Introduction to HA Technologies:

SSO/NSF with GR and/or NSR.

Ken Weissner / kweissne@cisco.com

Systems and Technology Architecture, Cisco Systems

That’s a lot of acronyms

Some definitions

- **HA** - High Availability

  High level terminology

- **SSO** - Stateful Switchover

  An operating mode where a dual processor router has transferred state information to a standby processor to allow the standby to pickup necessary router functions in the event of an active failure. Mostly refers to L2 information (PPP state, FIB ect) but some L3 applicability. In this operating mode both processors must run identical software versions.

- **NSF** - Non Stop Forwarding

  NSF refers to a routers ability to almost immediately start forwarding packets following an active processor failure. The FIB (Forwarding Information Base) is initially transferred and actively updated so that when a failure occurs, the router is able to forward packets while the control plane is rebuilt or refreshed.

That’s a lot of acronyms (cont)

- **GR** - Graceful Restart

  IETF specified mechanisms for interaction between routing protocol peers which allow the peer of a failing device to continue forwarding packets to that device, even though the neighbor relationship has been destroyed.

- **NSR** – Non Stop Routing

  A routing protocol operating mode where all information needed to fully maintain the neighbor relationship and all its relevant routing information is transferred (or “checkpointed”) to the standby processor. No additional communication or interaction with the routing protocol peer is needed in this mode.

  Some implementations allow the use of both GR and NSR for the same protocol, but single routing protocol session must be either GR or NSR.

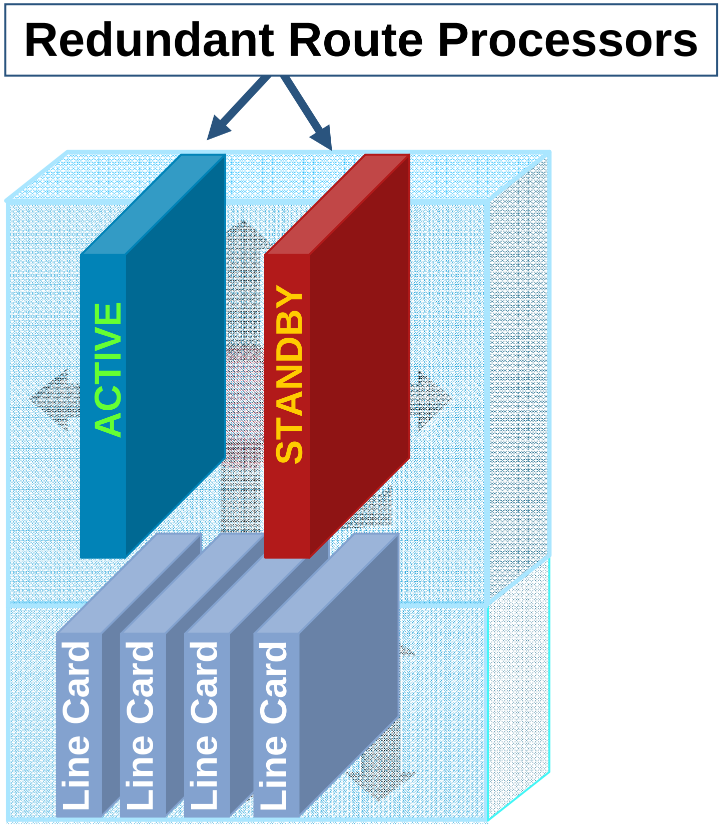
All of the above technologies combine to allow for an unplanned processor switchover to occur with very little interruption in availability.

