larger OSPF areas but is more applicable to WAN than Campus environments

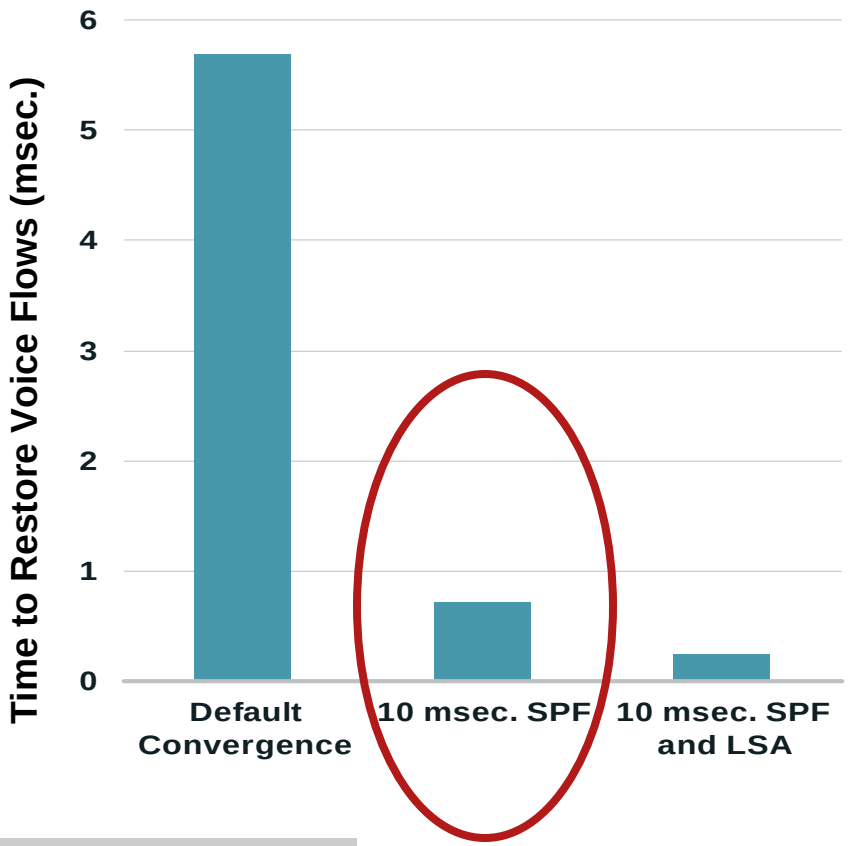



```
router ospf 100
 area 120 stub no-summary
 area 120 range 10.120.0.0 255.255.0.0 cost 10
 network 10.120.0.0 0.0.255.255 area 120
 network 10.122.0.0 0.0.255.255 area 0
```

OSPF Design Rules for HA Campus

OSPF SPF Throttling

- OSPF has an SPF throttling timer designed to dampen route recalculation (preserving CPU resources) when a link bounces
- 12.2S OSPF enhancements let us tune this timer to milliseconds; prior to 12.2S one second was the minimum
- After a failure, the router waits for the SPF timer to expire before recalculating a new route

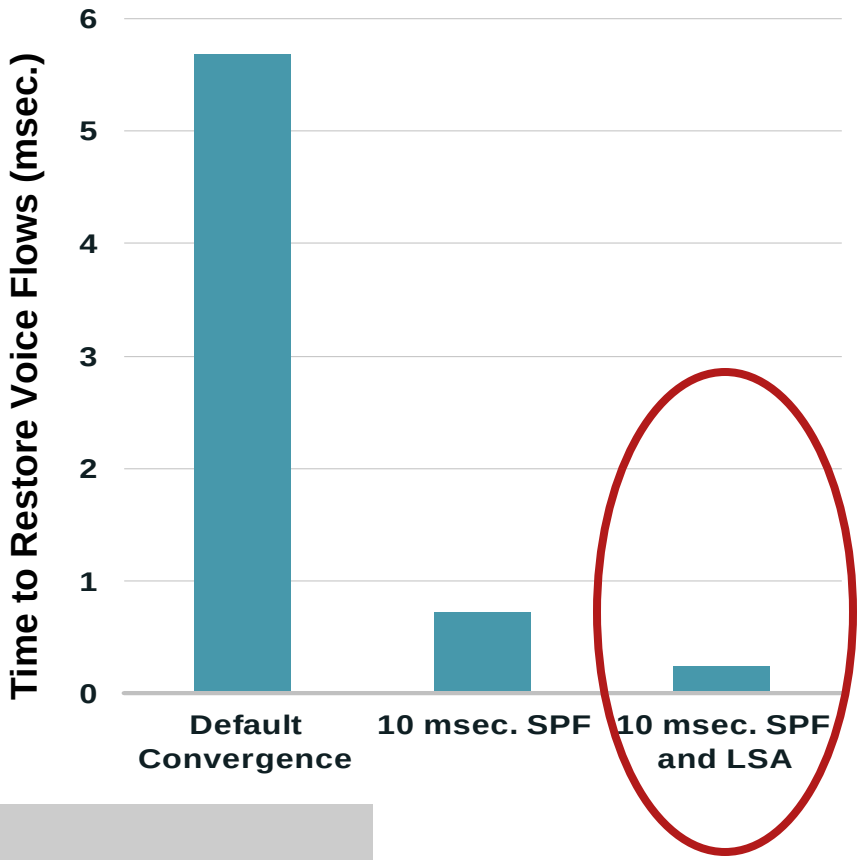



```
timers throttle spf 10 100 5000
timers throttle lsa all 10 100 5000
timers lsa arrival 80
```

OSPF Design Rules for HA Campus

OSPF LSA Throttling

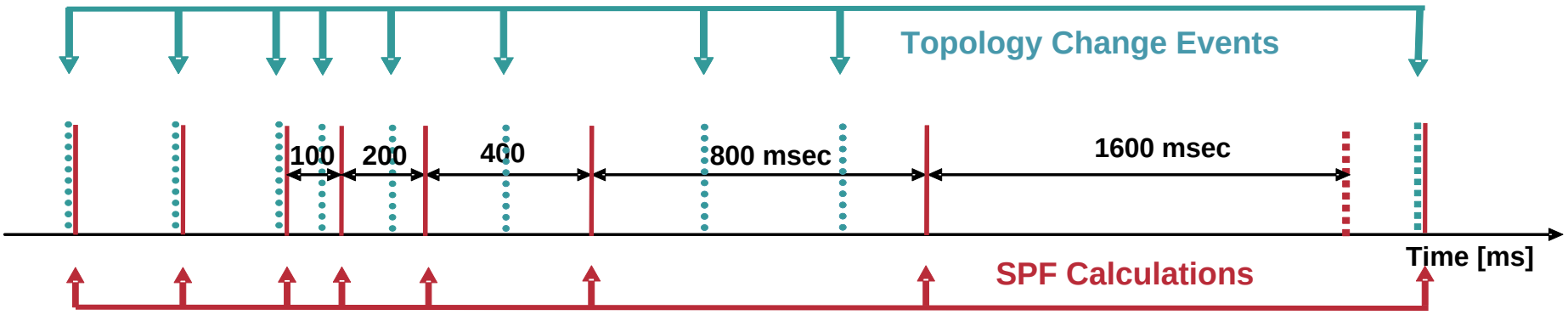
- By default, there is a 500ms delay before generating router and network LSA's; the wait is used to collect changes during a convergence event and minimize the number of LSA's sent
- Propagation of a new instance of the LSA is limited at the originator
- Acceptance of a new LSAs is limited by the receiver
- Make sure lsa-arrival < lsa-hold




```
timers throttle spf 10 100 5000
timers throttle lsa all 10 100 5000
timers lsa arrival 80
```

OSPF Design Rules for HA Campus

LSA/SPF Exponential Back-Off Throttle Mechanism



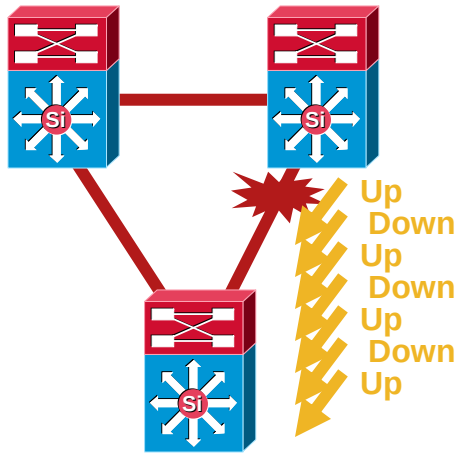

```
timers throttle spf <spf-start> <spf-hold> <spf-max-wait>
timers throttle lsa all <lsa-start> <lsa-hold> <lsa-max-wait>
```

- **Sub-second timers without risk**
  1. **spf-start or initial hold timer controls how long to wait prior to starting the SPF calculation**
  2. **If a new topology change event is received during the hold interval, the SPF calculation is delayed until the hold interval expires and the hold interval is temporarily doubled**
  3. **The hold interval can grow until the maximum period configured is reached**
  4. **After the expiration of any hold interval, the timer is reset**

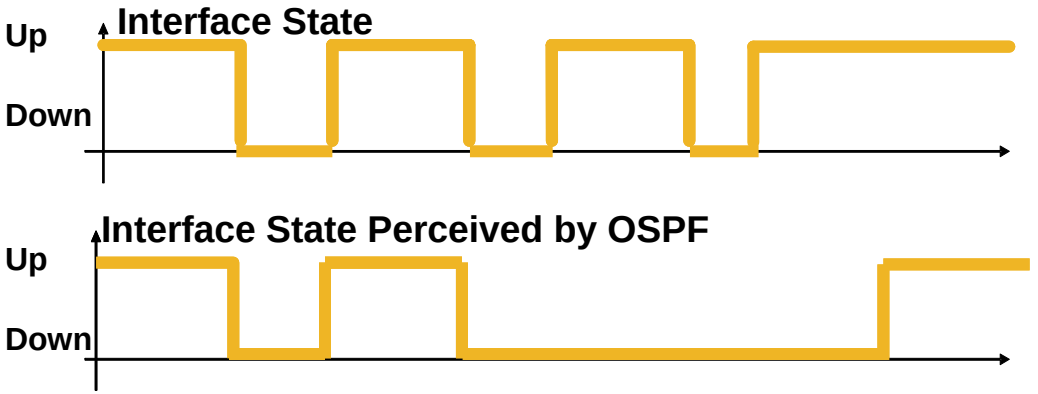
Routing Protocol Convergence

Event Detection and IP Event Dampening

- Prevents routing protocol churn caused by constant interface state changes
- Takes the concept of BGP route-flap dampening and applies it at the interface level, so all IP routing protocols can benefit
- Dampening is applied on a system: nothing is exchanged between routing protocols
- Supports all IP routing protocols
  - Static routing, RIP, EIGRP, OSPF, IS-IS, BGP
  - In addition, it supports HSRP and CLNS routing
  - Applies on physical interfaces and can't be applied on subinterfaces individually




```
interface GigabitEthernet1/1
 description Uplink to Distribution 1
 dampening
 ip address 10.120.0.205 255.255.255.254
 ip pim sparse-mode
 ip ospf network point-to-point
 ip ospf dead-interval minimal hello-multiplier 4
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
```

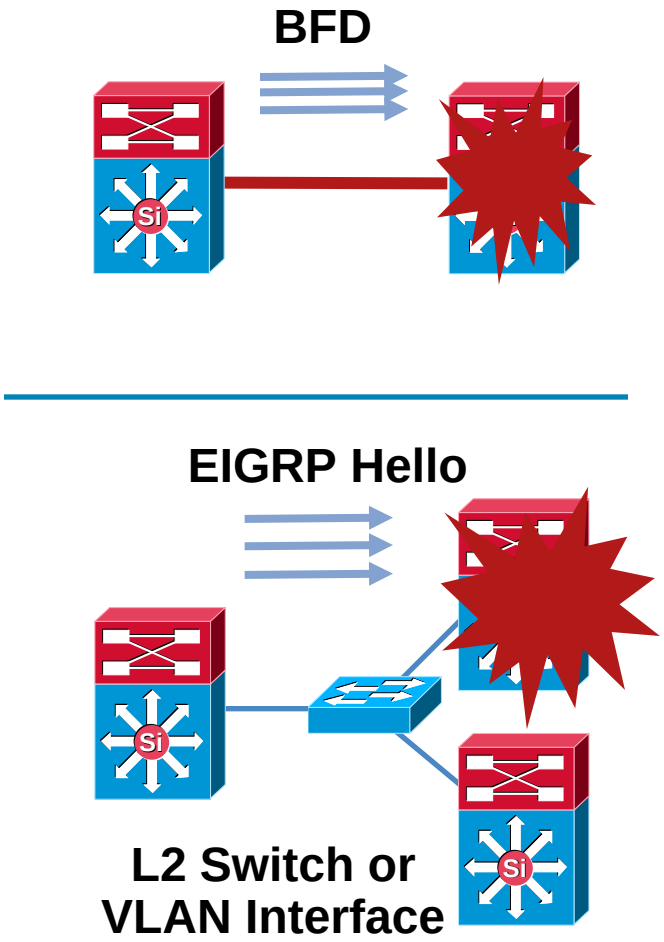


Routing Protocol Convergence

Improving Layer 3 Neighbour Failure Detection

- **EIGRP, OSPF, IS-IS, mBGP all have native hello/dead mechanisms**
- **Bidirectional Forwarding Detection (BFD)* provides a protocol independent mechanism**
  - **Negotiation of timers between peers**
  - **BFD control packets are encapsulated in UDP unicast datagrams, destination port 3784**
  - **Lightweight process, packets are not sequenced**

```
interface Vlan4
 dampening
 ip address 10.122.0.26 255.255.255.254
 bfd interval 100 min_rx 100 multiplier 3
 bfd neighbor 10.122.0.27

router eigrp 100
 bfd interface TenGigabitEthernet4/1
```

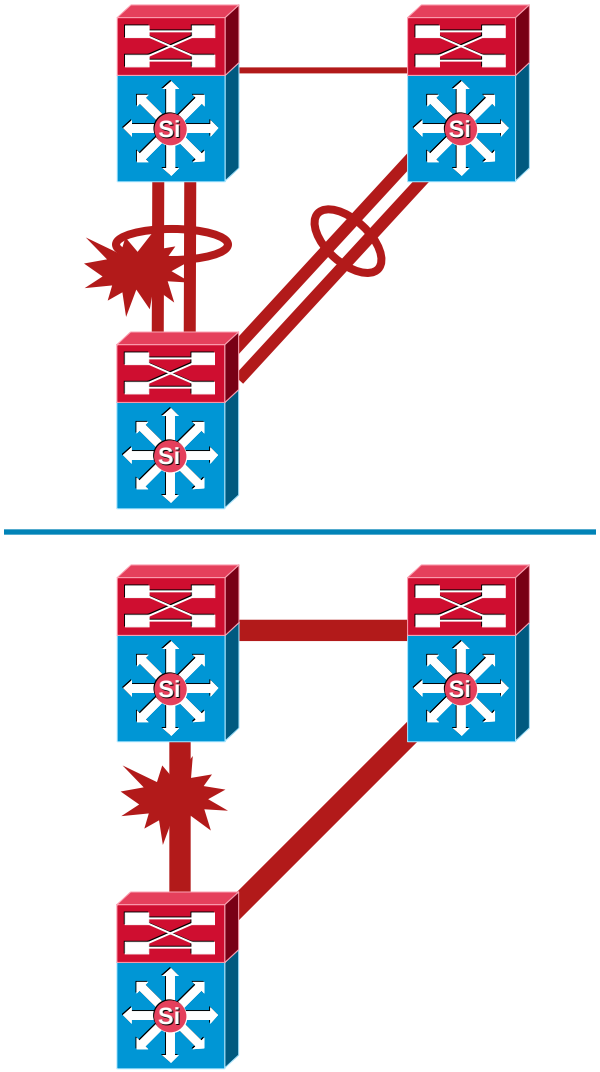


***Verify Cisco IOS Release Availability, ESE does not yet have specific configuration guidance**

Routing Protocol Convergence

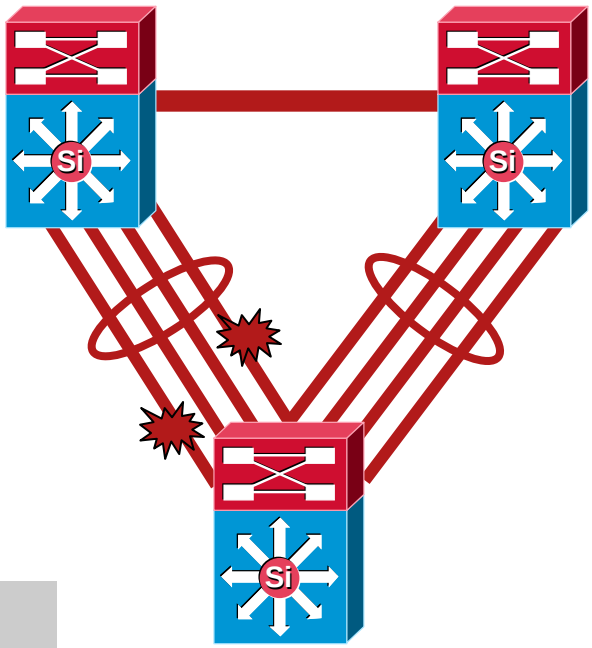
EtherChannel and L3 links

- L3 EtherChannel can provide for increased capacity
- However EtherChannel does not always provide for link redundancy in an L3 environment
- On single link failure in a bundle
  - OSPF running on a Cisco IOS based switch will reduce link cost and re-route traffic
  - OSPF running on a hybrid switch will not change link cost and may overload remaining links
  - EIGRP may not change link cost and may overload remaining links
- In an L3 environment single 10 Gigabit Links address both problems. Increased bandwidth without routing challenges



Routing Protocol Convergence
EtherChannel and L3 links

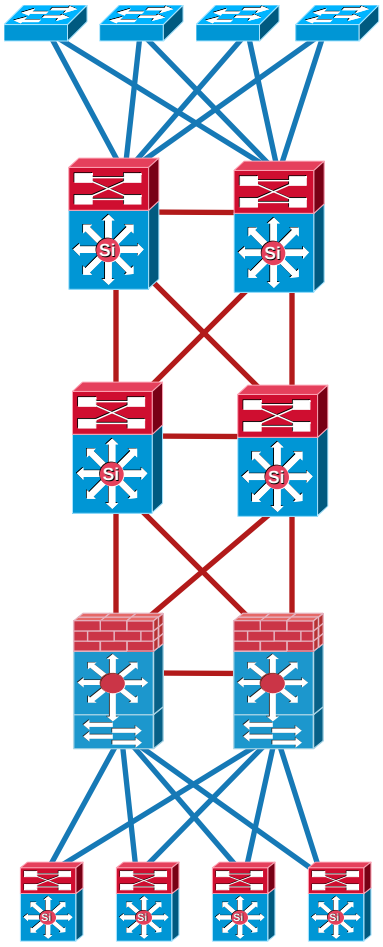
- By default the Port-Channel interface associated with a physical EtherChannel bundle remains up as long as ‘one’ of the physical links is up
- When using LACP as the channel protocol it is possible to define how many links need to be active for the Port-Channel interface to remain up
- Balance the need to re-route vs. the need for the network capacity




```
Sup720(config)# interface range gig 3/1 – 4
Sup720(config-if)#channel-protocol lacp
Sup720(config-if)#channel-group 5 mode on

Sup720(config)# interface port-channel 5
Sup720(config-if)#port-channel min-links 3
```

High Availability Campus Design
Agenda

- **Network Level Resiliency**
  - High Availability Design Principles
  - Redundancy in the Distribution Block
  - Redundancy and Routing Design
- **System Level Resiliency**
  - Integrated Hardware and Software Resiliency
    - NSF/SSO
    - ISSU & IOS Modularity
  - System Management Resiliency
    - GOLD & EEM
- **Hardening the Campus Network Design**



System Level Resiliency
Comprehensive Physical Redundancy

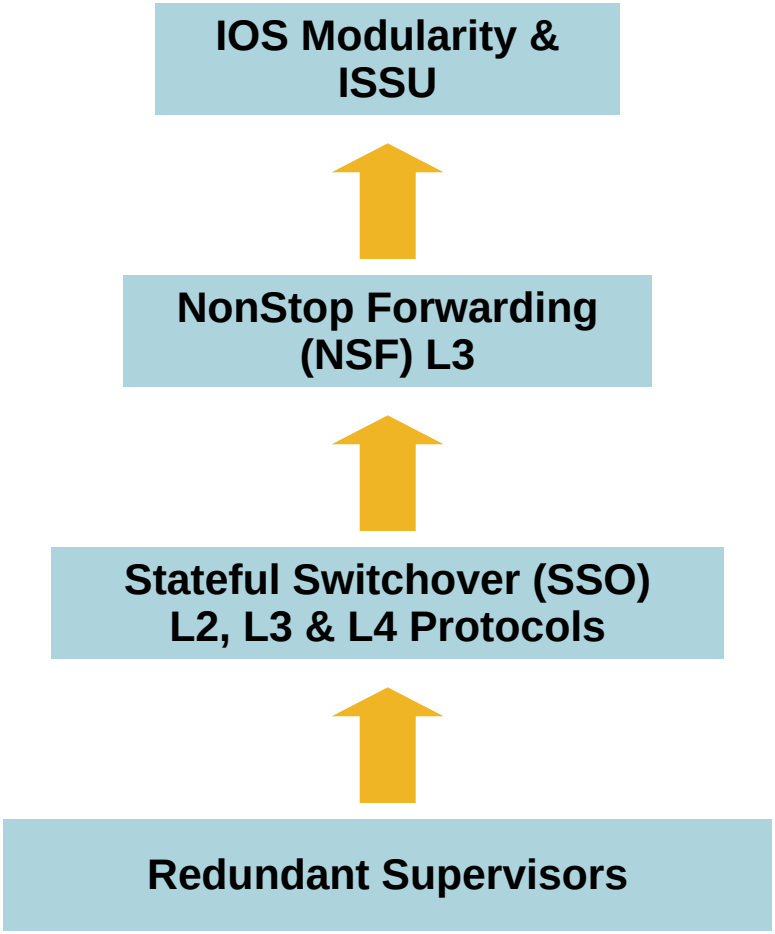
- **Catalyst 6500 and 4500 highly redundant Modular systems**
  - **Redundant hot swappable Supervisors**
  - **Redundant hot swappable Power Supplies**
  - **N+1 redundant fans with hot swappable fan trays**
  - **Hot swappable line cards**
  - **Passive data backplane**
  - **Redundant system clock modules**
- **Catalyst 3750/3750E StackwisePlus* technology**
  - **1:N Master redundancy**
  - **Hot swappable stack members**
  - **Hot swappable Power Supplies***

