NETWORKERS 2004

HIGH AVAILABILITY IN CAMPUS NETWORK DEPLOYMENTS

SESSION RST-2514

Agenda

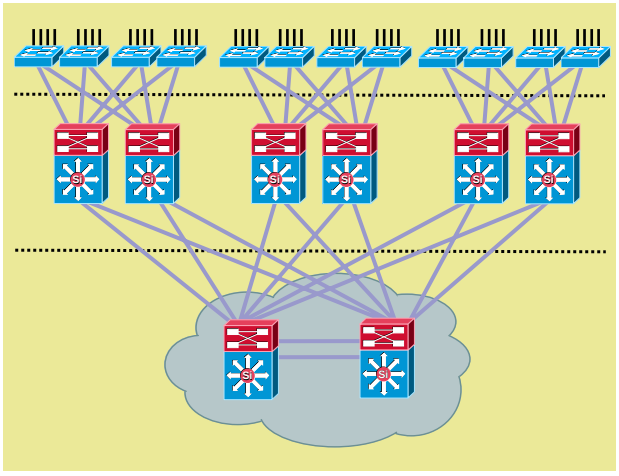
Cisco.com

- High Availability—Definition
- High Availability Features—Access
- High Availability Features—Distribution
- High Availability Features—Core
- High Availability—Summary

HIGH AVAILABILITY: DEFINITION

Multilayer Network Design: Three-Tier Model

Cisco.com



High Availability: What Is It??

Cisco.com

- **Classical definition:**
  - The proportion of time that a system can be used for productive work
- **Typical components/attributes of high availability:**
  - Mean Time Between Failure (MTBF): Average time taken to transition from an initial fault-free state to failure
  - Mean Time to Repair (MTTR): The average time required to diagnose a problem and repair it— whether manually or automatically by the system
  - Availability = MTBF / (MTBF + MTTR)
  - General target for most customers is 99.9% serviceable up time, which equates to a maximum of 4 hours and 23 minutes on average of downtime per year; this figure is not inclusive of scheduled maintenance

RST-2514
9684_05_2004_c2

5

High Availability: What Affects It??

Cisco.com

Common Causes of "Downtime" Include:

- Hardware failure
- Network failure
- Operating system error/failure
- Application error
- Human error
- Environmental issue (earthquakes, flood, storms…etc.)
- System overload
- Security breach or attack

RST-2514
9684_05_2004_c2

High Availability: Improving “Uptime”??

Cisco.com

Some Key Methods in Improving Availability:

- Design the network for high availability
- Minimize MTTR impact by optimizing the following processes:
  - Detect fault
  - Diagnose fault
  - Isolate fault
  - Repair system
- Reduce the risk of unscheduled downtime
- Improve network operations, procedures and processes
- Ensure that the appropriate network management tools are in place
- Measure and continuously improve network availability

RST-2514
9684_05_2004_c2

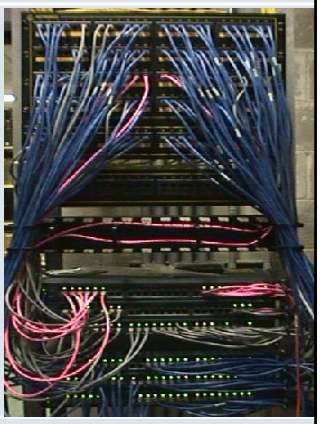
HIGH AVAILABILITY FEATURES: ACCESS

RST-2514
9684_05_2004_c2

HA Expectations of Access Switches

Cisco.com

- Support for redundant uplinks to distribution layer switches
- Power supply redundancy
- L2/L3 protocol cross-stack redundancy on stacked access switches
- Redundant supervisors in chassis-based access switches
- Access port/link/switch resiliency/redundancy features that ensure user uptime
- Going forward, it is expected that high availability support will become a differentiating factor for customers evaluating access switches



IEEE Spanning Tree Protocol Terminology

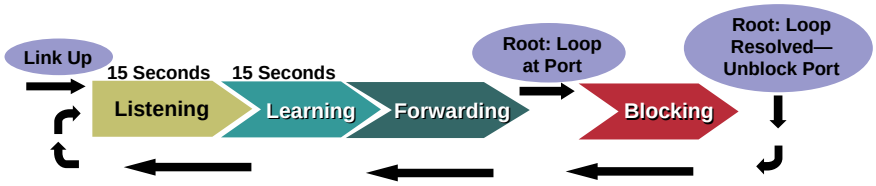
Cisco.com

- 802.1d—Spanning Tree Protocol (STP)
- 802.1w—Rapid Spanning Tree Protocol (RSTP)
- 802.1s—Multiple-instance Spanning Tree (MST)
- PVST—Per VLAN Spanning Tree
- Spanning Tree enhancement tools: PortFast, UplinkFast, BackboneFast, BPDUGuard, BPDUFilter, RootGuard, and LoopGuard