- **ISSU** – In Service Software Upgrade

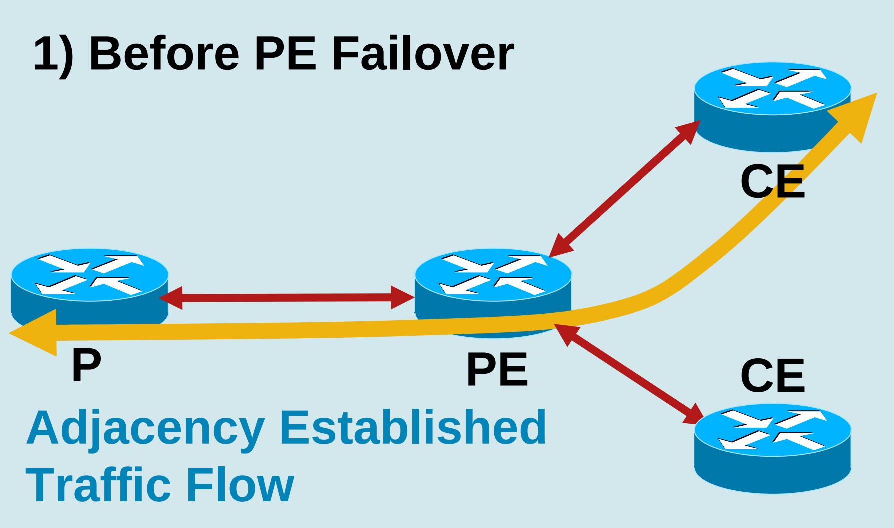
  A process which allows the complete upgrade of a dual processor router. This process uses the technologies of SSO/NSF with GR/NSR with the added functionality of a stateful switchover between different images.

Why SSO/NSF?

- Protection for CE devices connected to a single PE device
- Software or hardware failures induce processor switchover to maintain router availability
- Framework and infrastructure for supporting In Service Software Upgrades (ISSU)
- Leverages distributed nature of router