*** Discover the new C3560-E & C3750-E benefits at the World of Solutions floor**

System Level Resiliency

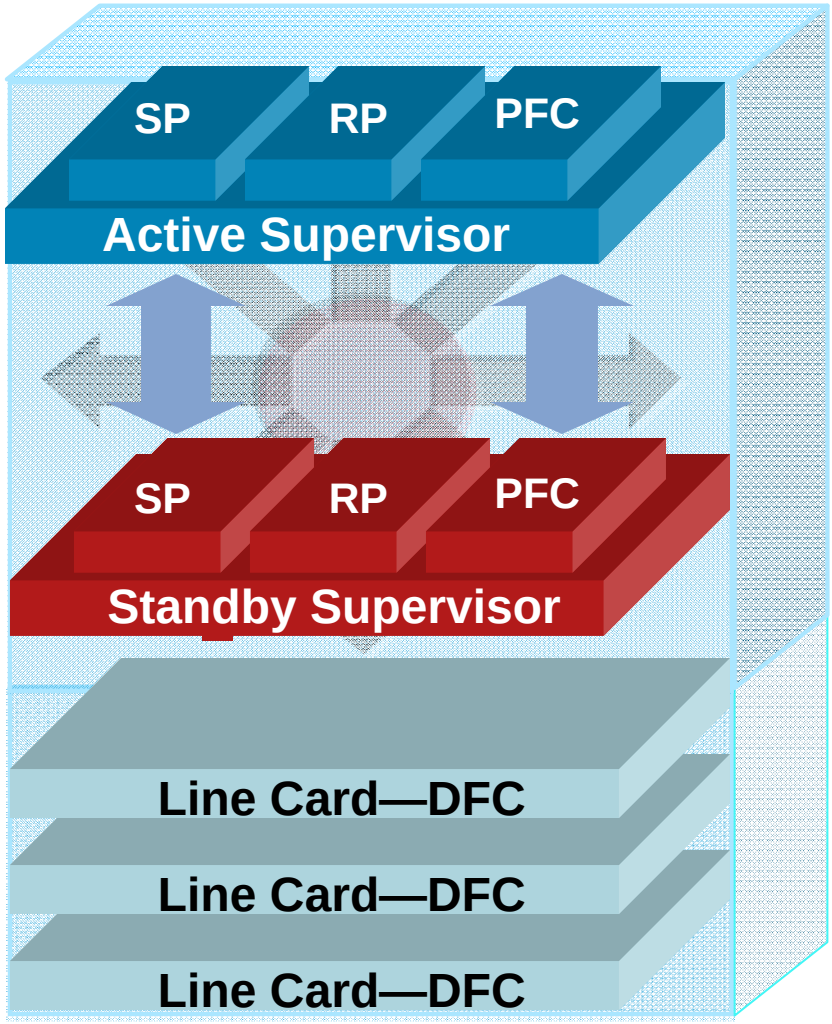
NSF/SSO, IOS Modularity and ISSU

- **Catalyst 4500 and 6500 Supervisor hardware redundancy (1+1) will leverage four key mechanisms to improve network resiliency and provide for enhanced operational change processes**
  - **SSO—Stateful Switchover**
  - **NSF—NonStop Forwarding**
  - **IOS Modularity**
  - **ISSU—In Service Software Upgrade**
- **Catalyst 3750 stack switch redundancy leverages two mechanisms to improve network resiliency**
  - **Stackwise and StackwisePlus**
  - **NSF supported as of 12.2(35)SE**



Supervisor Processor Redundancy

Stateful Switch Over (SSO)



- Active/standby supervisors run in synchronized mode
- Redundant supervisor is in ‘hot-standby’ mode
- Switch processors synchronize L2 port state information, (e.g., STP, 802.1x, 802.1q)
- Switching HW synchronizes L2/L3 FIB, NetFlow and ACL tables
- DFCs are populated with L2/L3 FIB, NetFlow, and ACL tables

Supervisor Processor Redundancy
Stateful Switch Over (SSO)

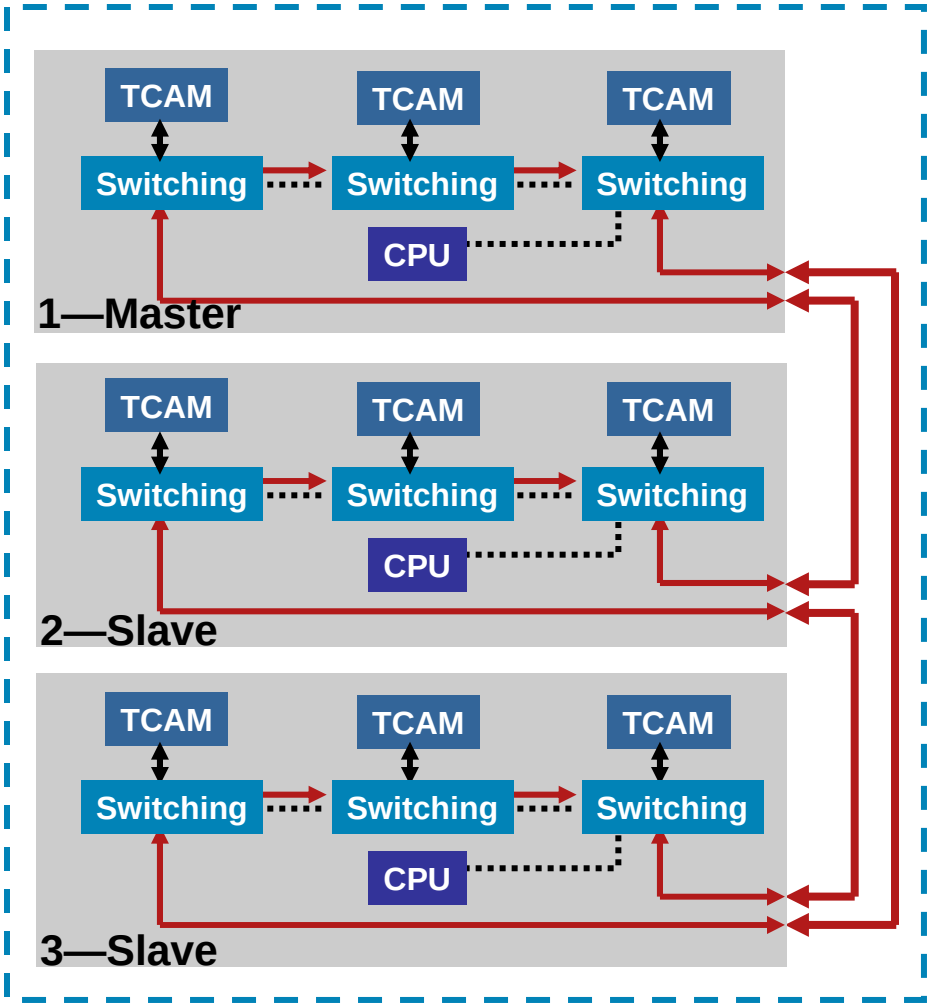
```
Switch(config)#redundancy

Switch(config-red)#mode ?
  rpr  Route Processor Redundancy

  sso  Stateful Switchover
```

```
Switch#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    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 Resiliency

Cisco Catalyst 3750 StackWise & StackwisePlus

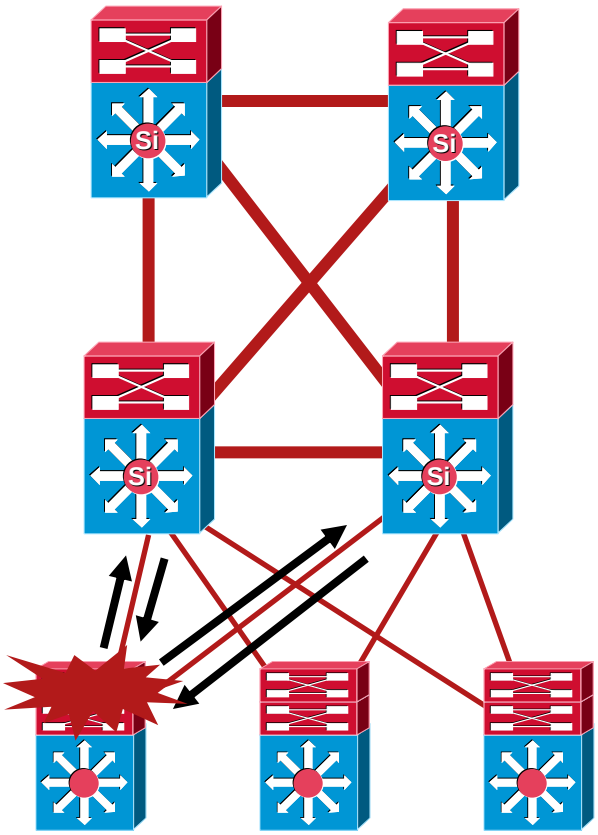
- **Centralized configuration and management**
- **Switching fabric extended via bidirectional self healing ring**
- **Each TCAM contains full FIB, ACL and QoS information**
- **Redundancy is provided via a combination of centralized and distributed feature replication**
  - **Certain functions are managed centrally on the stack master node (e.g. L3 is centrally managed)**
  - **Certain functions are replicated on all switches (e.g., Spanning Tree, Flex Links)**



System Resiliency

NSF Recovery (Routing Protocol Recovery)

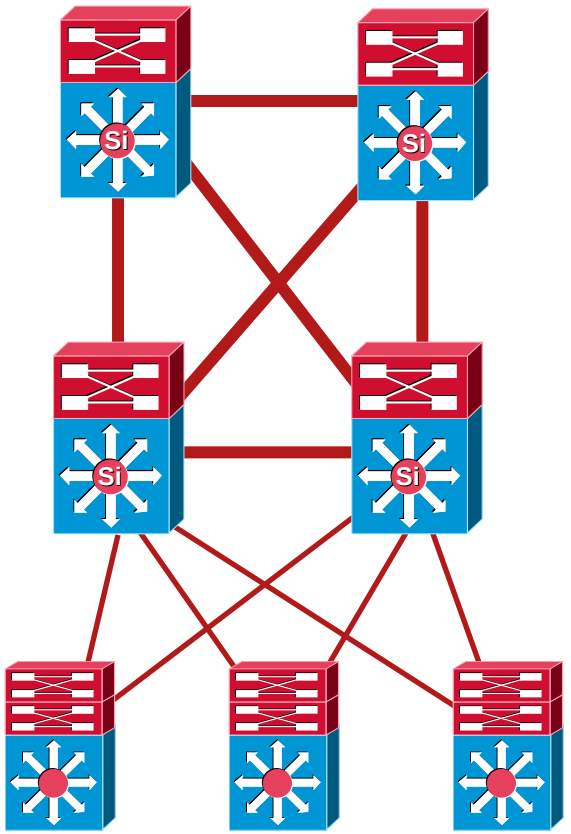
- **Non-Stop Forwarding (NSF) provides the capability for the routing protocols to gracefully restart after an SSO fail-over**
- **The newly active redundant supervisor continues forwarding traffic using the synchronized HW forwarding tables**
- **The NSF capable Routing Protocol requests a graceful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



No Route Flaps During Recovery

System 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

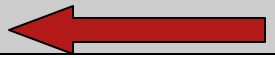
System Resiliency
NSF Configuration

```
Switch(config)#router eigrp 100
Switch(config-router)#nsf

Switch(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
```

```
Switch(config)#router ospf 100
Switch(config-router)#nsf

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

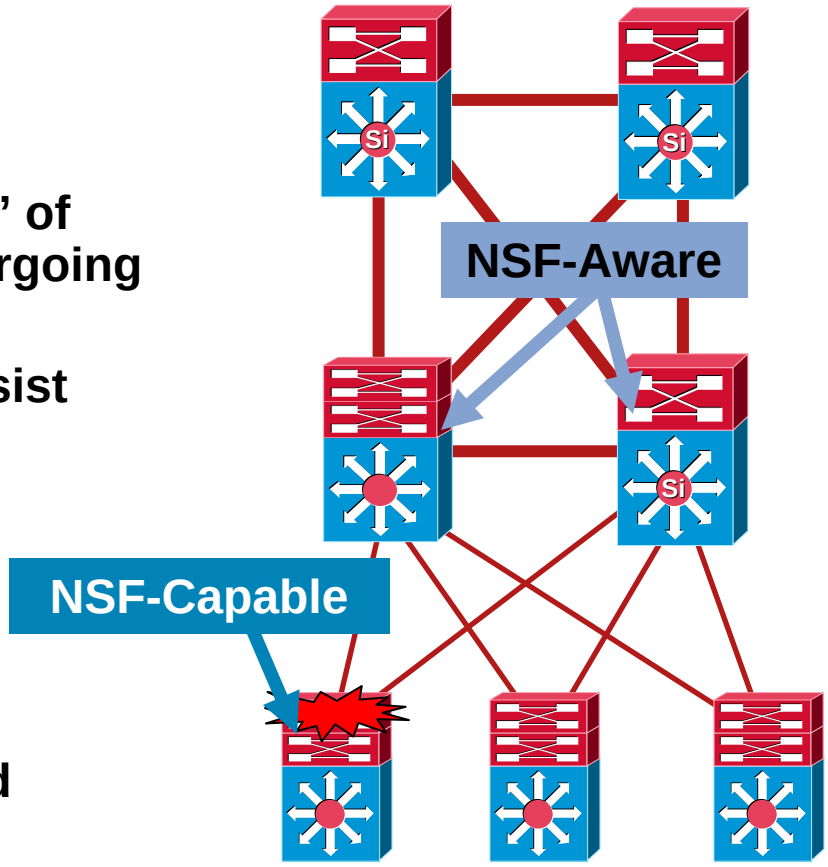
```
Switch(config)#router isis level2
Switch(config-router)#nsf cisco
        'or'
Switch(config)#router isis level2
Switch(config-router)#nsf ietf
```

```
Switch(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>

Switch(config-router)#bgp graceful-restart
```



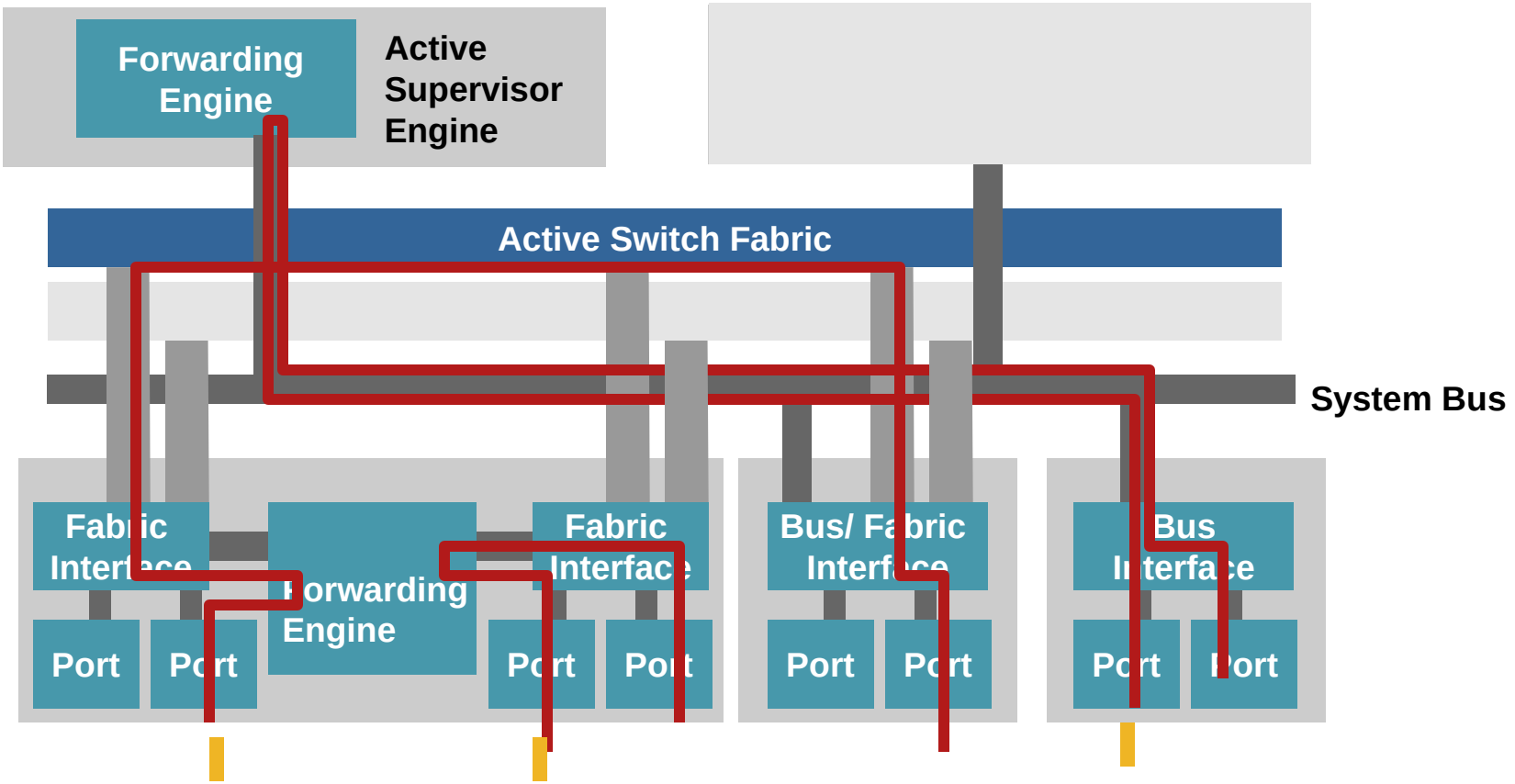
Design Considerations for NSF/SSO
NSF Capable vs. NSF Awareness

- **Two roles in NSF neighbor graceful restart**
  - **NSF Capable**
  - **NSF Aware**
- **An NSF-Capable router is 'capable' of continuous forwarding while undergoing a switchover**
- **An NSF-Aware router is able to assist NSF-Capable routers by:**
  - **Not resetting adjacency**
  - **Supplying routing information for verification after switchover**
- **NSF capable and NSF aware peers cooperate using Graceful Restart extensions to BGP, OSPF, ISIS and EIGRP protocols**



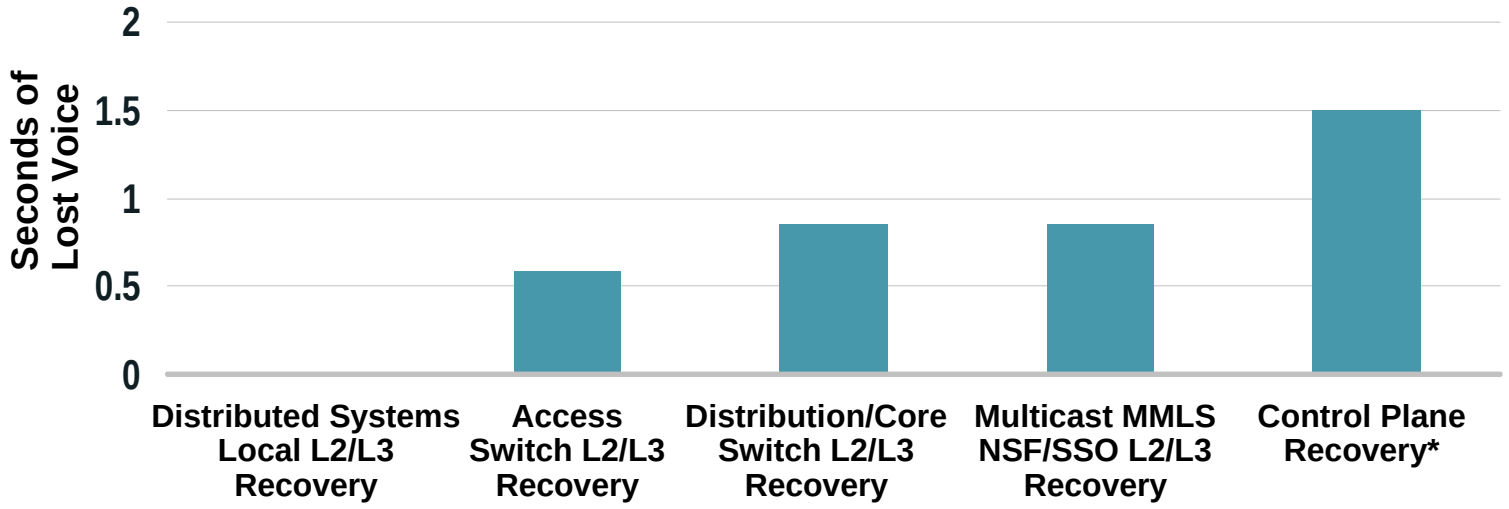
Design Considerations for NSF/SSO
Cisco Catalyst 6500 Line Card Interaction

- **Time to recover the data plane depends on how fast the forwarding engine, switch fabric and bus can be recovered**



Design Considerations for NSF/SSO
Cisco Catalyst 6500 Line Card Interaction

- **The time to recover the data plane depends on the traffic switching path**
  - **Access switches are generally classic systems that depend on the central forwarding engine and bus failover speed for data path recovery**
  - **Distribution and core switches are generally fabric-based systems that depend on the fabric failover for data path recovery. In addition, centralized system (no distributed forwarding engine DFC) depend on the central forwarding engine failover speed for data path recovery**
  - **Zero packet loss can achieved with distributed systems for traffic patterns where the fabric does not need to be traversed**

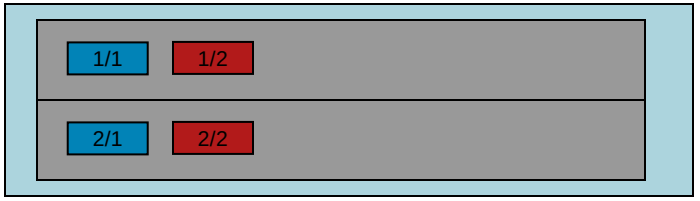


***The Time to Recover the Control Plane Is Calculated Based on the Ability of the Route Processor to Send ICMP Echo Reply Packets**

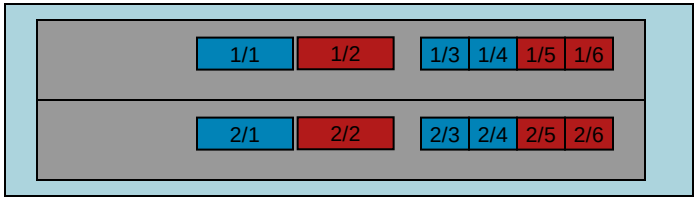
Design Considerations for NSF/SSO
Supervisor Uplinks

- **Cisco Catalyst 4500: supervisor uplink ports are active and forward traffic as long as the supervisor is fully inserted**
  - **Uplink ports do not go down when a supervisor is reset. There are restrictions on which ports can be active simultaneously in redundant systems**
- **Cisco Catalyst 6500: both the active supervisor and the standby supervisor uplink ports are active as long as the supervisors are up and running**
  - **Uplink ports go down when the supervisor is reset**

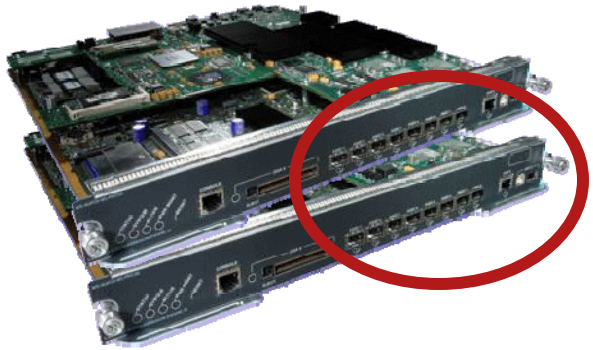
- **Catalyst 4500 Supervisor II+, Supervisor IV: 2 x GigE ports are active**



- **Catalyst 4500 Supervisor II+10GE: 2 x 10GE and 4 x GigE ports are active**

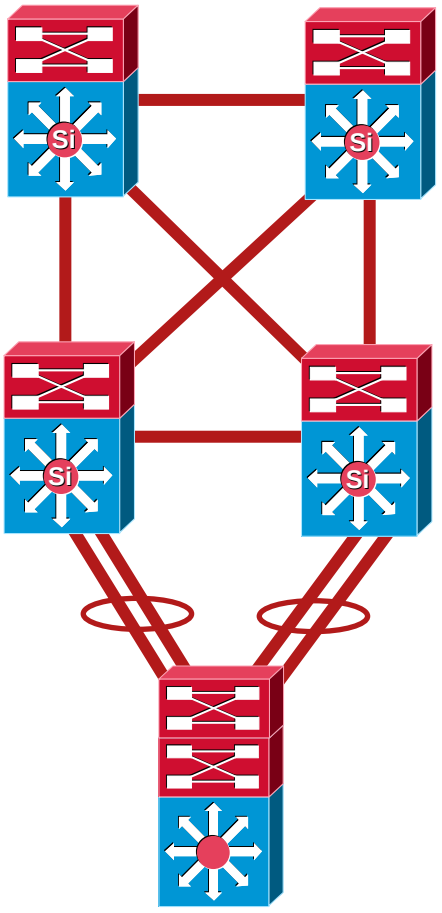


- **Catalyst 6500 Supervisors: all ports are active**



Design Considerations for NSF/SSO
Supervisor Uplinks and Design Recommendations

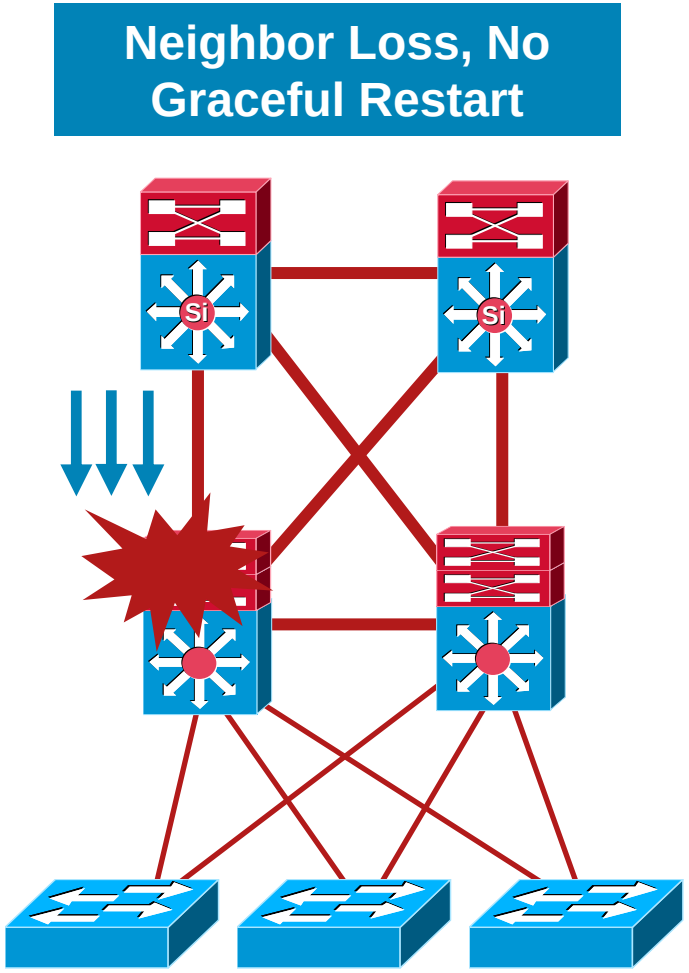
- The use of Catalyst 6500 Supervisor uplinks with NSF/SSO results in a more complex network recovery scenario
- Dual failure scenario
  - Supervisor Failure
  - Port Failure
- During recovery FIB is frozen but uplink port is gone
- PFC tries to forward traffic out a non-existent link → this leads to a 24 seconds worst case convergence time
- The problem is solved in the software release 12.2(18)SXF5
- Bundling Supervisor uplinks into Etherchannel links is still consider best practice design for Sup uplinks
- These recommendations do not apply to the Cisco Catalyst 4500



Design Considerations for NSF/SSO

NSF and Hello Timer Tuning?

