Cisco Catalyst 6500 High Availability: Deploying Redundant Supervisors for Maximum Uptime

Introduction

The Cisco® Catalyst® 6500 is deployed in the most critical parts of enterprise and service provider networks. Having such a vital position in the network, the Cisco Catalyst 6500 must provide the highest levels of availability. To achieve these levels of availability network engineers employ both network wide technologies as well as device level redundancy. This includes network designs with redundant switches, redundant paths using Cisco EtherChannel® technology, First Hop Redundancy Protocols, the Cisco Virtual Switching System (VSS) and of course redundant system components including power supplies, fans and Supervisor modules.

This paper discusses the Redundant Supervisor technologies for the Cisco Catalyst 6500. These technologies have evolved over time from nonstateful, Route Processor Redundancy mode (RPR) to the current Stateful Switchover (SSO) mode with Nonstop Forwarding (NSF).

The newest addition to the suite of Redundant Supervisor technologies is the In-Service Software Upgrade (ISSU) technology which enables redundant Supervisors to use the SSO redundancy mode even when running different versions of Cisco IOS® Software. The new ISSU versioning infrastructure allows for a streamlined software upgrade process with minimal downtime when performing full image software upgrades. The ISSU process can also be used to activate Maintenance Packs within Cisco IOS Software Modularity. The new ISSU infrastructure provides a significant improvement for full image software upgrades when performed with the Cisco Virtual Switching System (VSS).

This paper describes the NSF and SSO platform-specific details. Although it is not the primary goal of this paper, it is very important for readers to understand how to design a highly available network with NSF and SSO. For high-availability campus network design information, in-depth information about generic NSF with SSO operations and Multicast Multilayer Switching (MMLS) NSF with SSO is included.

Measuring Availability

Availability is used as a metric to describe the amount of time a system is available to its users. As long as the users can continue to use the system or application without any perceived degradation then the system is available. In other words, availability is always measured from the users' perspective.

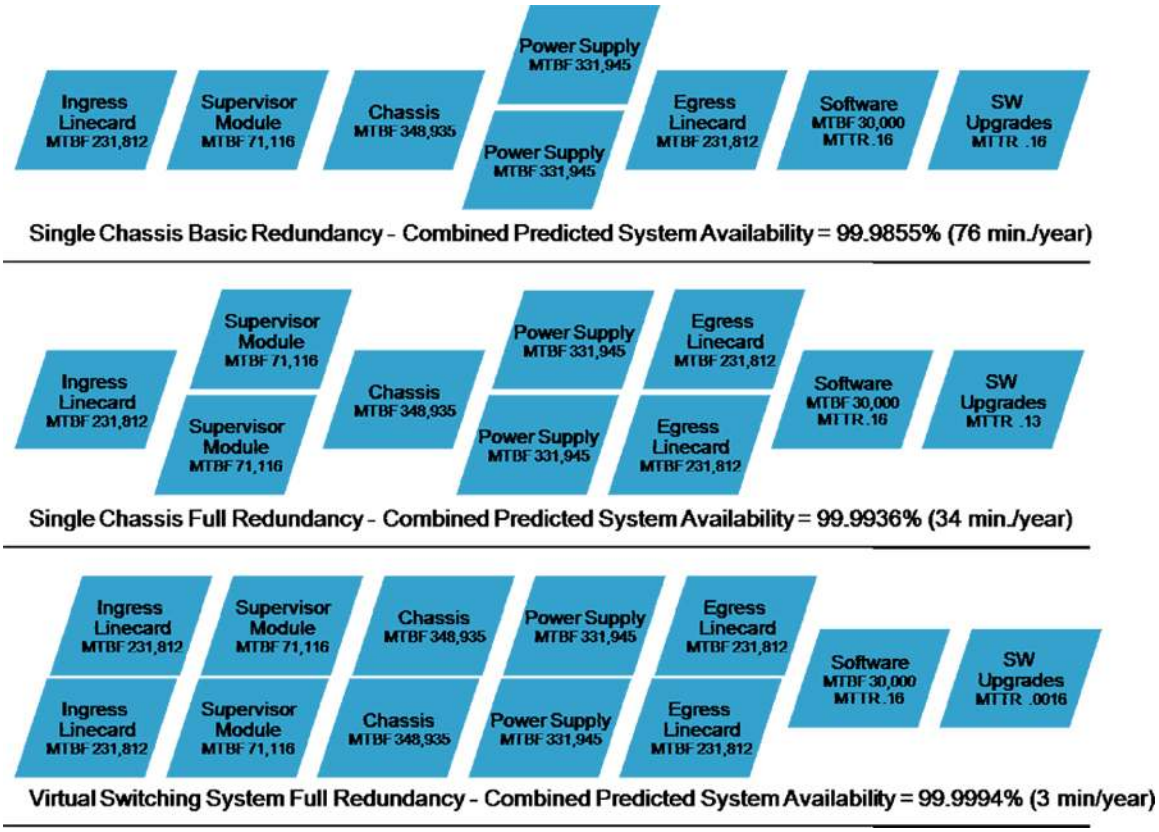
Availability ratings can be used to weigh different network designs and system configurations. For example, for a given amount of expense which network design provides the highest level of availability? Or how much redundancy is needed to achieve a certain level of availability?

To answer these types of questions different statistical models exist to obtain a predicted system availability rating. One method of calculating predicted system availability is to amortize the mean time to repair (MTTR) over the mean time between failures (MTBF) period. MTBF typically means the average amount time between component failures. The MTTR is an average amount of time required for the system to recover from a specific component failure. Therefore the predicted system availability would be expressed mathematically as a fraction of the MTBF divided by the sum of MTBF and MTTR.

Availability = MTBF/ (MTBF+MTRR)

Using MTBF values obtained from the manufacturer and MTTR values based on the specific environment one can calculate the *predicted system availability* rating for a given system and or network design. As an example, consider three different configurations, a single chassis with basic redundancy, a single chassis with full redundancy and finally the Virtual Switching System which uses redundant chassis. Figure 1 shows the different component MTBF values[1] and the overall system availability ratings. The overall system availability rating is derived using Reliability Block Diagram (RBD) where component availability ratings are combined in either a series or parallel calculation depending upon the level of redundancy. Nonredundant components are combined in series while redundant components are combined in parallel.

Figure 1. Redundant Configuration Comparison



Using the Reliability Block Diagrams above, the impact to the predicted system availability can be compared.

It is important to keep in mind that these are statistical models used to weigh different designs and configurations. These calculations do not include other types of failures such as outages due to high traffic rates, or human error. Furthermore the availability ratings are derived from a statistical average of a large number of devices, the larger the sample size the more accurate the rating.

Network engineers should also consider how long of an outage is tolerable by the users. If a Supervisor module fails, how long of an outage will users connected to that device experience? And is that acceptable? The answers to these questions are dependent upon the applications being used and the purpose of the network.

[1] MTBF vales used in Figure 1 are for example purposes only.

Switch Redundancy Components

The Cisco Catalyst 6500 Series switches are designed with redundant field replaceable components to achieve its maximum system availability. The following components in the Cisco Catalyst 6500 switches provide switch redundancy:

Supervisor engine: Every Cisco Catalyst 6500 chassis can support redundant supervisors. Supervisors operate in active and standby modes and support a variety of redundancy mechanisms for failover.

Switch fabric: The switch fabric provides a data path for fabric-enabled line cards and increases the available system bandwidth from the shared bus capacity of 32 Gbps to 720 Gbps for the Supervisor Engine 720. If a switch fabric fails, the redundant switch fabric (if present) takes over.

Power supplies: Every Cisco Catalyst 6500 chassis supports redundant power supplies.

Fan trays: All Cisco Catalyst 6500 chassis provide redundant fans within the fan tray assembly. The Cisco Catalyst 6509-V-E chassis also provides fan-tray redundancy.

Module online insertion and removal (OIR): New modules can be added while the system continues to operate. Line cards can be exchanged with like line cards without losing the configuration. When a module with a local forwarding engine (also referred to as distributed forwarding card) is inserted, the local forwarding-engine hardware tables are repopulated with the most current forwarding information.

Supervisor Redundancy

The supervisor engine that boots first becomes the active supervisor engine. The active supervisor is responsible for control-plane processing and data-plane forwarding. The second supervisor is the standby supervisor, which does not participate in the control or data-plane decisions. The active supervisor synchronizes configuration and protocol state information to the standby supervisor. As a result, the standby supervisor is ready to take over the active supervisor responsibilities if the active supervisor fails. This “take-over” process from the active supervisor to the standby supervisor is referred to as switchover.

While there is only one active supervisor engine at any given time, it is important to note that the user ports on the Standby Supervisor engine are operational. The ports on the Supervisor modules will follow the state of the Supervisor engine itself, for example is the standby Supervisor module were to perform a reload, then the ports on that Supervisor module will also follow go down and then back up as the module performs its initialization, just as any other IO module in the system would.

Supervisor Redundancy Operations

Supervisor redundancy technologies have evolved from providing basic stateless redundancy modes with only startup configuration synchronization ,to full stateful redundancy with startup and running configuration synchronization. Each of these redundancy modes of operation improves upon the previous mode.