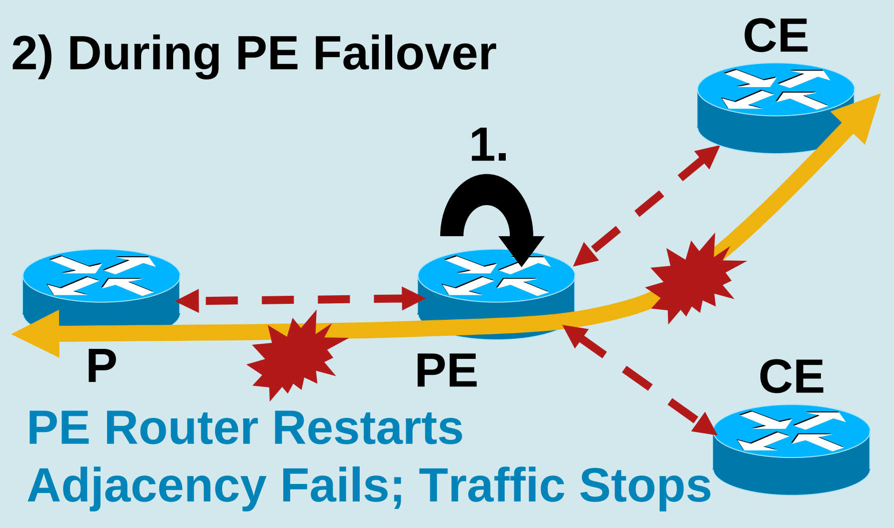


Networks Without NSF/SSO and Graceful Restart



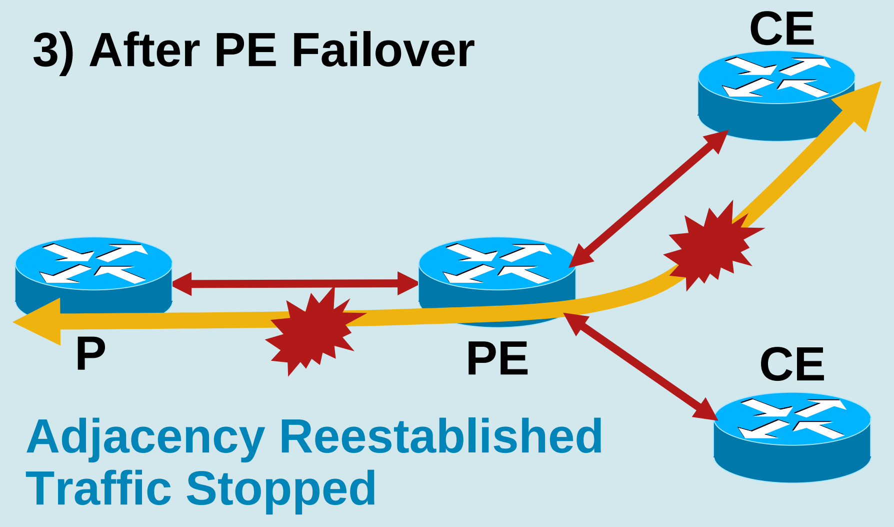
1) Before PE Failover

Adjacency Established
Traffic Flow



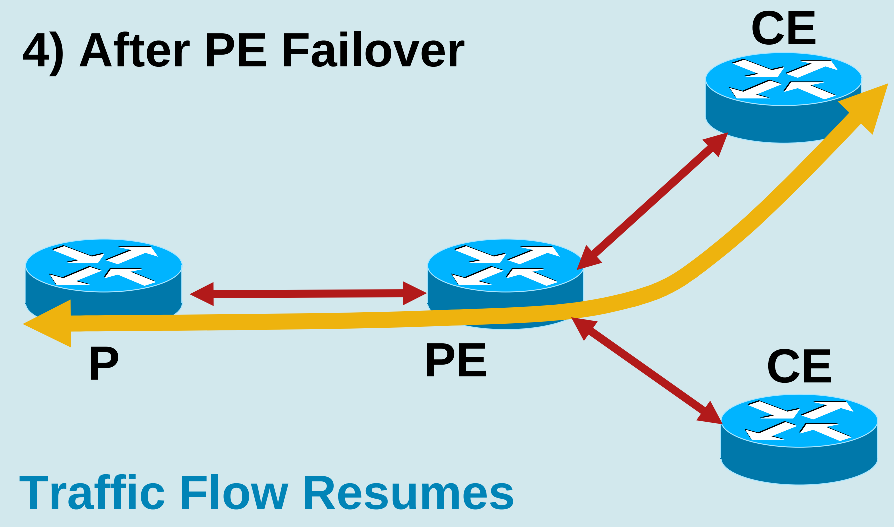
2) During PE Failover

PE Router Restarts
Adjacency Fails; Traffic Stops



3) After PE Failover

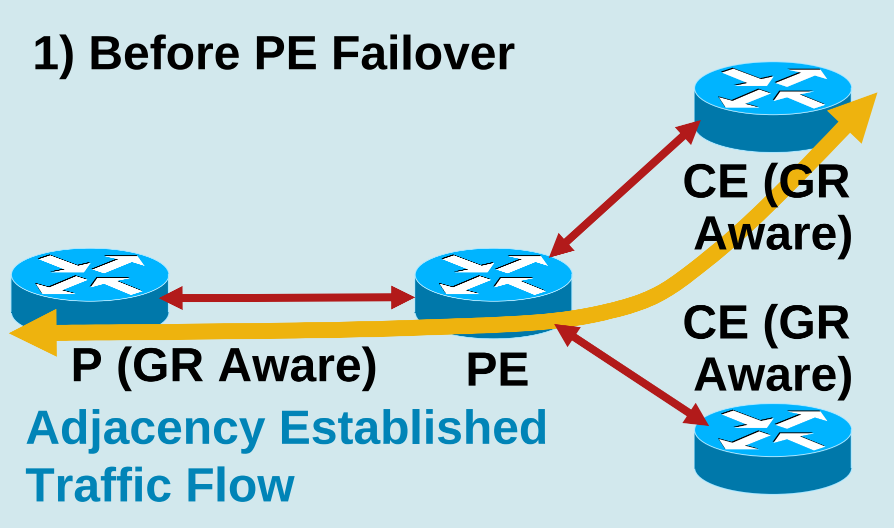
Adjacency Reestablished
Traffic Stopped



4) After PE Failover

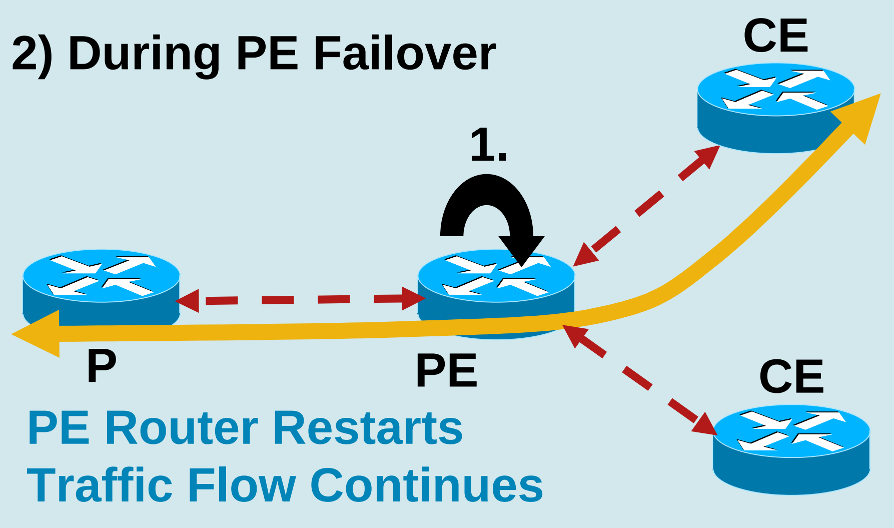
Traffic Flow Resumes

Networks with NSF/SSO and Graceful Restart



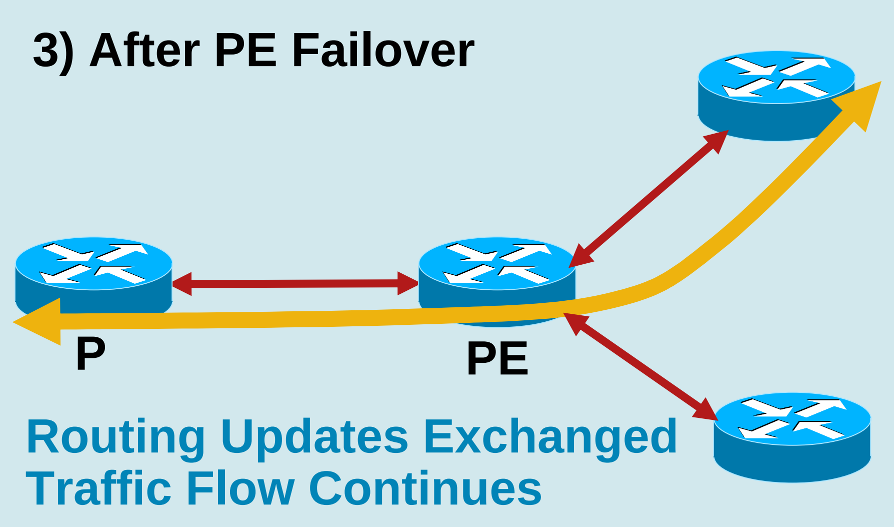
1) Before PE Failover

Adjacency Established
Traffic Flow



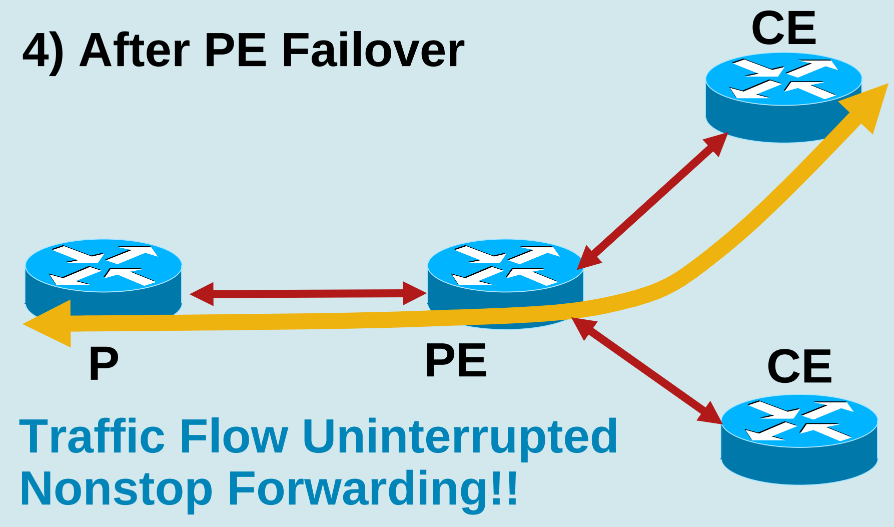
2) During PE Failover

PE Router Restarts
Traffic Flow Continues



3) After PE Failover

Routing Updates Exchanged
Traffic Flow Continues



4) After PE Failover

Traffic Flow Uninterrupted
Nonstop Forwarding!!

Graceful Restart IETF status

GR is the only feature that interacts with peer network devices, all other features (SSO/NSF/NSR) are internal to the router and therefore don’t require standards.

- Graceful Restart Mechanism for Label Distribution Protocol RFC 3478
- Graceful Restart Mechanism for BGP RFC 4724