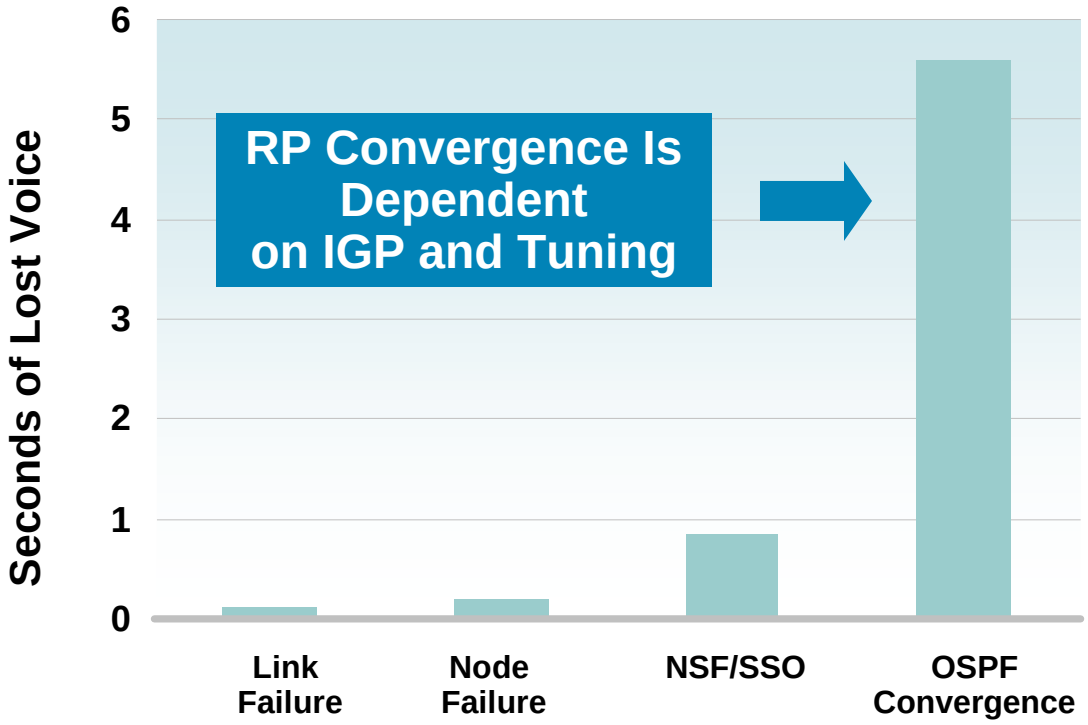
- NSF is intended to provide availability through route convergence avoidance
- Fast IGP timers are intended to provide availability through fast route convergence
- In an NSF environment dead timer must be greater than SSO Recovery + RP restart + time to send first hello
- Switches running Native IOS
  - OSPF 2/8 seconds for hello/dead
  - EIGRP 1/4 seconds for hello/hold
- Switches running Hybrid
  - OSPF 3/12 seconds for hello/dead
  - EIGRP 2/8 seconds for hello/hold

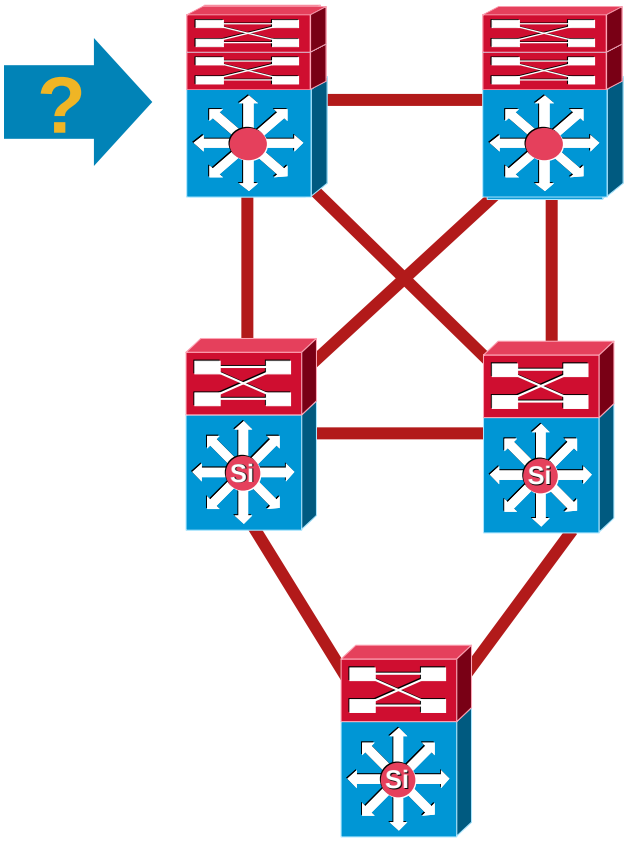


Design Considerations for NSF/SSO

Where Does It Make Sense?

- **Redundant topologies with equal cost paths provide sub-second convergence**
- **NSF/SSO provides superior availability in environments with non-redundant paths**

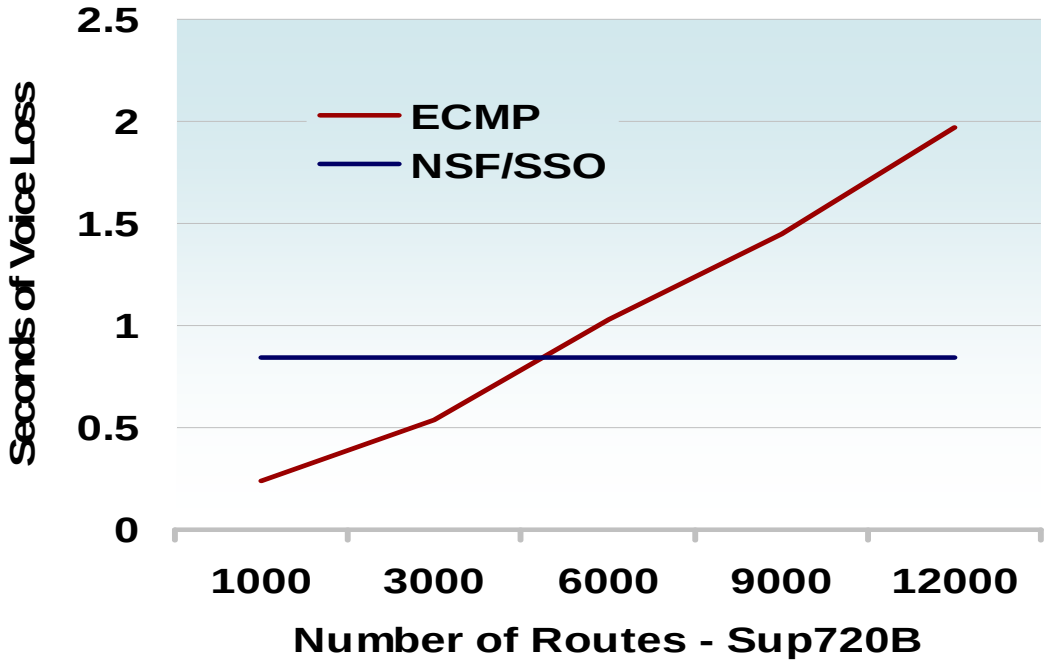


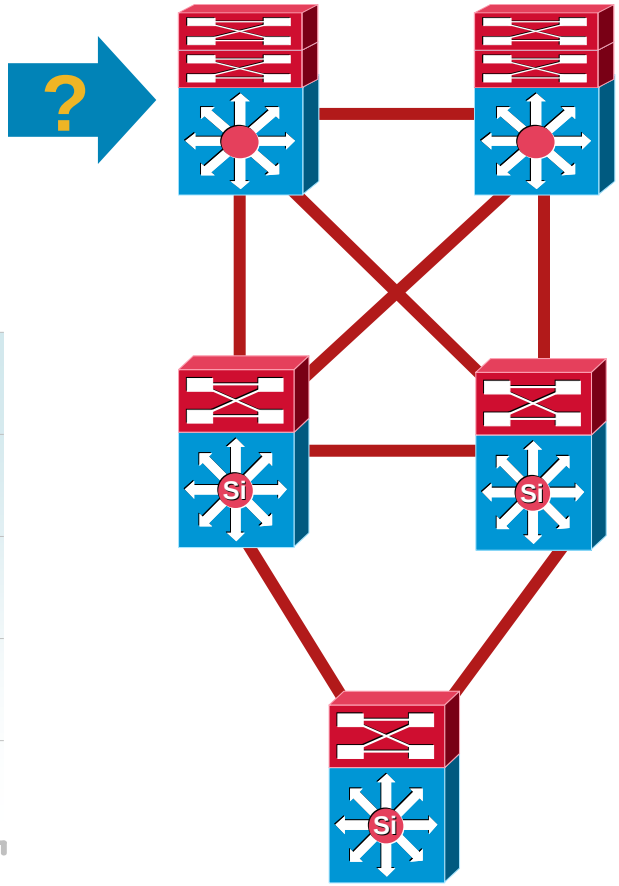


Design Considerations for NSF/SSO

Where Does It Make Sense?

- **Equal Cost Multipath (ECMP) recovery provides fastest convergence**
- **NSF/SSO provides consistent recovery independent of the number of routes**

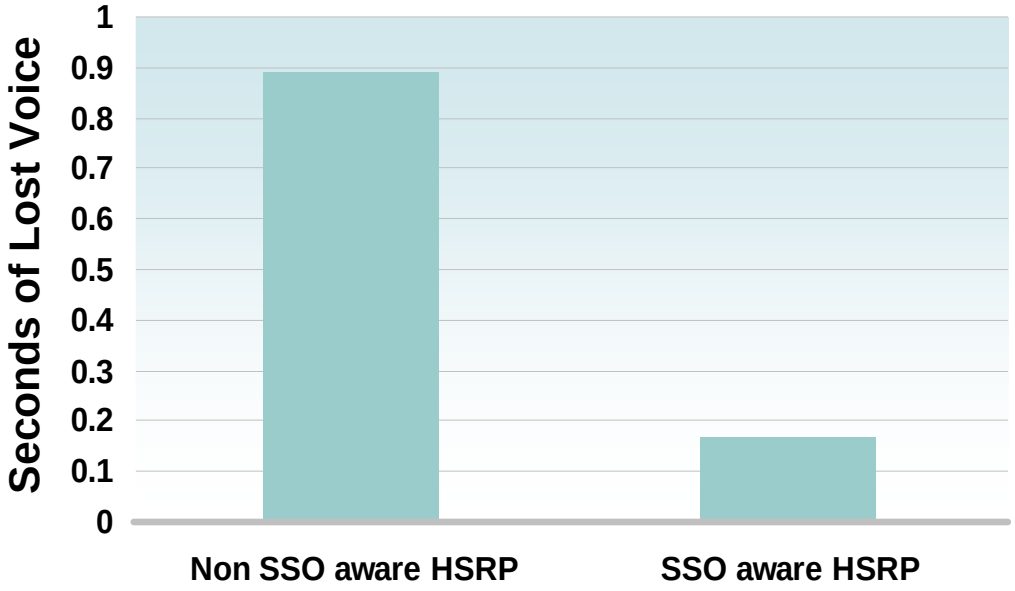


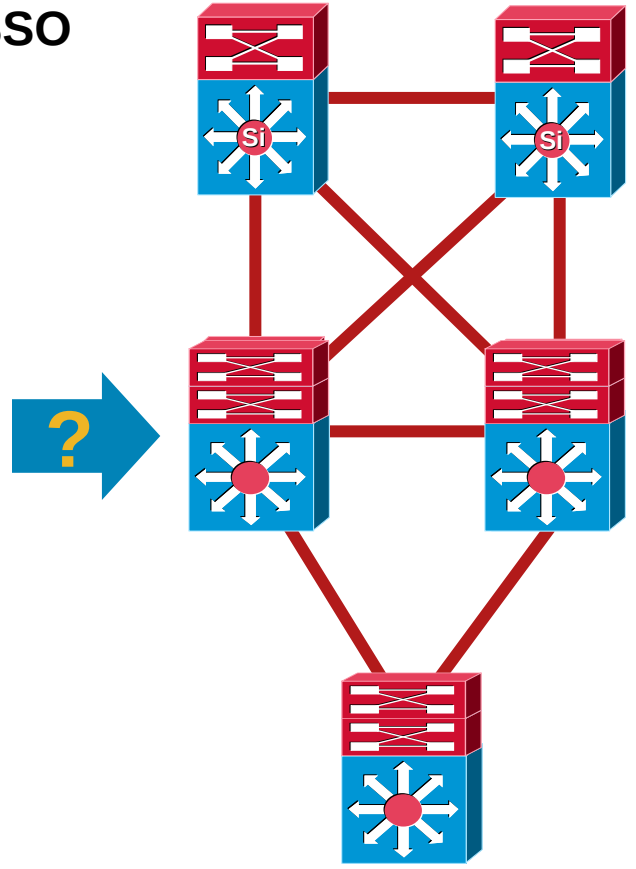


Design Considerations for NSF/SSO

Where Does It Make Sense?

- Not all IOS features are SSO aware
- As of 12.2(31)SG Catalyst 4500 supports SSO aware HSRP
- 6500 will support in Q107
- HSRP doesn't flap on Supervisor SSO switchover

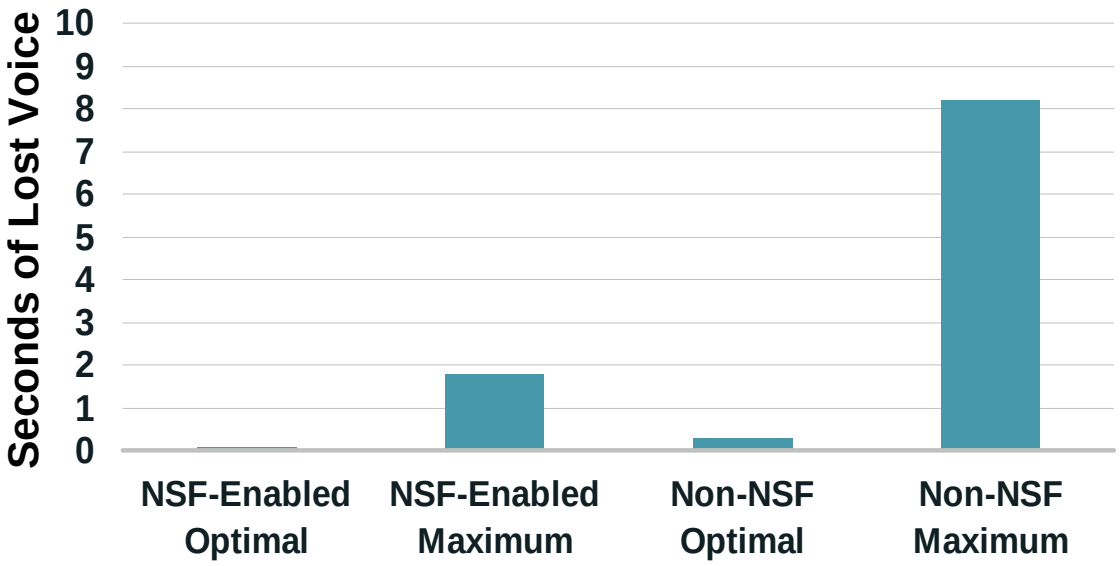


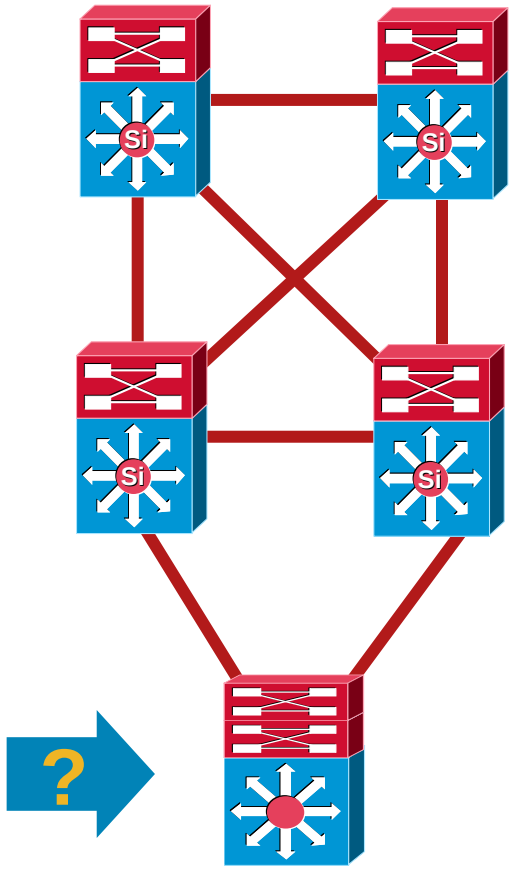


Design Considerations for NSF/SSO

Where Does It Make Sense?

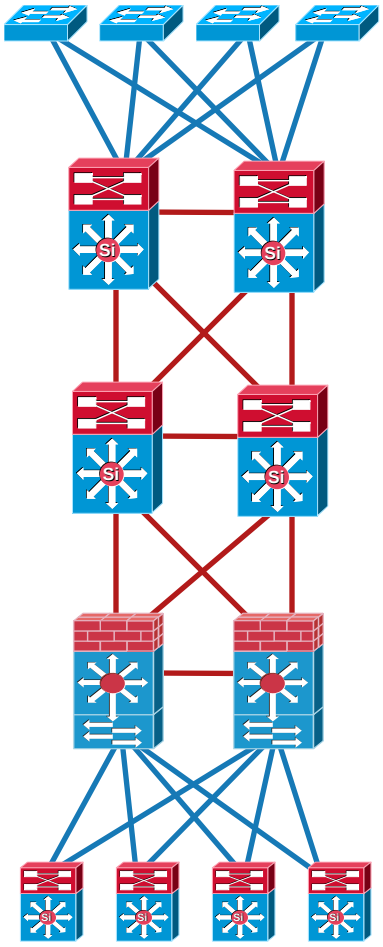
- **Access switch is the single point of failure in best practices HA campus design**
- **Supervisor failure is most common cause of access switch service outages**
- **Recommended design NSF/SSO provides for *sub 600 msec* recovery of voice and data traffic**





High Availability Campus Design
Agenda

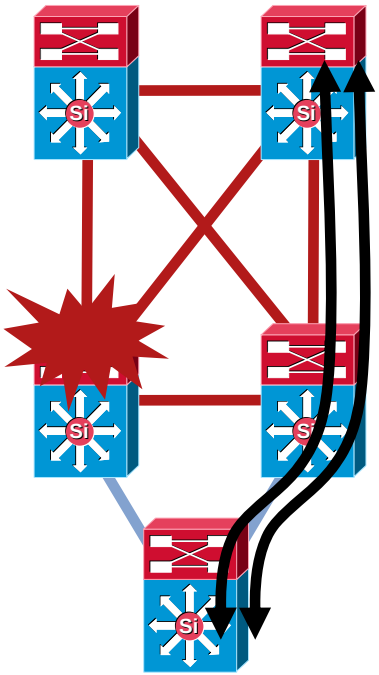
- **Network Level Resiliency**
  - High Availability Design Principles
  - Redundancy in the Distribution Block
  - Redundancy and Routing Design
- **System Level Resiliency**
  - Integrated Hardware and Software Resiliency
    - NSF/SSO
    - ISSU & IOS Modularity
  - System Management Resiliency
    - GOLD & EEM
- **Hardening the Campus Network Design**



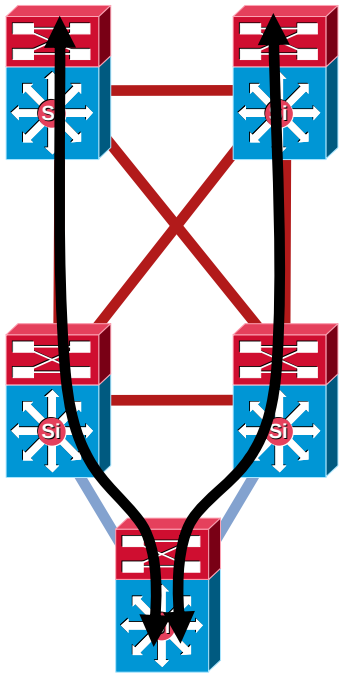
System Resiliency

IOS Modularity and In Service Software Upgrade

- In redundant topology standard maintenance practice is to shut down devices during upgrade and let the network converge
- IOS Modularity and ISSU provide the ability to patch or 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

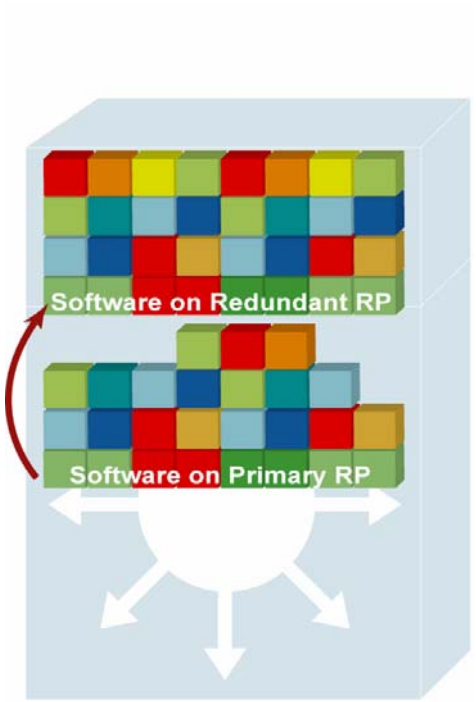


Scheduled Maintenance—Half Capacity

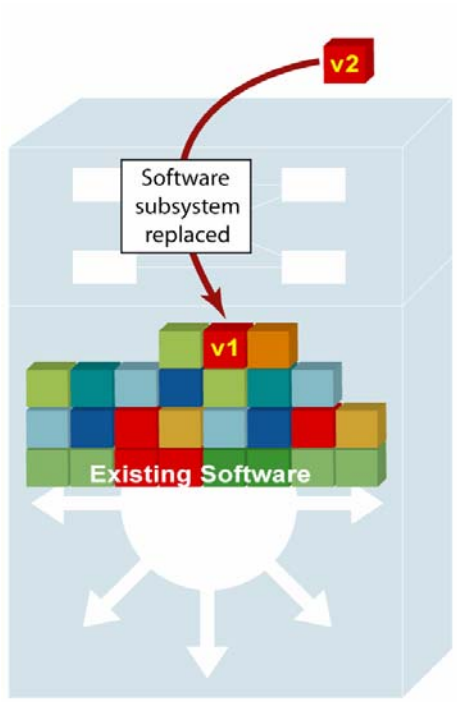


ISSU—All Paths and Switches Active During Upgrade

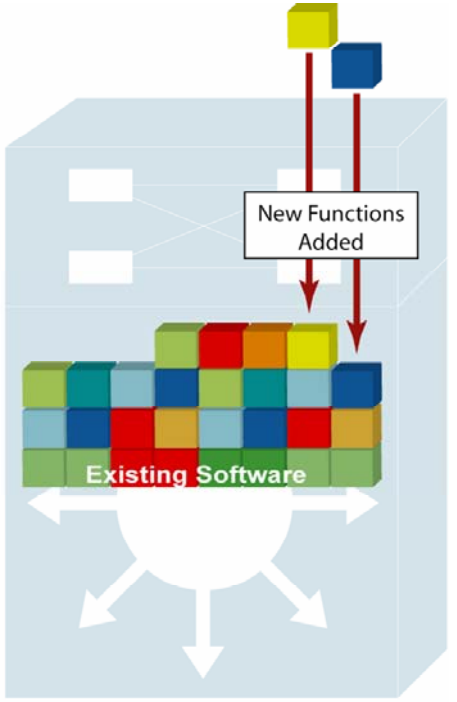
System Resiliency
In Service Software Upgrade (ISSU)