Route Processor Redundancy: RPR is the first redundancy mode of operation introduced in Cisco IOS Software. In RPR mode, the startup configuration and boot registers are synchronized between the active and standby supervisors, the standby software is not fully initialized. An important distinction of RPR mode is that the software images between the active and standby supervisors do not have to be the same version. Upon switchover, the standby supervisor becomes active automatically, but it must complete the boot process. In addition, all line cards are reloaded and the hardware is reprogrammed. The RPR switchover time is 1 or more minutes.

Route Processor Redundancy+: RPR+ is an enhancement to RPR in which the standby supervisor is completely booted and line cards do not reload upon switchover. The running configuration is synchronized between the active and the standby supervisors. All synchronization activities inherited from RPR are also performed. The synchronization is done before the switchover, and the information synchronized to the standby is used when the

standby becomes active to minimize the downtime. No link layer or control-plane information is synchronized between the active and the standby supervisors. Interfaces may bounce after switchover, and the hardware contents need to be reprogrammed. For both RPR+ mode and also SSO mode, the software images must of the same license type and same version. The RPR+ switchover time is 30 or more seconds.

Stateful Switchover: SSO expands the RPR+ capabilities to provide transparent failover of certain Layer 2 protocols and certain Cisco IOS Software applications when a supervisor switchover occurs. These protocols and Cisco IOS Software applications that are synchronized between the active and standby supervisor module are called *HA-aware applications*. Not all protocols and Cisco IOS Software applications are synchronized; these are referred to as *non-HA-aware applications*. Policy-feature-card (PFC) and distributed-forwarding-card (DFC) hardware tables are maintained across a switchover. This allows for transparent data-plane failover at Layer 2 and Layer 4. It is important to note that routing protocols are non-HA-aware applications and therefore software routing tables are not maintained across supervisor switchover events.SSO data-plane switchover time is 0 to 3 seconds.

Nonstop Forwarding with Stateful Switchover: NSF works in conjunction with SSO to help ensure Layer 3 integrity following a switchover. It allows a router experiencing the failure of an active supervisor to continue forwarding data packets along known routes while the routing protocol information is recovered and validated. Data-plane forwarding can continue to occur even though peering arrangements with neighbor routers have been lost on the restarting router. NSF relies on the separation of the control plane and the data plane during supervisor switchover. The data plane continues to forward packets based on pre-switchover Cisco Express Forwarding information. The control plane implements graceful restart routing protocol extensions to signal a supervisor restart to NSF-aware neighbor routers, reform its neighbor adjacencies, and rebuild its routing protocol database following a switchover. An *NSF-capable router* implements the NSF functionality and continues to forward data packets after a supervisor failure. An *NSF-aware router* understands the NSF graceful restart mechanisms: it does not tear down its neighbor relationships with the NSF-capable restarting router, and can help a neighboring NSF-capable router restart thus avoiding unnecessary route flaps and network instability. An NSF-capable router is also NSF-aware.

Multicast Multilayer Switching NSF with SSO enables the system to maintain multicast forwarding state in the PFC3 and DFC3 hardware during a supervisor-engine switchover, minimizing multicast service interruption. Prior to MMLS NSF with SSO, the multicast forwarding entries were not synchronized to the standby supervisor engine. The NSF with SSO switchover time is 0 to 3 seconds for Layer 2–4 unicast or multicast traffic. MMLS NSF with SSO is enabled by default when the redundancy mode is SSO.

Stateful Switchover (SSO)

SSO expands the synchronization capabilities of RPR+ to allow transparent failover at Layer 2 and Layer 4. Synchronization from the active to the standby supervisor is not limited to startup configuration, startup variables, and running configuration; it also applies to the SSO-aware applications and their runtime data. This dynamic data synchronization, referred to as check pointing, relies on the Cisco IOS Software Redundancy Facility and the Checkpoint Facility to initiate failovers and provide ordered and reliable communication between peer protocol processes on the active and standby supervisors. SSO bulk synchronization occurs at boot time. When a system is operational, configuration synchronization and state check pointing for various protocols happen as changes occur within the system.

Figure 2. Stateful Switchover Synchronization

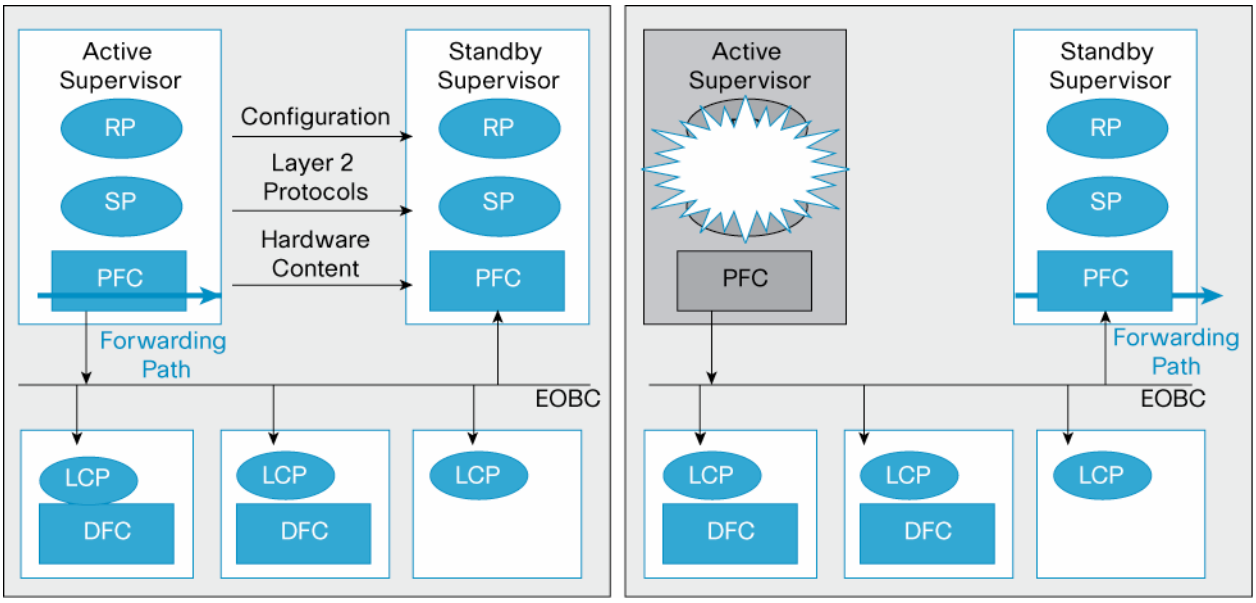


Figure 2 depicts SSO synchronization during normal operations. In SSO mode, Layer 2 protocols and PFC hardware contents are synchronized from the active supervisor to the standby supervisor. Upon switchover, Layer 2 traffic can be forwarded without disruption. On the figure, the RP is the route processor, the SP is the switch processor, and LCP is the linecard processor, PFC is the policy feature card, and DFC is the distributed forwarding card.

SSO synchronizes runtime data for Layer 2 dynamic protocols. As Layer 2 control-plane, configuration, or other network-related changes occur, the Cisco IOS Software Checkpoint Facility running between the peer processes on the active and standby supervisors communicates the changes. For example, the Spanning Tree Protocol database on the standby supervisor is kept up-to-date by check pointing both protocol information and port states from the active supervisor.

SSO also supports synchronization of hardware tables between the active and the standby supervisor. The Policy Feature Card (PFC) is a supervisor daughter-card that contains Application Specific Integrated Circuit (ASIC) responsible for hardware switching functions. The PFC contents are synchronized between the active and the standby supervisor. Every time new hardware table entries need to be downloaded to the PFC, entries are also downloaded to all other forwarding engines in the system. This allows the standby supervisor PFC to bear the same forwarding information as the active PFC and the Distributed Forwarding Cards (DFC). The standby supervisor PFC does not generate forwarding results and it is not a complete mirror of the PFC. Only new control driven hardware table contents are downloaded to the systems' forwarding engines. Control driven refers to updates driven by configuration or protocol updates on the route processor (RP) and switch processors (SP). Downloaded information includes Forwarding Information Base (FIB), adjacency, Access Control Lists (ACL), and Quality of Service (QoS) hardware tables contents. However, data driven hardware table contents are not the same on the active and standby supervisors. As a result, NetFlow table contents vary depending on local flow information. ACL, QoS, VLAN, NetFlow statistics and other local forwarding related statistics are not distributed to other forwarding engines included the standby's PFC. MAC address table operations are also local to forwarding engines even though MAC notification helps ensure consistency among forwarding engines.

Supervisor Switchover Operation

Upon supervisor fault detection, a series of steps occur before the standby supervisor completely takes over. This helps ensure that all modules in the system understand that a switchover has taken place and a new supervisor assumes the role of active supervisor. Figure 3 depicts the supervisor switchover operation in SSO redundancy mode.