- Restart Signaling for Intermediate System to Intermediate System (IS-IS) RFC 3847
- Graceful OSPF Restart RFC 3623
- draft-nguyen-ospf-restart-06

Graceful Restart Protocol Support

- RFC based support for BGP, OSPF, ISIS and LDP
- Additional support for EIGRP and draft-nguyen-ospf-restart-06 based OSPF GR.
- Two versions of OSPF GR are supported. Draft-nguyen-ospf-restart-06 was implemented prior to the existence of RFC3623, and is widely deployed in networks today.
- A router participating in GR is said to be “aware” or “capable”, with awareness being a subset of capability.

Graceful Restart Awareness and Capability

- A Graceful Restart capable device will announce its ability to perform graceful restart to the routing protocol peer. It will also initiate the Graceful Restart process when a route processor transition occurs and act as a graceful restart aware device.
- A Graceful Restart aware device has the components to be able to understand a peer router is transitioning, and will take the appropriate action when it detects the peer router is performing Graceful Restart (start timers, routes in holddown, ect)

  Awareness is also referred to running in “helper” mode

Graceful Restart Awareness and Capability Configuration

- Graceful Restart capability must always be enabled for all protocols. This is only necessary on routers with dual processors that will be performing switchovers.
- Graceful Restart awareness is on by default for non-TCP based interior routing protocols (OSPF,ISIS and EIGRP).   These protocols will start operating in GR mode as soon as one side is configured capable.
- TCP based protocols must enable GR on both sides of the session and the session must be reset to enable GR. The information enabling GR is sent in the Open message for these protocols.

Graceful Restart Concerns

Voiced at Nanog40 peering BOF

“With regards to BGP graceful restart, the problem we’ve seen with implementing it is that Cisco’s implementation of graceful restart assumes you have NSF (non-stop forwarding), and then tells your peers, “if I ever drop this BGP session, it’s because Im failing over from the primary to redundant supervisor, and will keep passing packets, so keep sending them my way”. That’s all well and good if that’s what actually happens. However, if the router really does go down, then the neighboring router continues sending traffic its was (blackholing it) for many minutes, rather than simply failing over to a working path.”