- Full image upgrade
- New features and patches



- Selective maintenance
- Patch a component

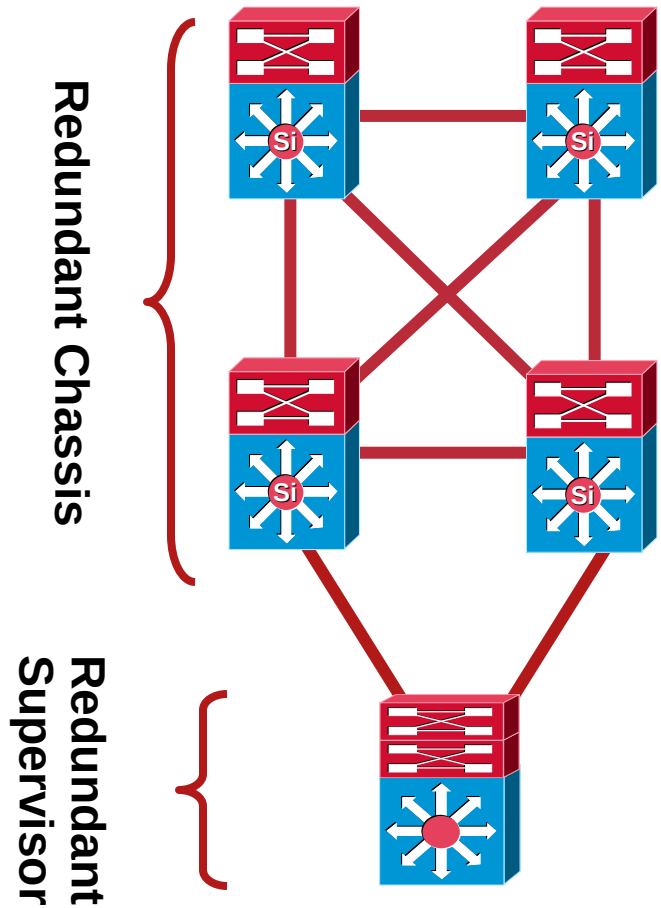


- Component Upgrade
- Add new features to existing base

System Resiliency

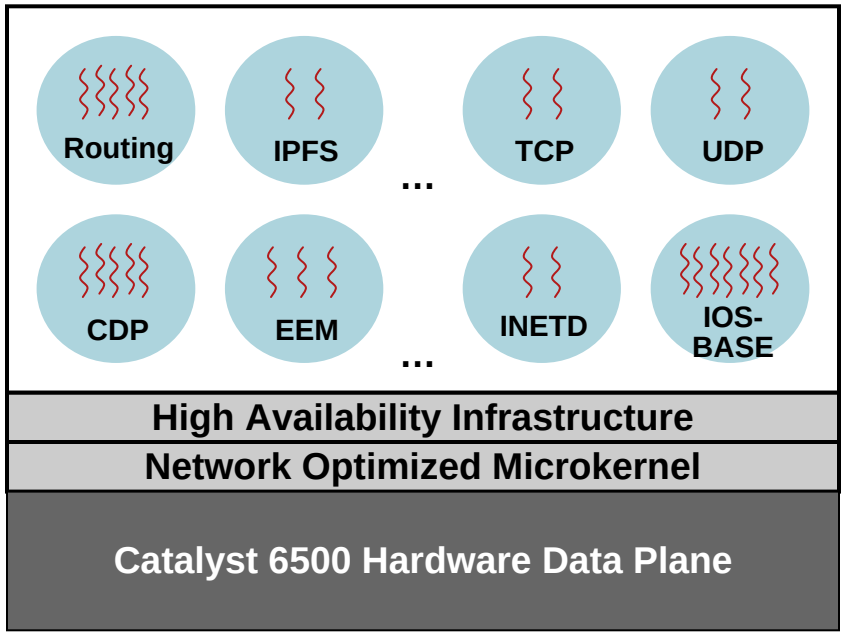
IOS Modularity and In Service Software Upgrade

- IOS Modularity provides for a mechanism to patch subsystem components in a single supervisor system and is well suited to dual switch configurations (e.g. typical Distribution and Core environments)
- Full image ISSU requires a dual supervisor environment and is well suited to single points of failure (e.g. Access layer environments)
- IOS Modularity is also supported in dual supervisor access environments



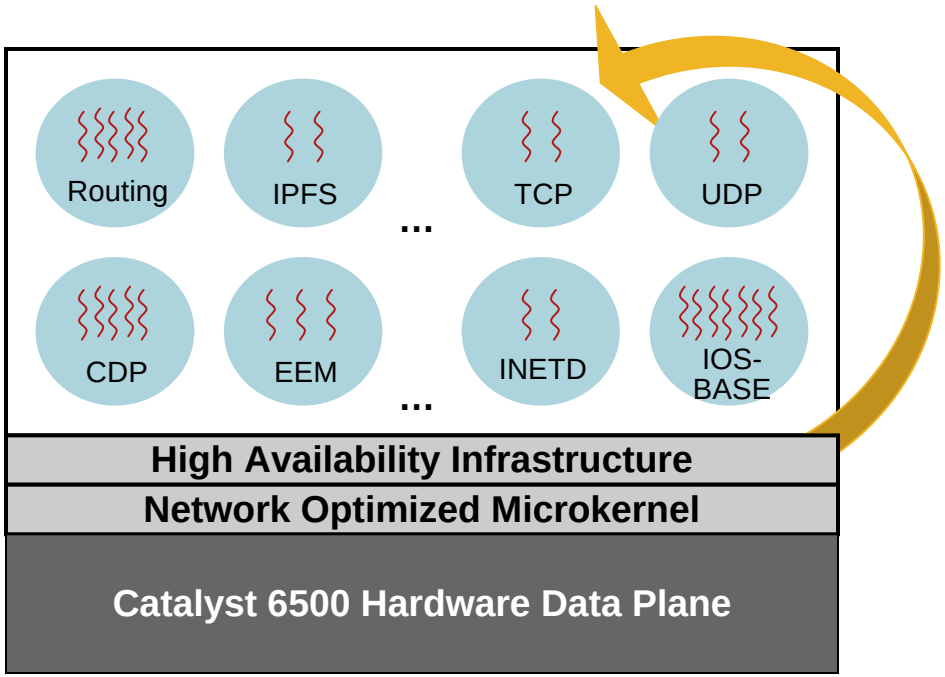
Cisco IOS Software Modularity
Catalyst 6500

- **Combines a network optimized microkernel with the feature subsystems and functions enterprise and metro Ethernet customers depend on:**
  - **20+ independent processes**
  - **Remaining feature subsystems live in Cisco IOS Base process**
  - **Retains support for Cisco IOS features**
- **Whole system benefits from integrated HA infrastructure which determines best action to take for improved resiliency**
- **Preserves Cisco Catalyst 6500 Series benefits:**
  - **Separate Control and Data Planes**
  - **NSF and GOLD**
  - **Hardware Acceleration**
  - **Scalability**



Cisco IOS Software Modularity Benefits

Minimize Unplanned Downtime



Traffic Forwarding Continues During Unplanned Process Restarts

If an Error Occurs in a Modular Process

- **HA subsystem determines the best recovery action**
  - **Restart a modular process**
  - **Switchover to standby supervisor**
  - **Remove the system from the network**
- **Process restarts with no impact on the data plane**
  - **Utilizes *Nonstop Forwarding (NSF)* even with a single Supervisor with NSF-Aware neighbors**
  - **State checkpointing allows quick process recovery**

Cisco IOS Software Modularity

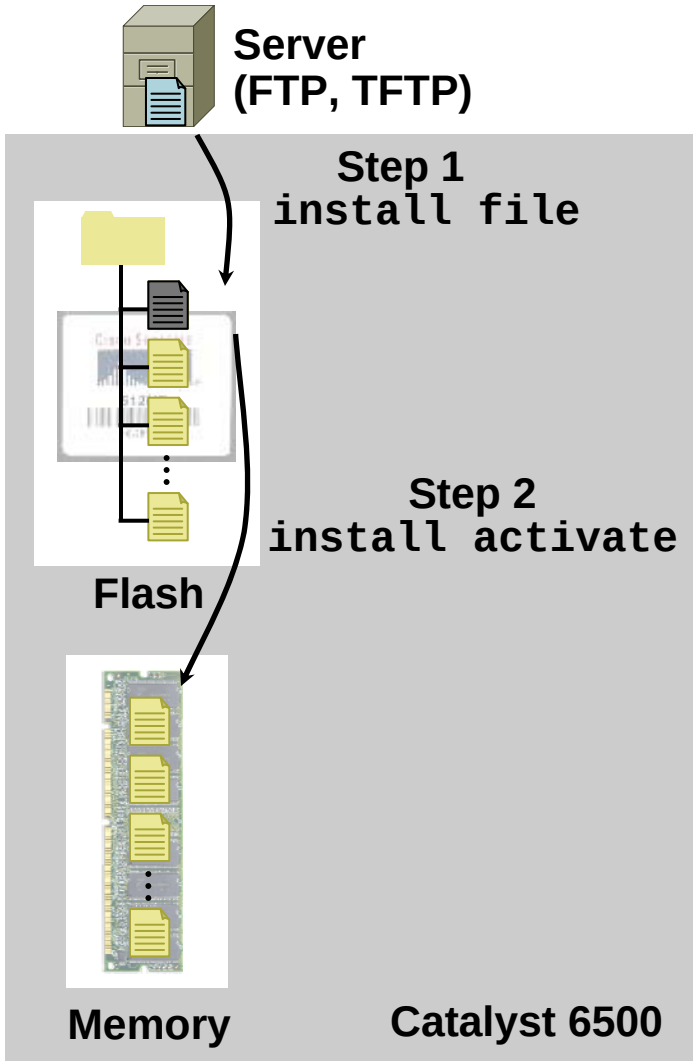
Subsystem ISSU – Software Patching

Patching is always a two steps process:

1. **Install the patch**
   - Does not change anything on the running version of code
   - Can be performed for multiple patches before next step
   - Verifies patch dependencies
2. **Activate the patch**
   - All patches that are pending for install are activated at the same time
   - Copy of previous code is retained for rollback purposes

Patches downloaded from CCO

http://www.cisco.com/go/pn



Cisco IOS Software Modularity

Subsystem ISSU – Software Patching

```
Swmod#install file tftp://172.16.1.1/images/demo/s72033-
XMA0001.122-18.SXF4 disk0:/sys
Address or name of remote host [172.16.1.1]?
Source filename [images/demo/s72033-XMA0001.122-18.SXF4]?
!!!!!!!!!!!!!!!!!!!
Verifying checksums of extracted files

Verifying installation compatibility
Gathering information for location s72033_rp - Slot 2
…
Activation will affect the following processes:
iprouting.iosproc
…
NOTE: The newly added patch is not yet active.
   Use 'install activate' to activate the patch
   in the currently running system.

[DONE]
```

Only Top Level Directory Needed

Indicates What Process(es) this Patch Will Affect

```
Swmod#install activate disk0:/sys
Determining processes to restart at location s72033_rp - Slot 2
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!...

The following processes will be restarted:
iprouting.iosproc

Do you want to continue with activating this change set...?
[yes/no]: yes
Proceeding with activation, writing installer meta-data ...

Updating more installer meta-data ...

Beginning process restarts ...
!!!!!!!!!!!!!!!!!!!!
Affected processes restarted.

[DONE]
```

User Confirmation Is Required

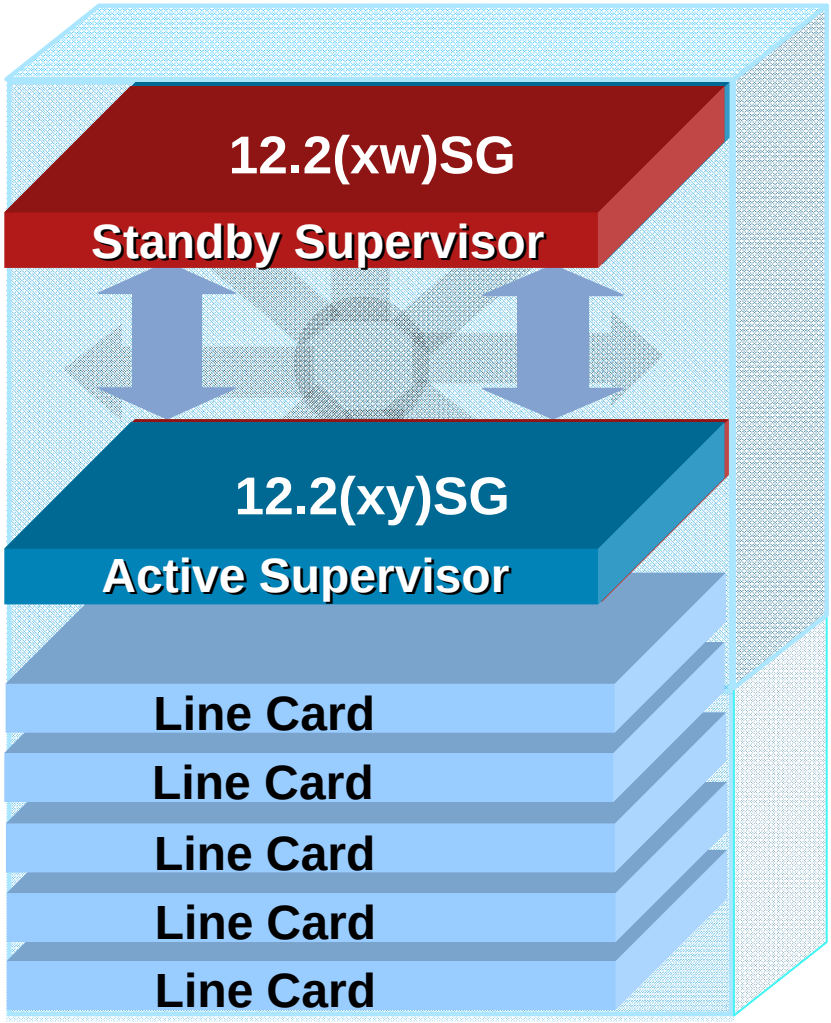
- **Step 1: Install the maintenance pack/patch**
- **Step 2: Activate the pending changes**

No packet loss during patching!

In Service Software Upgrade

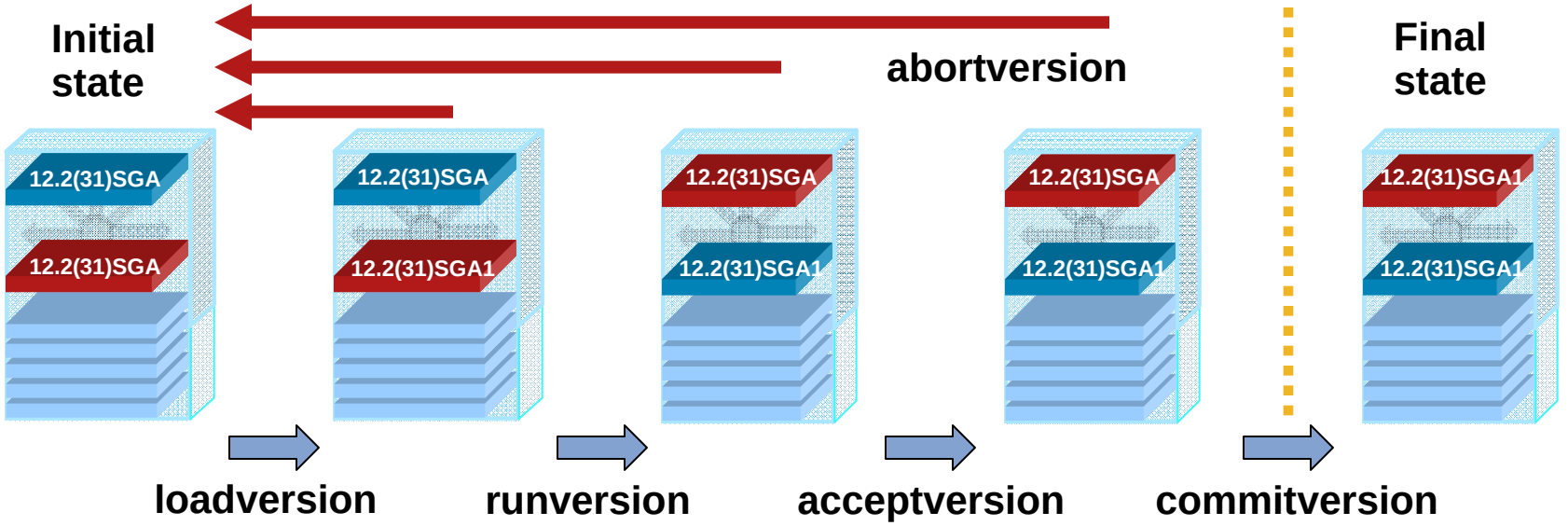
Catalyst 4500

- Full image 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)
- Network does not re-route and no active links are taken out of service



In Service Software Upgrade

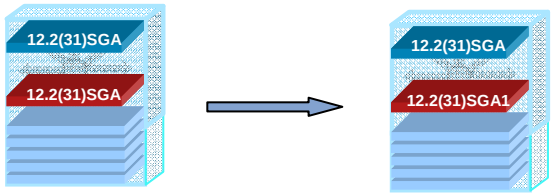
ISSU Stages



- **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



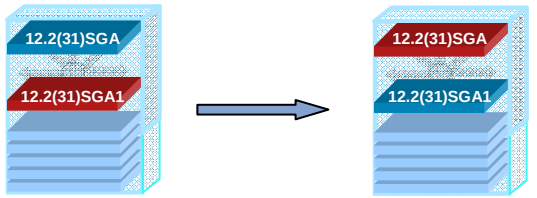

```
Switch#sh issu state
                        Slot = 1
                   RP State = Active
                  ISSU State = Init
                Boot Variable = bootflash:cat4500-entservicesk9-mz.122-31.SGA.bin,12;

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


Switch#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

```
Switch#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



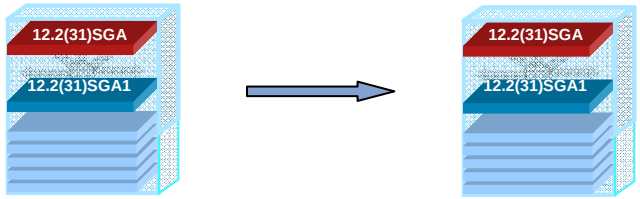

```
Switch#issu runversion 2 slavebootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin
This command will reload the Active unit.  Proceed ? [confirm]
```

SSO Failover to Redundant Supervisor running new image

```
Switch#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


Switch#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.
```

ISSU Upgrade Process

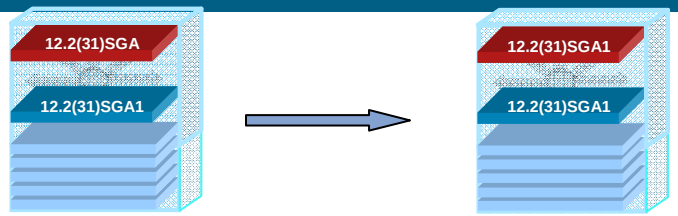
Step 3 - acceptversion




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

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

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



Switch#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



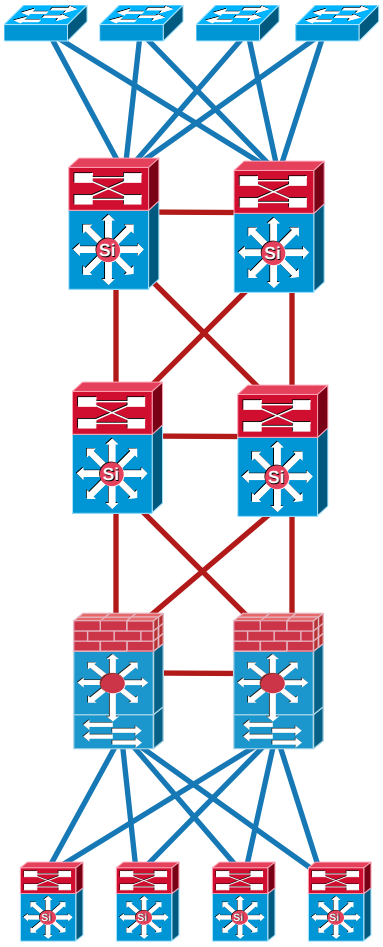

```
Switch#$issu commitversion 1 slavebootflash:cat4500-entservicesk9-mz.122-31.SGA1.bin
```

Standby Supervisor Reboots with new image

```
Switch#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;

Switch#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
```

High Availability Campus Design
Agenda

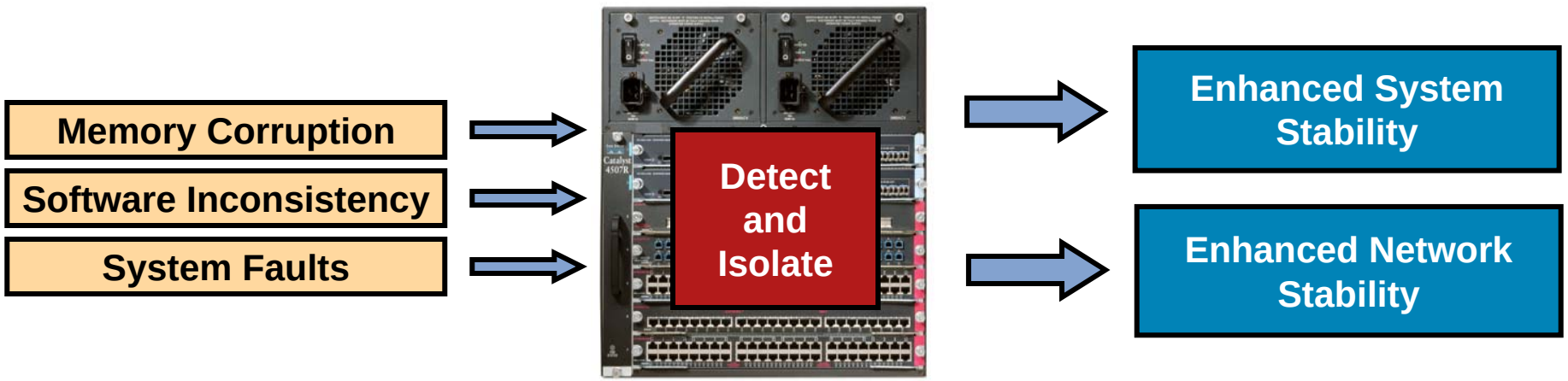
- **Network Level Resiliency**
  - High Availability Design Principles
  - Redundancy in the Distribution Block
  - Redundancy and Routing Design
- **System Level Resiliency**
  - Integrated Hardware and Software Resiliency
    - NSF/SSO
    - ISSU & IOS Modularity
  - System Management Resiliency
    - GOLD & EEM
- **Hardening the Campus Network Design**



Systems Resiliency

Proactive Fault Detection and Notification

Improved physical redundancy is not enough, intelligent system failure detection is key