Spanning Tree Protocol: 802.1d

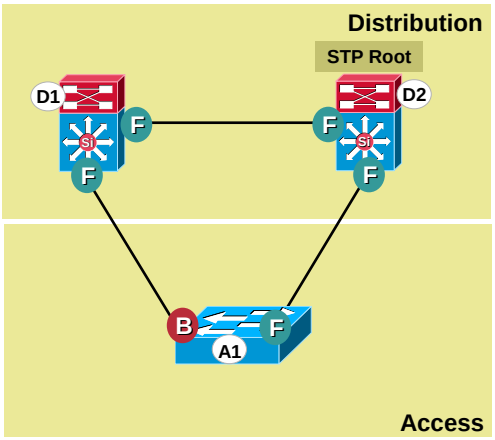
Cisco.com

- Spanning Tree Protocol implemented as a timer-based control mechanism for loops in the network
- A designated root node determines where loops are and which ports to disable in order to resolve loops
- Root sends Bridge Protocol Data Units (BPDU) every 2 seconds to determine topology map and network loops
- Each active port participates and is subject to a cyclic state machine involving several port “states”
- Minimum thirty second transition period when new node added or topology takes place



Spanning Tree Protocol: 802.1d Determining Port Operation Status

Cisco.com



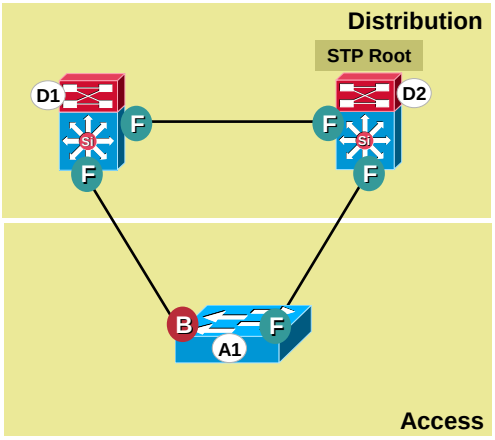
SCENARIO: A New Uplink Is Added Between D1 and A1—Forming a Potential Link

- Link is connected and ports come up—spanning tree senses topology change and begins recalculating the tree
- After tree is recalculated, affected ports go through a 30 second period of listening and learning as spanning tree is recalculated
- Finally, port at A1 for the new uplink is put in blocking state, while the port at the D1 side of the new uplink is placed in forwarding—preventing a loop condition

…Without 802.1d Spanning Tree—packets would continue looping the network

Spanning Tree Protocol: 802.1d How It Works

Cisco.com



SCENARIO: Switch D1 Broadcasts a Frame to Both A1 and D2

- Packet will not reach A1 directly because ingress interface has been blocked by spanning tree
- Switch D2 receives frame and propagates frame to all interfaces other than where frame was received; A1 receives the frame
- Switch A1 processes the frame, but does not propagate the frame further because the port which is connected to switch D1 is in a state of Blocked due to Spanning Tree determining that a loop condition would exist otherwise

…Without 802.1d Spanning Tree—packets would continue looping the network

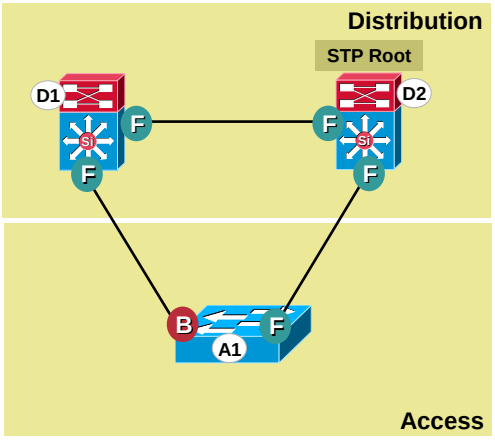
Rapid Spanning Tree Protocol: 802.1w

Cisco.com

- Commonly known as Rapid or Rapid Reconfiguration Spanning Tree (RST)
- Primary difference from 802.1d STP is that .1w not dependent on state timers—much faster convergence
- New states: Discarding (old blocking and listening states), learning, and forwarding
- Most topology changes caused by link state changes or node additions have potential to fully converge in 1 second instead of the 30 to 50 seconds common with 802.1d spanning tree

Rapid Spanning Tree Protocol: 802.1w Determining Port Operation Status

Cisco.com



SCENARIO: A New Uplink Is Added Between D1 and A1—Forming a Potential Link

- Link is connected and ports come up—rapid spanning tree senses topology change and begins recalculating the tree
- RSTP compliant devices in the network quickly determine (within 1 second possible) link/edge port types and where to block
- Result is port at A1 for the new uplink is put in blocking state, while the port at the D1 side of the new uplink is placed in forwarding—preventing a loop condition

…Without 802.1w Spanning Tree feature—convergence time would have been 30 seconds (Min) instead of 1+ seconds

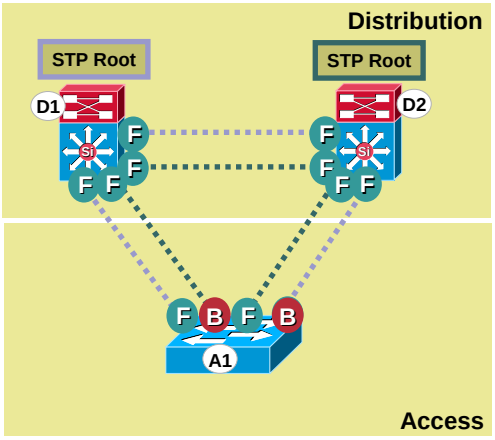
Multi-Instance Spanning Tree Protocol: 802.1s