Figure 3. SSO Operation

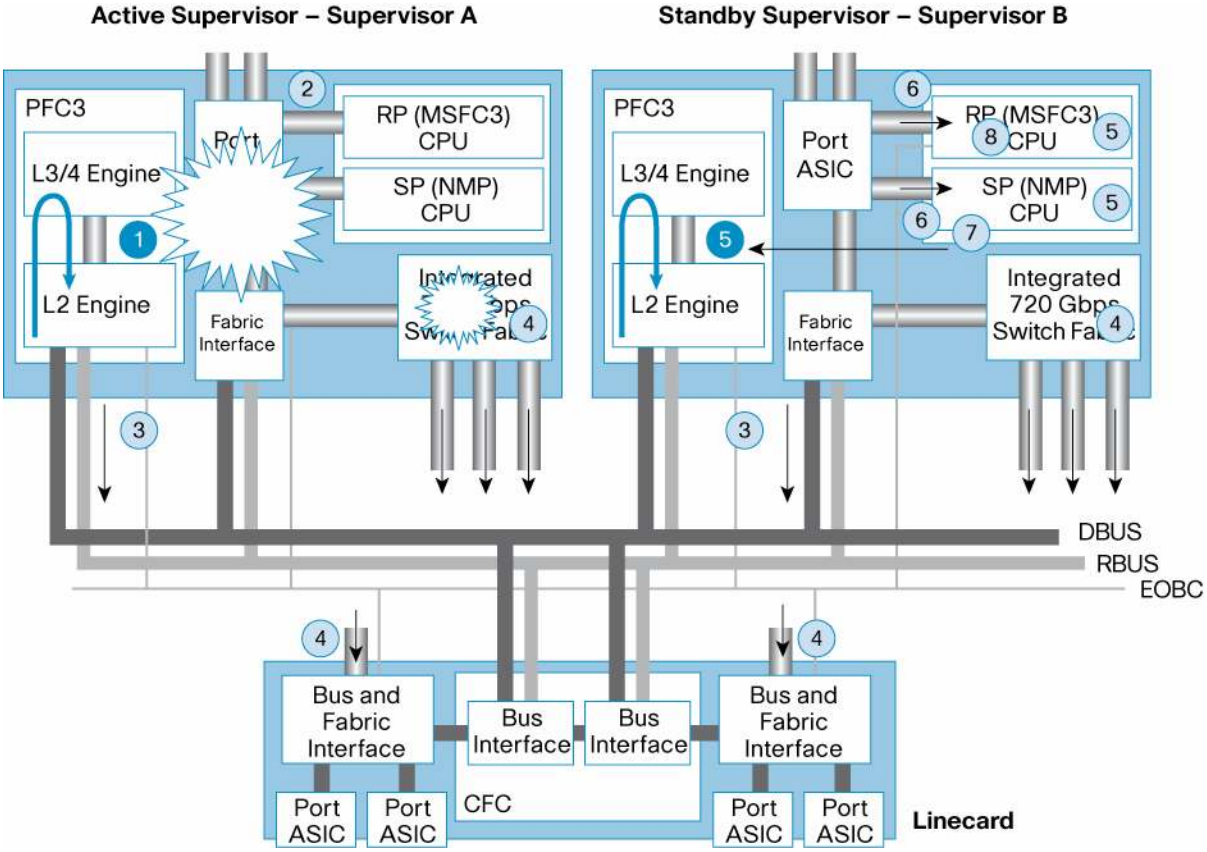


Figure 3 depicts the supervisor switchover operation in SSO redundancy mode. The numbers on the figure relate to switchover steps and are explained in details below.

Steps 1 through 8 appearing on Figure 3 are described below:

1. The active supervisor A is responsible for forwarding decisions. Hardware table synchronization and Layer 2 protocol state synchronization occur in the background.
2. A software or hardware fault is detected on the active supervisor. This fault could be detected by software exception conditions, Generic Online Diagnostics (GOLD) background checks, keep-alive failures between the Route Processor (RP) and the Switch Processor (SP), fabric switching module state changes on a Supervisor 720, or it could be the result of a user initiated switchover.
3. The failed supervisor A stops driving decisions on the bus. The forwarding engine on supervisor B is activated to drive forwarding results.
4. The active fabric access on supervisor A is disabled. The standby fabric on supervisor B assumes the role of the active fabric and signals the fabric enabled linecards to make the necessary changes to connect to the new fabric.

5. Supervisor B becomes the active supervisor. The data path is restored and data can be switched in hardware based on the latest synchronized PFC entries and hardly be affected by the switchover.
6. SSO-aware protocols are initialized and the CPU starts processing protocol and data packets. All other SSO-unaware protocols are then initialized.
7. Stale PFC entries or entries for non-SSO aware protocols are purged. Supported Layer 2 Control protocols and Layer 4 polices derived from QoS or ACL polices are not affected by a switchover.
8. Routing protocols need to restart: dynamic entries in the FIB and adjacency tables are flushed, thereby affecting the Layer 3 routed traffic. Static routes are maintained across a switchover because they are based on static configuration and are not dynamic.

Note that the steps listed above are generic and include bus and fabric synchronization. Beginning with Cisco IOS Software Release 12.2(33)SXH a new Hot-Sync Standby fabric feature was enabled to reduce switchover times associated with fabric synchronization.

The Hot-Sync Standby fabric feature reduces the switchover time by bringing the fabric channels on the standby Supervisor module to a fully synchronized state. The fully synchronized state includes sending clocking signals via the fabric channels. Even though the Standby Supervisors' switch fabric is in the Hot-Sync state, the fabric is still not used for user data traffic. This simply reduces the time for the Standby to take over as the Active Supervisor.

The Hot-Sync Standby fabric feature is enabled by default for all 6700-series line cards installed in an E-series chassis running Cisco IOS Software Release 12.2(33)SXH or newer. No configuration commands are need to enable this feature.

To verify the status of the Hot-Sync Standby fabric feature use the show fabric status CLI command. The output from the show fabric status command is given here in Figure 4, modules in slots 1-3 are 6700-series modules, slot 4 contains a 6500-series module and slots 5 and 6 contain Supervisor 720 modules. Note that the Hot-Sync Standby fabric feature is enabled even when a mix different IO modules types is installed, in this case the 6700-series and 6500-series IO modules are installed.

Figure 4. *show fabric status* Output

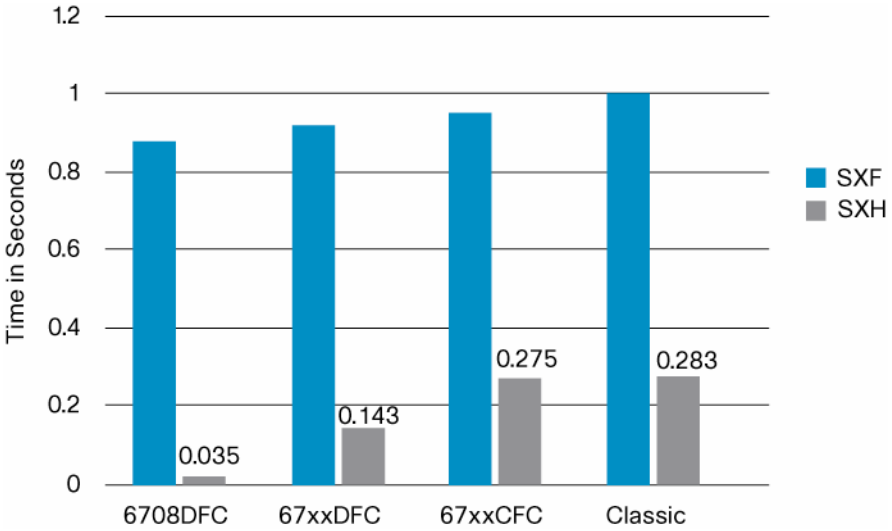
```
Switch#show fabric status
 slot  channel speed module   fabric   hotStandby  Standby  Standby
                    status   status      support  module   fabric
    1        0   20G     OK       OK       Y(hot)
    1        1   20G     OK       OK       Y(hot)
    2        0   20G     OK       OK       Y(hot)
    3        0   20G     OK       OK       Y(hot)
    3        1   20G     OK       OK       Y(hot)
    4        0    8G     OK       OK          N/A
    5        0   20G     OK       OK          N/A
    5        1   20G     OK       OK          N/A
    6        0   20G     OK       OK          N/A
    6        1   20G     OK       OK          N/A

Switch#
```

Even though SSO helps ensure stateful switchover for many protocols, some packet loss occurs in most circumstances during SSO switchover due to fabric and bus data plane reestablishment. The SSO switchover could cause data-plane traffic loss in the order of 0 to 3 seconds depending on the conditions.

As reference Figure 5 provides a graph showing data-plane switchover results for different line card types comparing switchover times with Cisco IOS Software 12.2.(18)SXF (without Hot-Sync Standby Fabric) and Cisco IOS Software 12.2(33)SXH (with Hot-Sync Standby fabric).

Figure 5. Average Data-Plane Switchover Time Comparison Between Cisco IOS Software Release 12.2(18)SXF and 12.2(33)SXH



The results in Figure 5 are an average for 10 test iterations and show an improvement for all IO Module types, even modules that do not employ Hot-Sync Standby fabric. The across the board improvements are due to the benefits achieved with the faster fabric switchover time as well as other switchover related improvements in the SXH software release.