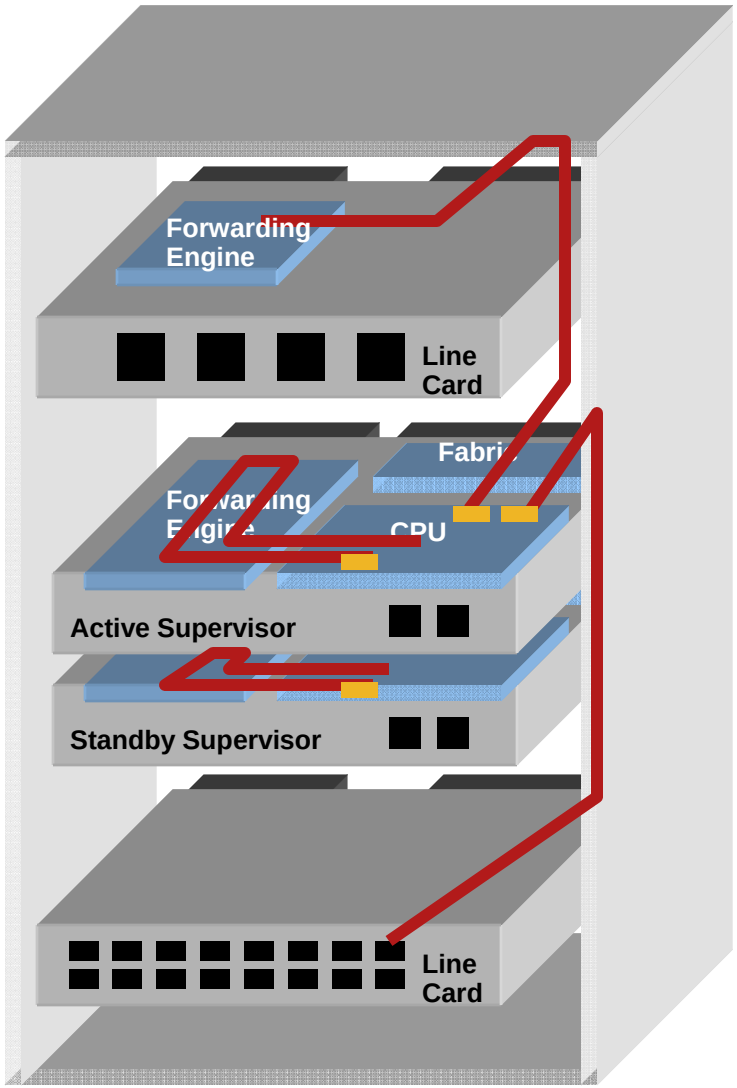


Power-on Diagnostics	Generic Online Diagnostics
Supervisor, Backplane, L2 ASIC, L3 ASIC, Memory, Port	HW/SW state, Memory LC module, Temperature, Power supply, Fan tray

Generic Online Diagnostics

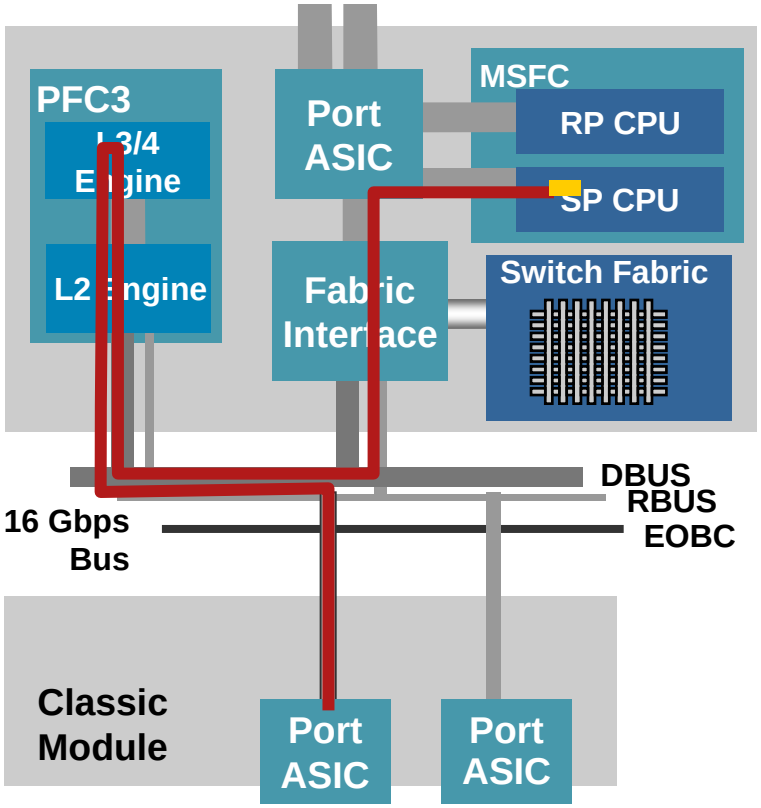
How Does GOLD Work?

- **GOLD: Check the health of hardware components and verify proper operation of the system data plane and control plane at run-time and boot-time**
- **Diagnostic packet switching tests verify that the system is operating correctly:**
  - **Is the supervisor control plane and forwarding plane functioning properly?**
  - **Is the standby supervisor ready to take over?**
  - **Are linecards forwarding packets properly?**
  - **Are all ports working?**
  - **Is the backplane connection working?**
- **Other types of diagnostics tests including memory and error correlation tests are also available**



Generic Online Diagnostics

An Example: LoopbackTest - Linecard Data Path Coverage



- **Test is disruptive for the tested port (subseconds)**
- **Verifies the tested port functionality and the datapath between the supervisor and the loopbacked port**
- **Newer linecards support non-disruptive loopback tests: ten consecutive failures are treated as fatal and will result in port being error-disabled**

```
Switch#diagnostic start module 1 test 2 port 1
Module 1: Running test(s) 2 may disrupt normal system operation
Do you want to continue? [no]: yes
13:50:01: %DIAG-SP-6-TEST_RUNNING: Module 1: Running TestLoopback{ID=2} ...
13:50:01: %DIAG-SP-6-TEST_OK: Module 1: TestLoopback{ID=2} has completed successfully
```

Generic Online Diagnostics

Diagnostic Operation

Boot-Up Diagnostics

```
Switch(config)#diagnostic bootup level complete
```

Run During System Bootup, Line Card OIR or Supervisor Switchover ***Makes Sure Faulty Hardware Is Taken out of Service***

Runtime Diagnostics

Health-Monitoring

```
Switch(config)#diagnostic monitor module 5 test 2
Switch(config)#diagnostic monitor interval module 5 test 2
00:00:15
```

Non-Disruptive Tests Run in the Background ***Serves as HA Trigger***

On-Demand

```
Switch#diagnostic start module 4 test 8
Module 4: Running test(s) 8 may disrupt normal
system operation
Do you want to continue? [no]: y
Switch#diagnostic stop module 4
```

All Diagnostics Tests Can Be Run on Demand, for ***Troubleshooting Purposes. It Can Also Be Used As A Pre-deployment Tool***

Scheduled

```
Switch(config)#diagnostic schedule module 4
test 1 port 3 on Jan 3 2005 23:32
Switch(config)#diagnostic schedule module 4
test 2 daily 14:45
```

Schedule Diagnostics Tests, for ***Verification and Troubleshooting Purposes***

Generic Online Diagnostics

Using Diagnostics as a Pre-Deployment Tool

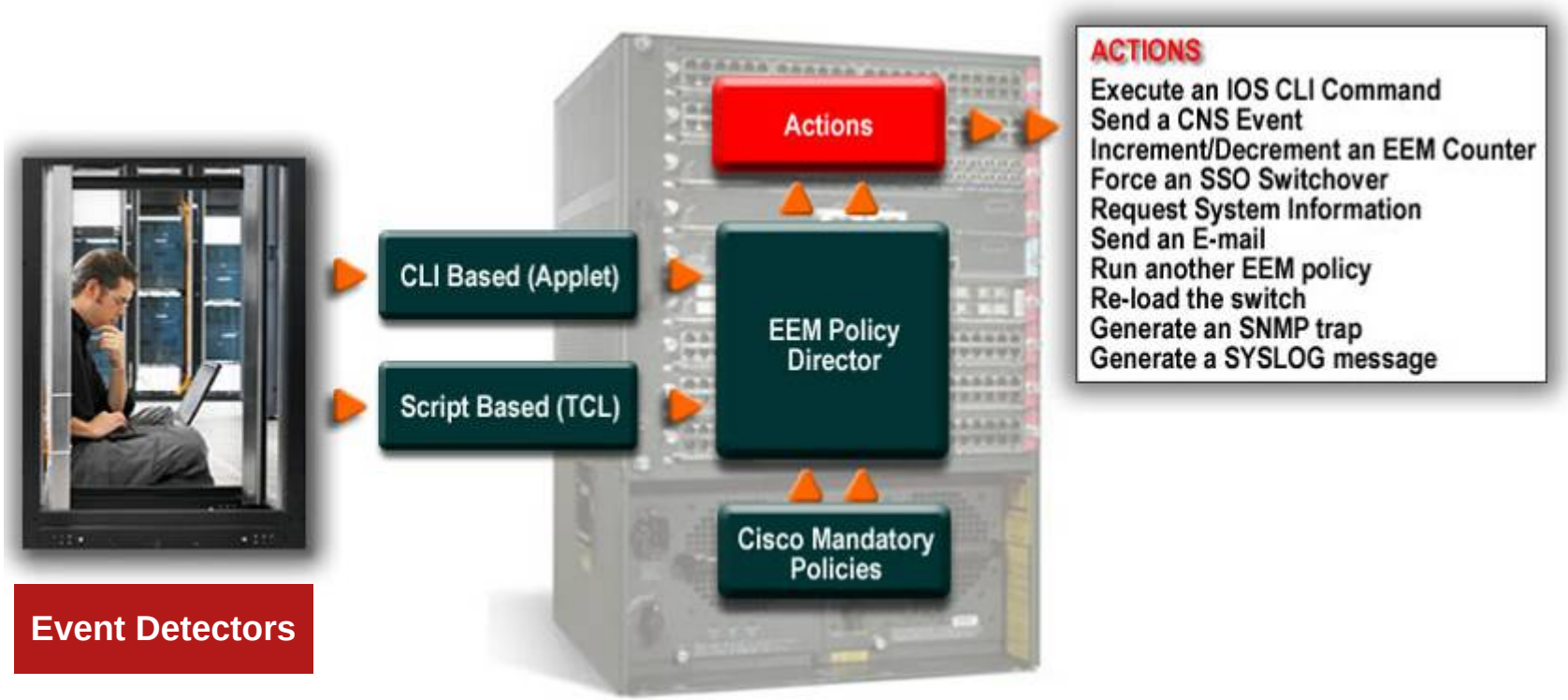
The Order in Which Tests Are Run Matters

- Run diagnostics first on linecards, then on supervisors
- Run packet switching tests first, run memory tests after

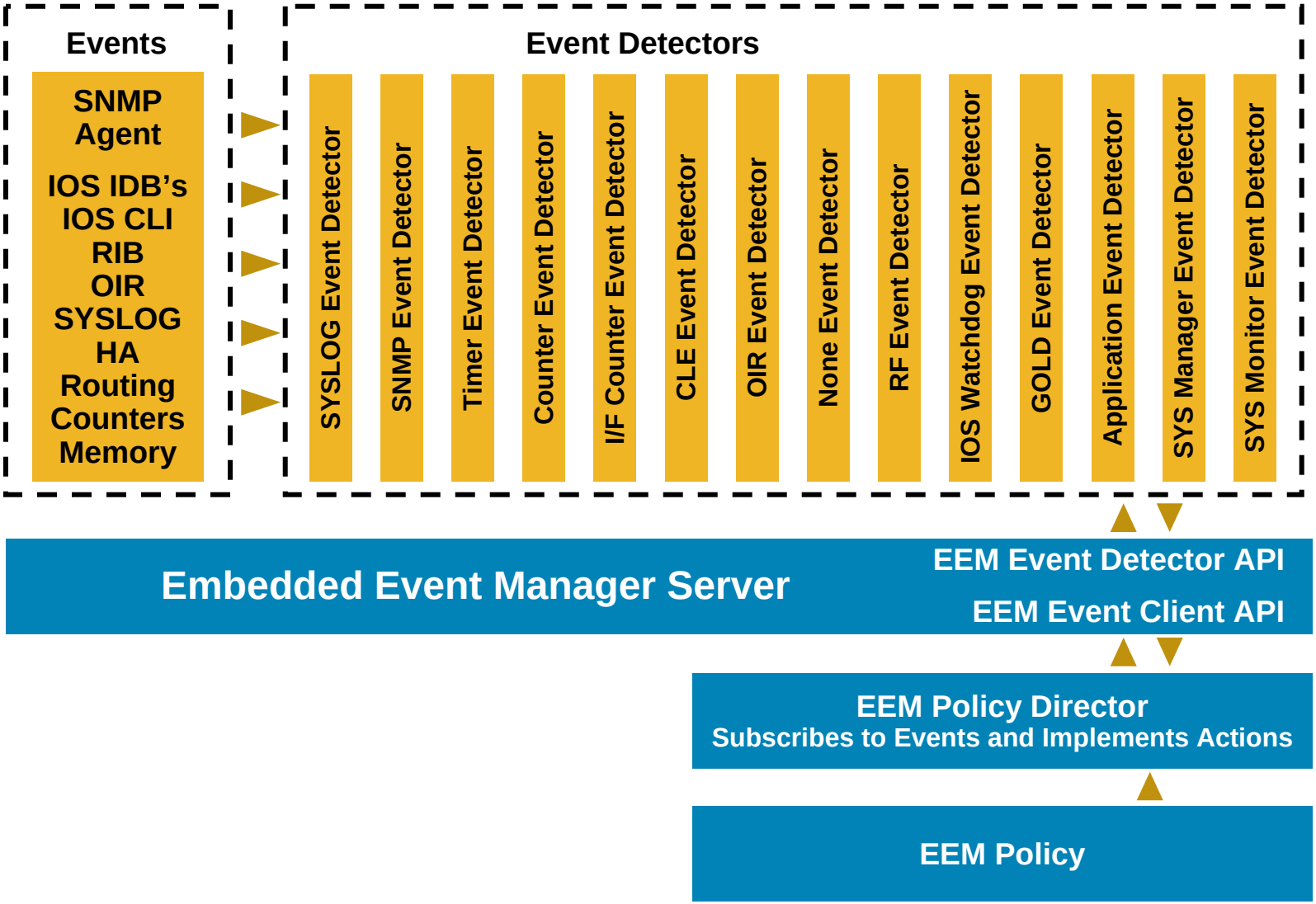
```
Switch#diagnostic start module 6 test all
Module 6: Running test(s) 8 will require resetting the line card after the test has
   completed
Module 6: Running test(s) 1-2,5-9 may disrupt normal system operation
Do you want to continue? [no]: yes
<snip>
*Mar 25 22:43:16: SP: ************************************************************
*Mar 25 22:43:16: SP: * WARNING:
*Mar 25 22:43:16: SP: * ASIC Memory test on module 6 may take up to 2hr 30min.
*Mar 25 22:43:16: SP: * During this time, please DO NOT perform any packet switching.
*Mar 25 22:43:16: SP: ************************************************************
<snip>
Switch#diagnostic start module 5 test all
Module 5: Running test(s) 27-30 will power-down line cards and standby supervisor should
   be power-down manually and supervisor should be reset after the test
Module 5: Running test(s) 26 will shut down the ports of all linecards and supervisor
   should be reset after the test
Module 5: Running test(s) 3,5,8-10,19,22-23,26-31 may disrupt normal system operation
Do you want to continue? [no]: yes
<snip>
```

Embedded Event Manager
Proactive Fault Detection and Notification

- **EEM is a Cisco IOS technology that runs on the control plane. It is a combination of processes designed to monitor key system parameters such as CPU utilization, interface errors, counters, SNMP and SYSLOG events, and act on specific events or thresholds/ counters that are exceeded**



Embedded Event Manager
Architecture

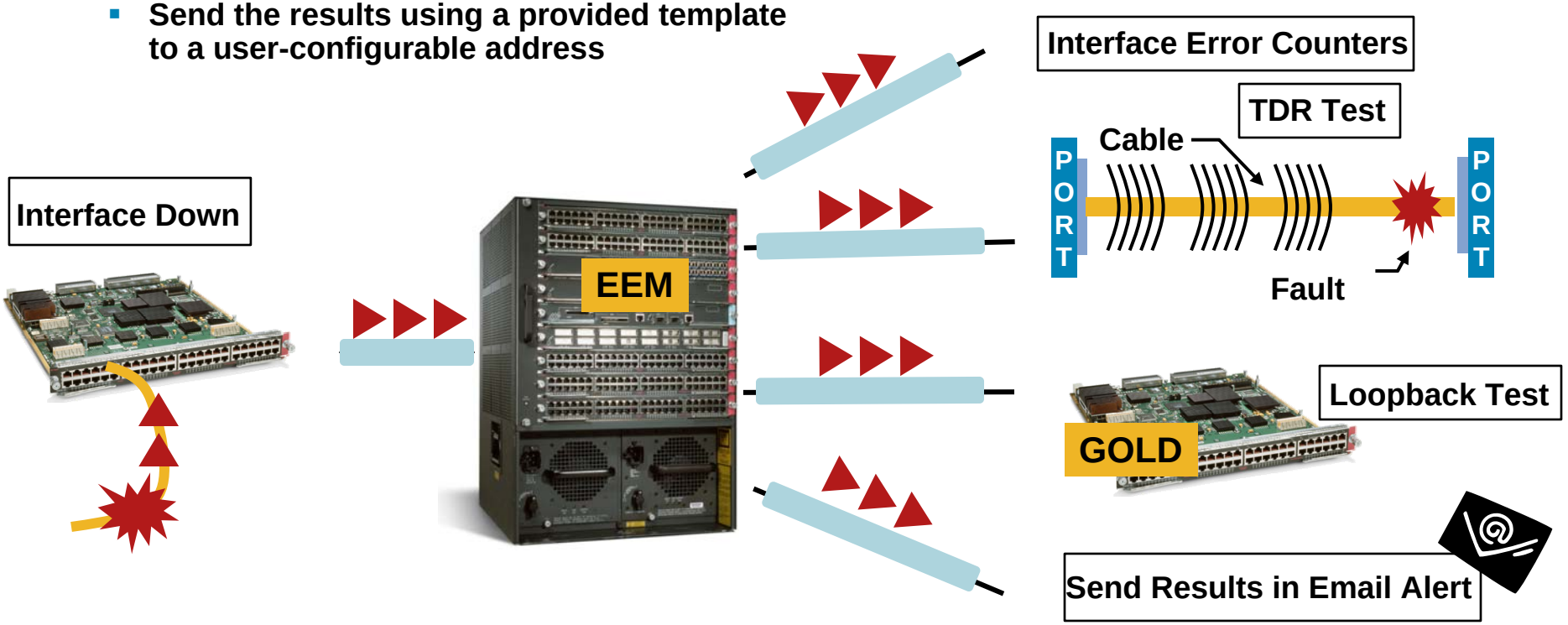


Embedded Event Manager

EEM Application Example

Upon Matching the Provided SYSLOG Message 'LINK-3-UPDOWN', the Switch Performs the Following Actions:

- Display error statistics for the link that has gone down
- Start a Time Domain Reflectometry (TDR) test
- Start a GOLD Loopback test
- Send the results using a provided template to a user-configurable address



Embedded Event Manager
Configuration Example

EEM Applet Example

```
event manager applet TEST
 event syslog pattern "%LINK-3-UPDOWN: Interface GigabitEthernet7/1" maxrun 20
 action 1.0 cli command “en”
 action 2.0 cli command "test cable-diagnostics tdr interface G7/1“
 action 3.0 cli command “diagnostic start module 7 test 2 port 1"
 action 4.0 mail server “x.x.x.x" to “email_id@x.com" from "Switch-1" subject "Urgent! Interface
   went down" body "G7/1 went down"
```

EEM TCL Script Example

```
event manager environment _email_server <IP_address>
event manager environment _email_to email_id@x.com
event manager environment _syslog_pattern .*UPDOWN.*state to down.*
event manager environment _email_from Switch1@mylab.com
event manager environment intchk_template disk1:/interfacecheck.template
event manager directory user policy disk1:/
event manager policy interfacecheck.tcl
```

```
::cisco::eem::event_register_syslog occurs 1 pattern $_syslog_pattern maxrun 90
# EEM policy to monitor for a specified syslog message.
# check if all the env variables we need exist
<snip>
namespace import ::cisco::eem::*
namespace import ::cisco::lib::*
# The Body of the code goes here
<snip>
```

Embedded Event Manager

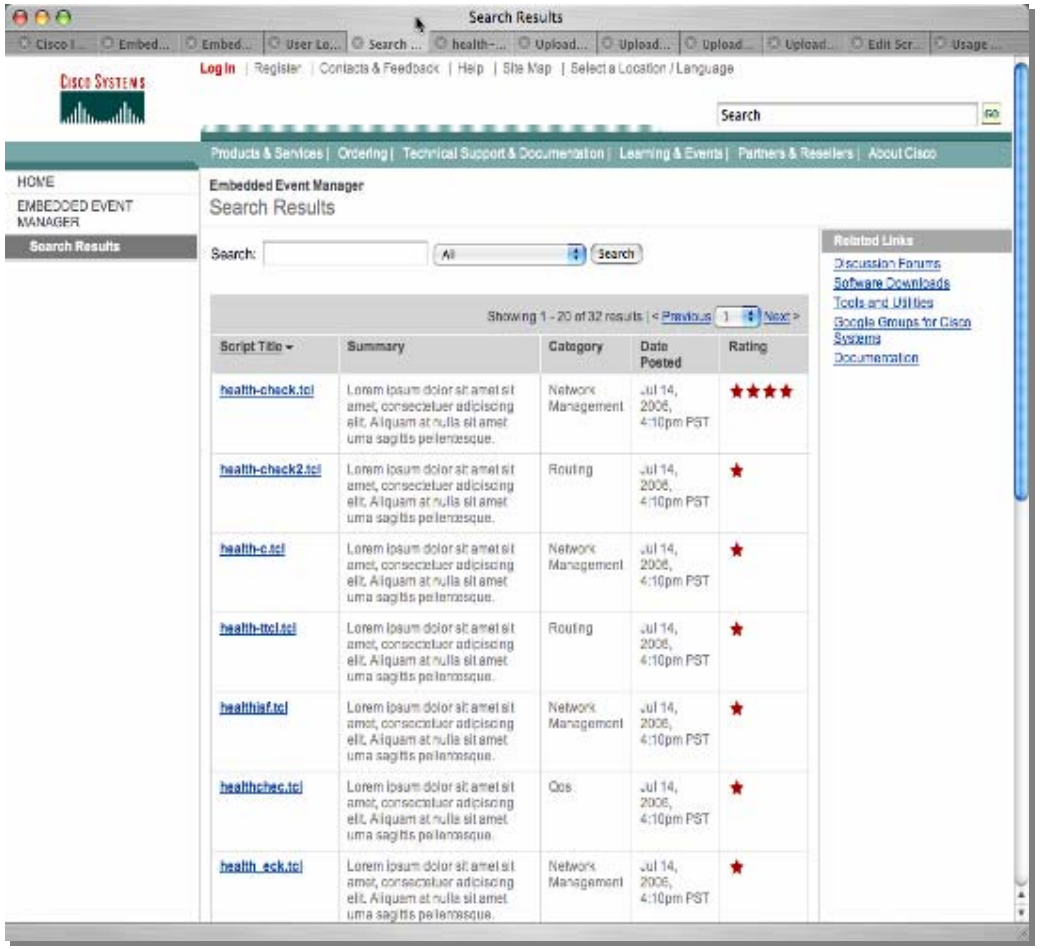
Embedded Event Manager (EEM) Scripting Community

- **Cisco IOS Embedded Event Manager (EEM)**
  - **Automation**
  - **Event driven scripts**
- **Cisco Beyond, an EEM scripting community**
  - **For customers, partners, and Cisco to share EEM scripts and get best-practice examples**