Cisco.com

- Supports multiple spanning trees in a single network—each supporting their own compliment of VLANs
- Each spanning tree utilizes a virtual tree determined by the VLAN membership footprint in the network; if a switch has a trunk or access port subscribed to a particular VLAN—that switch is now affected by the STP instance associated with that VLAN
- Useful for grouping many VLANs into a specific domain and others into a different domain

Multiple Instance Spanning Tree Protocol: 802.1s Determining Port Operation Status

Cisco.com



SCENARIO: A New Uplink Is Added Between D1 and A1—Forming a Potential Link; All Trunks Support VLANs 10, 20, 30, and 40

- Link is connected and ports come up—spanning tree senses topology change and begins recalculating instances of the tree
- Trees for both instances are recalculated
- The protocol determines which ports must be blocked to avoid loops for each instance

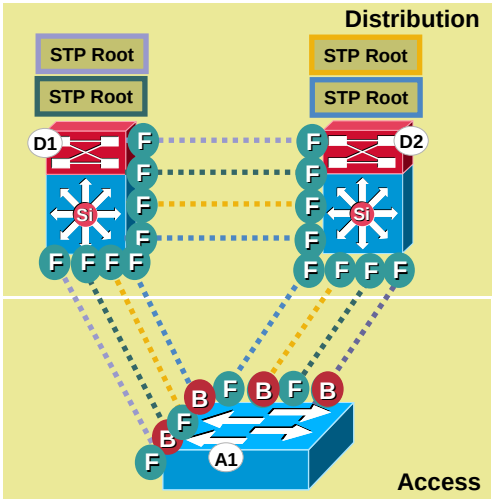
Per VLAN Spanning Tree

Cisco.com

- Per VLAN Spanning Tree (PVST) maintains a spanning tree instance for each VLAN configured in the network
- It allows a trunk (802.1q) to be forwarding for some VLANs while blocking for other VLANs
- Since PVST treats each VLAN as a separate network, it has the ability to load balance traffic (at Layer 2) by forwarding some VLANs on one trunk and other VLANs on another trunk without causing a spanning tree loop

Per VLAN Spanning Tree Protocol (PVST): Determining Port Operation Status

Cisco.com



SCENARIO: A New Uplink Is Added Between D1 and A1—Forming a Potential Link; All Trunks Support VLANs 10, 20, 30, and 40; Each VLAN Has Its Own STP Instance

- Link is connected and ports come up—spanning tree senses topology change and begins recalculating instances of the trees
- Trees for all four VLAN instances are recalculated
- The protocol determines which ports must be blocked to avoid loops for each instance

VLAN10 VLAN30
VLAN20 VLAN40

Spanning Tree Enhancement Tools

Cisco.com

…Various Spanning Tree Enhancements for Optimizing Convergence Performance and/or Improving Network Resiliency

- LoopGuard: Prevents loop conditions when port stops receiving BPDU's, but not effectively out of service
- BPDUGuard: When used with PortFast mode—feature prevents devices beyond the port to affect the STP topology
- PortFast: STP access port feature which allows port to bypass listening and learning states resulting in very rapid up times
- RootGuard: Allows specific enforcement of which devices can become STP root
- UplinkFast: Feature provides a link uptime of between 3–5 seconds when recovering from failure or newly implemented links
- BackboneFast: Feature allows ports to override timer and immediately move to listening state when sensing indirect connection failure in network

EtherChannel Protocol

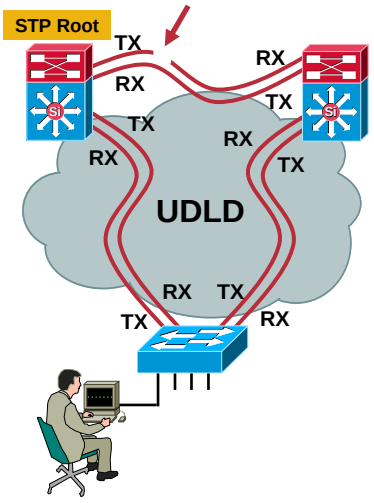
Cisco.com

- Based on the Port Aggregation Protocol (PAgP) and IEEE Link Aggregation Control Protocol (LACP: IEEE802.3ad) specifications—it allows multiple ports to be “bundled” together to produce higher bandwidth trunks
- Aggregated ports must all be the same speed, but can be any non-contiguous ports on any line card in the same chassis or switch in a stack
- EtherChannel® feature provides link redundancy supporting sub-second failover capabilities
- Ports within EtherChannel trunks are managed as a single entity
- Bandwidth of the EtherChannel link is the sum of the individual links’ bandwidth
- Traffic is load-shared between member links
- Outgoing link is selected by computing the hash value on the source and/or destination MAC (L2) and/or IP address
- Hash result is used to select the physical outgoing link in the EtherChannel
- Up to: 8 Fast Ethernet, 8 Gigabit, or 8 10Gigabit ports can be bundled together to form an EtherChannel

Uni-Directional Link Detection Protocol (UDLD)