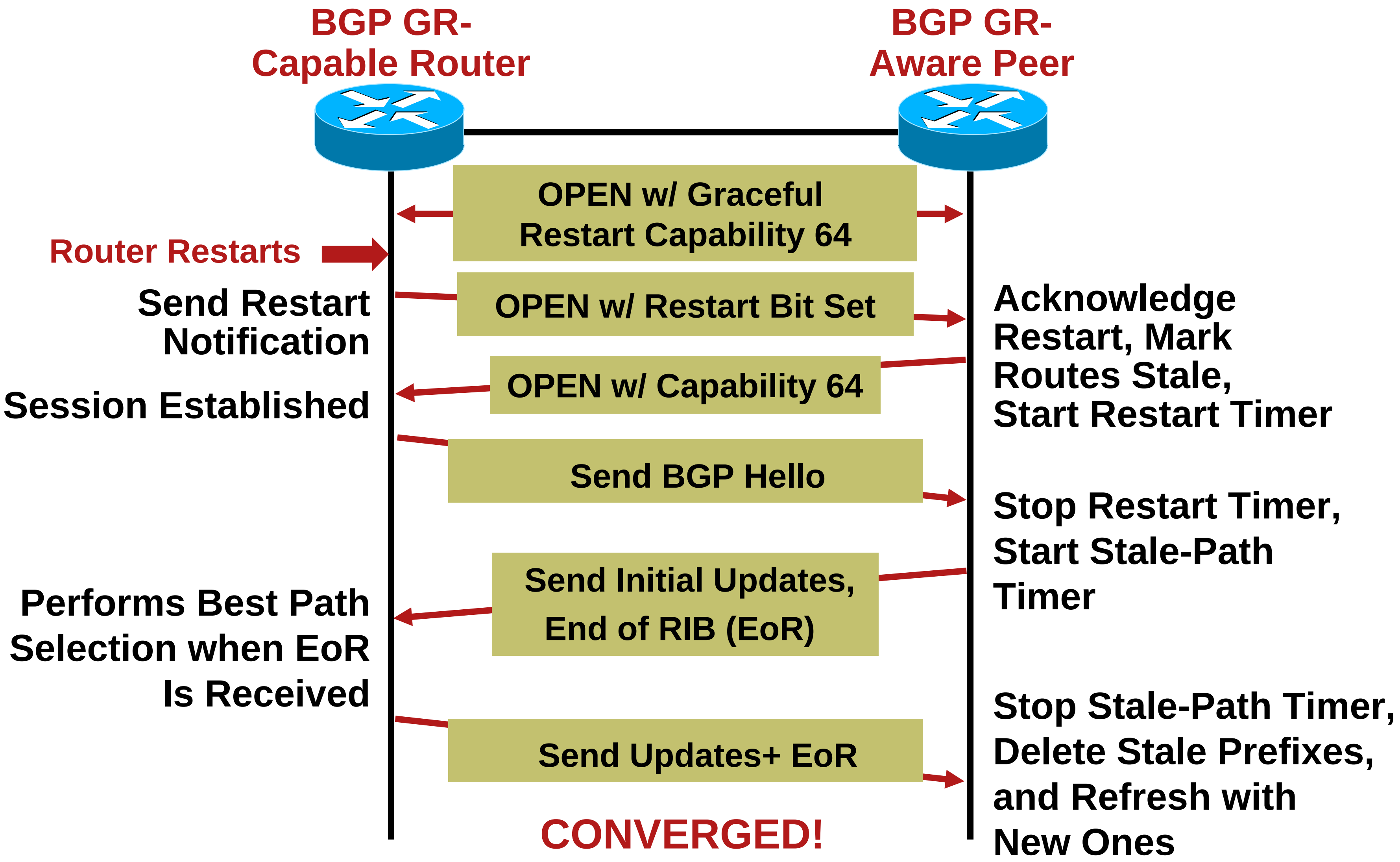
Graceful Restart Concerns Addressed

- Determining between the peer switching over and the peer going away (power off, reload) is key to deploying.
- NSF is not configurable, it is enabled by default when SSO is configured on the router. NSF is a function of checkpointing the FIB to the standby.
- Early on use of the term NSF by cisco (predating the generally accepted term of Graceful Restart) can cause confusion. GR and NSF are two very different functions.
- Today in IOS, to enable OSPF or EIGRP GR, the command “NSF” under the routing process is used, while other protocols (BGP and LDP) use more appropriate variants of the term “graceful restart”.