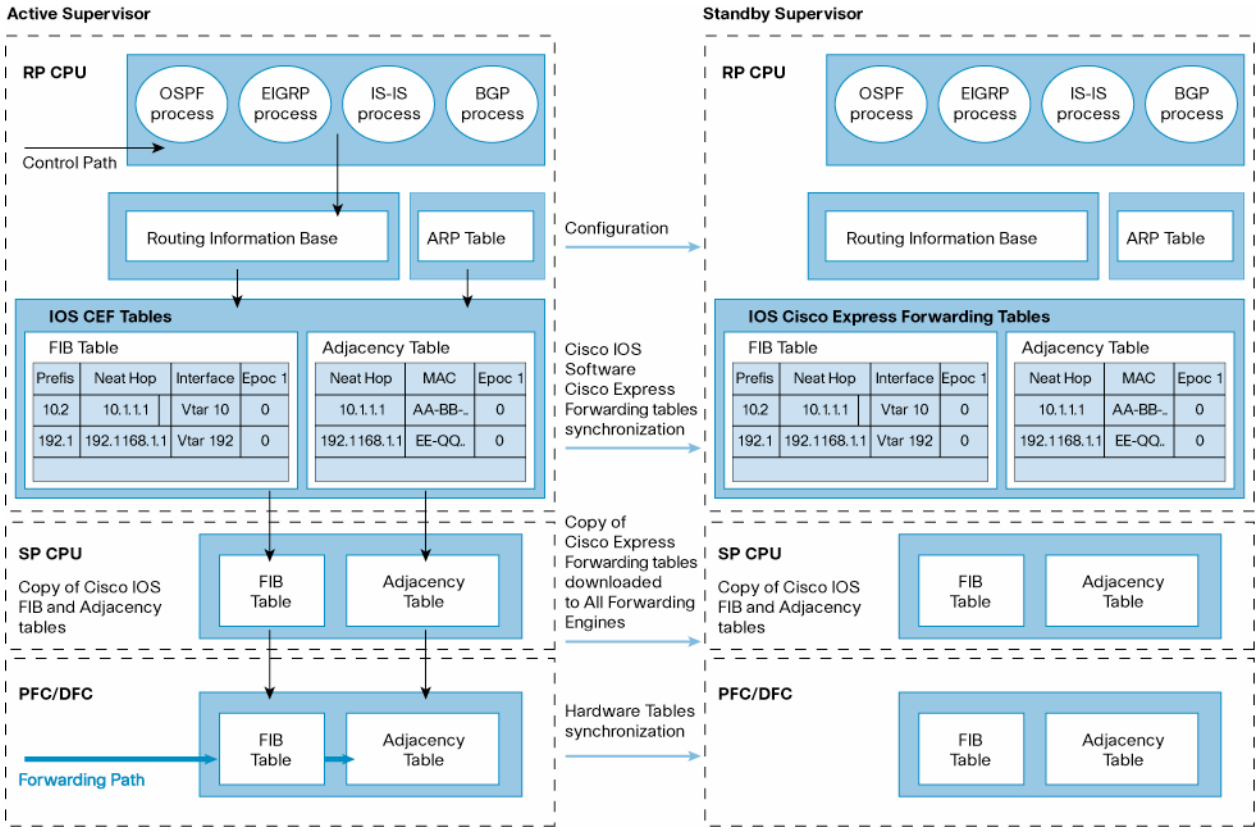
The 6708-10GE module improves significantly not only from the Hot-Sync Standby fabric feature but also because of a hardware based interrupt notification enhancement. The hardware based interrupt notification was initially available in the Cisco IOS Software 12.2(33)SXH release. In this case the newly active Supervisor module notifies the 6708-10GE module of the switchover event via a hardware-based signal. Previous to this enhancement the newly active Supervisor would notify the line cards of the switchover event via a message initiated in software and sent over an out of band communication channel. The hardware based interrupt notification technique is only available on the 6708-10GE and 6716-10GE line cards, and is planned for future generation Cisco Catalyst 6500 line cards.

Nonstop Forwarding with Stateful Switchover (NSF/SSO)

NSF/SSO Synchronization Operation

Layer 3 packet forwarding in a Cisco Catalyst 6500 is provided by a process called Cisco Express Forwarding (CEF). The CEF design maintains two tables: a Forwarding Information Base (FIB) and an adjacency table. The FIB table is a distilled version of the routing table, containing only information relevant to the forwarding process and not to particular routing protocols. For example, the routing protocols administrative distance is not relevant to the forwarding process. The adjacency table is a collection of next-hop rewrite information for adjacent nodes.

Figure 6. NSF/SSO Synchronization



In Figure 6, black steps are control plane-driven, whereas blue steps are data-driven. Green arrows show synchronization operations for the software Cisco Express Forwarding tables and the PFC Cisco Express Forwarding tables.

During normal operation, the system collects the routes calculated by each routing protocol into a common database called the Routing Information Base (RIB). When information for all routing protocols is present in the RIB, the RIB is scanned to determine the lowest-cost next-hop destination for each network and subnet. At that point, routing prefix and adjacency information for lowest-cost paths are populated to the Cisco Express Forwarding tables. As routing-protocol changes occur, the software Cisco Express Forwarding databases are check pointed from the active supervisor to the standby supervisor, and the Cisco Express Forwarding tables are downloaded to the hardware on all PFCs and DFCs present in the system, including the standby PFC. This helps ensure forwarding-table synchronization at the software and hardware level and helps ensure that post switchover data forwarding relies on the most accurate and up-to-date forwarding-table information.

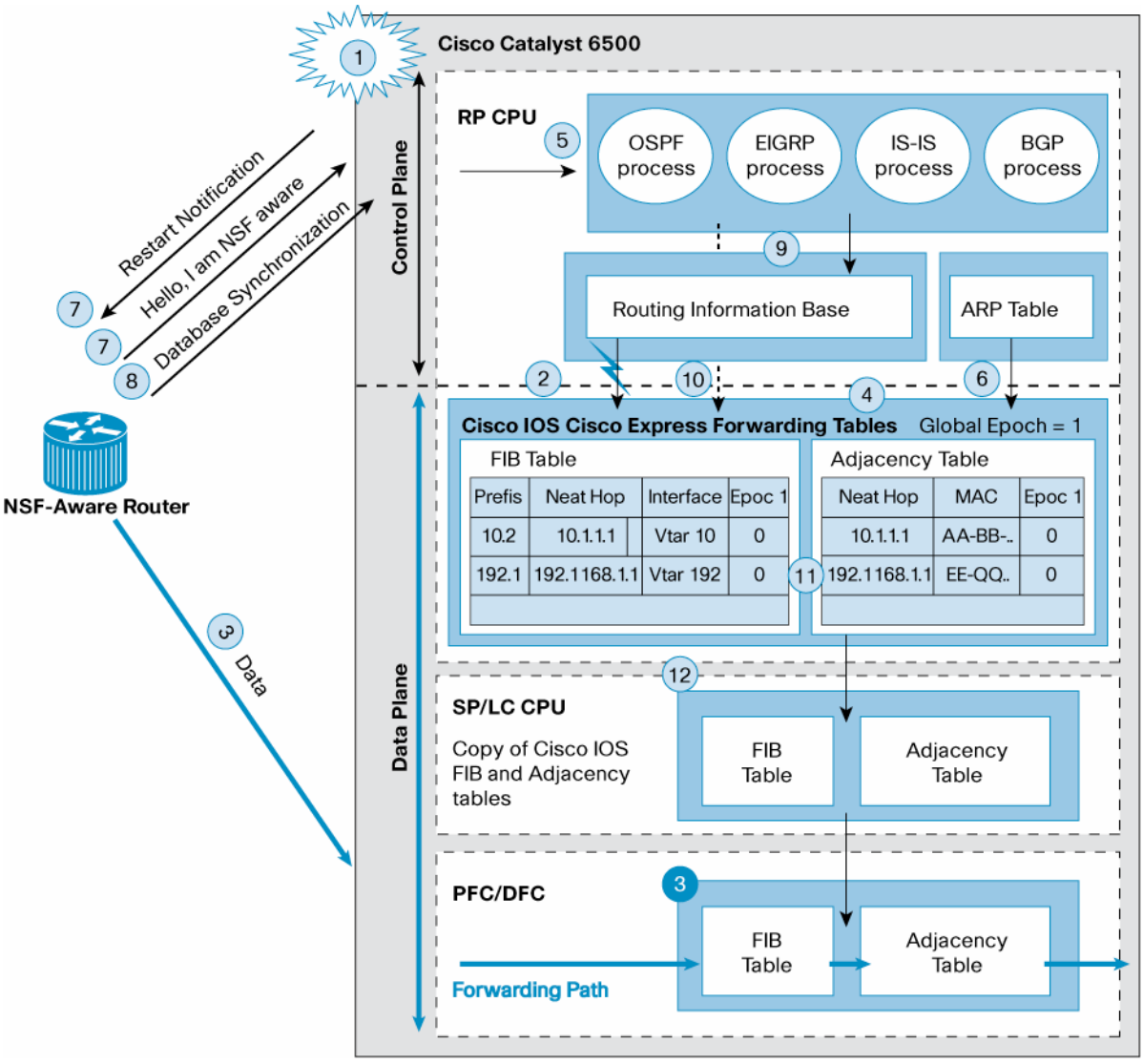
An epoch number per Cisco Express Forwarding entry is introduced in order to allow differentiation between old and new Cisco Express Forwarding entries. This is known as FIB and adjacency database versioning. Only software Cisco Express Forwarding tables keep track of the epoch number, and this version number does not impact the forwarding path. A “global epoch number” is incremented when a switchover occurs. The version number for the Cisco Express Forwarding entries is updated with the global epoch number when new routing information is populated after switchover on the newly active supervisor. When the routing protocols signal that they have converged, all FIB and adjacency entries that have version numbers older than the current epoch are cleared.