Cisco.com

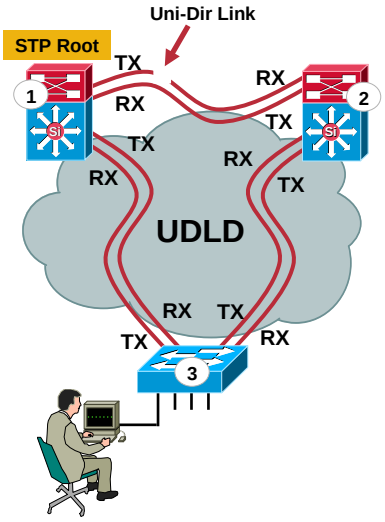
- Allows devices physically connected over copper or fiber cabling to monitor when/if a uni-directional link condition occurs
- Uni-directional links can result in several severe network conditions including spanning tree loops
- Keepalive sent every 15 seconds by default (configurable)



Uni-Directional Link Scenario

Cisco.com

- Spanning tree root (switch 1) transmits BPDUs
- Neighbor (switch 2) doesn't receive them and thinks the root is dead
  - Now claims it's the new root
- Lower switch (switch 3) opens up its blocked port
  - Now there is a loop in the network
- UDLD protocol detects problematic link between switch 1 and 2 and disables the offending port, resolving the loop condition



Switch Stacking and High Availability

Cisco.com

- Primary goal in switch stacking:
  - *Ability to connect a set of independent switches together so that they function as one single cohesive switch in the network*
- A switch when removed from the stack can work as an independent network entity; when connected in the stack— it acts as an integral part of the virtual switch
- Each switch in the stack is still able to make its own local (same switch) switching decisions
- If the acting master node goes offline because of an error or removal—another switch in the stack will automatically and expeditiously assume the master role; this includes all L2 and L3 tasks
- Switch management and configuration possible for all slave switches from the master node

Cisco StackWise Technology Introduction

Cisco.com

- Cisco StackWise technology is the stacking architecture at the heart of the Catalyst® 3750 Series
- Provides a 32Gbps stack interconnect that unifies up to nine individual switches to form a highly available virtual switching platform
- Switches in stack behave as a single, logical, convergence-optimized device
- Advanced hardware/software features provide unparalleled availability and resiliency

Cisco Catalyst 3750

Cisco StackWise Technology System Features

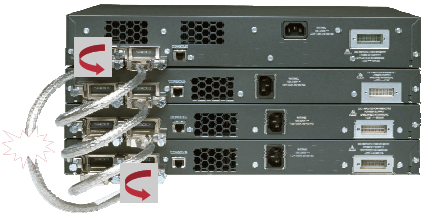
Cisco.com

- **Centralized management**
  - Unified CLI for entire stack
  - Single configuration file for the entire stack
- **Layer 2 distributed architecture**
  - Most Layer 2 features are port-centric
  - Optimal CPU utilization
- **Layer 3 centralized architecture**
  - Master switch is the Route Processor (RP)
  - Slave switches operate as Line Cards (LC) for DCEF
- **Address tables**
  - Each switch maintains all L2 addresses for the stack

Cisco StackWise Technology: Self-Healing Stack

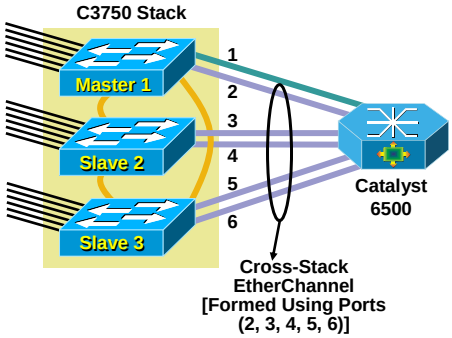
Cisco.com

- …What if stacking cable is removed or malfunctions?
- Port ASICs nearest the missing cable automatically form a loopback
- All intra-stack node messages/traffic are carried over remaining stack connection
- Once stack connection is recovered—loopbacks are automatically removed



Cisco StackWise Technology: Cross-Stack EtherChannels

Cisco.com



Link Failure (Msecs)	Slave Failure (Msecs)	Master Failure (Msecs)
20	660	682

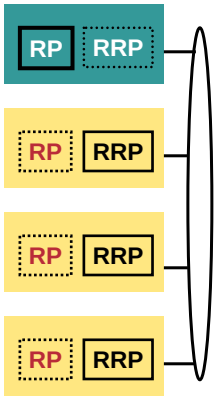
*Single VLAN Configuration

- Master switch owns all EtherChannels and hence the EtherChannel ports belong to the master's STP instance
- EtherChannel port's STP state needs to be synced across the stack to minimize the outage time on master failure
- Stateful Switchover (SSO) is done for EtherChannels STP state
- Because of SSO, when a slave is elected as the master—it has knowledge of the correct STP state of every EtherChannel

