Cisco Systems

poweredbycisco. networkers 2005

MPLS HIGH AVAILABILITY

SESSION RST-3108

Phil Harris

pharris@cisco.com

Recuerde siempre:

- Apagar su teléfono móvil/pager, o usar el modo “silencioso”.

- Completar la evaluación de esta sesión y entregarla a los asistentes de sala.

- Ser puntual para asistir a todas las actividades de entrenamiento, almuerzos y eventos sociales para un desarrollo óptimo de la agenda.

- Completar la evaluación general incluida en su mochila y entregarla el miércoles 8 de Junio en los mostradores de registración. Al entregarla recibirá un regalo recordatorio del evento.

Agenda

- High Availability Overview
- Device and Link Level Resiliency
- Protocol Level Resiliency
- Network Level Resiliency
- Operations and Management
- Summary

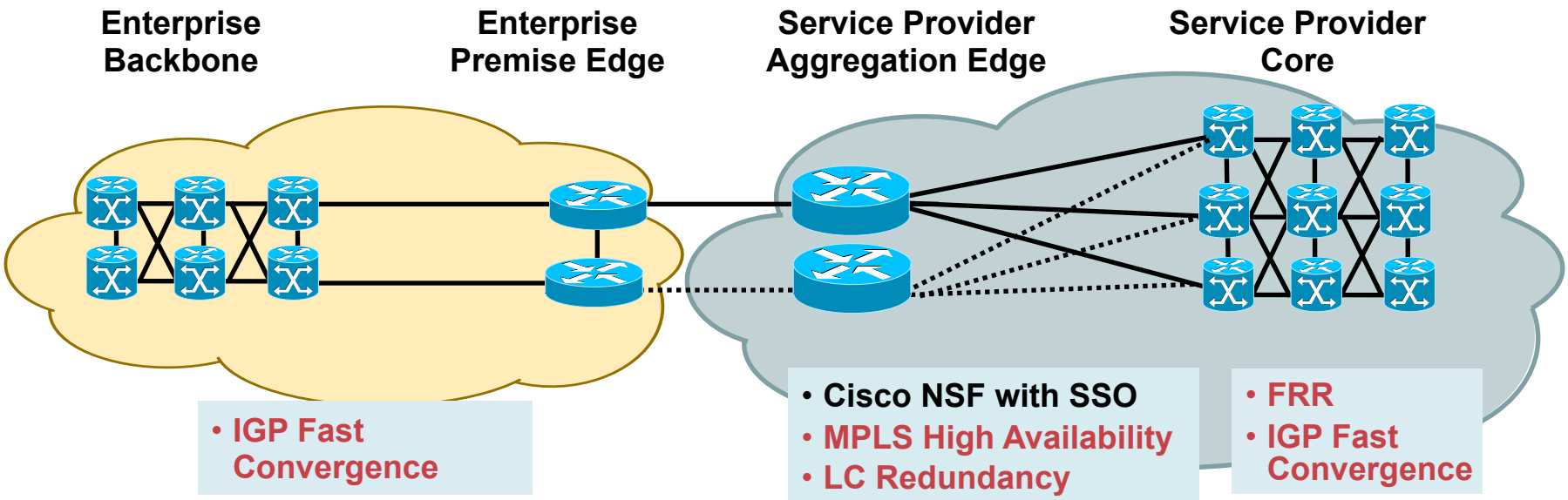
High Availability Definition

The Proportion of Time That a System/Network Can be Used for Productive Work

$$\text{Availability \%} = \frac{\text{MTBF}}{\text{MTBF+MTTR}}$$

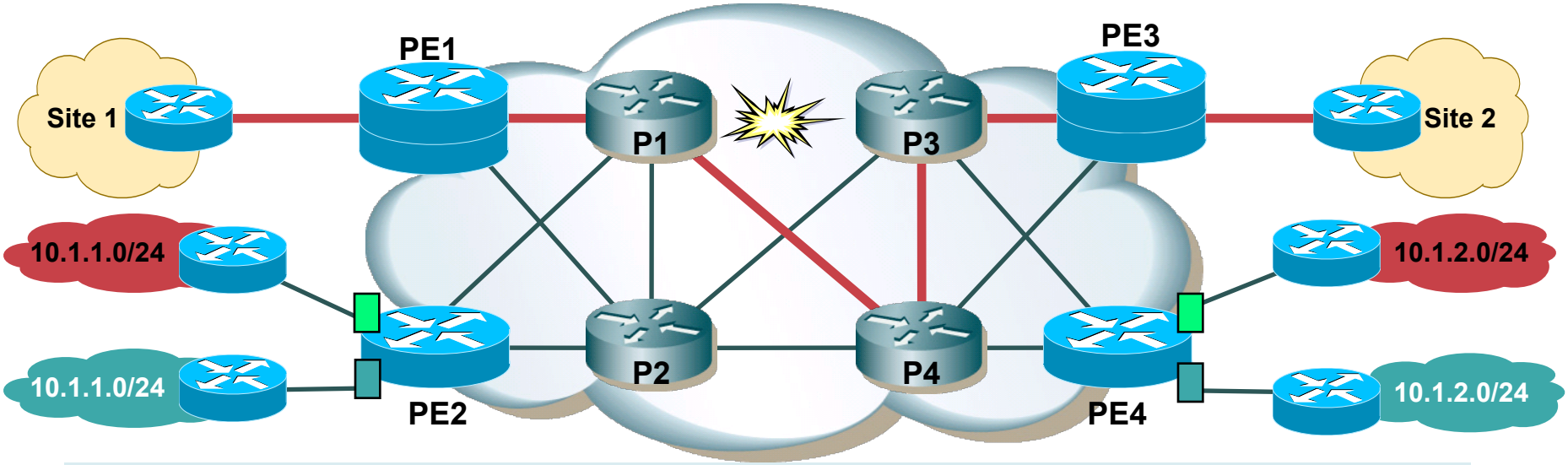
- **Common causes of "downtime" include:**
  - **Hardware failure**
  - **Network failure**
  - **Operating system error/failure**
  - **Application error**
  - **Human error**
  - **Security breach or attack**
  - **System overload**
  - **Power/Environment**

Improving Availability in MPLS Networks Common Goals



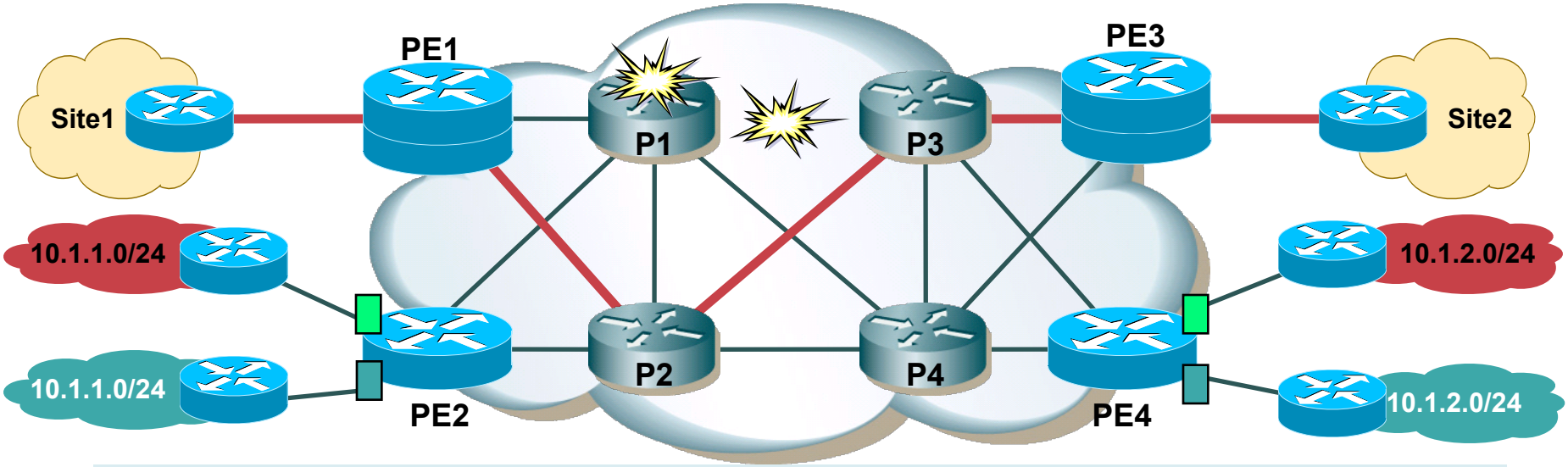
- Design the network for high availability
- Minimize MTTR impact by optimizing the following processes:
  detect fault - diagnose fault - isolate fault- repair system
- Reduce the risk of unscheduled downtime
- Improve network operations, procedures and processes
- Ensure that the appropriate network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

Basic HA Concepts in an MPLS Network



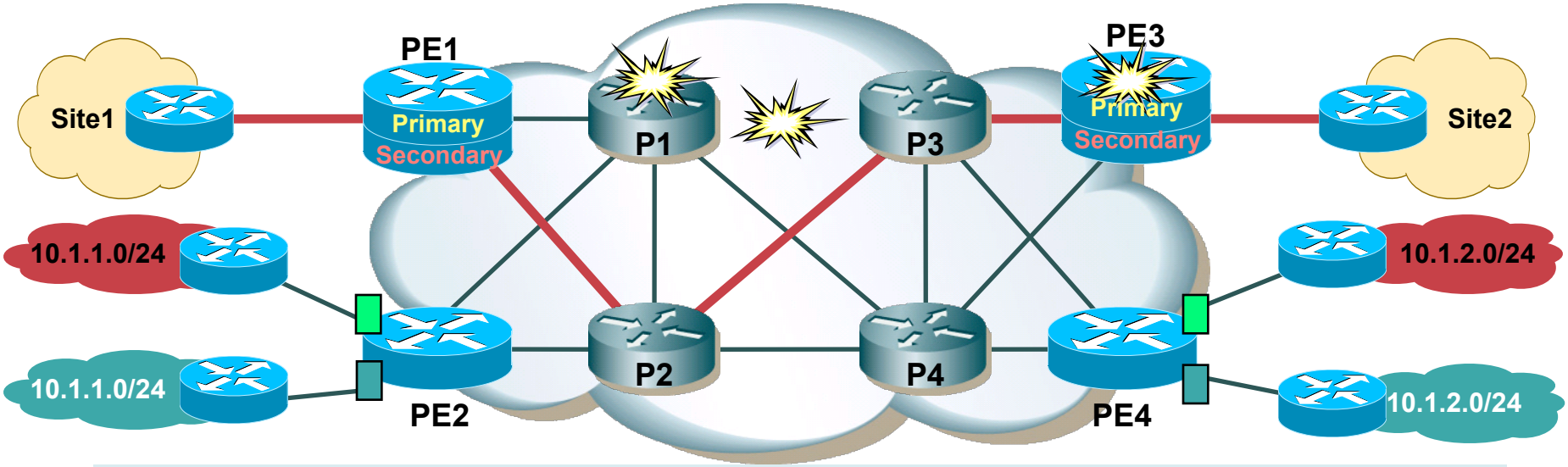
- **Link failures—solved using redundancy in network design and MPLS TE FRR and/or Fast IGP**
- **Node failures—solved using redundant nodes and meshing of connections**
  - **Also using MPLS TE FRR node protection**
- **Line card failures**
  - **Hardware redundancy—line card redundancy (1+1, 1:N, 1:1)**
- **PE router (RP) failures**
  - **Dual RP systems NSF/SSO**
  - **HA software provides seamless transition…**

Basic HA Concepts in an MPLS Network



- Link failures—solved using redundancy in network design and MPLS TE FRR and/or Fast IGP
- Node failures—solved using redundant nodes and meshing of connections
  - Also using MPLS TE FRR node protection
- Line card failures
  - Hardware redundancy—line card redundancy
- PE router (RP) failures
  - Dual RP systems NSF/SSO
  - HA software provides seamless transition…

Basic HA Concepts in an MPLS Network



- Link failures—solved using redundancy in network design and MPLS TE FRR and/or Fast IGP
- Node failures—solved using redundant nodes and meshing of connections
  - Also using MPLS TE FRR node protection
- Line card failures
  - Hardware redundancy—line card redundancy (1+1, 1:N, 1:1)
- **PE router (RP) failures**
  - **Dual RP systems NSF/SSO**
  - **HA software provides seamless transition…**

Agenda

- High Availability Overview
- **Device and Link Level Resiliency**
  - Platform Redundancy (LC, RP, Power, Fan etc)
  - RPR
  - RPR+
  - SSO/NSF
- Protocol Level Resiliency
- Network Level Resiliency
- Operations and Management
- Summary

DEVICE AND COMPONENT LEVEL RESILIENCY

Processor Redundancy Model Introduction

- **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 re-converging Layer 3 protocols

RP Failures

Problem:

- With RP failures, need to wait for
  - Standby RP to initialize
  - Configuration synch up
- With RPR+:
  - Full software image pre-initialized on standby RP
  - Line cards stay up with NO reload/reinitialization
  - Both startup and running configs are synced to the standby RP
  - No link flaps for HDLC (POS) and Ethernet
  - Recovery in 5–30 seconds
  - No MPLS specific changes

Comparison of Redundancy Modes

Redundancy Mode	Startup Config Synch	Running Config Synch	Line Cards Reset	Link Flap	Traffic Loss
RPR	Yes	NO	Yes	Yes	Yes
RPR+	Yes	Yes	No	No * (HDLC/ Eth)	Yes <RPR
SSO	Yes	Yes	NO	No ** (HDLC/ Eth/FR/ATM/ PPP)	NO

*** Session state (i.e. Frame Relay, PPP, ATM) is lost during RP switchover. Resulting in "dropped calls" and time to re-establish connections.**

**** Passes state information from the Active RP to the Standby RP. Resulting in maintaining sessions during a RP switchover.**

Redundancy Configuration

Cisco.com

```
HA-Router(config)#redundancy

HA-Router(config-red)#mode ?
  rpr       Route Processor Redundancy
  rpr-plus  Route Processor Redundancy Plus
  sso       Stateful Switchover (Hot Standby)

HA-Router(config-red)#mode sso
```

```
HA-Router#show redundancy
Active  GRP in slot 0:
Standby GRP in slot 9:
Preferred GRP:  none
Operating Redundancy Mode: SSO
Auto synch: startup-config running-
  config
switchover timer 3 seconds [default]
```

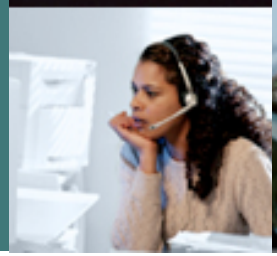
```
HA-Router#sh redundancy states
       my state = 13 -ACTIVE
     peer state = 8  -STANDBY HOT
           Mode = Duplex
        Unit ID = 7
```

Agenda

- High Availability Overview
- Device and Link Level Resiliency
- Protocol Level Resiliency
  - NSF/SSO
    - IGP
    - LDP
    - BGP VPNv4
  - Graceful Restart
    - IGP
    - LDP
    - BGP VPNv4
  - LDP/IGP Synch
  - LDP Session Protection
  - IP Event Dampening
- Network Level Resiliency
- Operations and Management
- Summary

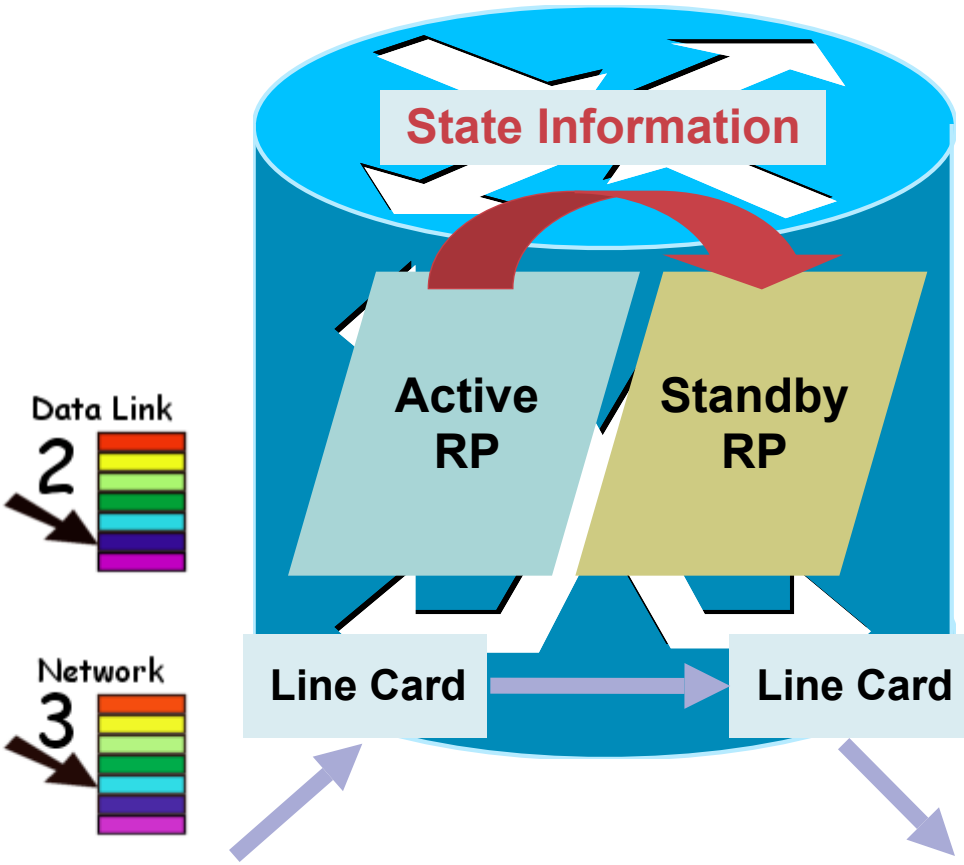
NON-STOP FORWARDING

GRACEFUL RESTART AND STATEFUL SWITCHOVER



Cisco Non-Stop Forwarding with Stateful Switchover (NSF/SSO)

- Standby route processor (RP) takes control of router after a hardware or software fault on the active RP
- SSO allows standby RP to take immediate control and maintain connectivity protocols
- NSF continues to forward packets until route convergence is complete
- GR (graceful restart) reestablishes the routing information bases without churning the network



NSF/SSO Design Goals

- No Link Flap
- CEF table sync to the standby RP
- LC CEF Table not cleared on switchover
- Packet forwarding during switchover while routing is converging on Standby RP
- Maintain peer relationship, no adjacency flapping with peers
- Limit restart to be local event, not network wide

Ultimate Goal: Achieve 0% Packet Loss

OSPF/ISIS NSF

OSPF/ISIS NSF CLI


```
router ospf <proc>
      nsf [enforce global]
```

NSF will abort per-interface where non-NSF-aware neighbors are discovered. Use “enforce global” to abort NSF for the entire OSPF process.

Debug command, enabled at router prompt

```
debug ip ospf nsf
```

Allows user to trace NSF-related activities

```
router isis [tag]
       nsf [cisco/ietf]
```

enable/disable isis nsf , OFF by default

```
 nsf interval  xxx
```

minimal interval between two restarts (default =5min)

```
nsf holdtime [manual
<seconds> | adjacency]
```

IETF version only, Time NSF will wait for the LSP database to synchronize before generating and flooding its own LSP with the overload-bit set

OSPF Configuration Example(s)

Restarting-Router

```
HA-Router#show ip ospf
 Routing Process "ospf 1" with ID 200.200.200.3
 Supports only single TOS(TOS0) routes
 Supports opaque LSA
 Supports Link-local Signaling (LLS)
 Non-Stop Forwarding enabled, last NSF restart 00:25:00 ago
          (took 32 secs)
```

ISIS Configuration Examples

NSF Cisco mode:

```
HA-Router#show isis nsf
NSF is ENABLED, mode 'cisco'
RP is ACTIVE, standby ready, bulk sync complete
NSF interval timer expired (NSF restart enabled)
Checkpointing enabled, no errors
Local state: ACTIVE,  Peer state: STANDBY HOT,  Mode: SSO
```

```
HA-Router#show clns neighbor detail
System Id  Interface   SNPA        State Holdtime Type Protocol
chi-ar1    PO2/0       *HDLC*      Up    21       L2   IS-IS
  Area Address(es): 10
  IP Address(es):  172.1.1.21*
  Uptime: 01:52:02
  NSF capable
```

LDP

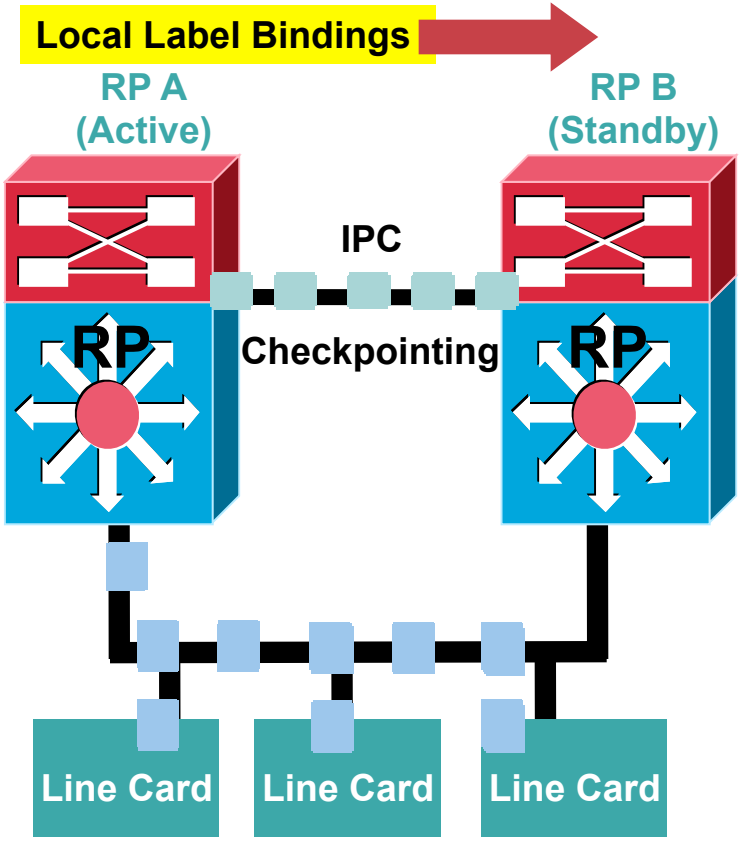
Non Stop Forwarding Stateful Switch Over Graceful Restart (NSF/SSO/GR)

LDP NSF and HA Key Elements

- LDP NSF/SSO/GR is also known as LDP NSF
- LDP NSF allows a route processor to recover from disruption in LDP control plane without losing its MPLS forwarding state
- LDP NSF works with LDP sessions between directly connected peers as well as with targeted sessions
- LDP HA Key Elements
  - Checkpointing local label bindings
    - On devices with route processor redundancy
  - LDP graceful restart capability
    - On participating PEs, RRs, and P routers

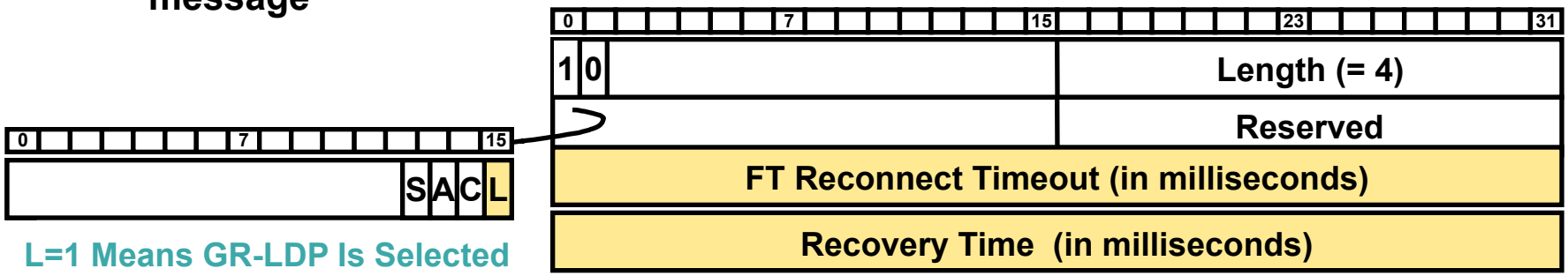
LDP HA Key Elements
Checkpointing Local Label Bindings

- The checkpointing function is enabled by default
- The checkpointing function copies active RP's LDP local label bindings to the backup RP
- For the first round, all the labels are copied from active to back up RP
- Periodic incremental updates are done to reflect new routes that have been learned or routes that have been removed and/or when labels are allocated or freed
- Checkpointing stops during control plane disruptions, GR, and recovery
- Label bindings on backup RP are marked checkpointed
- This marking is removed when it becomes active RP

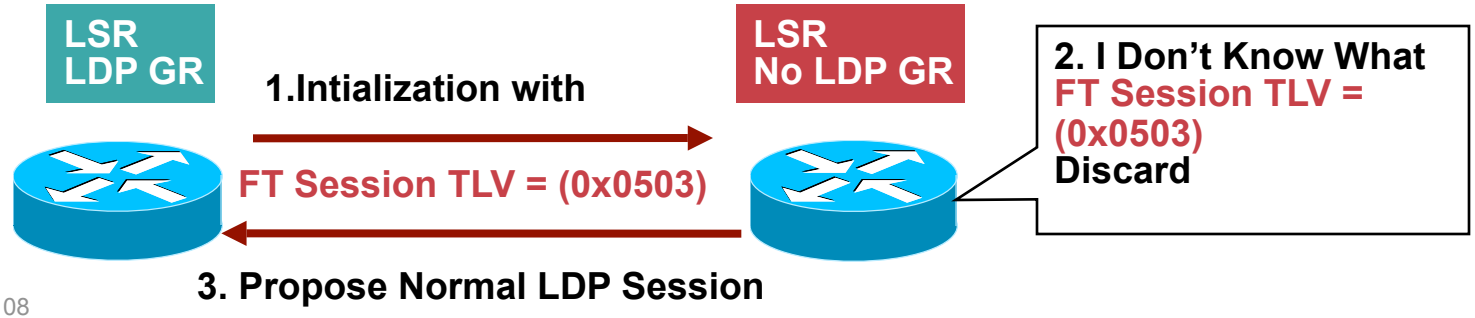


LDP Graceful Restart Mechanism

- Described in RFC3478
- MPLS LDP GR must be enabled before an LDP session is established on all the LSRs
- The LSR sends the LDP init message to a neighbor to establish an LDP session
- The Fault Tolerant (FT) session TLV is included in the LDP initialization message