Supervisor Switchover Operation

The separation between the data plane and Layer 3 control plane is critical for the correct function of NSF upon switchover. Whereas the control plane builds a new routing protocol database and restarts peering agreements, the data plane relies on preswitchover forwarding-table synchronization to continue forwarding traffic. The following section assumes the presence of an NSF-aware neighbor. Without the help of NSF-aware neighbors, NSF-capable systems cannot rebuild their database nor maintain their neighbor adjacencies across a switchover. (Note that the Cisco Intermediate System-to-Intermediate System [IS-IS] NSF implementation does not require any NSF-aware neighbor.)

The same switchover operations as described in Figure 3 occur. However, reinitialization of the NSF-capable routing protocol does not cause route flaps. Figure 6 describes the generic routing protocol NSF with SSO operations that take place. Figure 7 depicts an NSF-aware neighbor router and an NSF-capable Cisco Catalyst 6500. The Cisco Catalyst 6500 newly active supervisor is represented along with NSF with SSO operation steps. This figure does not represent the failing former active supervisor. Note that the steps applying to the supervisor switch processor (SP) and policy feature card (PFC) apply also to the line-card (LC) processor and DFCs. Black steps are control-plane driven, whereas blue steps are data-driven.

Figure 7. NSF/SSO Operation: NSF-Aware Router and NSF-Capable Cisco Catalyst 6500



The Cisco Catalyst 6500 newly active supervisor is represented along with NSF/SSO operation steps. Figure 7 does not represent the failing former active supervisor. Note that the steps applying to the supervisor switch processor (SP) and policy feature card (PFC) apply also to the linecard (LC) processor and distributed forwarding cards (DFC). Black steps are control plane driven, whereas blue steps are data driven.

Figure 7 steps 1 through 12 are described as follows. All these steps occur on the “newly active” supervisor.

1. Switchover is triggered.
2. Routing-protocol processes are informed of the supervisor failover. In order to provide control- and data-plane separation, the FIB is detached from the RIB until the routing protocol converges.
3. Packet forwarding continues based on last-known FIB and adjacency entries while the standby takes over.
4. The global epoch number is incremented: if the preswitchover global epoch was 0, it is incremented to 1.
5. The supervisor starts processing control-plane traffic.
6. The software adjacency table is populated with the preswitchover Address Resolution Protocol (ARP) table contents. Updated Cisco Express Forwarding entries receive the new global epoch number. The epoch number is available only in the route processor software Cisco Express Forwarding entries. It is not present in the hardware table. New adjacency entries are downloaded to the hardware.
7. The routing protocol-specific neighbor and adjacency reacquisition occurs: the restarting NSF-capable router notifies its neighbor that the adjacency is being reacquired and that the NSF-aware neighbor should not reinitialize the neighbor relationship. Upon receiving the restart indication, protocol-specific procedures occur to allow adjacencies to be maintained. In most cases, the restart indication consists of setting a restart flag in hello packets and sending hello packets at a shorter interval for the duration of the recovery process. NSF-aware neighbors might also indicate their NSF awareness to restarting routers. Non-NSF-aware neighbors ignore the restart indication and bring down the adjacency. Note also that the current NSF implementation does not support multiple NSF-capable neighbor restarts at once.
8. The routing protocol-specific database synchronization occurs: routing protocol processes rebuild their database using database information from NSF-aware neighbors.
9. When the routing databases are synchronized, distance-vector, path-vector, or shortest-path-first (SPF) algorithm computations determine the best route for specific prefix destinations. The RIB is repopulated with new routing entries. The corresponding Cisco Express Forwarding entries are updated.
10. As the software Cisco Express Forwarding databases are populated with updated information, updated entries receive the global epoch number to indicate that they have been refreshed. Corresponding FIB entries and hardware entries are updated.
11. Each routing protocol notifies Cisco Express Forwarding that it has converged. After all of them have converged, the last routing protocol flushes the stale route and adjacency information: software Cisco Express Forwarding entries with an epoch number not corresponding to the current global epoch number are flushed. Corresponding FIB and adjacency hardware entries are also flushed.
12. The Cisco IOS Software Cisco Express Forwarding tables on the route processor and the forwarding tables on the switch processor and PFC and DFCs are now synchronized.

Software Upgrades

Beginning with Cisco IOS Software Release 12.2(33)SXI the Cisco Catalyst 6500 supports a new software upgrade procedure called Enhanced Fast Software Upgrade, or EFSU. The EFSU process uses redundant Supervisor technologies to streamline the process of performing software version updates. The new EFSU feature supports full image Cisco IOS Software upgrades and downgrades as well as the ability to activate Maintenance Pack updates for Cisco IOS Software Modularity. The EFSU procedure is supported in standalone chassis configurations with redundant Supervisors and in the Virtual Switching System.

The key new technology within the EFSU feature is the In Service Software Upgrade Versioning Infrastructure, or ISSU versioning infrastructure. The ISSU versioning infrastructure provides a framework for the active and standby supervisor modules to establish SSO redundancy mode while running different software revisions. Prior to the availability of the ISSU versioning infrastructure, two supervisor modules would have to run the same software version in order to establish the SSO redundancy mode.

As discussed in the SSO section of this paper, there are various software processes and applications that utilize the Redundancy Facility and the Check Pointing Facility to synchronize application states and data between two peer endpoints (typically the endpoints are two Supervisor modules, but they could also be IO Modules running a Cisco IOS Software image). Software process and applications which use the Redundancy Facility and Checkpoint Facility are referred to as SSO-aware applications.

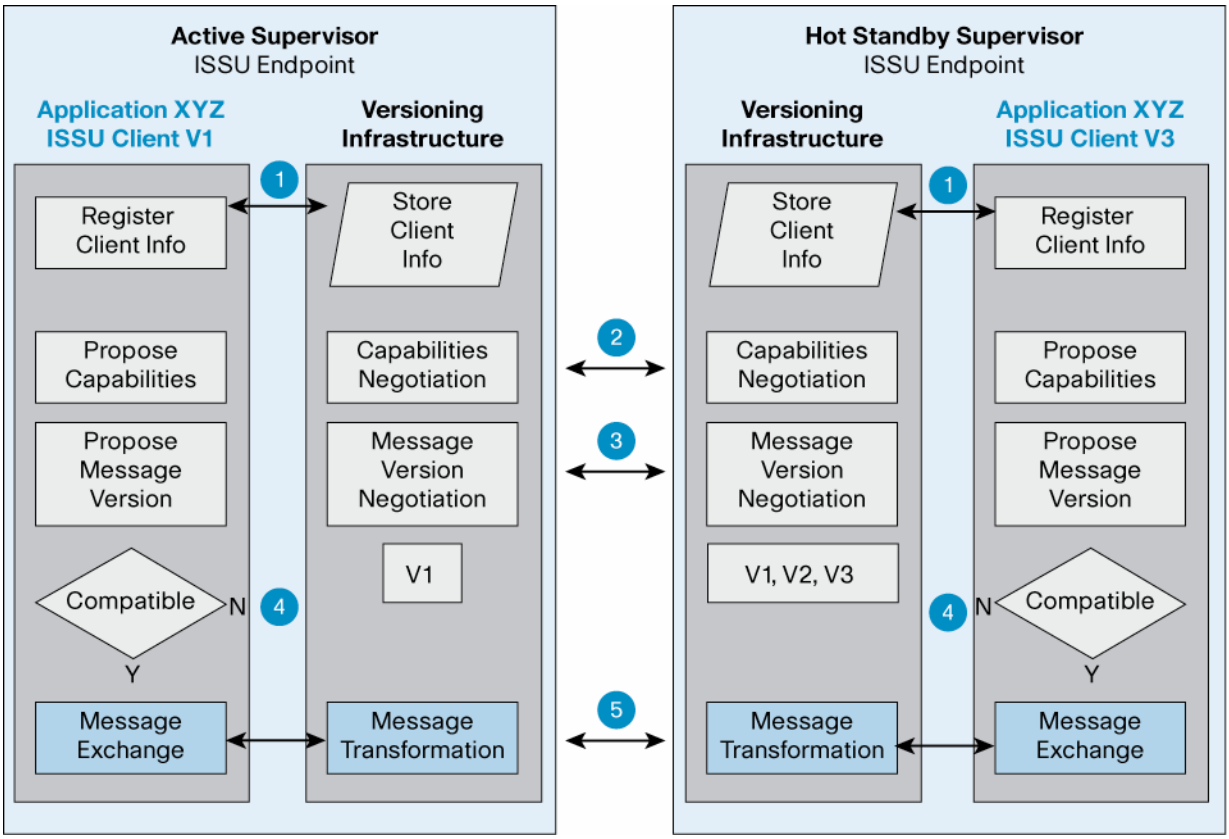
When two Supervisor modules are running the same software version the peer SSO-aware applications perform synchronization with a known set of message types and capabilities. When two different software versions are involved the SSO synchronization process is complicated by the fact the peer applications may have different feature capabilities hence different message types and data structures. The ISSU versioning infrastructure was developed to address this challenge.