Graceful Restart Concerns Addressed

- For BGP GR, there is a “restart” timer which gives a window for which the initiation of GR is allowed to happen. If the BGP peer does not come back up within this time, the GR is aborted and all stale routes are flushed. The default is 120 seconds, but given network conditions and test, this number can be reduced to reduce black holing for non-switchover events.
- Other conditions will abort GR as well. If the link is POS point to point, and the peer router reloads, the interface will go down and GR will abort.

Graceful Restart BGP Operation Summary



BGP GR-
Capable Router

BGP GR-
Aware Peer

OPEN w/ Graceful
Restart Capability 64

Router Restarts

Send Restart
Notification

OPEN w/ Restart Bit Set

Acknowledge
Restart, Mark
Routes Stale,
Start Restart Timer

OPEN w/ Capability 64

Session Established

Send BGP Hello

Stop Restart Timer,
Start Stale-Path
Timer

Send Initial Updates,
End of RIB (EoR)

Performs Best Path
Selection when EoR
Is Received

Send Updates+ EoR

Stop Stale-Path Timer,
Delete Stale Prefixes,
and Refresh with
New Ones

CONVERGED!

RFC3623 vs. draft-nguyen-ospf-restart-06 GR

RFC 3623

- Uses “Grace LSA” to signal capability
- Uses LSDB resync as defined in RFC2328
- Terminated if a routing topology chance occurs during GR (configurable)
- GR terminated if there is one or more GR unaware peers on a broadcast domain

draft-nguyen-ospf-restart-06

- Uses a “Restart-Signaling” bit in the LLS fields of hello packets
- Uses an “Out of band” LSDB resynchronization
- GR process uninterrupted until complete.
- GR continues even if unaware peers (configurable)

Two very similar ways of accomplishing the same thing

Non-Stop Routing

- Cisco currently supports NSR for ISIS and BGP in IOS
- “In box” solution that required no additional communication with routing protocol peer.
- For interior protocols, NSR acts on the entire routing process. For BGP, NSR is enabled on a per-peer basis.
- Increases workload on router due to checkpointing routing information in addition to forwarding information to the standby processor.
- “Hybrid” solution helps with scalability of BGP NSR deployments

Hybrid BGP NSR

- The most compelling reason to run NSR on BGP sessions is to avoid having to worry if the peer has GR enabled or not. Perhaps the peer has older code, or even has the right code but does not have GR enabled.
- Relatively, a CE device will have much less routes than a route-reflector. The hybrid solution recommends running GR to the route-reflectors and NSR to non-GR capable CE devices to reduce the checkpointing load on the router.
- Operator is more likely to have the right software on the route-reflector (or the ability to change it) to support BGP GR.

Per Peer GR/NSR config for BGP

- Currently, BGP GR is globally enabled for all peers in the routing process.
- Where BGP NSR is available, it is configured on a per-peer basis.
- If GR is enabled globally and an individual session is configured for NSR, and that session receives GR signaling from its peer, GR will be used over NSR.
- Per peer GR config on horizon

Deployment Considerations

- From a high level, you need to protect the interfaces (SSO), the forwarding plane (NSF) and the control plane (GR or NSR).
- Many requests in the past to turn these features on one at a time, yet for a switchover to work, all need to be enabled and operating properly.
- Enabling SSO also enables NSF (FIB checkpointing)
- Each routing protocol peer needs to be examined to ensure that both its capability has been enabled and that its peer has awareness enabled.

Routing Protocol Timer Manipulation

- Lower than default routing protocol timers are common for faster detection of failures.
- When a switchover occurs, packet forwarding picks up almost immediately, but it takes a bit more time for the new processor to start sending routing protocol control packets. Delay is independent of use of GR or NSR.
- Typically not a problem for BGP, even with 10/30 timers we can get the first packet out well within 10 seconds.
- Using timers such as 1/5 for OSPF can be more problematic.
- Testing under your “real-world” conditions is essential, as the time to first packet can depend on config and platform.

Q and A