EEM and Cisco Beyond

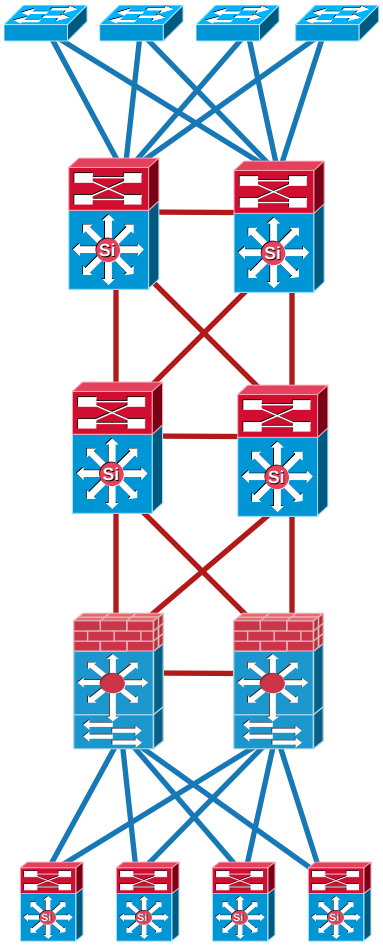
http://cisco.com/go/eem

http://forums.cisco.com/eforum/servlet/EEM?page=main

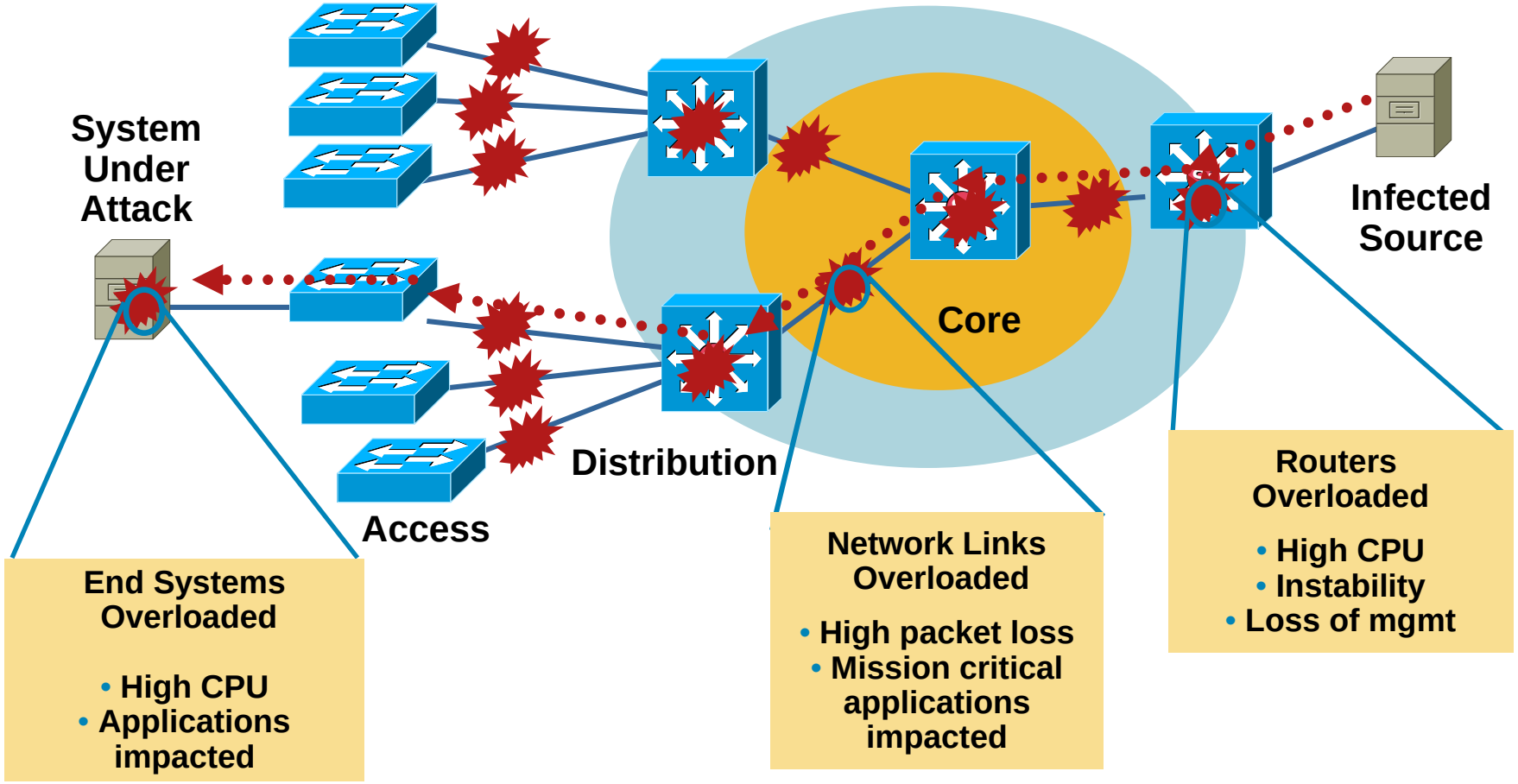


High Availability Campus Design
Agenda

- **Network Level Resiliency**
  - High Availability Design Principles
  - Redundancy in the Distribution Block
  - Redundancy and Routing Design
- **System Level Resiliency**
  - Integrated Hardware and Software Resiliency
    - NSF/SSO
    - ISSU & IOS Modularity
  - System Management Resiliency
    - GOLD & EEM
- **Hardening the Campus Network Design**

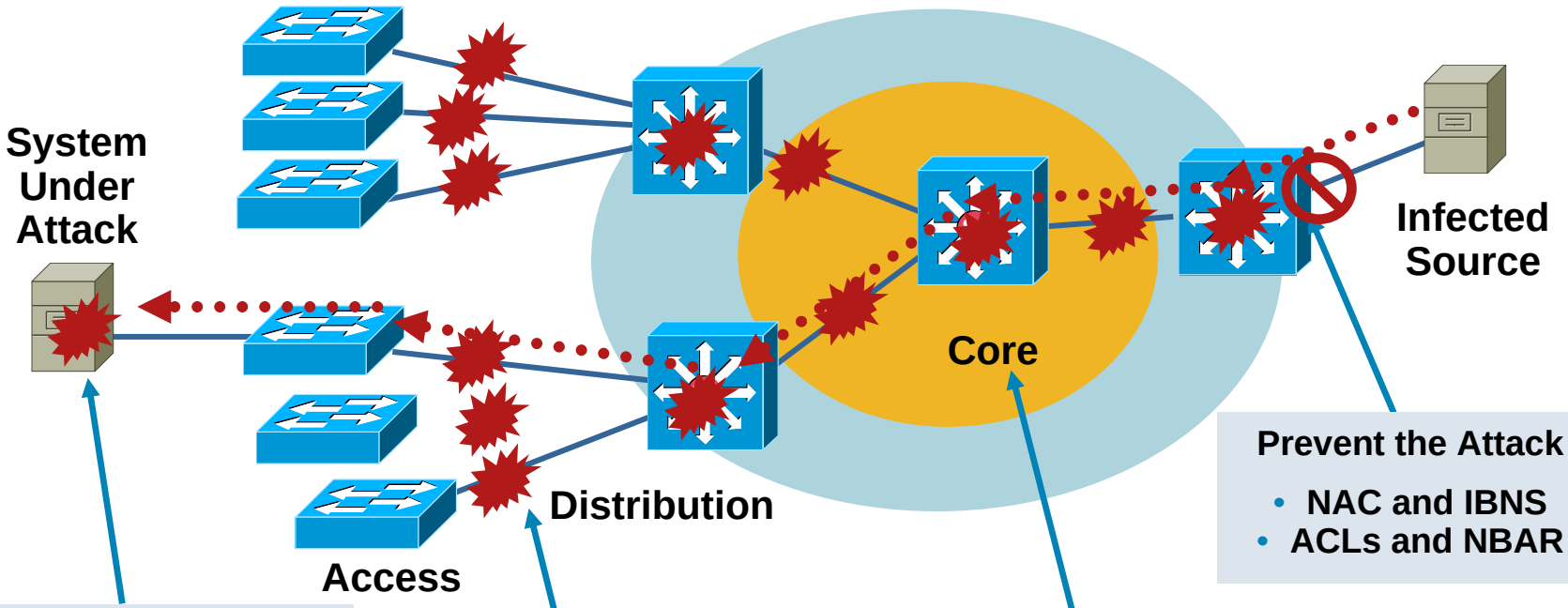


Impact of an Internet Worm
Direct and Collateral Damage



Availability of Networking Resources Impacted by the Propagation of the Worm

Mitigating the Impact
Preventing and Limiting the Pain



Protect the End Systems
- Cisco Security Agent

Protect the Links
- QoS
- Scavenger Class

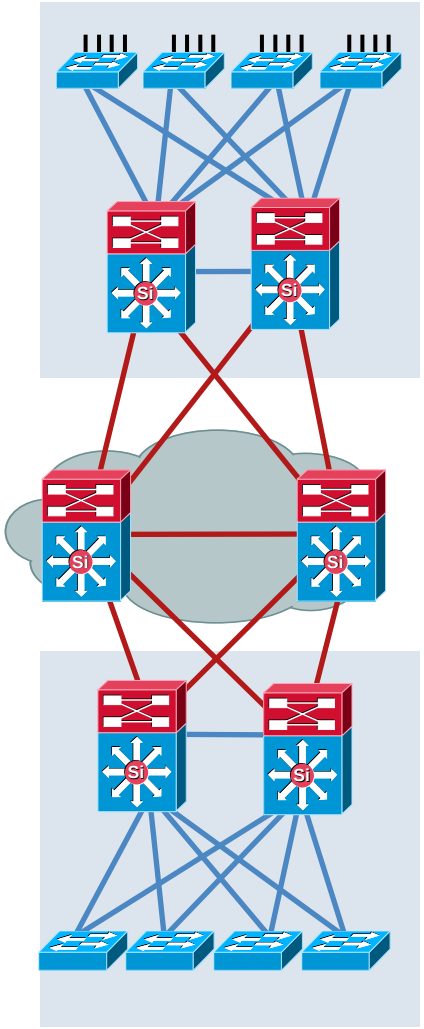
Protect the Switches
- CEF
- Rate Limiters
- CoPP

Prevent the Attack
- NAC and IBNS
- ACLs and NBAR

Allow the Network to Do What You Designed It to Do but Not What You Didn't

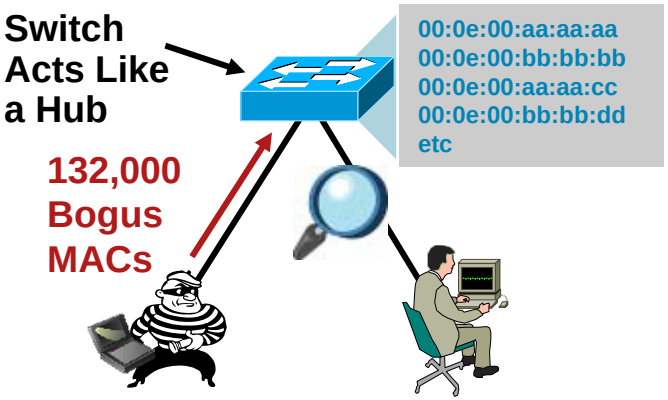
Worms Are Only One Problem
Other Sources of Pain

- **Internet worms are not the only type of network anomaly**
- **Multiple things can either go wrong or be happening that you want to prevent and/or mitigate**
  - **Spanning Tree Loops**
  - **NICs spewing garbage**
  - **Distributed Denial of Service (DDoS)**
  - **TCP Splicing, ICMP Reset attacks**
  - **Man-in-the-Middle (M-in-M) attacks**
  - **…**

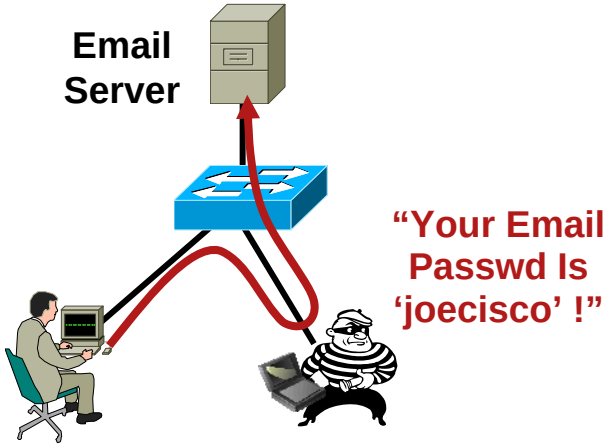


Hardening the Edge
CISF's, BPDU Guard & Port Security

Port Security



IP Source Guard



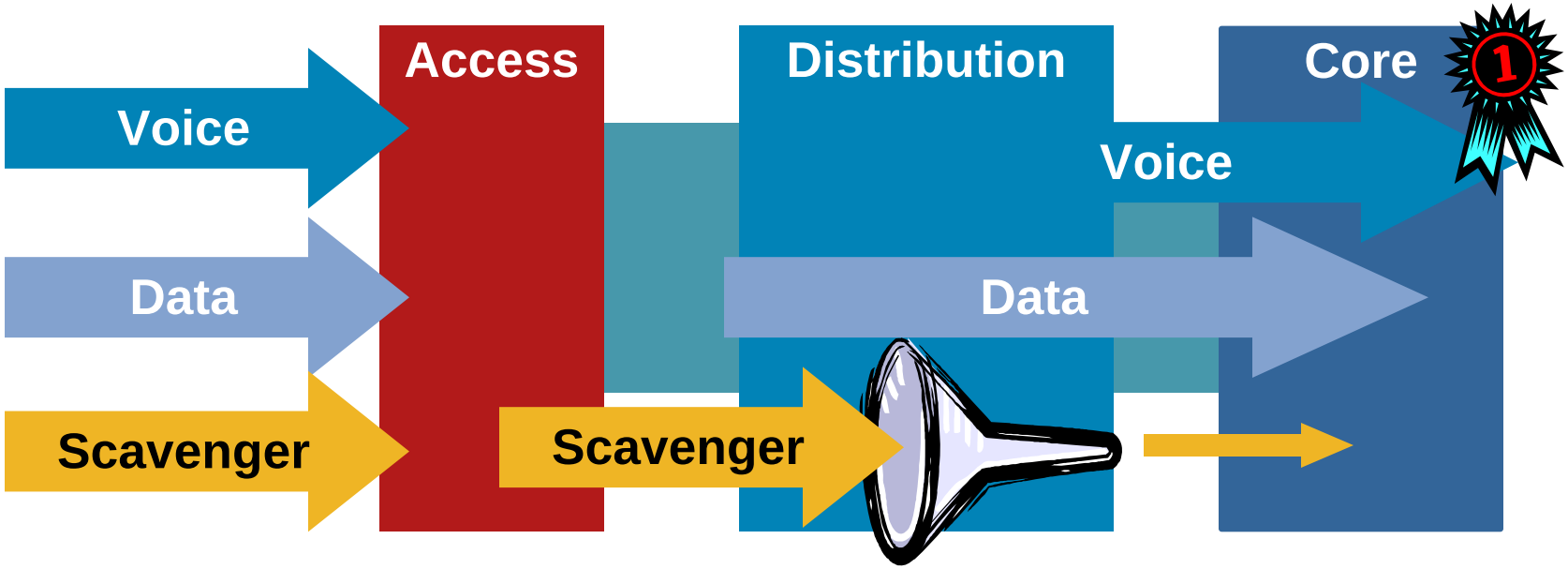
- **Plugging all of the Layer 2 security holes also serves to prevent a whole suite of other attack vectors**
- **Port security and BPDU Guard mitigate against spanning tree loops**
- **In addition to preventing M-i-M attacks IP source guard prevents**
  - **DDoS attacks which utilize a spoofed source address, e.g. TCP SYN Floods, Smurf**
  - **TCP splicing and RST attacks**

BRKCAM-2008 - Understanding and Preventing Layer 2 Attacks

Harden the Network Links

Protect the Relevant Traffic

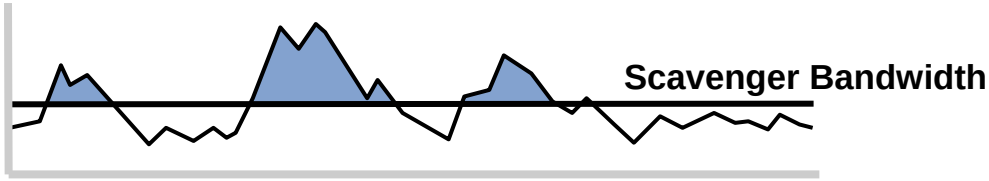
- QoS does more than just protect voice and video
- For “best-effort” traffic an implied “good faith” commitment that there are at least some network resources available is assumed
- Need to identify and potentially punish out of profile traffic (potential worms, DDOS, etc.)
- Scavenger class is an Internet-2 Draft Specification → CS1/CoS1



Harden the Network Links

Scavenger-Class QoS

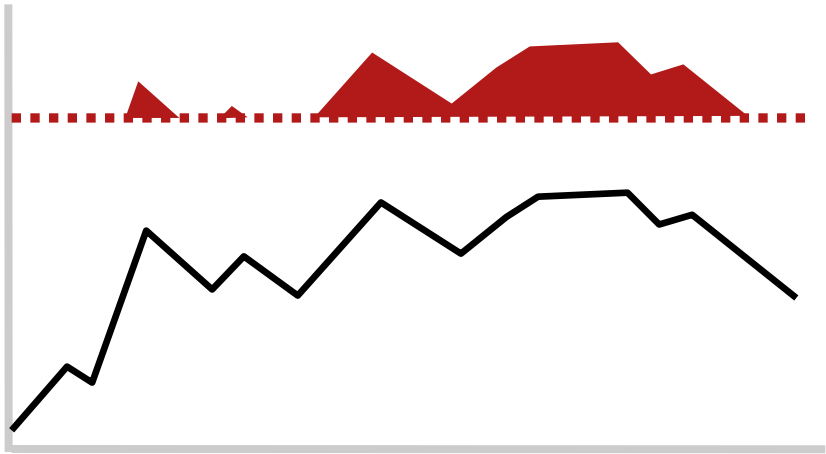
- All end systems generate traffic spikes
- Sustained traffic loads beyond ‘normal’ from each source device are considered suspect and marked as scavenger
- First order anomaly detection—no direct action taken



Network Entry Points

- During ‘abnormal’ worm traffic conditions traffic marked as Scavenger is aggressively dropped—second order detection
- Priority queuing ensuring low latency and jitter for VoIP
- Stations not generating abnormal traffic volumes continue to receive network service

- During ‘normal’ traffic conditions network is operating within designed capacity

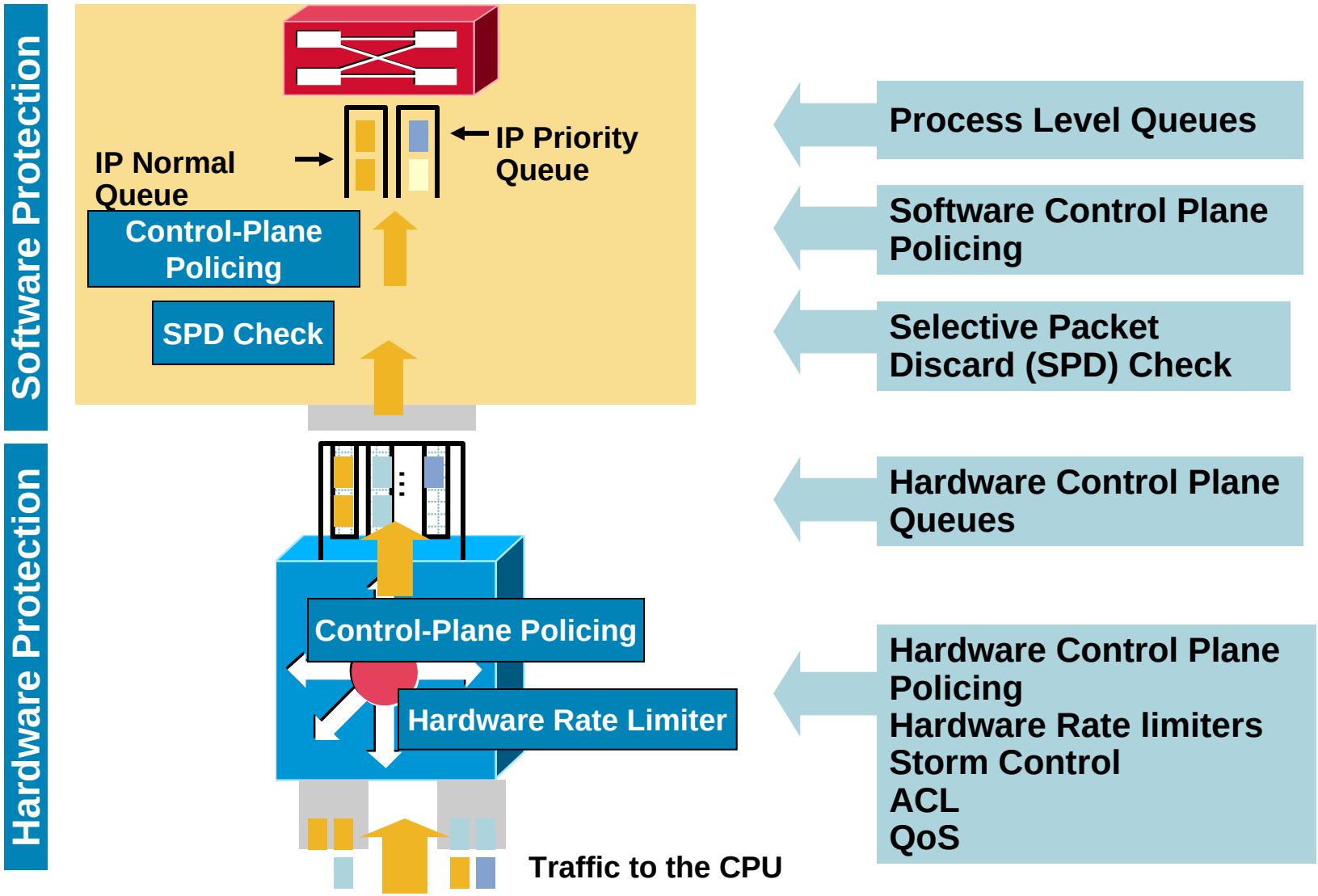


Aggregation Points

BRKCAM-3006 - Advanced Campus QoS Design

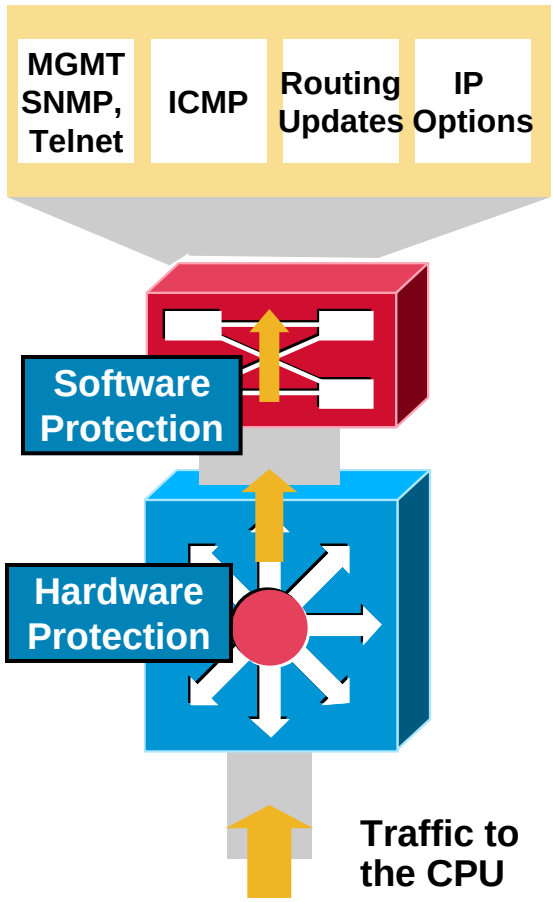
Control Plane Protection

Hardening the Switches



Hardening the Switches
Control Plane Protection

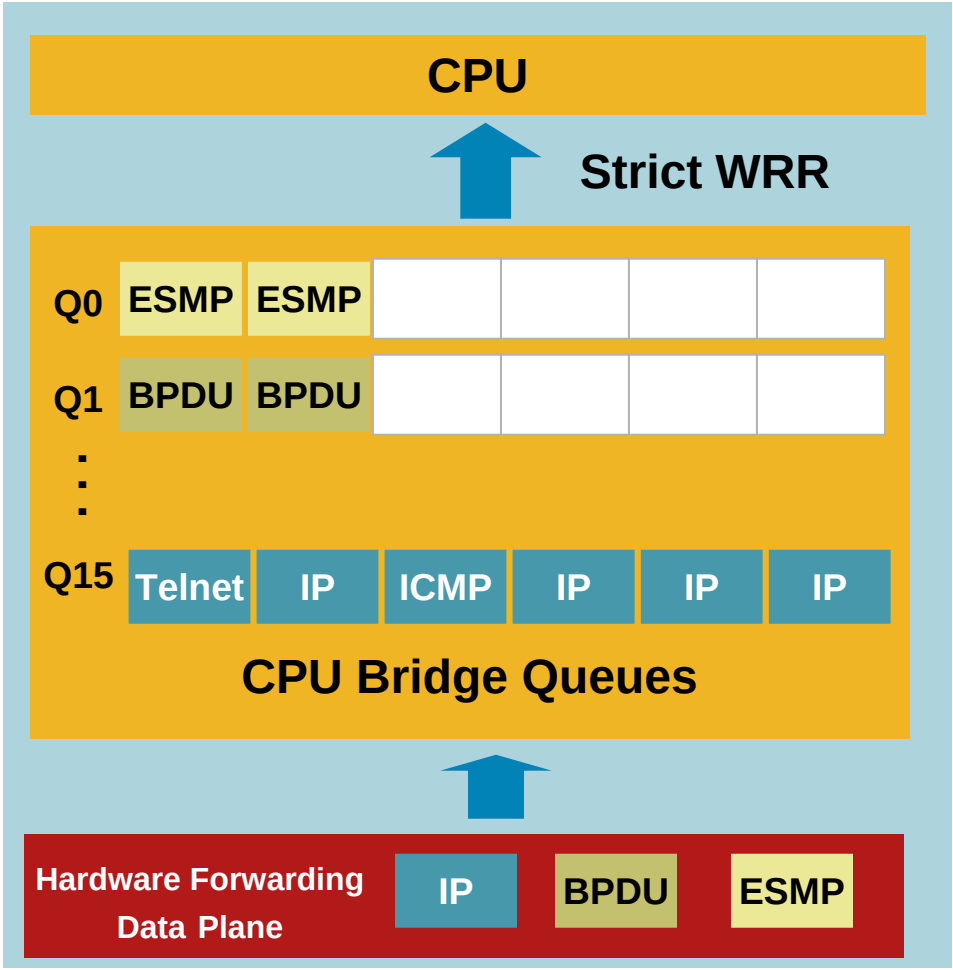
- CEF protects against system overload due to flow flooding
- System CPU still has to be able to process certain traffic
  - BPDUs, CDP, EIGRP, OSPF
  - Telnet, SSH, SNMP
  - ARP, ICMP, IGMP
- System needs to provide throttling on CPU-bound traffic
  - IOS Based SW Rate Limiters
  - Hardware Rate Limiters and CPU queuing
  - Hardware and Software Control Plane Policing (CoPP)