The ISSU versioning infrastructure provides the following services between two different Cisco IOS Software endpoints:

- Verifys the compatibility level between two software versions or image files. This is done by validating a compatibility matrix to set the correct redundancy mode for the two versions involved with the upgrade. If the two software versions are compatible then the upgrade procedure will proceed using the SSO mode, if the versions are deemed incompatible then the procedure can continue using RPR mode.
- Provides the infrastructure for SSO-aware Cisco IOS Software applications to register the messages they support on a given session also which versions of these messages they support.
- Provides the infrastructure for the Clients to register the transformation functions (a complete set of upgrade and downgrade functions is required)

Each SSO-aware application becomes an ISSU client and uses the ISSU infrastructure to synchronize information with its peer application on another Cisco IOS Software endpoint. A high level depiction of the ISSU client interaction with the ISSU versioning infrastructure is provided in Figure 8.

Figure 8. ISSU Versioning Infrastructure



ISSU client interaction between two Supervisor modules:

1. The ISSU client application registers with the ISSU versioning infrastructure with key information to uniquely identify the application
2. The ISSU endpoints communicate with their peer application and negotiate common capabilities including message types and message lengths
3. The endpoints communicate the different message version they are capable of supporting
4. Each endpoint compares the information exchanged to determine if they are capable of supporting a common message format
5. If the two ISSU clients agree to a common message format for the given application then the applications can proceed with message exchange to synchronize the application states and data. If the applications are not compatible then synchronization will not occur.

The ISSU client interaction depicted in Figure 8 occurs individually for all the SSO-aware applications.

Certain SSO-aware applications are deemed SSO base level applications meaning these applications are critical and required to be compatible in order for the two endpoints to reach SSO redundancy mode. Other applications may be SSO-aware but compatibility may not be required for the endpoints to reach SSO redundancy mode, these applications are deemed non-base-level applications.

With the designation of SSO base-level applications and non-base-level applications. A given software release can be described as fully compatible, base-level only compatible or incompatible in relation to another software release. If all of the SSO-aware applications are compliant then the image is deemed fully compatible. If one or more non-base-level applications are not compliant then the image is deemed base-level only compatible. Finally if one or

more base-level applications are not compliant then the image would not support SSO redundancy mode and would therefore be deemed incompatible.

Overtime as more and more ISSU capable software versions are released the number of combinations for possible software upgrades also increases. Cisco is committing to provide ISSU compatibility, where ISSU compatibility is possible (theoretically certain applications may change functionality to the extent that message transformation will not be feasible), for ISSU base-level applications in software releases occurring within an 18 month time frame.

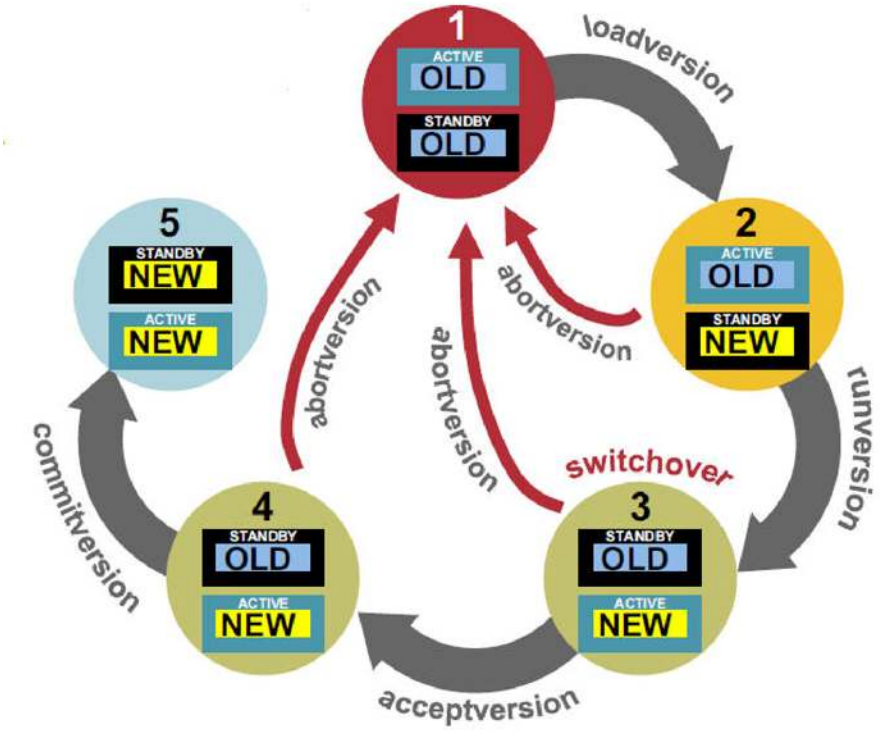
For example, consider an EFSU upgrade between version N and version N+3; as long as the release dates for the two software versions occur within an 18 month time frame then the images should be at a minimum base level compatible.

It is important to note that in the event two software images are ISSU incompatible; the EFSU upgrade process can still be used. In this case, instead of using SSO mode, the upgrade will continue using RPR mode.

EFSU Upgrade Process

In addition to the ISSU versioning infrastructure, the upgrade process itself is simplified into 5 distinct stages with 4 upgrade steps providing minimal data disruption with less complexity. The EFSU upgrade process uses the same ISSU CLI syntax as the ISSU process supported on other Cisco switching and routing platforms including the Cisco Catalyst 4500 platform.

Figure 9. ISSU Upgrade Process



The major steps associated with an EFSU upgrade are illustrated Figure 9 and described below:

1. **ISSU Loadversion**: This step loads the new image onto the standby supervisor module and, if supported, will predownload the new software image for an IO module into the available memory of the IO module. The Standby Supervisor will use the ISSU versioning infrastructure to reach the SSO redundancy mode.
2. **ISSU Runversion:** Here the ISSU process initiates a Supervisor switchover using either SSO redundancy mode or RPR redundancy depending up on the ISSU compatibility between the two versions. The switchover is initiated by the previously active Supervisor reload, the Standby moves to the newly Active role. As the newly active Supervisor takes over the IO modules must also reload to run the same image as the newly active Supervisor. The downtime associated with the reload of the IO module is discussed further in this section.
3. **ISSU Acceptversion:** This is an optional step that can be used to stop the automatic rollback timer which begins counting to zero at the ISSU runversion step. If the timer reaches zero without an ISSU acceptversion or ISSU commitversion command being executed, the ISSU process will initiate an abortversion sequence which will revert the system back to software versions in place before the upgrade process began. The rollback timer is a safeguard incase a problem occurs during the runversion stage where connectivity is lost to the switch. By default the rollback timer starts counting backwards from 45 minutes, the timer is configurable.

   a. At this point the newly active Supervisor and IO modules are running the new software image, while the previously active Supervisor remains configured to load the old software image and will reach SSO redundancy mode provided the images are ISSU compatible. The user can then verify certain features and functionality in the new software image until they are satisfied they want keep this new image. Any functionality unique to the new image however, will not be enabled at this point. The user must finalize the upgrade process first before any unique features are active.

   b. Once the ISSU acceptversion command is entered, deactivating the rollback timer, the system can remain in this state indefinitely to verify functionality.
4. **ISSU commitversion:** With this command the standby supervisor is configured to load the new image and then reloads with the new image. The Supervisor will initialize itself as the standby once again since the peer Supervisor is already in the active role.
5. **ISSU abortversion:** This is an optional step available at any time between the loadversion and commitversion steps, where the system will revert back to the original software image and exit the ISSU process.

The downtime experienced by user traffic during an EFSU upgrade differs somewhat depending upon whether the process occurs on a standalone Cisco Catalyst 6500 chassis versus a Virtual Switching System pair of chassis, versus applying a Maintenance Pack to a system running Cisco IOS Software Modularity. The reason for the differences is primarily due to how the IO module software is updated, if updated at all, during the EFSU process. To understand this further each scenario will be discussed individually.