L=1 Means GR-LDP Is Selected
FT Reconnect = 0 Means LSR Is Not NSF Capable
FT Recovery Time = 0 Means LSR Was Unable to Preserve MPLS Forwarding State Across Restart



LDP Graceful Restart

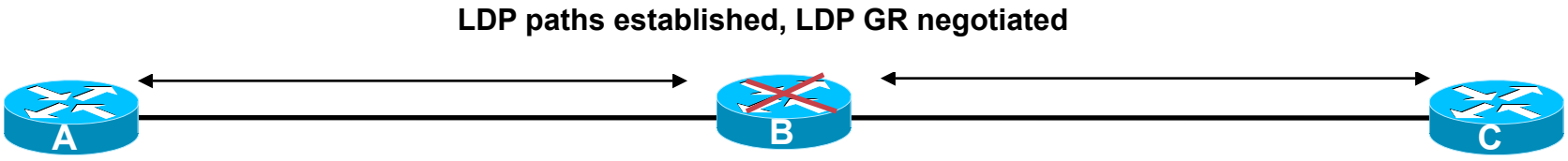
- **LDP GR supports both failure cases:**
  1. **LDP restarts**
  2. **LDP session resets**
- **LDP Graceful Restart Process**

  If LDP GR supported on LSRs: negotiate restart capabilities

  Retain old forwarding plane info—LDP bindings—during (fault tolerant reconnect) timeout

  Restart/recover

LDP Graceful Restart Operation



- When RP fails on LSR “B”, communication between peers is lost:
  - LSR “B” encounters a LDP restart, while LSR “A” and LSR “C” encounter an LDP session reset
  - LSR “A” and LSR “C” mark all the label bindings from LSR “B” as stale, but continue to use the same bindings for MPLS forwarding
  - LSR “A” and LSR “C” attempt to re-establish an LDP session with LSR “B”
  - LSR “B” restarts and marks all of its forwarding entries as stale
  - LSR “A” and LSR “C” re-establish LDP sessions with LSR “B”, but keep their stale label bindings
  - All routers re-advertise their label binding info

LDP Graceful Restart Configuration

LDP GR Must be Enabled on All the LSRs to Take Full Advantage of LDP GR in a Network

1. Enable GR Globally (must do before enabling LDP)

```
mpls ldp graceful-restart
```

```
00:23:58: LDP GR: Received FT Sess TLV from 192.168.1.2:0  (fl
    0x1, rs 0x0, rconn 0, rcov 0)
00:23:58: LDP GR: searching for down nbr record (192.168.1.2:0,
    192.168.1.21)
00:23:58: LDP GR: Added FT Sess TLV (Rconn 0, Rcov 0) to INIT msg
    to 192.168.1.2:0
00:23:58: LDP GR: GR session 192.168.1.2:0:: allocated instance, 1
00:23:58: LDP GR: GR session 192.168.1.2:0:: established
00:23:58: %LDP-5-NBRCHG: LDP Neighbor 192.168.1.2:0 is UP
```

LDP Graceful Restart Troubleshooting

Cisco.com

Show mpls ldp neighbor	Display LDP/TDP neighbor info
Show mpls ldp discovery	Display status of the LDP discovery process
show mpls ldp neighbor graceful-restart	Verify all LDP sessions are configured for LDP GR
Show mpls bindings	Look at LIB table, make sure active and backup processor has identical copies of the local label bindings
Show mpls ldp checkpoint	To display local checkpoint info on the active RP
Clear mpls ldp checkpoint	Clear checkpoint info in LIB on the Active RP and delete all LIB entries learned by checkpointing on the Standby RP
Debug mpls ldp checkpoint Debug mpls ldp grace-restart	For debugging sessions

MPLS LDP GR Verification

Shows a Summary of the Global LDP Restart States

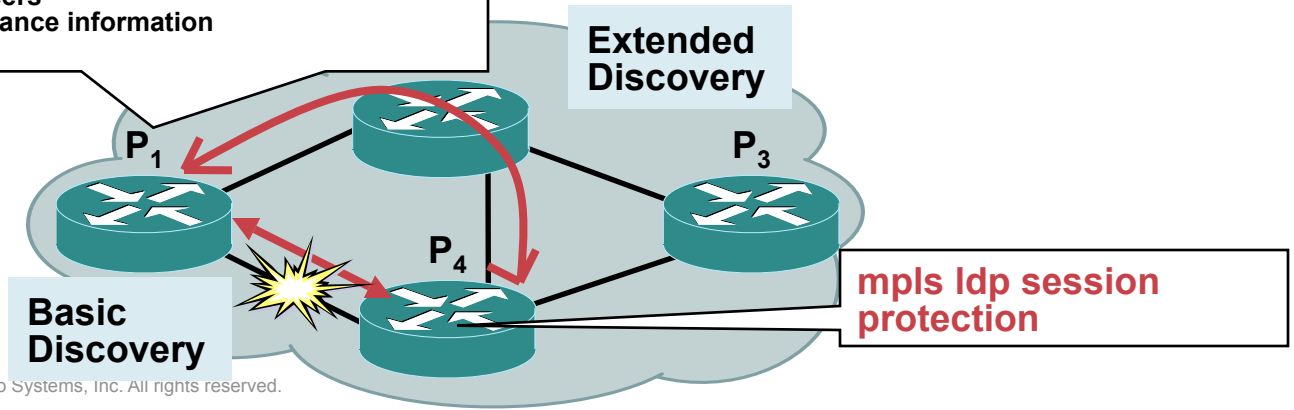
```
pe1#sh mpls ldp graceful-restart
LDP Graceful Restart is enabled
Neighbor Liveness Timer: 120 seconds
Max Recovery Time: 120 seconds
Down Neighbor Database (0 records):
Graceful Restart-enabled Sessions:
  VRF Default-IP-Routing-Table:
    Peer LDP Ident: 192.168.1.2:0, State: estab
```

LDP SESSION PROTECTION

LDP Session Protection

- There are two discovery mechanisms for LDP peers
  1. LDP Basic Discovery—Discovery of directly connected neighbors via link hello's
  2. LDP Extended Discovery—Discovery of non-directly connected neighbors via Targeted Hello's
- If Session Protection is enabled, the establishment of a directly connected LDP session triggers Extended Discovery with neighbors
- With targeted hello's, session stays up even when the link goes down
- No need to re-establish an LDP session with the link neighbor and relearn prefix label bindings when the link recovers

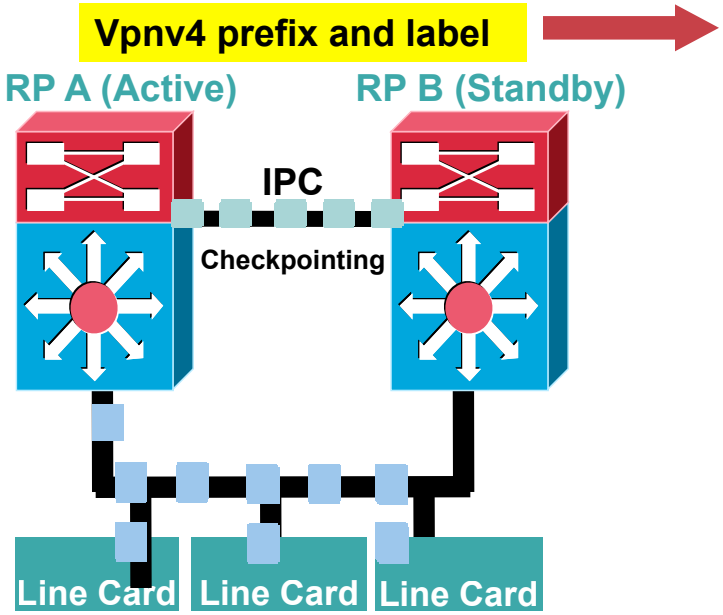
```
p1(config)#mpls ldp session protection ?
  duration  Period to sustain session protection after loss of link discovery
  for       Access-list to specify LDP peers
  vrf       VRF Routing/Forwarding instance information
  <cr>
```



BGP VPNv4 NSF/SSO/GR

MPLS VPNv4 HA Elements
MPLS VPN Checkpointing

Active RP checkpoints the Following Information to the Backup RP After the MPLS Forwarding is Updated: {<VRFID>, <prefix>, <mask>, <local label>}



Router#show ip bgp vpnv4 all labels

Network	Next Hop	In label/Out label
Route Distinguisher: 100:1 (vpn1)		
12.12.12.12/32	0.0.0.0	16/aggregate(vpn1)
135.0.0.0/8	0.0.0.0	17/aggregate(vpn1)
Route Distinguisher: 609:1 (vpn0)		
13.13.13.13/32	0.0.0.0	18/aggregate(vpn0)

Router#show ip bgp vpnv4 all labels

Network	Masklen	In label
Route Distinguisher: (dec) 0001000001		
12.12.12.12/32		16
135.0.0.0/8		17
Route Distinguisher: (dec) 002970001		
13.13.13.13/32		18

MPLS VPNv4 HA Elements
BGP VPNv4 GR Mechanism

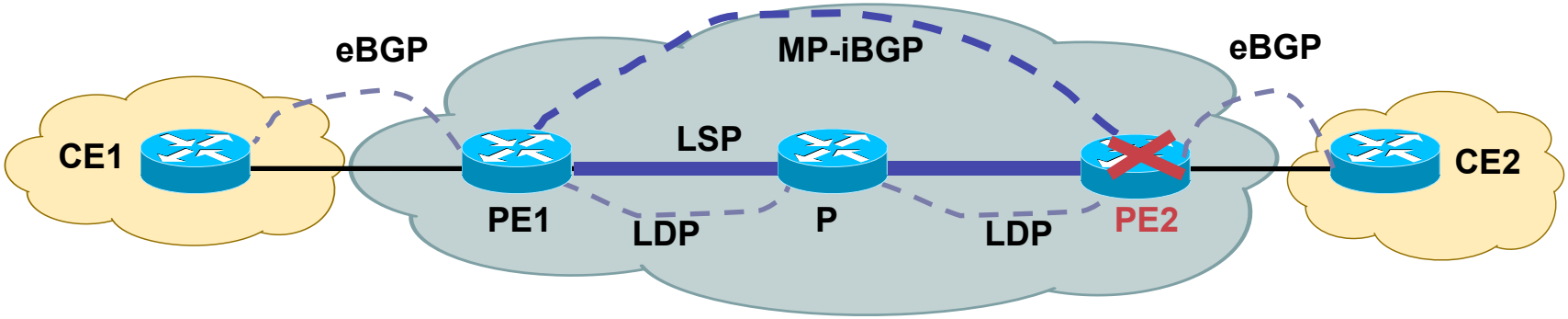
- **MPLS BGP GR mechanism defines preservation of forwarding state across BGP restart**

  draft-ietf-mpls-bgp-mpls-restart

- **A new graceful restart capability is carried in BGP open message**

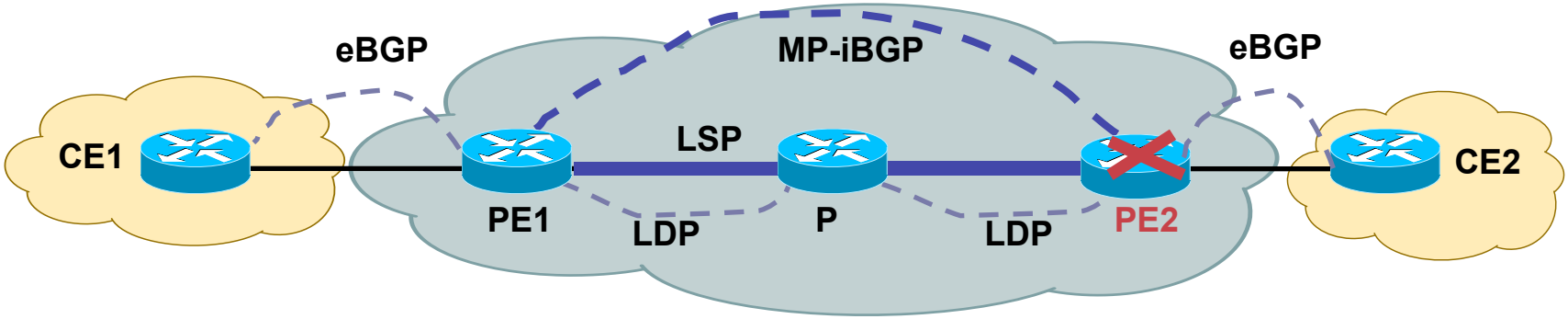
  A BGP update message with no reachable NLRI and empty withdrawn NLRI is specified as an End-of-RIB marker