Cisco StackWise Technology: IP Unicast Routing High Availability

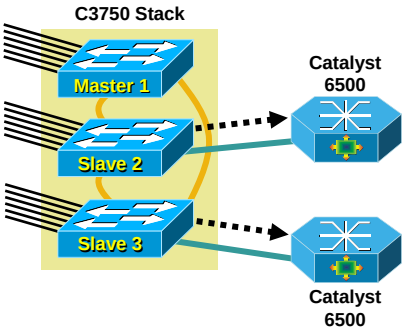
Cisco.com

- Every switch in the stack is capable of becoming an Route Processor (RP)
- Stack is L3 capable when 2 or more switches exist
- High availability paradigm is RPR+ with Fib level NSF
- Master is Route Processor (RP), and all slaves are Redundant Route Processors (RRP) in the stack
- RRP can symbolically be viewed as a redundant line card in a chassis but with the ability to transition into an RP, when given an external stimuli (when switch is elected as the new master)
- Router MAC address for the stack is the RP switch's MAC address
- All switches have interfaces and forward traffic
- RP (master) runs the routing protocols and owns all Layer 3 interfaces (IP stack active on RP)



Cisco StackWise Technology: Cross-Stack Equal Cost Routes

Cisco.com



- Catalyst 3750 supports Cross Stack Equal Cost Routes (CS-ECR)
- IP flows are dynamically load balanced between CS-ECR entities
- Hashes are computed on the source and destination IP addresses
- Hash is used as an index into CS-ECR table of the route to select the outgoing adjacency
- CS-ECR provides switch and link level redundancy

HIGH AVAILABILITY FEATURES: DISTRIBUTION

Processor Redundancy Model Introduction

Cisco.com

- **RPR—Route Processor Redundancy**
  - **Links reset on processor failure**
- **RPR+—Route Processor Redundancy Plus**
  - **Faster switchover than RPR, however links still reset**
- **SSO—State full Switchover**
  - **Maintain selective layer 2 states between Route Processors (RP) in a scalable and efficient manner**
- **NSF—Non-Stop Forwarding**
  - **Continue forwarding packets and prevent route flaps while reconverging Layer 3 protocols**

Route Processor Redundancy (RPR)

Cisco.com

- One supervisor active
- Other supervisor suspended (standby) during bootup
- Console on standby processor is not available after initialize state
- Algorithm to decide active supervisor
  - If there is only one supervisor in the chassis, then it is active
  - If there are two supervisors in the chassis, upon power-up, whoever manages to get the hardware lock first becomes active and other becomes standby; no special treatment for supervisor in slot-1 or slot-2
  - If both supervisors attempt to get the lock at the same time then supervisor in slot-1 has higher probability of becoming active
  - If there is already an active supervisor in the chassis, upon insertion, second supervisor will become standby

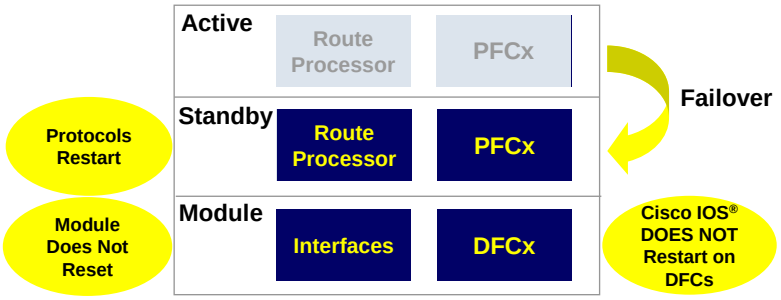
Route Processor Redundancy Plus (RPR+)

Cisco.com

- With RPR+, the secondary supervisor is fully booted, so switchover time is reduced significantly
- RPR+ keeps loaded software image at initialization state (halfway booted) to speed switchover and resumption of traffic forwarding/routing
- RPR+ maintains saved and running configuration images on both the Active and Standby supervisors
- With RPR+, the standby supervisor has full runtime knowledge of configured VLAN, security, and L2/L3 port/protocol parameters

Route Processor Redundancy Plus
Cisco IOS

Cisco.com



Accelerated Switchover, Minimal Interface Impact

- Active supervisor manages the system
- Standby supervisor completely booted
- Supervisor switchover further accelerated over RPR
- Interfaces available 30 seconds after supervisor switchover (no line card reload required)

Stateful Switchover (SSO)

Cisco.com

- High-availability feature which maintains all Layer 2 link and protocol information on both Active and Standby route processors
- In the event of software or hardware malfunction involving one of the route processors, SSO intelligently determines which should be active
- Switchover happens without the rebooting of line cards; L2 protocols stay up so that there is no physical loss of link connectivity

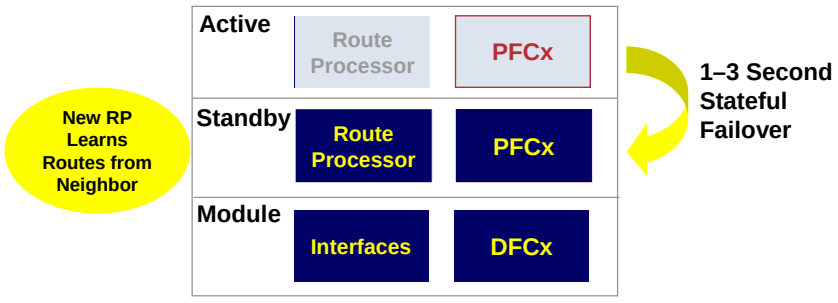
Non-Stop Forwarding (NSF)

Cisco.com

- A high-availability feature which enables routers to continuously forward IP traffic immediately following a route processor switchover
- Continually updates/maintains Layer 3 routing/forwarding information on the backup processor—ensuring continuous IP and routing protocol information processing
- Ensures application reliability
- Combined usage of SSO with NSF results in a highly available routing/switching architecture which can provide zero protocol state and packet loss in the event of processor switchover