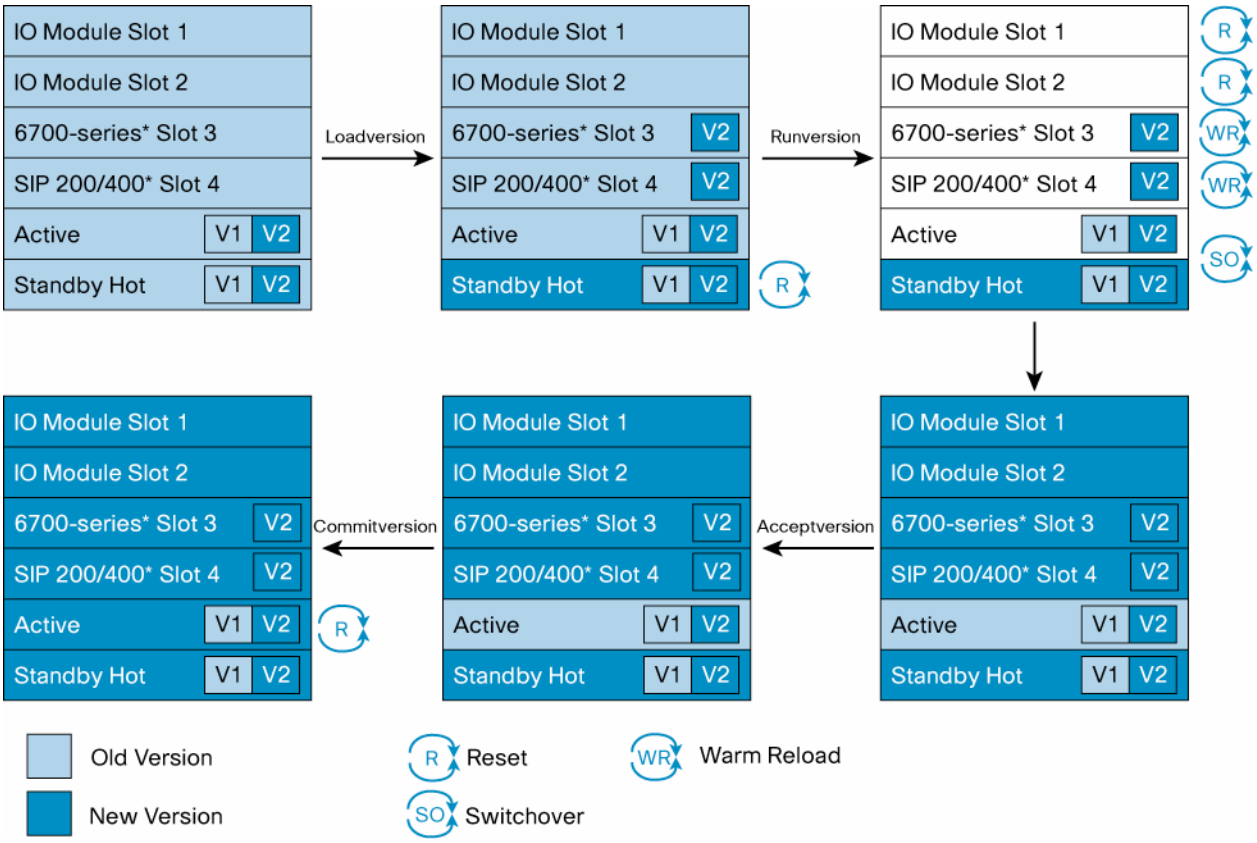
Enhanced Fast Software Upgrade for Standalone Cisco Catalyst 6500 Chassis

Performing a full image EFSU upgrade on a standalone chassis not only updates the software image for the Supervisor module, but this can also mean software on the IO modules will need to be updated as well. IO Modules may run their own software image in the form of a Cisco IOS Software image, a firmware image or even both.

The IO module software image is embedded within the Supervisor module's Cisco IOS Software image. Each time an IO module boots, it downloads its software image from the Cisco IOS Software image running on the active Supervisor module. Therefore the EFSU process must include the software update on the Supervisor module as well as the IO modules.

In order for an IO module to load an new software image the IO module must perform a reload, therefore when the system's active Supervisor switches-over during the ISSU runversion phase, the IO modules will need to reload and download the new IO module software image from the newly active Supervisor. Therefore user data traffic is interrupted while the IO modules reload. The EFSU upgrade process is illustrated in Figure 10 and shows the status of the different IO modules and Supervisor modules as the EFSU process occurs.

Figure 10. EFSU Upgrade Process for a Standalone Chassis



The downtime associated with the IO module reload is affected by multiple factors including the file size of the IO module software, the diagnostics level configured for the IO module, and the interface specific configurations applied. Certain IO modules can pre-download their software image during the ISSU Loadversion stage; this would in turn reduce the downtime associated with the reload during the ISSU Runversion stage.

The pre-download functionality is supported only with 6700-series IO modules or the SIP 200, SIP 400 and SIP 600 WAN modules. In addition, the module must have at least 256MB of available memory. When an IO module is capable of pre-downloading its software image, the associated reload is called a warm-reload as depicted in Figure 10.

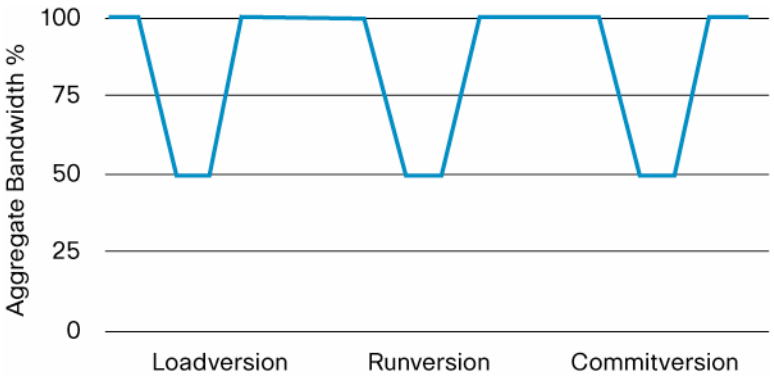
Enhanced Fast Software Upgrade for the Cisco Virtual Switching System

Cisco Virtual Switching Systems (VSS), performing full image software updates using the EFSU process will experience significant improvements over the previous upgrade process. Not only is the process simplified using the ISSU CLI but the outage time is reduced to sub-second levels for dual homed connected devices.

The primary difference between performing an EFSU upgrade on a standalone system versus a VSS is that with VSS the IO modules themselves become ISSU endpoints, messages are transformed between the IO module and the active Supervisor using the ISSU versioning infrastructure, allowing the IO modules to run a different version of software version.

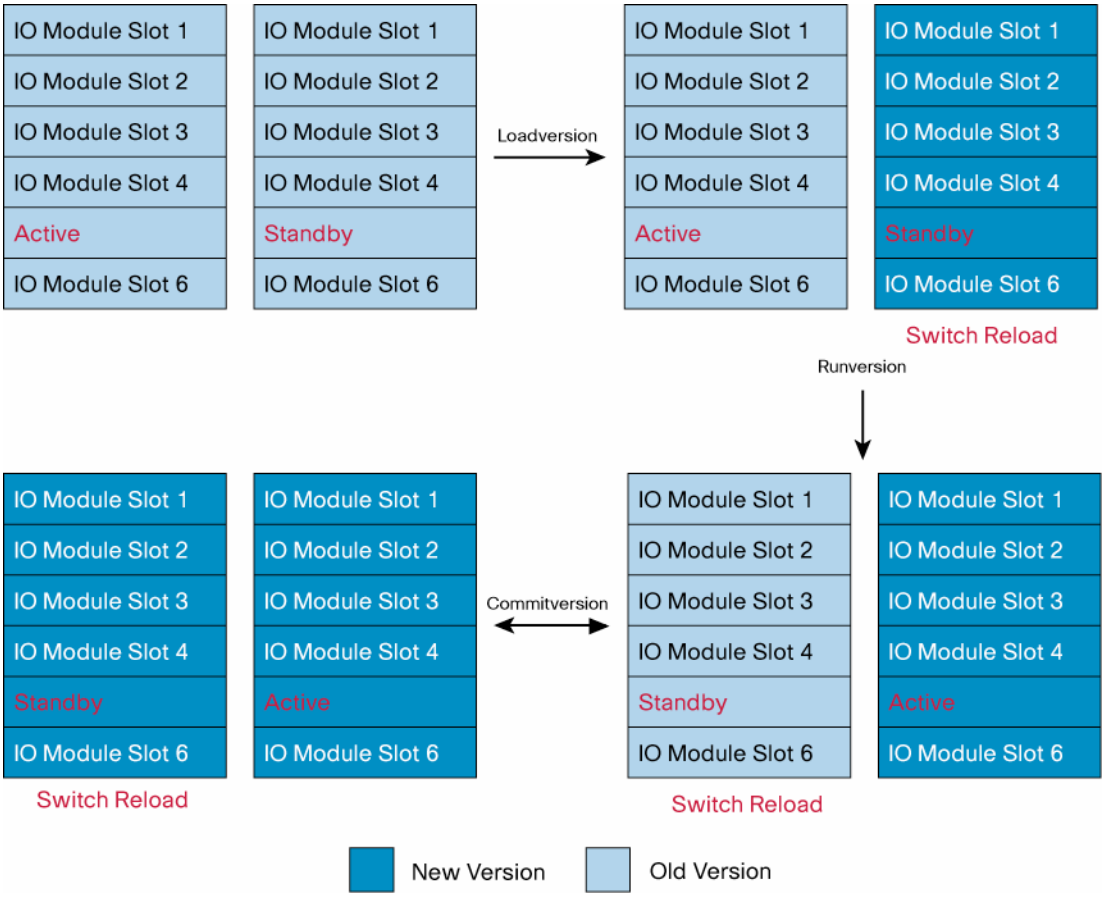
The EFSU process also manages the IO module reloads so that at any given time only the IO modules from one of the two physical chassis is ever reloading, This helps ensure that devices connected to the VSS in a dual-homed fashion maintain connectivity to the network. In other words the overall capacity of the VSS may decrease by 50 percent as one of the two physical chassis performs a reload, but network connectivity is always maintained.

Figure 11. VSS Aggregate Bandwidth During an EFSU



A brief traffic outage may occur as traffic flows are moved off the chassis performing the reload. The transference of these flows occurs on the VSS and in the device connected to the VSS, be it with Etherchannel technology or with layer 3 routed interfaces. This switchover time is typically less than 200 milliseconds. Further enhancements are planned that will allow devices connected via Etherchannel to reduce this switchover time to zero using a graceful switchover mechanism supported with the Cisco Port Aggregation Protocol.

Figure 12. EFSU Process for a Virtual Switching System



In- Service Patching with EFSU

Cisco Catalyst 6500 systems running Cisco IOS Software Modularity now uses EFSU to apply Maintenance Packs. Since its inception Cisco IOS Software Modularity has provided an Install utility to install and activate Maintenance Packs; however these are applied individually per Supervisor module. With EFSU In-Service Patching, redundant Supervisor technologies can now be used to minimize downtime and streamline the activation process.