MPLS VPN: BGP Graceful Restart



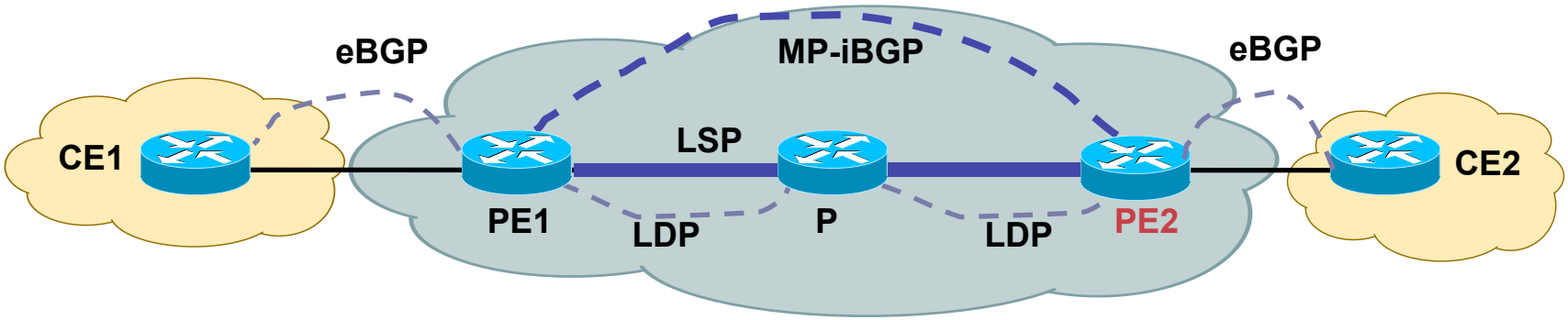
- PE1 detects PE2's failure, retains its last Adj-RIB-In and forwarding state learned from PE2; PE1 will delete this state if session is not re-established within the restart time-out
- The BGP session between PE1 and PE2 goes down
- PE1 marks the entries in its BGP table which it learned from PE2 as stale but does not delete the routes from its VRF tables; hence it continues to forward traffic to PE2
- PE2 switches over to back up RP; continues to forward traffic using the backed up info

MPLS VPN: BGP Graceful Restart
Procedure (Cont.)



- PE2 re-establishes TCP session with PE1 (hopefully within Restart Time—180 seconds default)
- PE1 sends BGP Updates from Adj-RIBs-Out to PE2 along with the label mapping; on completion, sends End-of-RIB marker
- PE2 runs decision process after receiving End-of-RIB markers from all BGP peers, updates its Loc-RIB, FIB, Adj_RIBs-Out and advertise its routes to PE1; on completion sends End-of-RIB marker

MPLS VPN: BGP Graceful Restart
Procedure (Cont.)



- Suppose PE2 had bound an in label L1 to a FEC, it picks the same label (if checkpointed) or allocates a new one and advertises it to PE1 for this route
- PE2 updates its Adj-RIBs-In, on receipt of End-of-RIB marker, deletes stale entries, runs its decision process, updates its Loc-RIB, FIB, and Adj-RIBs-Out; removes stale state if the labels are the same
- Back to normal operation

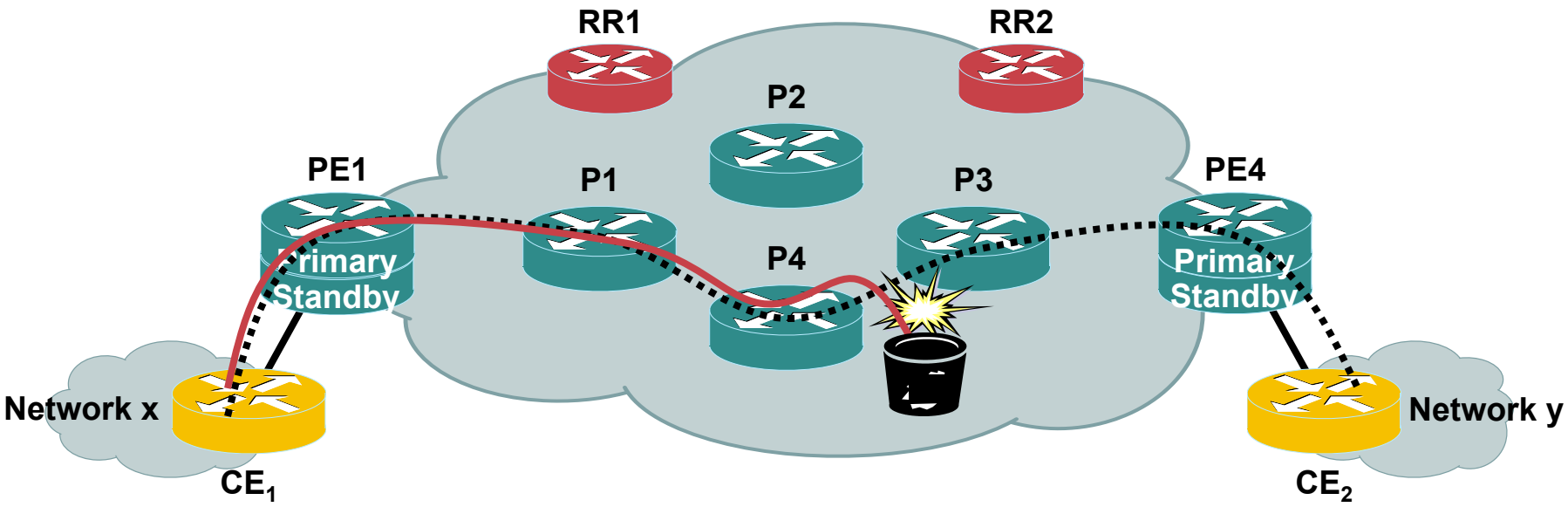
BGP VPNv4 Graceful Restart Troubleshooting

Show ip bgp labels	Displays information about MPLS labels from eBGP route table
Show ip bgp vpnv4 all labels	Displays info on the active and standby RPs when they are configured for MPLS VPN NSF
Debug ip bgp vpnv4 checkpoint	Displays the events for the VRF checkpointing system between the active and standby RP
Debug ip bgp vpnv4 nsf	Look at LIB table, make sure active and backup processor has identical copies of the local label bindings
Show mpls ldp checkpoint	Displays the NSF events for the VRF tableid synch subsystem between the active and standby RPs

LDP/IGP SYNCH

LDP Down but IGP Still Up

1. Data stream between CE1 and CE2 traversing SP cloud
2. LDP Adjacency goes down between P4 and P3 but IGP still points to P3 as the best path



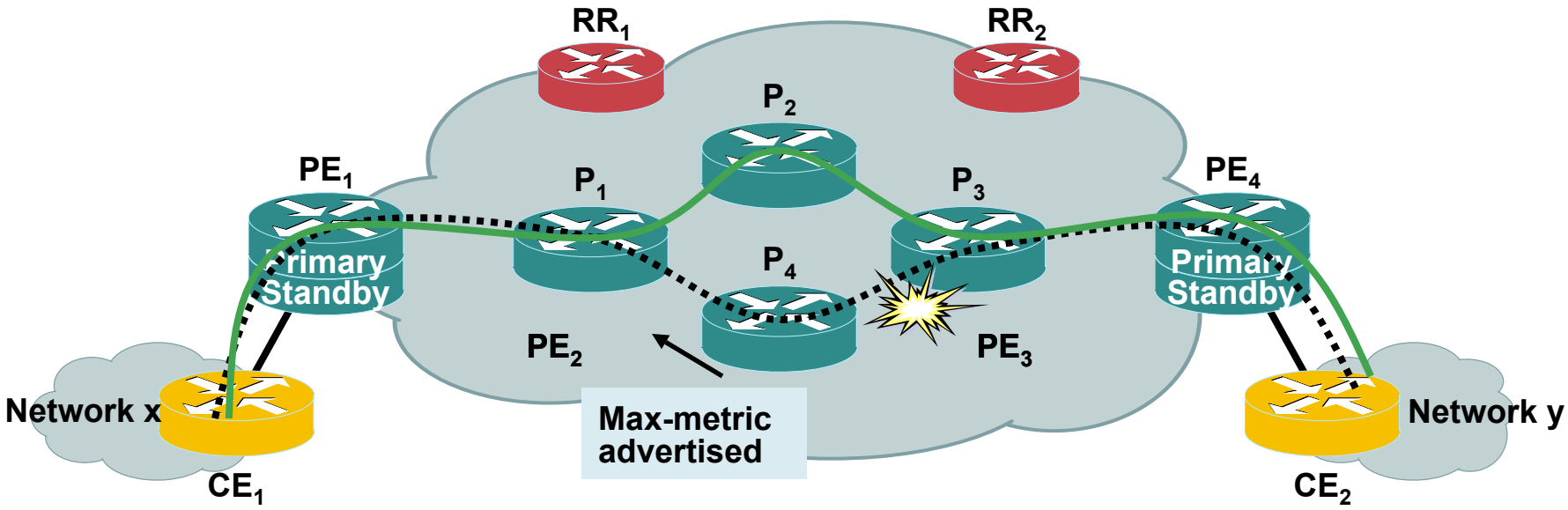
IGP-LDP Synch

- IGP-LDP Synch features synchronizes the state of IGP and LDP between two routers
- If an LDP adjacency between two routers goes down then these routers would advertise a max-metric for the link connecting them via IGP
- This way upstream router can choose an alternate path if available
- Following configuration needs to be enabled on the routers

```
P4(config)# router ospf 1
 P4(config-router)# mpls ldp sync
```

LDP down but IGP Still Up with IGP-LDP Synch Feature

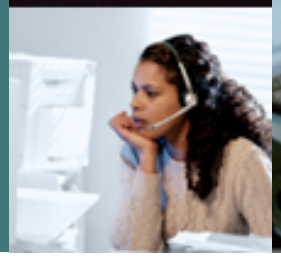
1. Data stream between CE1 and CE2 traversing SP cloud
2. LDP Adjacency goes down between P4 and P3 but IGP still points to P3 as the best path
3. P1 chooses P2 as the preferred IGP path because of the lower metric



LDP/IGP Verification

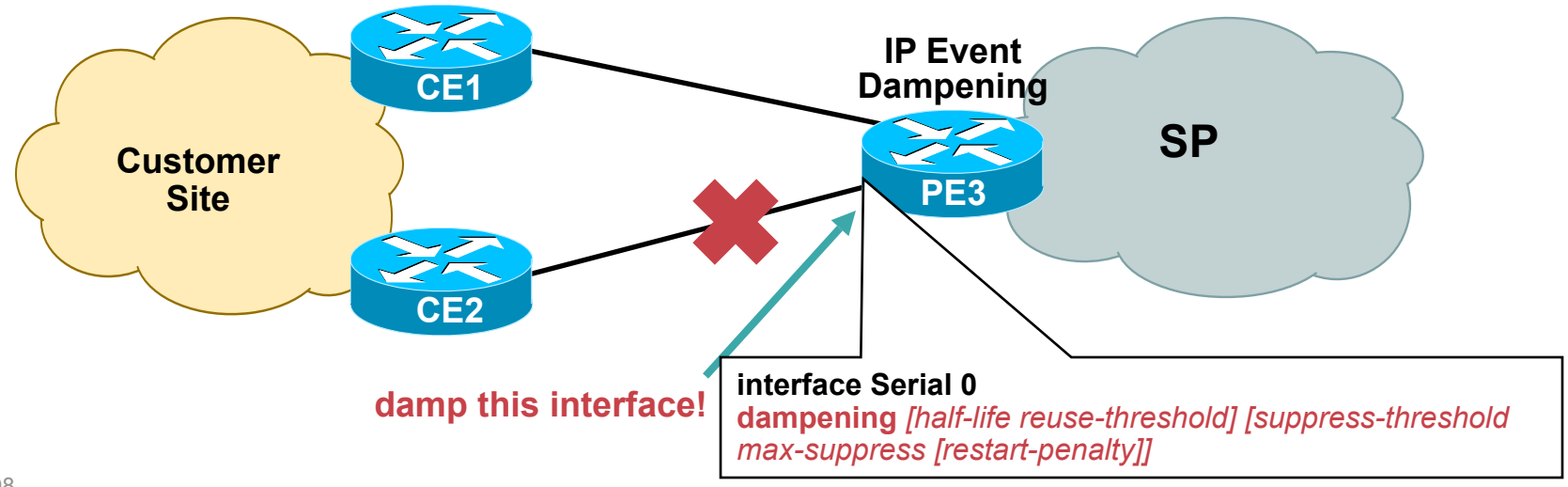

```
P4#show mpls ldp igp sync
    Ethernet0/0:
        LDP configured;  SYNC enabled.
        SYNC status: sync achieved; peer reachable.
        IGP holddown time: 250 milliseconds.
        Peer LDP Ident: 130.0.0.1:0
        IGP enabled: OSPF 1
    Ethernet1/0:
        LDP configured;  SYNC enabled.
        SYNC status: sync not achieved; peer reachable.
        IGP holddown time: 250 milliseconds.
        IGP enabled: OSPF 1
```