Non-Stop Forwarding (SSO + NSF)
Cisco IOS

Cisco.com



Non-Stop Forwarding

- No impact to ATM, PPP, Frame Relay sessions
- Required NSF-aware routing protocols (BGP, OSPF, IS-IS, EIGRP) to recover without route thrash
- Supported on the Sup2 and Sup720
- Planned for Q3 CY ’04 in native Cisco IOS and 2H CY ’04 in hybrid

Processor Redundancy and High Availability: Summary

Cisco.com

HA Mode	Line Card State	Standby RP State	Traffic Outage	Line Card Reset	Route Flap
RPR	H/W Reset	Startup Configuration Synchronized	90+ Secs	Yes	Yes
RPR+	S/W Reinitialized	Standby Processor Fully Booted. Startup/Current Configurations Kept Synchronized	30+ Secs	No	Yes
NSF/ SSO	Reconciliation	Application Runtime State Preserved	~1–3 Secs	No	No

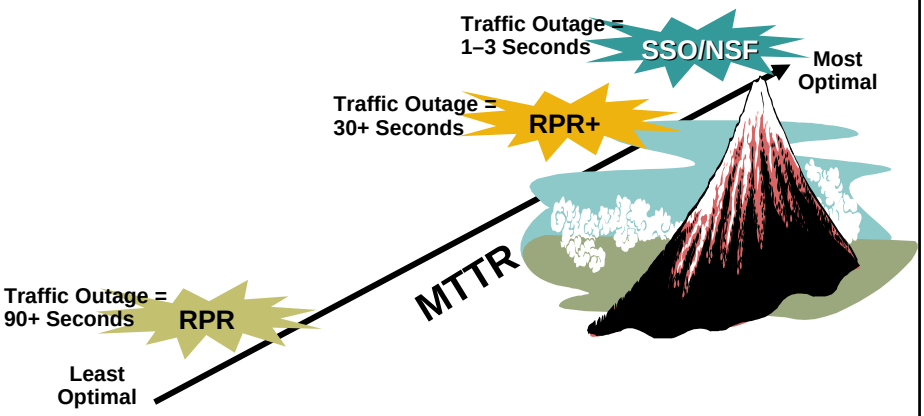
NSF Configuration Example: OSPF

Cisco.com

```
Sup720(config)#router ospf 1234
Sup720(config-router)# nsf
!enable NSF for OSPF routing process 1234
Sup720(config-router)# timers nsf flush <seconds>
Sup720(config-router)# timers nsf wait <seconds>
```

Processor Models and High Availability: Minimizing MTTR

Cisco.com



- Mean Time To Repair (MTTR) value is closest to optimal as processor switchover interruption time and associated effects are minimized

Switching Fabric Redundancy Models

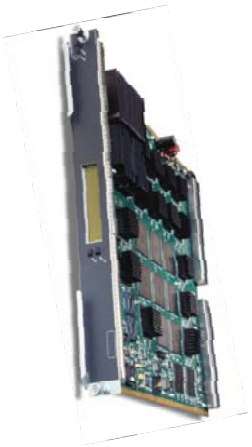
Cisco.com

- Modular fabric model
  - Switching fabric reside on system line modules
- Active backplane model
  - Switching fabric is resident on backplane

Switching Fabric Redundancy Models: Modular Fabric Model

Cisco.com

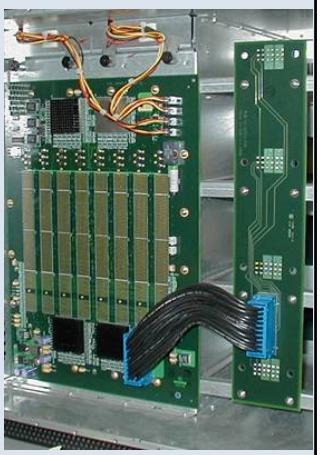
- Switching fabric located on either dedicated fabric module or integrated on supervisor module
- Under certain architectural conditions, may be upgraded with new modules providing greater fabric capacity
- Removable and easily replaced in the event of failure
- Ideally functions using a redundancy model which ensures full fabric capacity even after redundant fabric failover
- Optimally serviceable



Switching Fabric Redundancy Models: Fixed Backplane Model

Cisco.com

- Switching fabric typically is permanently affixed to system backplane or some other nonremovable assembly
- Switching fabric failure generally requires chassis replacement
- Typically not upgradeable and may not offer redundancy
- Not optimally serviceable