Control Plane Protection

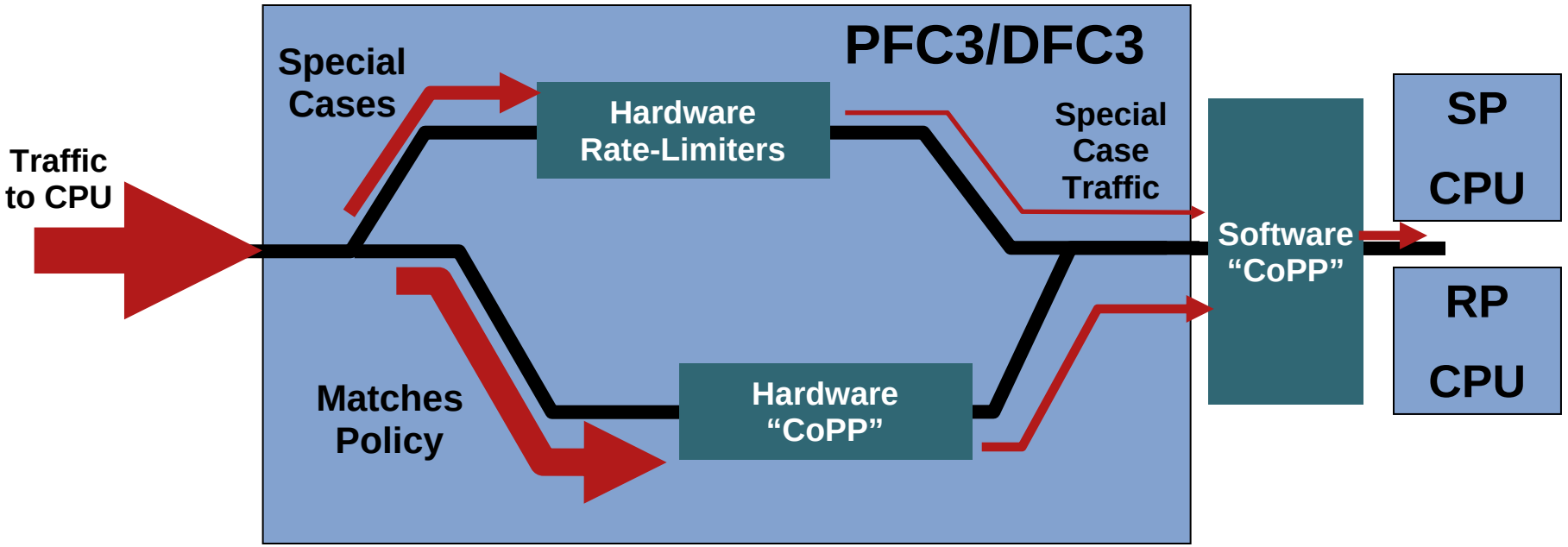
Cisco Catalyst 4500 and 3750 Multiple CPU Queues

- **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**
- **These 16 processor queues are not configurable**
  - STP, OSPF and inter-CPU packets on separate queues
- **The 3750 stack ring reserves bandwidth for priority traffic**
  - Bandwidth reservations on the ring ensure the CPU communication is not effected by data traffic



Control Plane Protection

Catalyst 6500 Multi-level HW and SW Protection



- Traffic punted to the Switch Processor and Route Processor are managed via a series of dedicated Hardware Rate Limiters
- In addition to the Hardware Rate Limiters Hardware QoS policiers are applied to ingress traffic that is not managed by the Rate Limiters

Cisco Catalyst 6500 Control Plane Protection

PFC3 Hardware Rate Limiters Support

Unicast Rate Limiters	
CEF Receive	Traffic Destined to the Router
CEF Glean	ARP Packets
CEF No Route	Packets with Not Route in the FIB
ICMP Redirect	Packets that Require ICMP Redirects
IP Errors	Packet with IP Checksum or Length Errors
ICMP No Route	ICMP Unreachables for Unroutable Packets
ICMP ACL Drop	ICMP Uncreachables for Admin Deny Packets
RPF Failure	Packets that Fail uRPF Check
L3 Security	CBAC, Auth-Proxy, and IPSEC Traffic
ACL Input	NAT, TCP Int, Reflexive ACLs, Log on ACLs
ACL Output	NAT, TCP Int, Reflexive ACLs, Log on ACLs
VACL Logging	CLI Notification of VACL Denied Packets
IP Options	Unicast Traffic with IP Options Set
Capture	Used with Optimized ACL Logging

Multicast Rate Limiters	
Multicast FIB-Miss	Packets with No mroute in the FIB
IGMP	IGMP Packets (Actually Layer 2)
Partial Shortcut	Partial Shortcut Entries
Directly Connected	Local Multicast on Connected Interface
IP Options	Multicast Traffic with IP Options Set
V6 Directly Connect	Packets with No Mroute in the FIB
V6*, G M Bridge	IGMP Packets
V6*, G Bridge	Partial Shortcut Entries
V6 S, G Bridge	Partial Shortcut Entries
V6 Route Control	Partial Shortcut Entries
V6 Default Route	Multicast Traffic with IP Options Set
V6 Second Drop	Mulicast Traffic with IP Options Set

Shared Across the Ten Hardware Revocation Lists

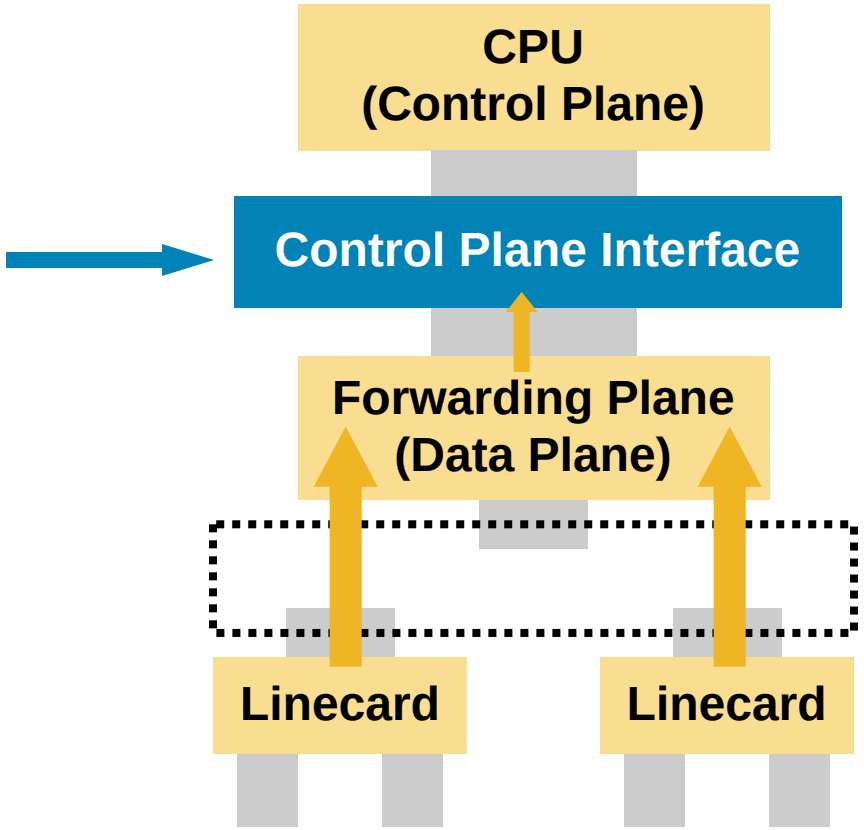
Layer 2 Rate Limiters	
L2PT	L2PT Encapsulation/Decapsulation
PDU	Layer 2 PDUs

General Rate Limiters	
MTU Failure	Packets Requiring Fragmentation
TTL Failure	Packets with TTL<=1

Control Plane Protection

Control Plane Policing (CoPP)

- Control Plane Policing applies Catalyst Hardware QoS policies to traffic punted to the CPU
- New Logical Control Plane Interface
  - Provides ability to Rate Limit Total traffic volume destined to Control Plane
- Hardware based CoPP is supported on the Cisco Catalyst 6500 and 4500 in Hardware
- Catalyst 6500 also supports a second tier of Software CoPP

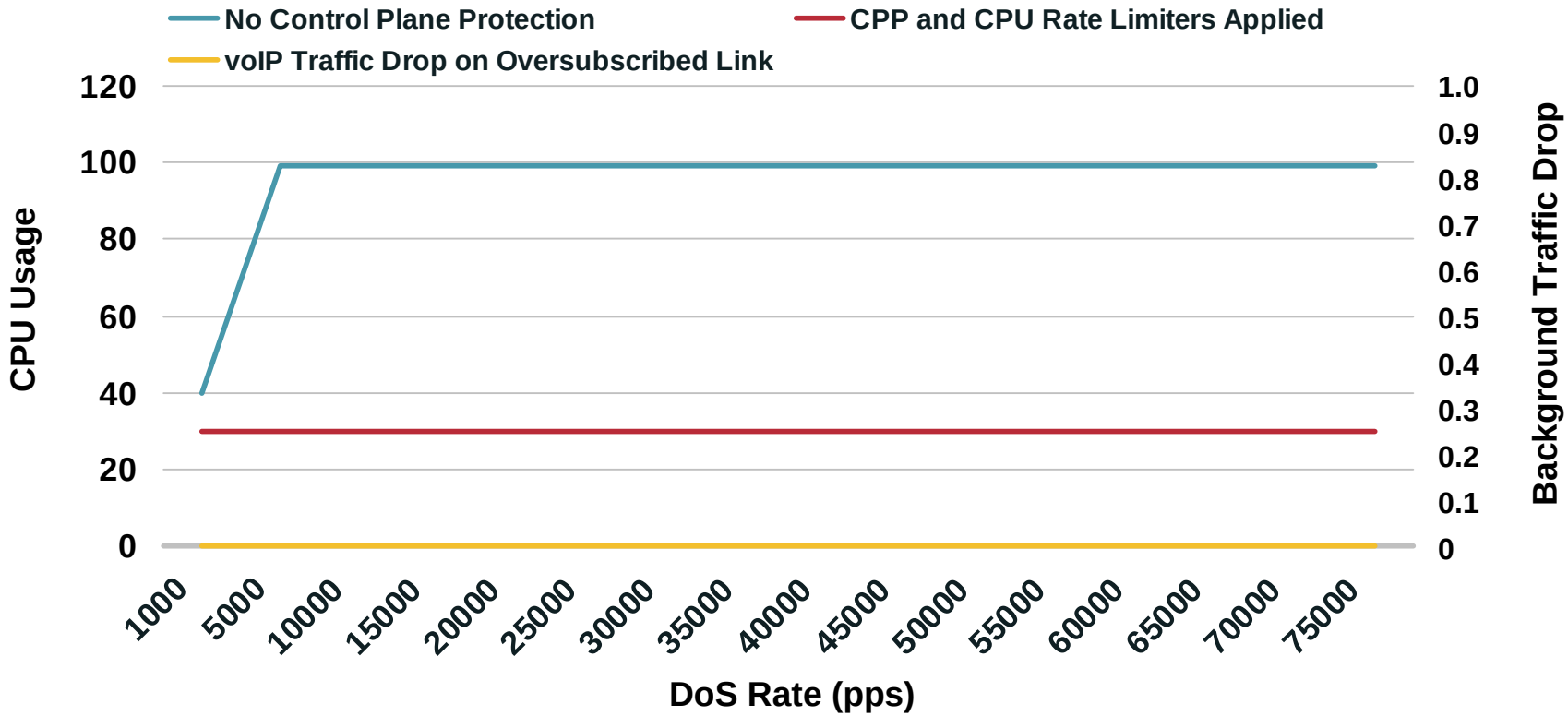


BRKCAM -3006 - Advanced Campus QoS Design

Mitigating the Impact: CoPP

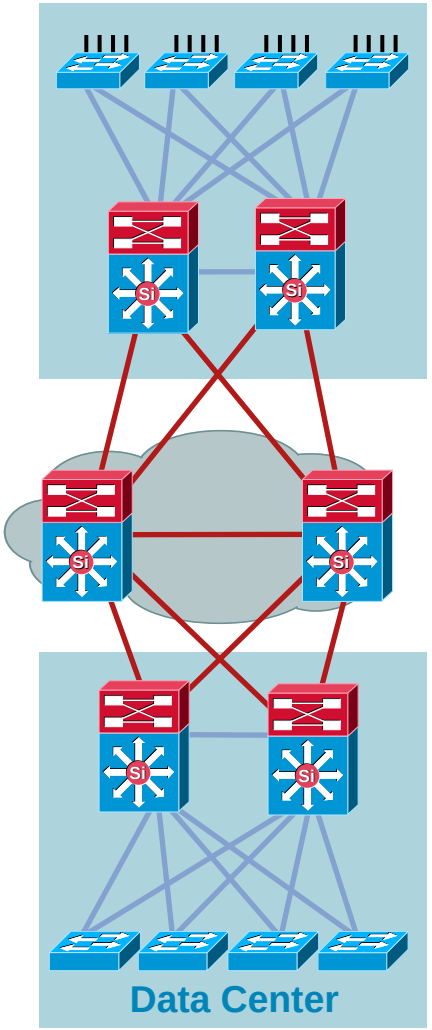
CoPP and Rate Limiters Compliment CEF

- Multiple concurrent attacks (multicast ttl=1, multicast partial shortcuts, unicast IP options, unicast fragments to receive adjacency, unicast TCP SYN flood to receive adjacency)
- CPU kept within acceptable bounds with no loss of mission critical traffic



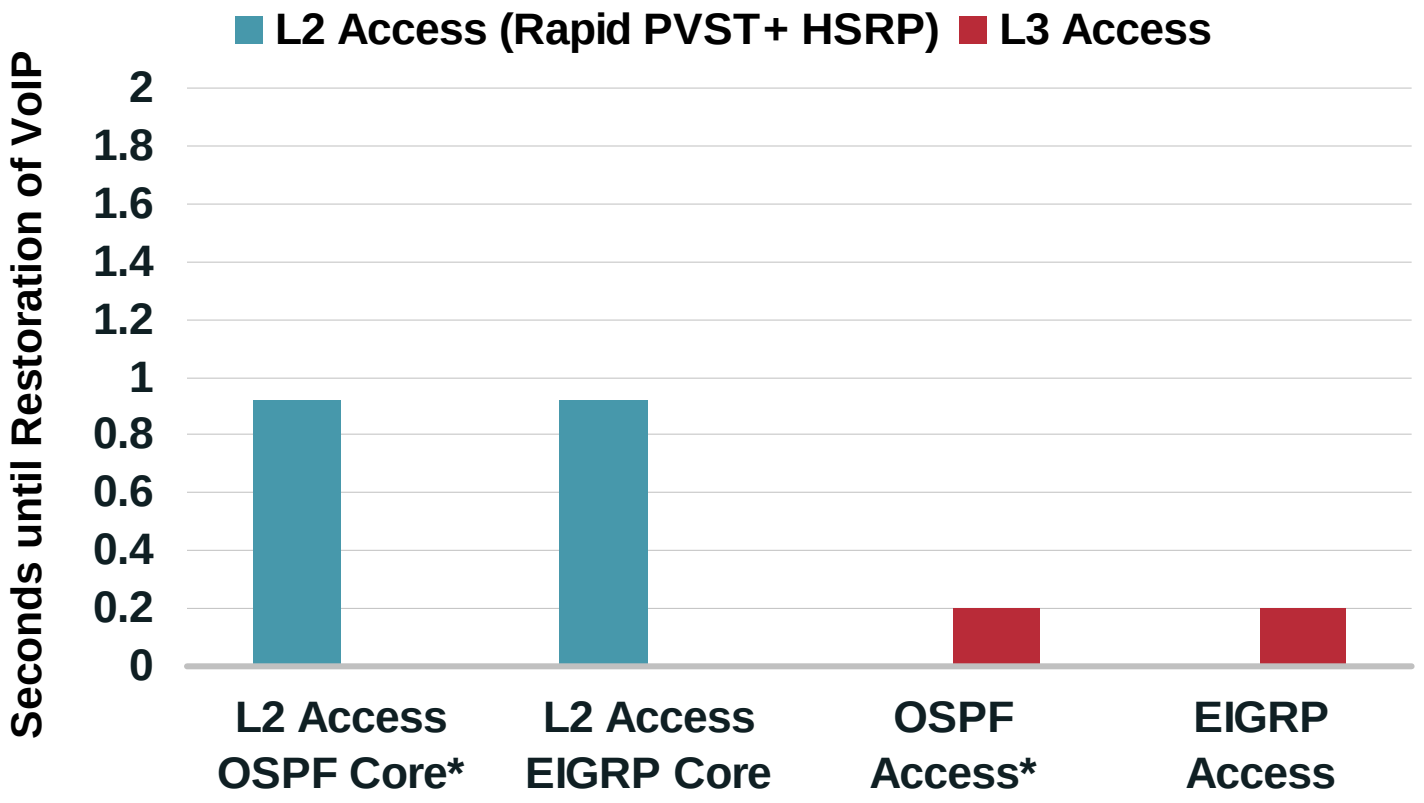
Resilient Network Design
Stick to Your Principles

- **Develop an architecture and stick to it**
  - **Ease operational support**
  - **Consistent deployment**
- **Balance OPeX and CapEX**
  - **Remember you will have to live with this for a long time**
  - **Requirements will change**
- **Plan for evolution**
  - **The one thing that doesn't change is that there will be change**
- **Understand change**
  - **How your environments are changing**
  - **How the network equipment is evolving to meet that change**