IP EVENT DAMPENING

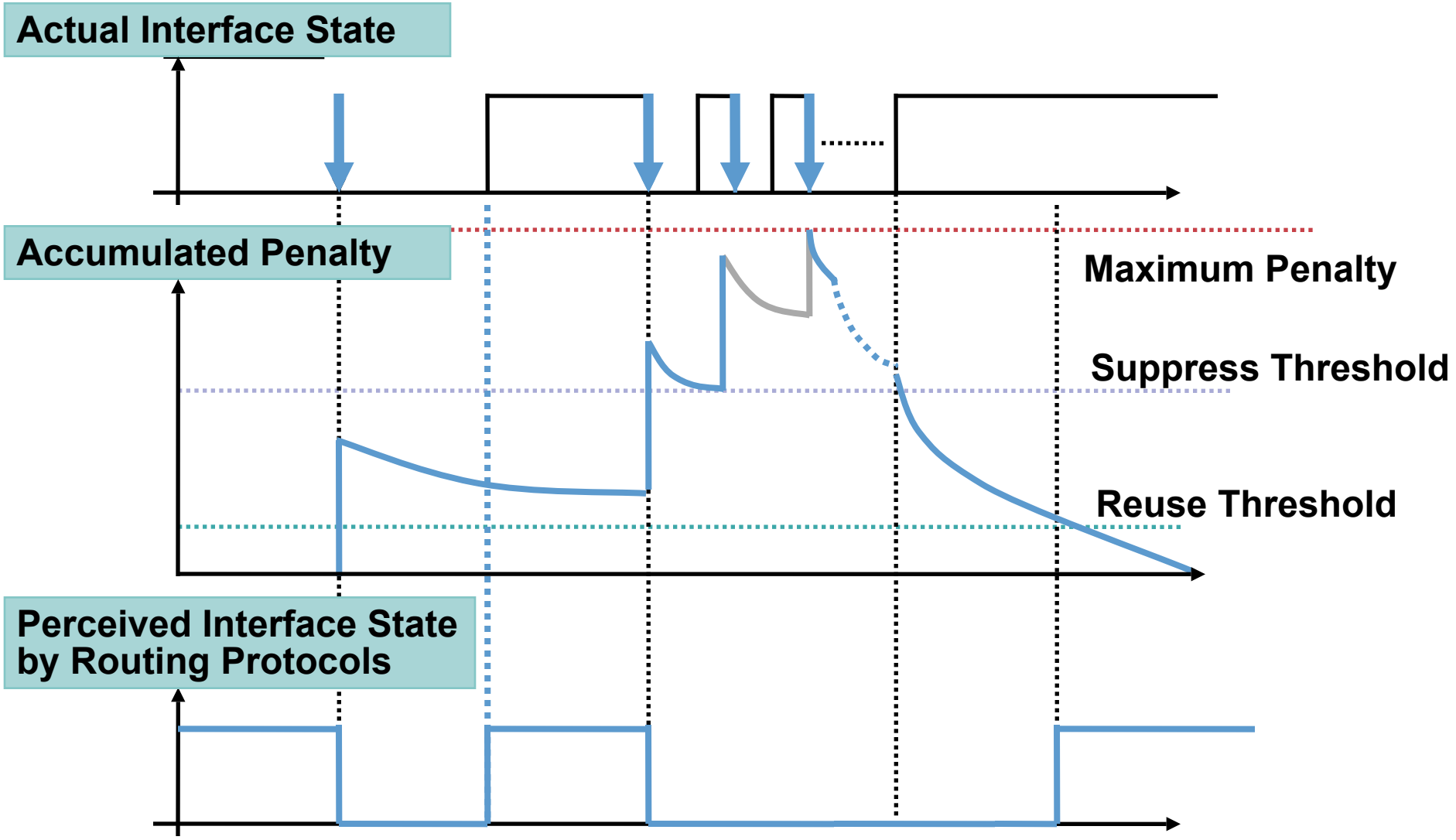


IP Event Dampening

- A flapping link can cause major problems many hops away.
- Even if you are using good network design techniques, like summarization, link flaps can still cause a major portion of your network to converge with each flap
- IP event dampening catches the problem at its source, the flapping interface
- Takes the concept of BGP route-flap dampening and applies it at the interface level, so all IP-routing protocols can benefit



IP Event Dampening: Algorithm Illustration



Agenda

- High Availability Overview
- Device and Link Level Resiliency
- Protocol Level Resiliency
- **Network Level Resiliency**
  - **BGP NHT**
  - **PW Redundancy**
  - **MPLS Fast ReRoute**
- Operations and Management
- Summary

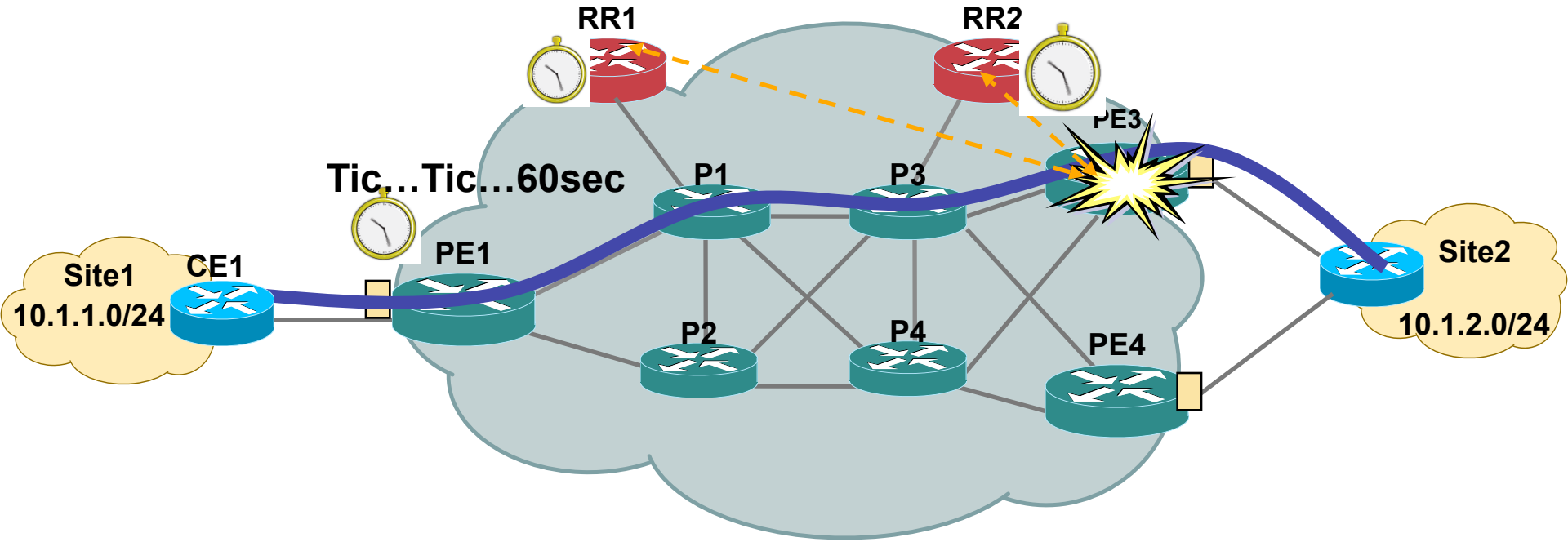
BGP Next Hop Tracking (NHT)

BGP Next-Hop Tracking

- Makes the next-hop failure detection event-driven instead of timer-driven
- Next-hop tracking (NHT) feature allows to track BGP next-hops in the RIB
- If the RIB entry changes, then the client such as BGP is notified
- Allows for new path selection for BGP routes as soon as the notification is received
- On/off knob as well as configuration option on how long to wait before starting new path selection

Behavior Without NHT

Traffic Loss for up to 60 Secs Due to BGP Scanner Interval

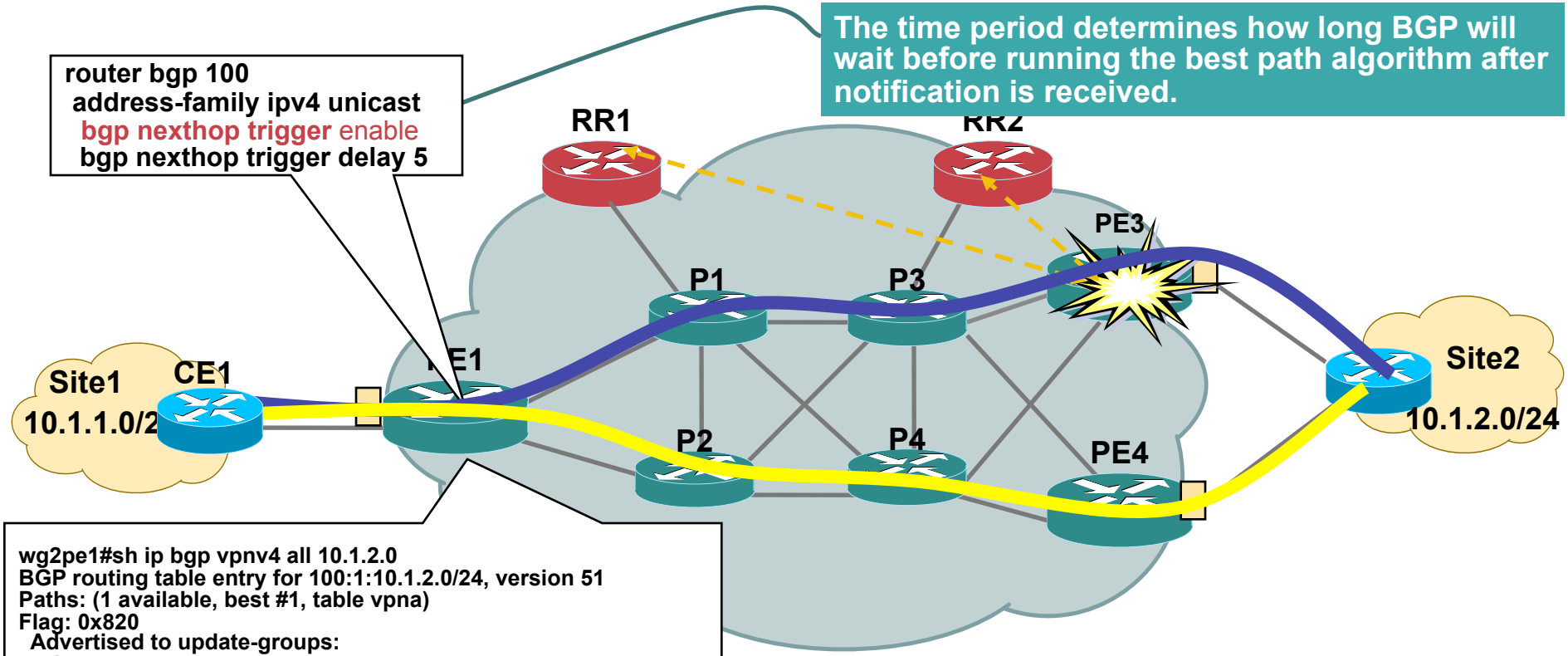


Behavior with NHT Enabled

Potential Time Saving Is Up to 60 Secs

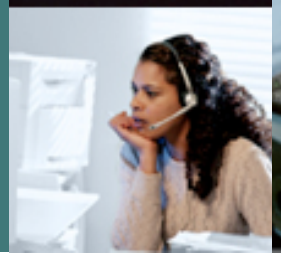
The time period determines how long BGP will wait before running the best path algorithm after notification is received.

```
router bgp 100
 address-family ipv4 unicast
  bgp nexthop trigger enable
  bgp nexthop trigger delay 5
```