Power Supply Redundancy

Cisco.com

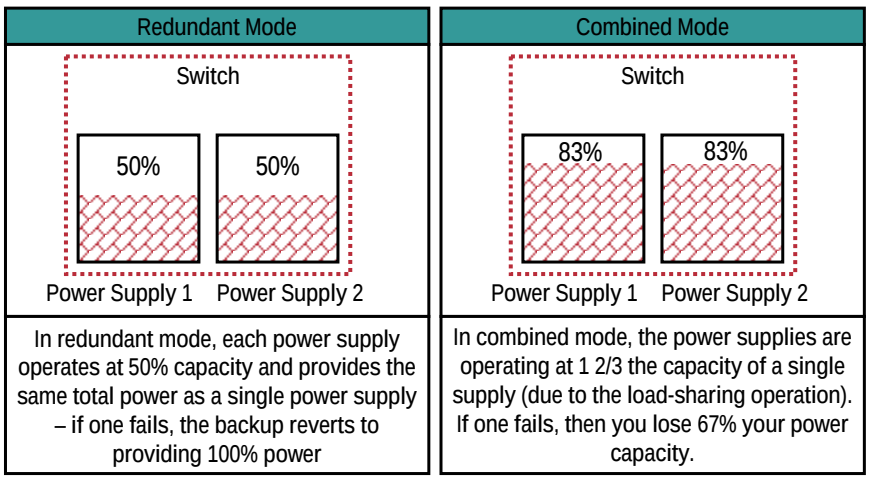
Two Types of Redundancy-Modes Are Supported: Redundant and Combined Modes

- **Redundant mode**
  - **Only one power supply provides power; the other, if present, provides backup if primary supply fails**
  - **Often results in reduced power capacity due to wasted potential on part of standby supply**
- **Combined mode**
  - **Both the power supplies provide power**
  - **System has no standby redundancy, but loss of one power supply usually only results in a minimal reduction of power budget**
  - **Allows system to use all available wattage for system needs**

Catalyst 6500 Series
Understanding Power Redundancy

Cisco.com

The 6500 can utilize two power supplies to work in either combined or redundant mode



Catalyst 6500 PoE
Insufficient Power Scenarios

Cisco.com

- If the available power in the system decreases, the system might not be able to power the same modules and devices in a particular configuration. The power management software is forced to power down certain modules and Powered Devices (PDs).
- The order that modules and devices are powered down is as follows:
  - For modules that have PD's attached, the PD's are powered down before any linecard is powered down.
  - The PD's are powered down beginning from the highest to the lowest port number (e.g. port 48 down to port 1).
  - Linecards are powered down from the bottom slots to the top slots until the system is under the power budget.
  - Supervisors, switch fabrics, and service modules are skipped and are subsequently the last to be powered down in order to maintain system integrity.

Catalyst 6500 Series
Calculating Power Capacity

Cisco.com

To help aid in the configuration and sizing of PoE on Catalyst Switches, Cisco has built an online Power Calculator. This tool is available on CCO for customers and partners to input their specific configurations. The Tool recommends the power supply and the total number of PD's that each configuration can support.

www.cisco.com/go/powercalculator

First Hop Redundancy Protocols

Cisco.com

- **Hot Standby Router Protocol (HSRP)**
  - **Cisco informational RFC 2281 (March 1998)**
- **Virtual Router Redundancy Protocol (VRRP)**
  - **IETF standard RFC 2338 (April 1998)**
- **Gateway Load Balancing Protocol (GLBP)**
  - **Cisco designed, load sharing, patent pending**

HSRP

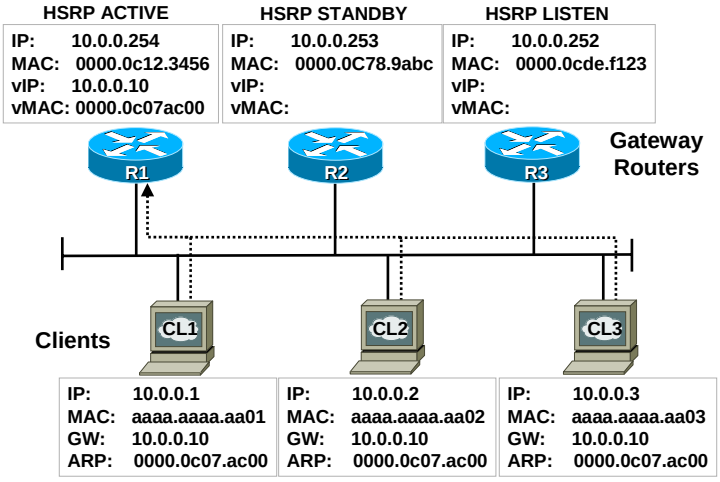
Cisco.com

- **A group of routers function as one virtual router by sharing ONE virtual IP address and one virtual MAC address**
- **One (active) router performs packet forwarding for local hosts**
- **The rest of the routers provide “hot standby” in case the active router fails**
- **Standby routers stay idle as far as packet forwarding from the client side is concerned**
- **Up to 255 HSRP groups are supported**