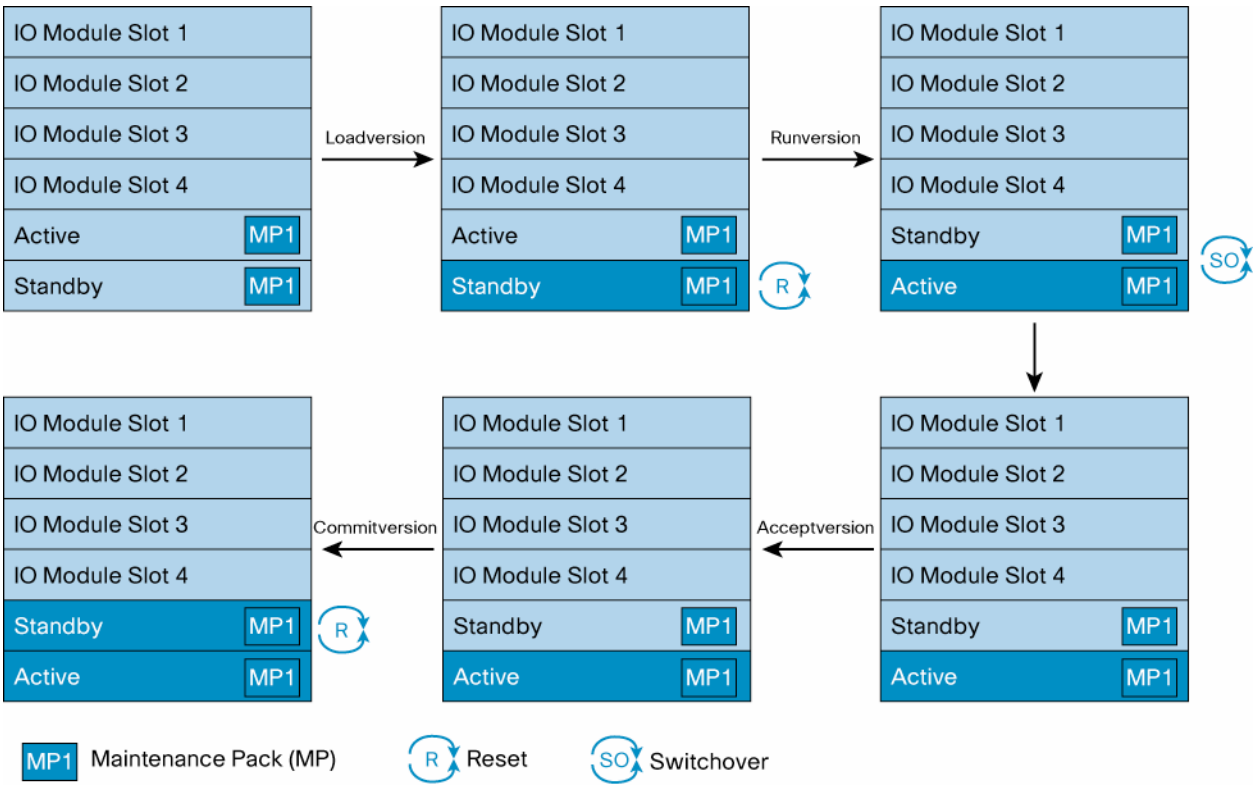
Applying a Maintenance Pack (MP) with Cisco IOS Software Modularity is a primarily a two step process, first installing the Maintenance Pack and second activating the Maintenance Pack. With In-Service patching the MP is still installed using the Install utility. The Install utility performs verification and compatibility checks between the MP and the base software version and also copies the specific patch files into the appropriate location in the file system. The MP activation step is where the EFSU process comes in.

Once the MP has been installed successfully using the Install utility, the MP still needs to be activated. The activation process will restart the appropriate subsystem or subsystems and load the software provided by the MP. Using the EFSU process one can simply specify the target directory where the MP has been installed. Using the EFSU process, redundant Supervisor technologies can be use to activate the MP with minimal downtime.

The In-Service Patching process with EFSU occurs much the same way as an EFSU full image upgrade, the same ISSU CLI still applies. The primary difference is that in this case only the specific subfiles associated with the MP are being updated, not the entire image.

Consider a scenario where the MP contains a very specific fix that is applied to only a single subsystem. In this case if the subsystem did not affect any software on the IO modules, then the MP could be activated without requiring any of the IO modules to reload. In affect only the Supervisor modules would be affected by the MP activation. With the ISSU versioning infrastructure in place the Supervisor modules can still achieve SSO redundancy mode even with different MP levels, thus allow for minimal impact during the upgrade process. Figure 13 illustrates the In-Service Patching process.

Figure 13. EFSU Process for In-Service Patching



Supervisor Fault Detection

Redundant Supervisor modules provide an obvious benefit over nonredundant configurations. A critical part of this solution is to quickly and intelligently determine when to switchover to the redundant Supervisor. It's one thing to detect when the active Supervisor fails completely and severely, but what about if only a certain subsystem failed and the Supervisor remains active. The more intelligent way to handle these types of event is to detect them during runtime and switchover to the healthy redundant Supervisor. Detecting hardware and software faults is function of the Generic Online Diagnostics (GOLD) feature in the Cisco Catalyst 6500.

GOLD defines a common architecture for diagnostic operation on Cisco platforms. GOLD works together with platform-specific online diagnostics to help ensure that a system booting up and a live system are healthy. Given that most of the intelligence of a Cisco Catalyst 6500 is hardware-based, it is very important to make sure the hardware functions are tested regularly. Fault-detection diagnostics mechanisms are enabled on most modules in a Cisco Catalyst 6500 system, including the active and standby supervisors. Diagnostic test results can be used to make switchover decisions. With online diagnostics being integrated on the Cisco Catalyst 6500, switchover triggers are not limited to software crashes or keep-alive mechanisms. Instead, switchovers can be triggered when the supervisor control and data paths are inconsistent or faulty, or when runtime diagnostics detect a malfunctioning piece of hardware. In addition to helping trigger switchover decisions, GOLD regularly monitors the standby supervisor to make sure that it is ready to take over if the need to switchover occurs. GOLD also integrates a feature that allows scheduling of switchovers: an administrator can schedule a switchover at a specific time through an online diagnostics command-line interface (CLI).

GOLD detects the following problems to make supervisor switchover decisions:

- Faulty hardware components including Application Specific Integrated Circuits (ASICs)
- Faulty connectors

- Failed interfaces
- Memory errors
- Inconsistencies between the data plane and the control plane

Simple Network Management Protocol (SNMP) and statistics

Statistics

The various statistics maintained by an active supervisor are not synchronized to the redundant supervisor because they may change often and the degree of synchronization they require is substantial. A network-management system should be used to poll affected statistics regularly to maintain accurate statistics.

SNMP

Simple Network Management Protocol (SNMP) data is synchronized between redundant supervisors when the supervisor is operating in SSO mode. This is done to make sure that the standby and the active supervisor are indistinguishable from a network-management perspective. Some of the SNMP objects that are synchronized include interface-related features such as ifindex and SNMP configuration.

The Cisco High-Availability MIB, CISCO-RF_MIB, reports redundancy information to an administrator. This information includes identification of the primary and secondary supervisors, current redundancy state, the reason for the last switchover that occurred, and when the last switchover occurred. When a switchover occurs, the ciscoRFSwactNotif notification is used to signal a switchover.

In addition to using the Cisco High-Availability MIB, syslog messages and SNMP traps are sent to notify the administrator of any component failure.

SNMP data synchronization is not available in the RPR and RPR+ modes of operations.

For more information about SSO SNMP support, refer to:
http://www.cisco.com/en/US/docs/ios/12_0s/feature/guide/ssomibs3.html

Supervisor SSO support with Services Modules

It is important for service modules to continue working through an NSF with SSO supervisor failover event. Many of the service modules have specific high-availability mechanisms in place today to allow intrachassis or interchassis module-to-module switchover. Supervisor NSF with SSO support with services modules complements the high-availability mechanism of each of these services modules by minimizing the impact of a supervisor failover.

Each of the properties pertaining to SSO with standard switching modules holds true for services modules: the services modules do not reboot, the services modules interfaces stay up, and the service modules are not affected by a supervisor switchover except for the short period corresponding to line-card synchronization.

Supervisor SSO support with WAN Modules

Optical services modules (OSMs) and FlexWAN modules are supported with redundant supervisor engines and continue working through an NSF with SSO supervisor failover event.

References

Cisco Catalyst 6500 High Availability

Configuring supervisor-engine redundancy using NSF with SSO:
http://www.cisco.com/en/US/partner/docs/switches/lan/catalyst6500/ios/12.2SX/configuration/guide/nsfsso.html

Release notes for Cisco IOS Software Release 12.2SX:
http://www.cisco.com/en/US/partner/products/hw/switches/ps708/prod_release_notes_list.html

Stateful Switchover

SNMP SSO: http://www.cisco.com/en/US/docs/ios/12_0s/feature/guide/ssomibs3.html

Nonstop Forwarding with Stateful Switchover

NSF: http://www.cisco.com/en/US/docs/ios/12_2s/feature/guide/fsnsf20s.html

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-532835-00 04/09