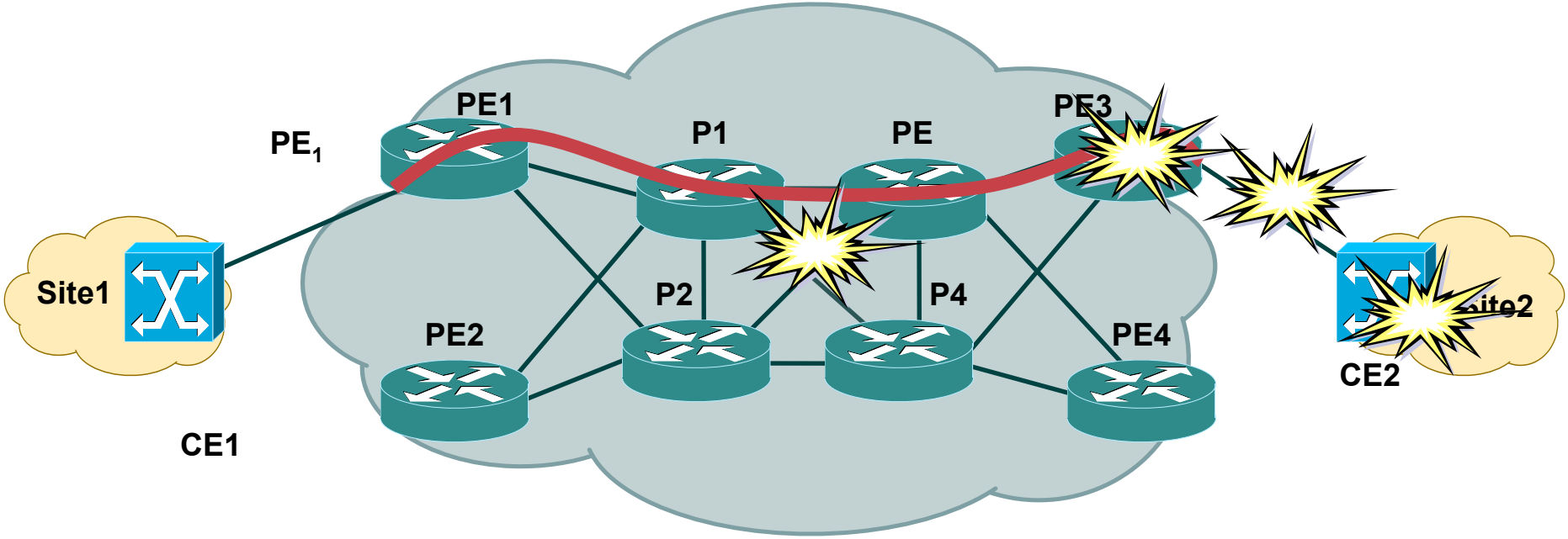



```
wg2pe1#sh ip bgp vpnv4 all 10.1.2.0
BGP routing table entry for 100:1:10.1.2.0/24, version 51
Paths: (1 available, best #1, table vpna)
Flag: 0x820
  Advertised to update-groups:
     1
  Local
    192.168.1.4 (metric 193) from 192.168.1.2 (192.168.1.2)
      Origin incomplete, metric 0, localpref 100, valid, internal, best
      Extended Community: RT:100:1
      Originator: 192.168.1.4, Cluster list: 192.168.1.2,
      mpls labels in/out nolabel/32
```

Psuedo Wire (PW) Redundancy

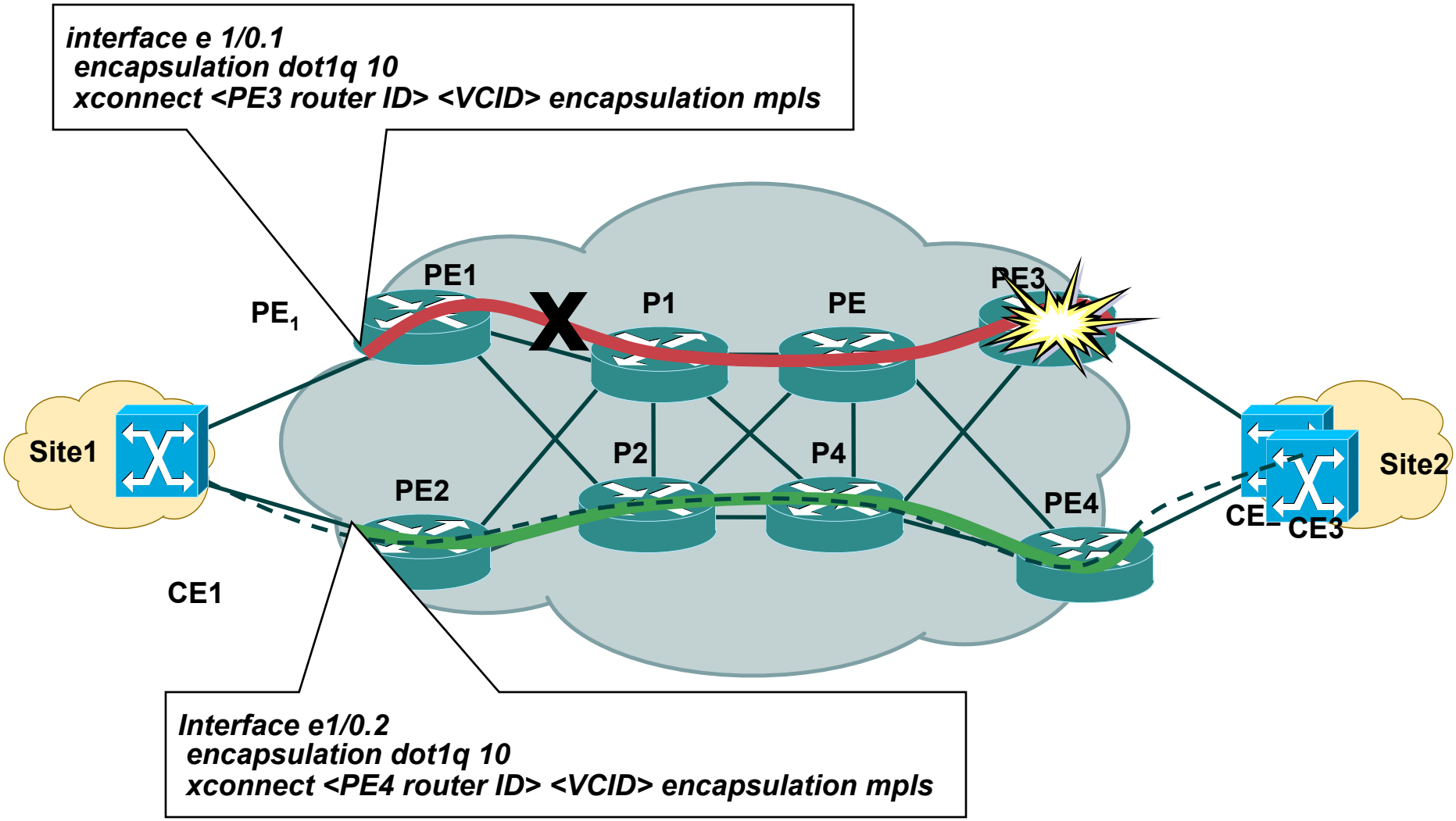


PW High Availability

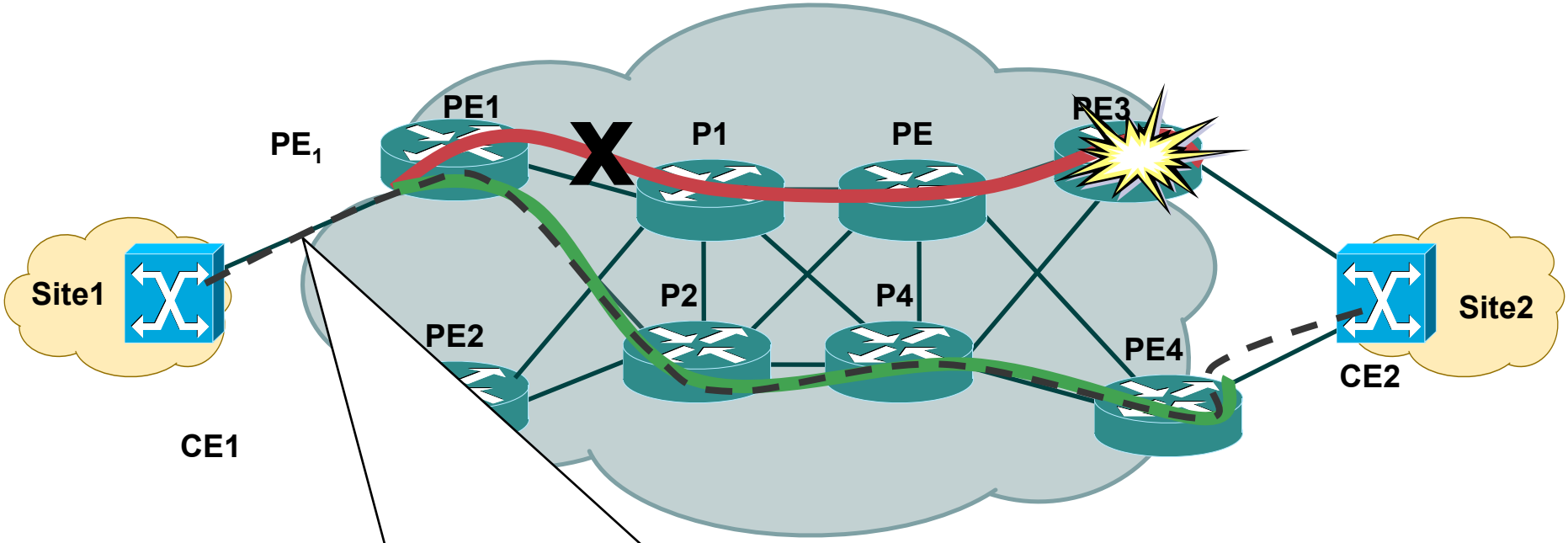


- Failure in the provider core mitigated with link redundancy and FRR
- PE router failure–PE diversity
- Attachment Circuit failure–need pair of attachment Ckts end-to-end
- CE Router failure–redundant CEs

PW High Availability



PW High Availability




```
pe1(config)#int e 0/0.1
pe1(config-subif)#encapsulation dot1q 10
pe1(config-subif)# xconnect <PE3 router ID> <VCID> encapsulation mpls
pe1(config-subif-xconn)#backup peer <PE4 router ID> <VCID>
```

PW Redundancy Configuration

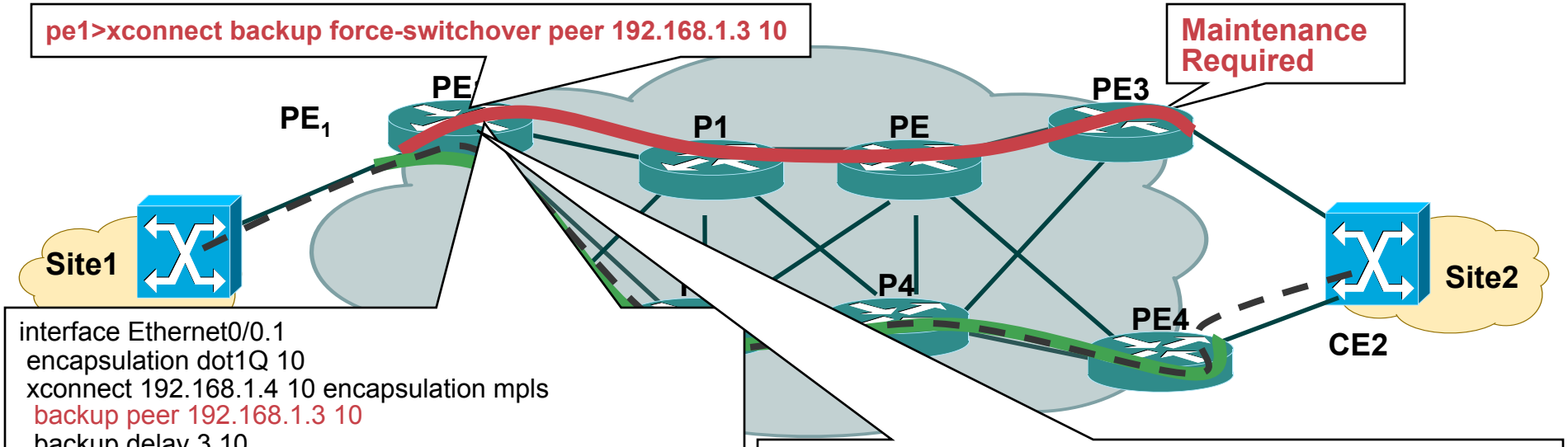
xconnect <ip-addr> <vcid> pw-class <name>

backup peer <ip-addr> <vcid> pw-class <x> *priority* <value>

backup delay <enable-delay> <disable-delay | never>

Default values: <pw-class> = if not specified, then this is inherited from the parent xconnect

<priority> = 1 (highest priority) <enable-delay> = 0 <disable-delay> = 0




```
interface Ethernet0/0.1
 encapsulation dot1Q 10
 xconnect 192.168.1.4 10 encapsulation mpls
  backup peer 192.168.1.3 10
  backup delay 3 10
```

pe1#sh mpls l2transport vc 10

Local intf	Local circuit	Dest address	VC ID	Status
Et0/0.1	Eth VLAN 20	192.168.1.4	10	UP
Et0/0.1	Eth VLAN 20	192.168.1.3	10	DOWN

pe1#sh mpls l2transport vc 10

Local intf	Local circuit	Dest address	VC ID	Status
Et0/0.1	Eth VLAN 20	192.168.1.4	10	DOWN
Et0/0.1	Eth VLAN 20	192.168.1.3	10	UP