First-Hop Redundancy with HSRP

Cisco.com

R1: Active, Forwarding Traffic; R2, R3: Hot Standby, Idle



VRRP

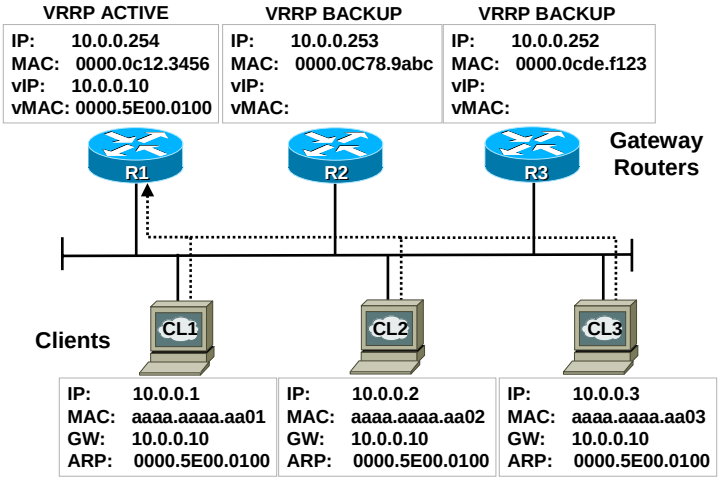
Cisco.com

- Very similar to HSRP
- A group of routers function as one virtual router by sharing ONE virtual IP address and one virtual MAC address
- One (master) router performs packet forwarding for local hosts
- The rest of the routers act as “back up” in case the master router fails
- Backup routers stay idle as far as packet forwarding from the client side is concerned

First-Hop Redundancy with VRRP

Cisco.com

R1: Master, Forwarding Traffic; R2, R3: Backup



GLBP

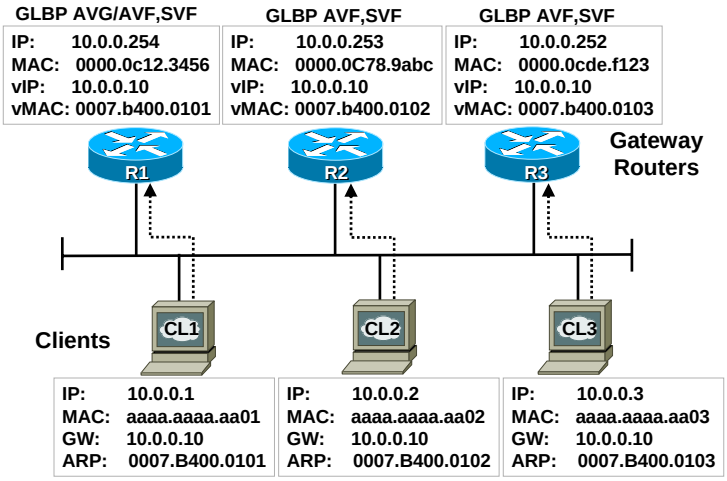
Cisco.com

- A group of routers function as one virtual router by sharing one virtual IP address but using multiple virtual MAC addresses for traffic forwarding
- Traffic is shared over multiple upstream links, improving throughput and reducing congestion when no failure state exists
- Allows traffic from a single common subnet to go through multiple redundant gateways using a single virtual IP address
- Improved Tracking Capabilities (Interface or Route)

First-Hop Redundancy with GLBP

Cisco.com

R1: AVG; R1, R2, R3: All Forward Traffic



HIGH AVAILABILITY FEATURES: CORE

High Availability in the Core: Quick Overview

Cisco.com

- Layer 3 routing protocol redundancy
  - BGP convergence optimization
  - Routing protocol Non-Stop Forwarding (NSF) awareness
  - Incremental Shortest Path First (SPF) technologies
  - IP event dampening
  - Stateful network address translation
- Network level redundancy
  - Multicast sub-second convergence
  - MPLS traffic engineering fast reroute

HIGH AVAILABILITY: SUMMARY

High-Availability Summary: Key Areas of Interest

Cisco.com

- **Device technologies**
  - Route processors
  - Switching fabrics
  - Stacking technologies
  - Power supplies
  - Trunk ports
- **Network technologies**
  - Routing protocols
  - Router redundancy protocol
  - Spanning Tree Protocols

High Availability Across Cisco Catalyst Portfolio

Cisco.com

Catalyst 6500

- Non-Stop Forwarding (IOS)
- Hitless Software Upgrades (CatOS)
- Stateful SwitchOver and Services (IOS)
- Maintenance of State Data for Services

Catalyst 4500

- Redundant Supervisors (Catalyst 4507R/4710R)
- Sub-Minute Supervisor Failover (RPR)
- Redundant Clocking
- Stateful Switch Over

Catalyst 3750

Catalyst 3550/ Catalyst 2950

- Bi-Directional Stack Failover
- Layer 3 Failover
- Cross-Stack EtherChannel
- Optional Redundant External PSUs
- Self-Healing Stack
- Switch Clustering
- Optional Redundant External PSUs

Catalyst Chassis Foundation

- Full HW Redundancy
- Hot Swappable, Load Balancing Integrated PSUs
- Hot Swappable Modules
- MISTP
- Multi-Module EtherChannel
- Multiple Spanning Tree
- Hot Swappable Supervisors in Redundant Configuration

Cisco High Availability Foundation

- High MTBF
- HSRP
- PortFast
- UplinkFast
- BackboneFast
- RootGuard
- LoopGuard
- BPDUGuard
- 802.1w
- 802.1s
- L2 Load Balancing (PVST+)
- L3 Load Balancing (Equal Cost Routing)
- UDLD
- GLBP
- EtherChannel Resiliency
- PAgP

Complete Your Online Session Evaluation!

Cisco.com

WHAT: Complete an online session evaluation and your name will be entered into a daily drawing

WHY: Win fabulous prizes! Give us your feedback!

WHERE: Go to the Internet stations located throughout the Convention Center

HOW: Winners will be posted on the onsite Networkers Website; four winners per day