Resilient Campus Design

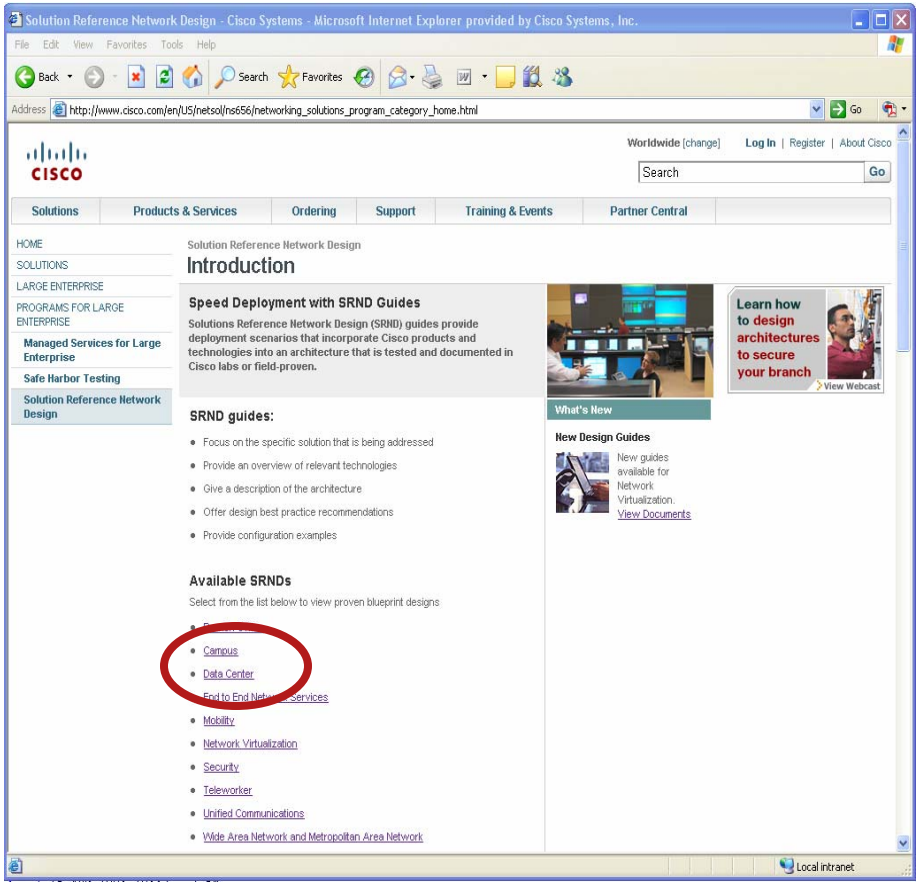
This Is What You Can Expect



Worst Case Convergence for Any Campus Failure Event

***OSPF Results Require Sub-Second Timers**

Campus Design Guidance
Where to go for more information



http://www.cisco.com/go/srnd

Meet the Experts
Campus and Wireless Evolution

- Mark Montanez
  Corporate Dev Consulting Engineer

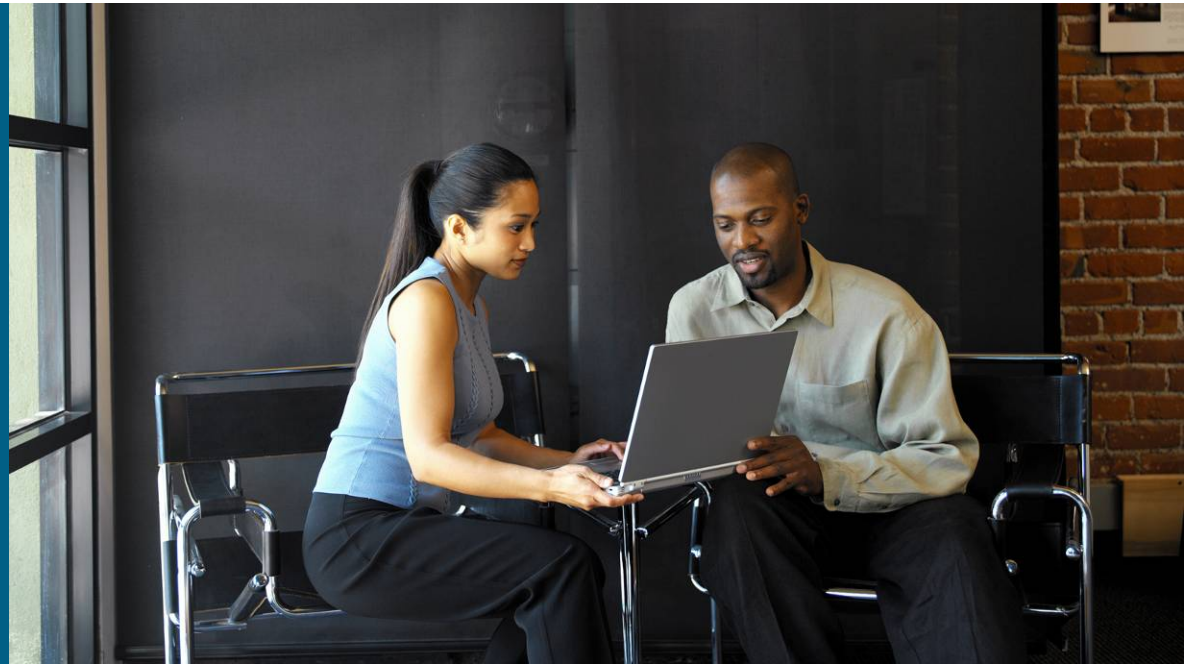
- Tim Szigeti
  Technical Leader

- Sujit Ghosh
  Technical Mktg Eng

- Victor Moreno
  Technical Leader

- Mike Herbert
  Technical Leader

Questions



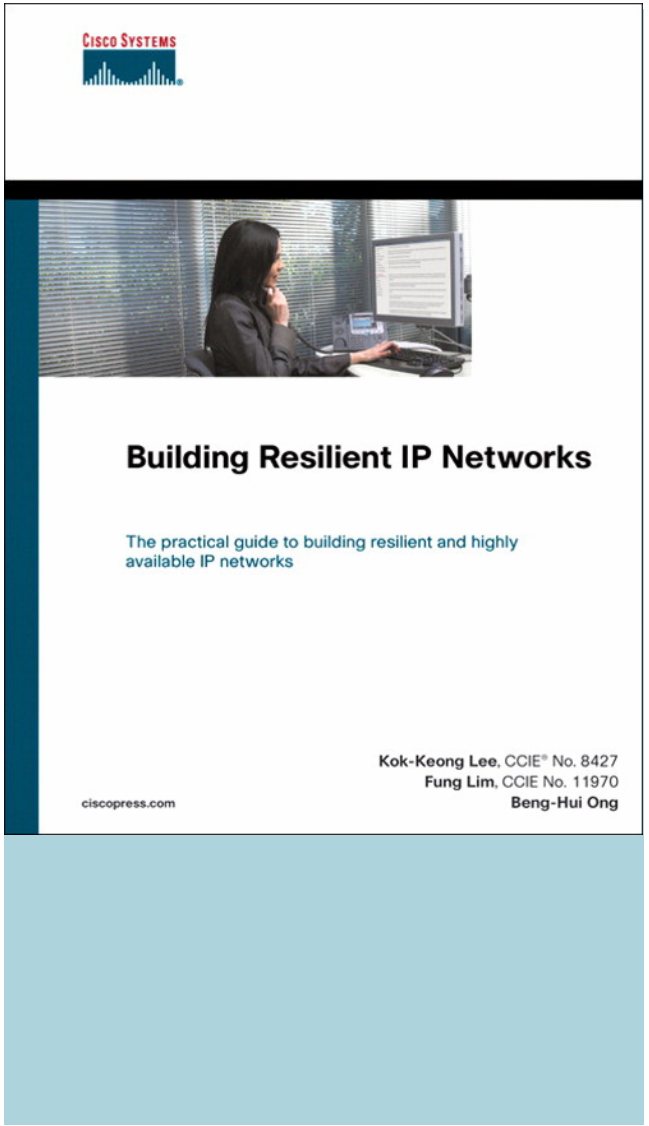
Recommended Reading

- Continue your Cisco Networkers learning experience with further reading from Cisco Press

  End-to-End QoS Network Design: Quality of Service in LANs, WANs, and VPNs, ISBN:1-58705-176-1

  Building Resilient IP Networks, ISBN: 1-58705-215-6

  Top-Down Network Design, Second Ed., ISBN: 1-58705-152-4



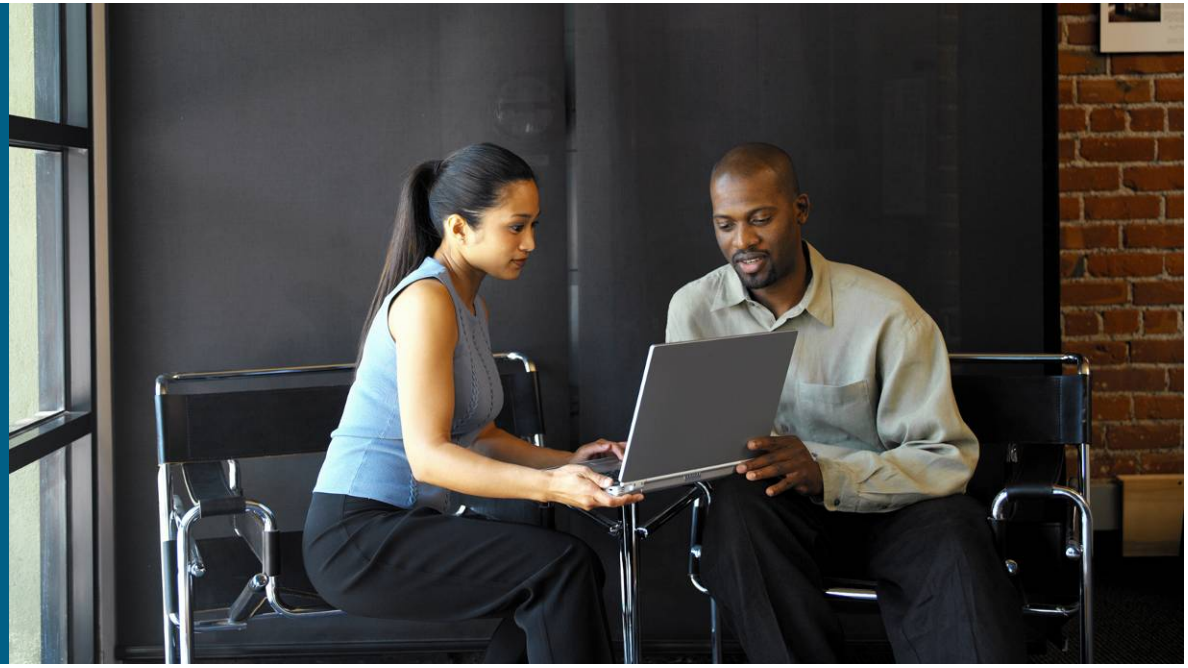
Available Onsite at the Cisco Company Store

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
- Drawings will be held in the World of Solutions

  ???

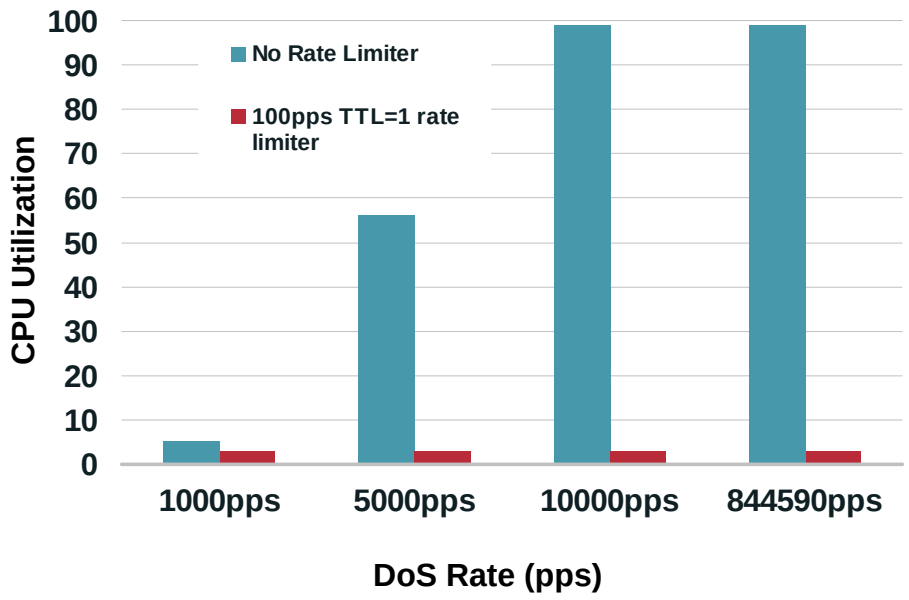
Appendix A: Control Plane Policing Configuration



Cisco Catalyst 6500 Control Plane Protection
Hardware Rate Limiters Deployment

- Use all eight Layer-3 rate limiters
- Consider most likely attack vectors for the network environment
  - Enable the rate limiters, which are most likely to be used
- Do not waste a rate-limiter on VACL logging, if it is not happening
  - No mls rate-limit unicast acl vacl-log
- Disable redirects and save a rate limiter
  - Hardware forwarding platform reduces need for redirect efficiency
- MTU limiter is not required if all interfaces have same MTU
- Configure PDU Layer 2 rate limiter with care
  - Remember that revocation lists do not discriminate between “good” frames and “bad” frames

```
mls rate-limit multicast ipv4 fib-miss 10000 10
mls rate-limit multicast ipv4 igmp 5000 10
mls rate-limit multicast ipv4 ip-options 10 1
mls rate-limit multicast ipv4 partial 10000 10
mls rate-limit unicast cef glean 1000 10
mls rate-limit unicast acl input 500 10
mls rate-limit unicast acl output 500 10
mls rate-limit unicast ip options 10 1
mls rate-limit unicast ip rpf-failure 500 10
mls rate-limit unicast ip icmp unreachable no-route 500 10
mls rate-limit unicast ip icmp unreachable acl-drop 500 10
mls rate-limit unicast ip errors 500 10
mls rate-limit all ttl-failure 500 10
```



Cisco Catalyst 6500 Control Plane Protection

Hardware Rate Limiters Deployment

```
Switch#show mls rate
 Sharing Codes: S - static, D - dynamic
 Codes dynamic sharing: H - owner (head) of the group, g - guest of the group

   Rate Limiter Type       Status     Packets/s   Burst  Sharing
 ---------------------   ----------   ---------   -----  -------
         MCAST NON RPF   Off                -       -     -
        MCAST DFLT ADJ   On             10000      10  Not sharing
      MCAST DIRECT CON   Off                -       -     -
        ACL BRIDGED IN   On                 500      10  Group:1 S
       ACL BRIDGED OUT   On                 500      10  Group:1 S
           IP FEATURES   Off                -       -     -
          ACL VACL LOG   Off                -       -     -
           CEF RECEIVE   Off                -       -     -
             CEF GLEAN   On                1000      10  Not sharing
      MCAST PARTIAL SC   On               10000      10  Not sharing
        IP RPF FAILURE   On                 500      10  Group:0 S
           TTL FAILURE   On                 500      10  Not sharing
 ICMP UNREAC. NO-ROUTE   On                 500      10  Group:0 S
 ICMP UNREAC. ACL-DROP   On                 500      10  Group:0 S
          ICMP REDIRECT   Off                -       -     -
           MTU FAILURE   Off                -       -     -
       MCAST IP OPTION   On                  10       1  Group:3 S
       UCAST IP OPTION   On                  10       1  Group:2 S
           LAYER_2 PDU   Off                -       -     -
            LAYER_2 PT   Off                -       -     -
             IP ERRORS   On                 500      10  Group:0 S
           CAPTURE PKT   Off                -       -     -
            MCAST IGMP   On                1000      10  Not sharing
 MCAST IPv6 DIRECT CON   Off                -       -     -
 MCAST IPv6 ROUTE CNTL   Off                -       -     -
 MCAST IPv6 *G M BRIDG   Off                -       -     -
  MCAST IPv6 SG BRIDGE   Off                -       -     -
  MCAST IPv6 DFLT DROP   Off                -       -     -
 MCAST IPv6 SECOND. DR   Off                -       -     -
  MCAST IPv6 *G BRIDGE   Off                -       -     -
        MCAST IPv6 MLD   Off                -       -     -
  IP ADMIS. ON L2 PORT   Off                -       -     -
```

Cisco Catalyst 6500 Control Plane Protection
Hardware Rate Limiters Deployment

- **Configure PDU Layer 2 rate limiter with care**
  - **Remember that revocation lists do not discriminate between "good" frames and "bad" frames**
- **Layer 2 HWRL (L2PDU, L2PT, IGMP) are not supported in truncated mode**
  - **Truncated mode occurs with a mix of fabric and classic cards or an all-classic chassis with dual Sups**
  - **If the system is running in truncated mode, the following error message will be seen when configuring Layer 2 HWRLs:**
  - **04:23:12: %MLS_RATE-4-NOT_SUPPORTED: This functionality is not configurable.**
- **2 x [Layer 2] and 8 x [General/Unicast/Multicast] are configurable—choose carefully. There is no performance penalty for using all ten HWRLs.**
- **When a packet matches both HW CoPP and HWRL, the packet undergoes HWRL policy and skips HW CoPP—(see slides in CoPP deployment guidelines)**
  - **Be extra careful with the CEF receive and the ACL rate limiters since they will overlap with CoPP.**
- **Configure the CEF receive rate limiter with care**
  - **Given that the CEF receive rate limiter matches all traffic destined to the Route Process ("good" frames and "bad" frames) and takes precedence over CoPP, it is best to only use CoPP instead**
- **Configure the CEF glean rate limiter with care**
  - **If there is an output ACL configured on the ingress VLAN, it will be applied before the rate limiter**
- **TTL=1 rate limiter is not affecting control plane traffic (using 224.0.0/24)**

Cisco Catalyst 6500 Control Plane Protection
CoPP Deployment—Step 1

- **Step 1: Identify traffic of interest and classify it into multiple traffic classes:**
  - **BGP**
  - **IGP (EIGRP, OSPF, ISIS)**
  - **Management (telnet, TACACS, ssh, SNMP, NTP)**
  - **Reporting (SAA)**
  - **Monitoring (ICMP)**
  - **Critical applications (HSRP, DHCP)**
  - **Undesirable**
  - **Default**

```
ip access-list extended coppacl-bgp
permit tcp host 192.168.1.1 host 10.1.1.1 eq bgp
permit tcp host 192.168.1.1 eq bgp host 10.1.1.1
!
ip access-list extended coppacl-igp
permit ospf any host 224.0.0.5
permit ospf any host 224.0.0.6
permit ospf any any
!
ip access-list extended coppacl-management
permit tcp host 10.2.1.1 host 10.1.1.1 established
permit tcp 10.2.1.0 0.0.0.255 host 10.1.1.1 eq 22
permit tcp 10.86.183.0 0.0.0.255 any eq telnet
permit udp host 10.2.2.2 host 10.1.1.1 eq snmp
permit udp host 10.2.2.3 host 10.1.1.1 eq ntp
!
ip access-list extended coppacl-reporting
permit icmp host 10.2.2.4 host 10.1.1.1 echo
!
ip access-list extended coppacl-monitoring
permit icmp any any ttl-exceeded
permit icmp any any port-unreachable
permit icmp any any echo-reply
permit icmp any any echo
!
ip access-list extended coppacl-critical-app
permit ip any host 224.0.0.1
permit udp host 0.0.0.0 host 255.255.255.255 eq bootps
permit udp host 10.2.2.8 eq bootps any eq bootps
!
ip access-list extended coppacl-undesirable
permit udp any any eq 1434
```

Cisco Catalyst 6500 Control Plane Protection
CoPP Deployment—Step 2

- **Step 2: Associate the identified traffic with a class, and permit the traffic in each class**
  - **Must enable QoS globally, else CoPP will not be applied in hardware**
  - **Always apply a policing action for each class since the switch will ignore a class that does not have a corresponding policing action (for example "police 31500000 conform-action transmit exceed-action drop"). Alternatively, both conform-action and exceed-action could be set to transmit, but doing so will allocate a default policer as opposed to a dedicated policer with its own hardware counters.**
  - **HW CoPP classes are limited to one match per class-map**
  - **No properly functional class-default support prior to 12.2(18)SXE—non-IP traffic will likely end up being caught in the transit/catch-all class in software (not hardware).**

```
mls qos
```

```
class-map match-all copp-bgp
 match access-group name coppacl-bgp
class-map match-all copp-igp
 match access-group name coppacl-igp
class-map match-all copp-management
 match access-group name coppacl-management
class-map match-all copp-reporting
 match access-group name coppacl-reporting
class-map match-all copp-monitoring
 match access-group name coppacl-monitoring
class-map match-all copp-critical-app
 match access-group name coppacl-critical-app
class-map match-all copp-undesirable
 match access-group name coppacl-undesirable
```

```
policy-map copp-policy
 class copp-bgp
  police 30000000 conform-action transmit exceed-action drop
 class copp-igp
  police 30000000 conform-action transmit exceed-action drop
 class copp-management
  police 30000000 conform-action transmit exceed-action drop
 class copp-reporting
  police 30000000 conform-action transmit exceed-action drop
 class copp-monitoring
  police 30000000 conform-action transmit exceed-action drop
 class copp-critical-app
  police 30000000 conform-action transmit exceed-action drop
 class copp-undesirable
  police 30000000 conform-action transmit exceed-action drop
 class class-default
  police 30000000 conform-action transmit exceed-action drop
```

```
control-plane
 service-policy input copp-policy
```

Cisco Catalyst 6500 Control Plane Protection
CoPP Deployment—Step 3

- **Step 3: Adjust classification, and apply liberal CoPP policies for each class of traffic**
  - **show policy-map control-plane displays dynamic information for monitoring control plane policy. Statistics include rate information and number of packets/ bytes confirmed or exceeding each traffic class**
  - **CoPP rates on Sup720 are bps—pps is not possible. However, HWRL rates are in pps**

```
Switch# show policy-map control-plane
 Control Plane Interface
 Service-policy input: copp-policy
<snip>
  Hardware Counters:
  class-map: copp-monitoring (match-all)
   Match: access-group name coppacl-monitoring
   police :
    30000000 bps 937000 limit 937000 extended limit
   Earl in slot 5 :
    0 bytes
    5 minute offered rate 0 bps
    aggregate-forwarded 0 bytes action: transmit
    exceeded 0 bytes action: drop
    aggregate-forward 0 bps exceed 0 bps
   Earl in slot 7 :
    112512 bytes
    5 minute offered rate 3056 bps
    aggregate-forwarded 112512 bytes action: transmit
    exceeded 0 bytes action: drop
    aggregate-forward 90008 bps exceed 0 bps

 Software Counters:
  Class-map: copp-monitoring (match-all)
   1036 packets, 128464 bytes
   5 minute offered rate 4000 bps, drop rate 0 bps
   Match: access-group name coppacl-monitoring
   police:
     cir 30000000 bps, bc 937500 bytes
    conformed 1036 packets, 128464 bytes; action: transmit
    exceeded 0 packets, 0 bytes; action: drop
    conformed 4000 bps, exceed 0 bps
<snip>
```

Cisco Catalyst 6500 Control Plane Protection
CoPP Deployment—Step 3 (Cont.)

- **Step 3: Adjust Classification, and Apply liberal CoPP policies for each class of traffic**
  - **show ip access-lists provides packet count statistics per ACE. Absence of any hits on an entry indicate lack of traffic matching the ACE criteria—the rule might be rewritten**
  - **Hardware ACL hit counters are available in PFC3B/BXL for security ACL TCAM only (not QoS ACL TCAM)**

```
Switch#sh access-list
Extended IP access list coppacl-bgp
  10 permit tcp host 192.168.1.1 host 10.1.1.1 eq bgp
  20 permit tcp host 192.168.1.1 eq bgp host 10.1.1.1
Extended IP access list coppacl-critical-app
  10 permit ip any host 224.0.0.1
  20 permit udp host 0.0.0.0 host 255.255.255.255 eq bootps
  30 permit udp host 10.2.2.8 eq bootps any eq bootps
Extended IP access list coppacl-igp
  10 permit ospf any host 224.0.0.5 (64062 matches)
  20 permit ospf any host 224.0.0.6
  30 permit ospf any any (17239 matches)
Extended IP access list coppacl-management
  10 permit tcp host 10.2.1.1 host 10.1.1.1 established
  20 permit tcp 10.2.1.0 0.0.0.255 host 10.1.1.1 eq 22
  30 permit tcp 10.86.183.0 0.0.0.255 any eq telnet
  40 permit udp host 10.2.2.2 host 10.1.1.1 eq snmp
  50 permit udp host 10.2.2.3 host 10.1.1.1 eq ntp
Extended IP access list coppacl-monitoring
  10 permit icmp any any ttl-exceeded (120 matches)
  20 permit icmp any any port-unreachable
  30 permit icmp any any echo-reply (17273 matches)
  40 permit icmp any any echo (5 matches)
Extended IP access list coppacl-reporting
  10 permit icmp host 10.2.2.4 host 10.1.1.1 echo
Extended IP access list coppacl-undesirable
  10 permit udp any any eq 1434
```

Cisco Catalyst 6500 Control Plane Protection
CoPP Deployment—Step 4

- **Step 4: Fine tune the control plane policy**

  Narrow the ACL permit statements to only allow known authorized source addresses and depending on class defined, apply appropriate policy

  Routing protocol traffic—no rate limit or very conservative rate limit

  Management traffic—conservative rate limit

  Reporting traffic—conservative rate limit

  Monitoring traffic—conservative rate limit

  Critical traffic—conservative rate limit

  Default traffic—low rate limit

  Undesirable traffic—drop

```
policy-map copp-policy
 class coppclass-bgp
  police 15000000 conform-action transmit exceed-action drop
 class coppclass-igp
  police 15000000 conform-action transmit exceed-action drop
 class coppclass-management
  police 2560000 conform-action transmit exceed-action drop
 class coppclass-reporting
  police 1000000 conform-action transmit exceed-action drop
 class coppclass-monitoring
  police 1000000 conform-action transmit exceed-action drop
 class coppclass-critical-app
  police 7500000 conform-action transmit exceed-action drop
 class coppclass-undesirable
  police 32000 conform-action transmit exceed-action drop
 class class-default
  police 1000000 conform-action transmit exceed-action drop
```

Cisco Catalyst 6500 Control Plane Protection

CoPP Deployment Considerations - Summary

Key deployment considerations:

- No HW CoPP processing unless “mls qos” is enabled: this enables also port-level QoS mechanisms
- HW CoPP will ignore a class that does not have a corresponding policing action
- HW CoPP decisions are per forwarding engines
  - SW CoPP for the aggregate traffic
- HW CoPP does not support IP/ARP broadcast/multicast traffic
  - Use multicast HWRL/Dynamic ARP Inspection or “mls qos protocol arp”/Storm Control in conjunction
  - Remember, software CoPP will still match multicast and broadcast traffic, so you MUST classify these packets in CoPP policies

Cisco Catalyst 6500 Control Plane Protection

CoPP Deployment Considerations - Summary

Other considerations:

- No HW CoPP for PFC3A with egress QoS policy
- HW CoPP processing only for packets where HW FIB or HW ingress ACL determines punting. HW egress ACL punts do not pass through CoPP.
- HW CoPP classes can only match what IP ACLs can handle in hardware
- HW CoPP supports only IPv4 and IPv6 (starting 12.2(18)SXE) unicast traffic
  - No support for ARP ACLs, MAC ACLs…
- CoPP rates on Sup720 are bps—pps is not possible. However, HWRL rates are in pps
- CoPP is supported in ingress only (no support for silent mode)
- Not supported today:
  - –SNMP support for CoPP
  - –ACL Log keyword support for CoPP
  - –SP/RP inband SPAN support

http://www.cisco.com/en/US/products/hw/switches/ps708/products_white_paper0900aecd802ca5d6.shtml

Control Plane Policing

Cisco Catalyst 4500 CoPP Deployment

```
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

Cisco Catalyst 4500 CoPP Deployment (cont.)

```
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
```

Define specific policing limits and define any special traffic types