Verifying PW Redundancy


```
pe1#sh mpls l2transport vc 10

Local intf     Local circuit          Dest address    VC ID      Status
-------------  ---------------------- --------------- ---------- ----------
Et0/0.1        Eth VLAN 20            192.168.1.4     10         UP
Et0/0.1        Eth VLAN 20            192.168.1.3     10         DOWN
```

```
pe1#show mpls l2 vc 10 detail

Local interface: Et1/0.1 up, line protocol up, Eth VLAN 10 up
  Destination address: 192.168.1.4, VC ID: 10, VC status: up
    Preferred path: not configured
    Default path: active
    Next hop: point2point
    Output interface: Se7/0, imposed label stack {23 26}
  Create time: 00:06:05, last status change time: 00:06:03
  Signaling protocol: LDP, peer 192.168.1.4:0 up
    MPLS VC labels: local 33, remote 26
    Group ID: local 0, remote 0
    MTU: local 1500, remote 1500
    Remote interface description:
```

```
Local interface: Et1/0.1 up, line protocol up, Eth VLAN 10 up
  Destination address: 192.168.1.3, VC ID: 10, VC status: down
  Create time: 00:06:05, last status change time: 00:06:05
  Signaling protocol: LDP, peer 192.168.1.3:0 up
    MPLS VC labels: local unassigned, remote 26
    Group ID: local unknown, remote 0
    MTU: local unknown, remote 1500
    Remote interface description:
  Sequencing: receive disabled, send disabled
  VC statistics:
    packet totals: receive 0, send 0
    byte totals:   receive 0, send 0
    packet drops:  receive 0, seq error 0, send 0
```

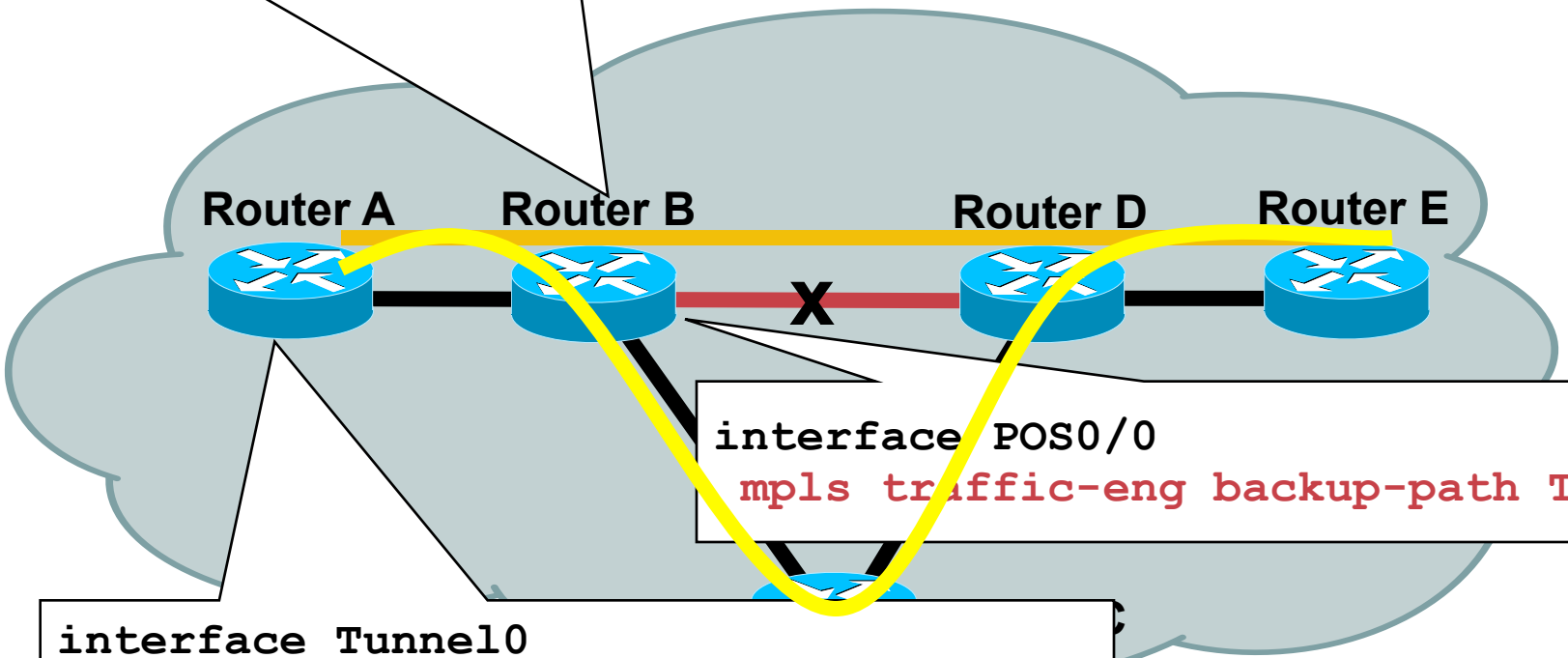
MPLS Fast Re-Route (FRR)

Protection

- Mechanism to minimize packet loss during a failure
- Pre-provisioned protection tunnels that carry traffic when a protected link or node goes down
- MPLS TE protection also known as FAST REROUTE (FRR)
- FRR protects against LINK FAILURE
  - For example, Fibre cut, Carrier Loss, ADM failure
- FRR protects against NODE FAILURE
  - For example, power failure, hardware crash, maintenance
- FRR protects against LSP path failures

Link Protection Configuration


```
interface Tunnel0
 tunnel destination Router D
 … explicit-path R2-R3-R4
 no tunnel mpls traffic-eng autoroute announce
```

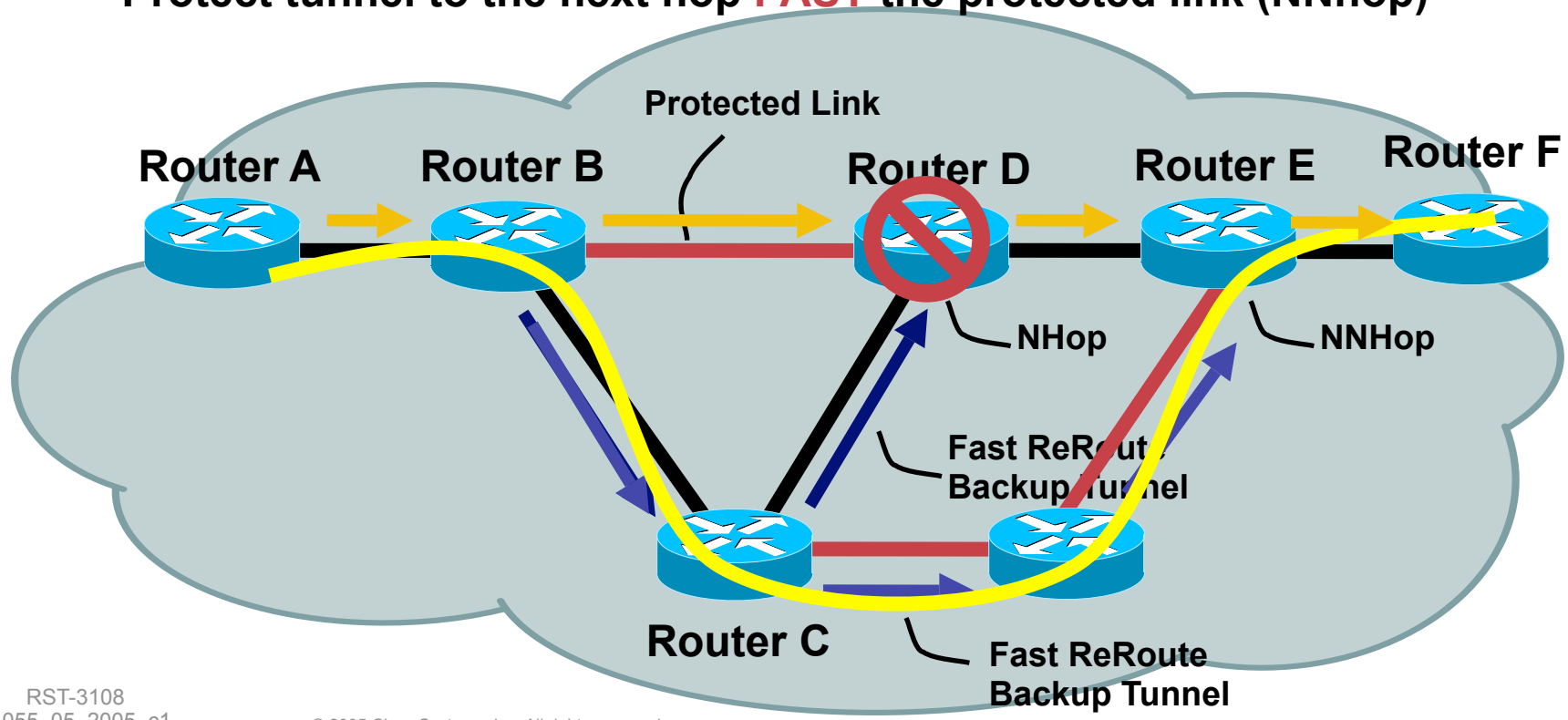



```
interface POS0/0
 mpls traffic-eng backup-path Tunnel0
```

```
interface Tunnel0
 tunnel destination Router E
 .. etc ...
 tunnel mpls traffic-eng fast-reroute
```

Node Protection

- What if Router D failed?
- Link protection would not help as the backup tunnel terminates on Router D (which is the NHop of the protected link)
- Protect tunnel to the next hop PAST the protected link (NNhop)



Node Protection

- Node protection still has the same convergence properties as link protection
- Deciding where to place your backup tunnels is a much harder problem to solve on a large-scale
- For small-scale protection, link may be better
- Auto-tunnel and auto-mesh can help with this
- Configuration is identical to link protection, except where you terminate the backup tunnel (NNHop vs. NHop)

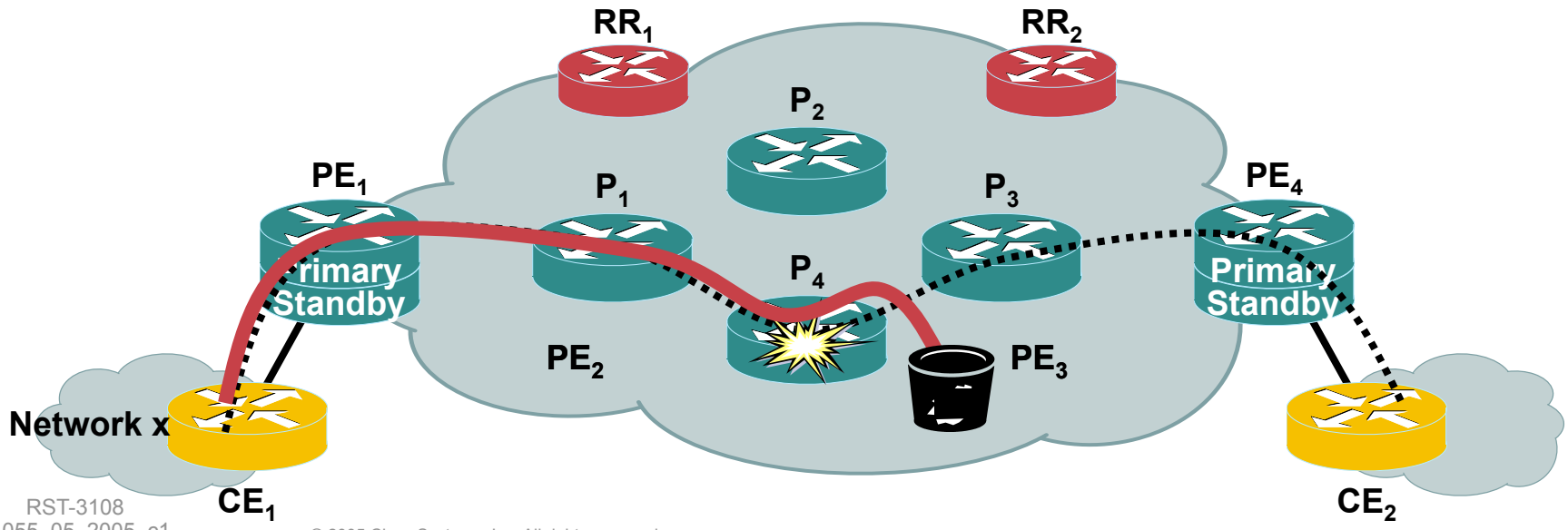
Agenda

- High Availability Overview
- Device and Link Level Resiliency
- Protocol Level Resiliency
- Network Level Resiliency
- Operations and Management
  - MPLS LDP AutoConfig
  - BFD
  - LSP Ping/Trace
- Summary

LDP AUTO CONFIG

LDP Autoconfig

- Today "[no] mpls ip" configured on each interface participating in MPLS. (Same global command to disable MPLS)
- LDP not enabled on certain interface by mistake could cause VPN (..and potentially internet) traffic to be black holed
- These configuration errors can reduce the availability of the network
- By using Auto-config feature LDP can be automatically configured on links for which a specified IGP has been enabled



LDP Autoconfig

- **Following configuration enables mpls on all the interfaces that are in OSPF area3**

```
P4(config)# router ospf 1
 P4(config-router)# network 133.0.0.0 0.0.255.255 area 3
 P4(config-router)# network 133.1.0.0 0.0.255.255 area 3
 P4(config-router)# mpls ldp area 3
```

- **One can disable mpls for individual interfaces under area3**

```
P4(config)# interface pos1/0
 P4(config-if)# no mpls ldp autoconfig
```

LDP Autoconfig Verification

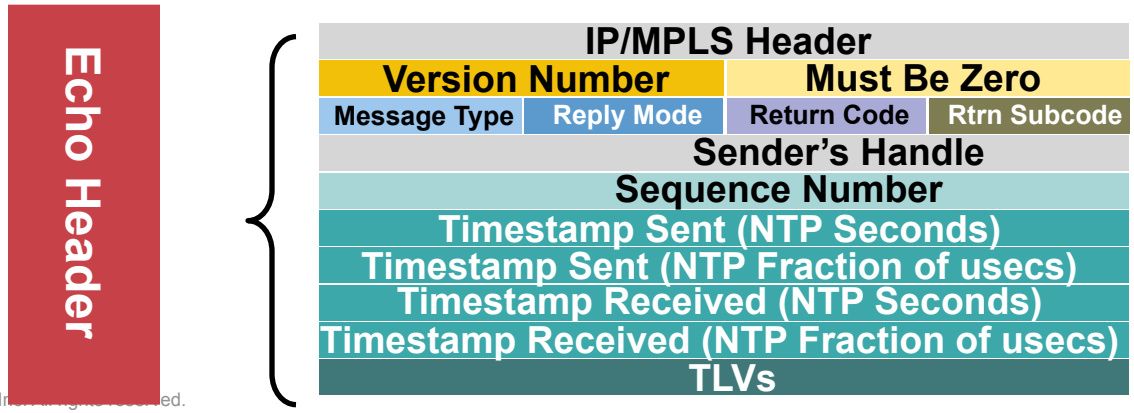

```
P4#sh mpls ldp discovery detail
  Local LDP Identifier:
    11.11.11.11:0
    Discovery Sources:
    Interfaces:
        Ethernet0/1 (ldp): xmit/recv
            Hello interval: 5000 ms; Transport
  IP addr: 11.11.11.11
            Enabled: mpls ip, IGP triggered;
            LDP Id: 10.10.10.10:0
             Src IP addr: 130.0.0.4; Transport
  IP addr: 10.10.10.10
```

```
P4#sh mpls interfaces e0/1 detail
  Interface Ethernet0/1:
      IP labeling enabled (ldp); Enabled:
    mpls ip, IGP triggered
      LSP Tunnel labeling not enabled
      BGP labeling not enabled
      MPLS operational
      Fast Switching Vectors:
        IP to MPLS Fast Switching Vector
        MPLS Turbo Vector
      MTU = 1500
```

MPLS PING/TRACE

Reducing MTTR Using MPLS Ping/Trace

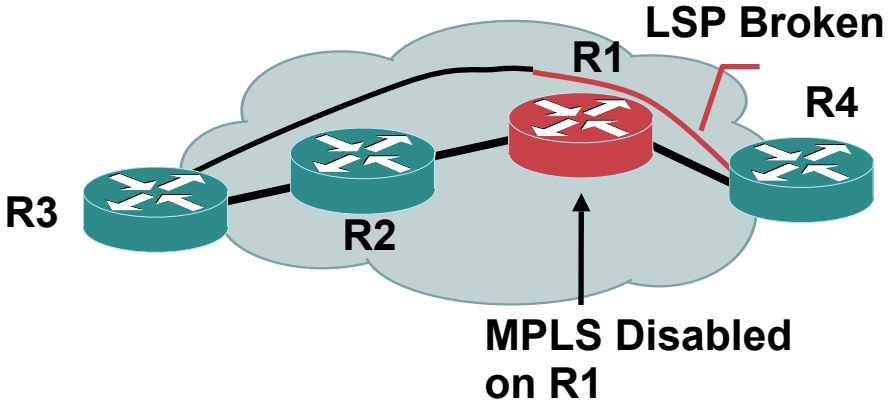
- MPLS introduces separation of control plane and data plane which has resulted in increased time to troubleshoot problems in MPLS networks
- Traditional tools such as IP Ping/Trace are not able to do fault detection/isolation in a short time
- The new LSP(MPLS) Ping/Trace, like the traditional IP Ping, is based on echo request and echo reply. They greatly reduce MTTR by faster fault detection/isolation
- LSP Ping/Trace, unlike IP Ping/Trace, relies on IPv4(or IPv6) UDP packets with port 3503



LSP Ping Configuration

R3#ping mpls ?

ipv4	**Target specified as an IPv4 address**
pseudowire	**Target VC specified as an IPv4 address and VC ID**
traffic-eng	**Target specified as TE tunnel interface**



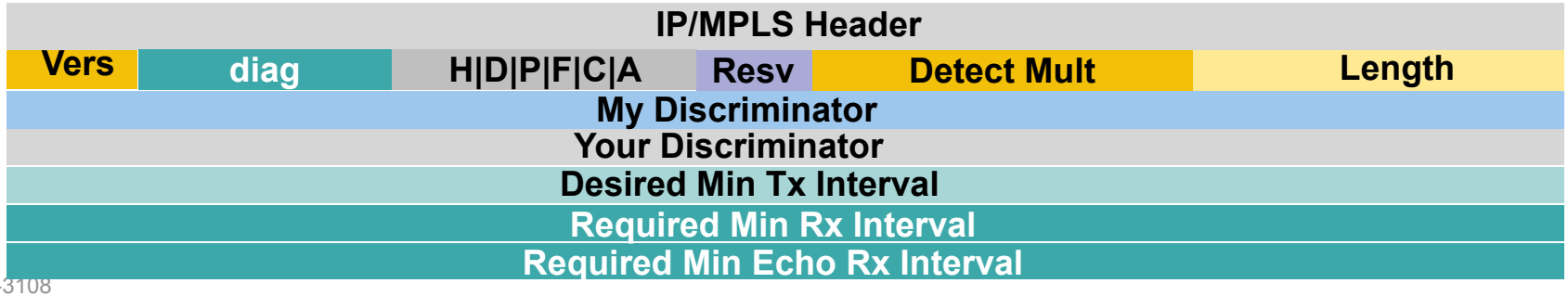
- **If a regular ping is done from R3 to R4, it would be successful. But an LSP ping would fail**
- **The response would come from R1**

```
R3#ping mpls ipv4 10.200.0.4/32 verbose
Sending 5, 100-byte MPLS Echos to 10.200.0.4/32,
     timeout is 2 seconds, send interval is 0 msec:

Codes: '!' - success, 'Q' - request not transmitted,
       '.' - timeout, 'U' - unreachable,
       'R' - downstream router but not target

Type escape sequence to abort.
U   10.200.21.1, return code 4
U   10.200.21.1, return code 4
U   10.200.21.1, return code 4
U   10.200.21.1, return code 4
U   10.200.21.1, return code 4

Success rate is 0 percent (0/5)
```

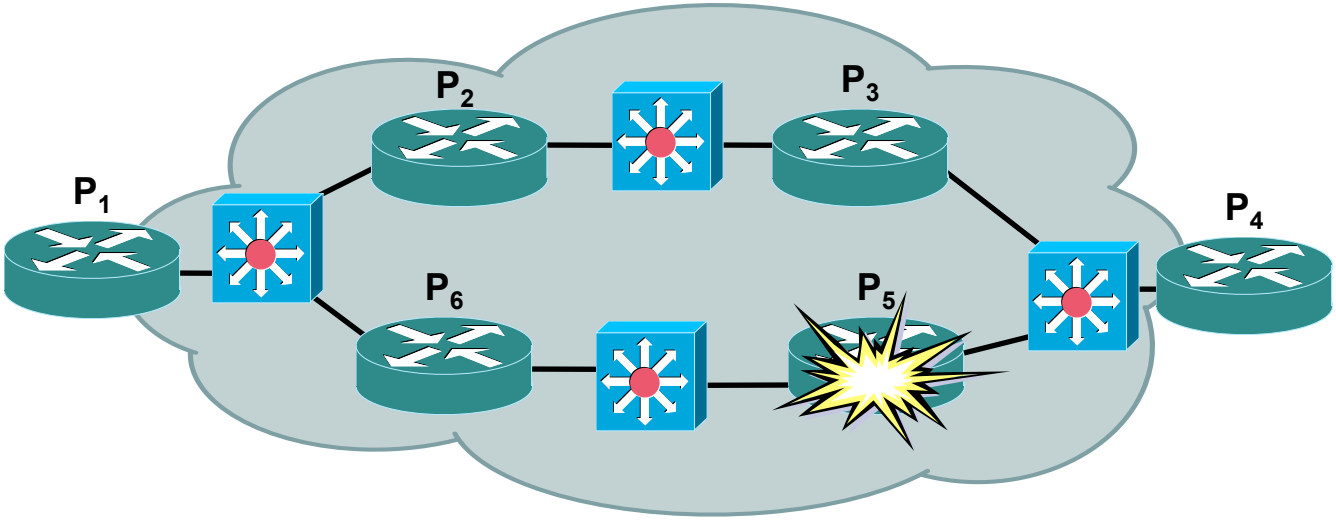
Bi-Direction Forwarding Detection (BFD)

BFD Protocol Overview

- Described in draft-katz-ward-bfd-02.txt and draft-katz-ward-bfd-v4v6-1hop-01.txt (Single Hop application)
- Similar in concept to HELLO or heartbeat-type protocol based on 3-way handshake—BFD is Protocol Independent
- Neighbors exchange unicast hello packets at negotiated regular intervals
- A neighbor is declared down when expected hello packets don't show up
- Two different methods
  - Control (Async) Mode
  - Demand Mode
  - Echo Function
- BFD payload packets are sent using encapsulation of each protocol/ connection you want to monitor (IPv4, IPv6, 802.3, etc)
- BFD control packets will be encapsulated in UDP datagram



BFD Example



- In normal scenarios if P5 goes down, P4 and P6 would rely on their IGP/BGP timers to detect the failures
- With BFD failure can be detected in less than a second

```
P4# router ospf 1
P4(config-router)# bfd all-interfaces
```

If you don't want to enable on all the interface you can use

```
P4(config-router)# bfd interface Gig 4/0
```

Following configuration may be needed on on an interface

```
[no] bfd interval <50-999> min_rx <1-999> multiplier <3-50>
```

BFD Verification CLI


```
P4#show bfd neighbor detail
Cleanup timer hits: 0
Pseudo pre-emptive process count: 115857 min/max/avg: 8/8/8 last: 4 ms ago
OurAddr      NeighAddr     LD/RD RH  Holdown(mult)  State     Int
1.1.1.1      1.1.1.2         1/1   1   150  (3 )          Up        Et2/1
Local Diag: 0, Demand mode: 0, Poll bit: 0
MinTxInt: 50000, MinRxInt: 50000, Multiplier: 3
Received MinRxInt: 50000, Received Multiplier: 3
Holdown (hits): 150(25), Hello (hits): 50(12265)
Rx Count: 12284, Rx Interval (ms) min/max/avg: 40/64/49 last: 0 ms ago Tx Count: 12401, Tx Interval
    (ms) min/max/avg: 40/64/49 last: 12 ms ago Registered protocols: Unknown
Uptime: 00:11:53…….
```

Print debugging information about BFD packets sent and received. The optional access-list is used to filter based on neighbor IP address

debug bfd [access-list <list>]

Print debugging information about BFD state transitions

debug bfd state

Agenda

- High Availability Overview
- Device and Link Level Resiliency
- Protocol Level Resiliency
- Network Level Resiliency
- Operations and management
- **Summary**

Summary

- Cisco is enhancing its portfolio to add features for improved full HA solution
- MPLS HA features provide stateful switchover and NSF capability for VPN, LDP, AToM, etc.
- Need IP HA enabled to support MPLS HA
  - GR must be enabled for all participating RPs (OSPF, BGP, IS-IS) on P, PE, and CE routers
- For MPLS VPN HA: LDP HA, BGP HA is required
- AToM NSF/SSO is exactly the same as directed LDP
  - AToM application will do checkpointing for local labels only
- Higher availability achieved with fast convergence (e.g. Fast IGP/FRR, NHT, LDP/IGP synch etc)
- Operational efficiency improved with new features like LDP Auto-config, BFD and other OAM tools

Complete Your Online Session Evaluation!

Por favor, complete el formulario de evaluación.

Muchas gracias.

Session ID: RST-3108

MPLS HIGH AVAILABILITY