

# Designing Networks for High Availability



Paresh Khatri (paresh.khatri@alcatel-lucent.com.au)

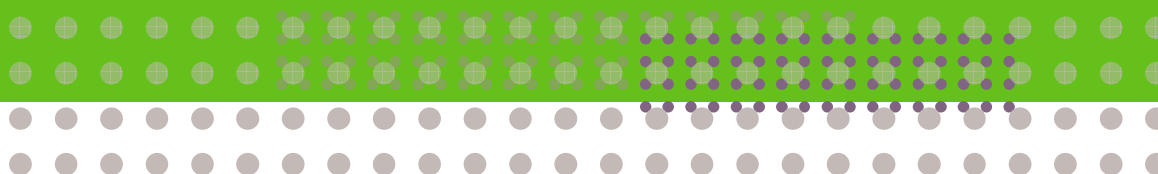
July 2009

## Agenda

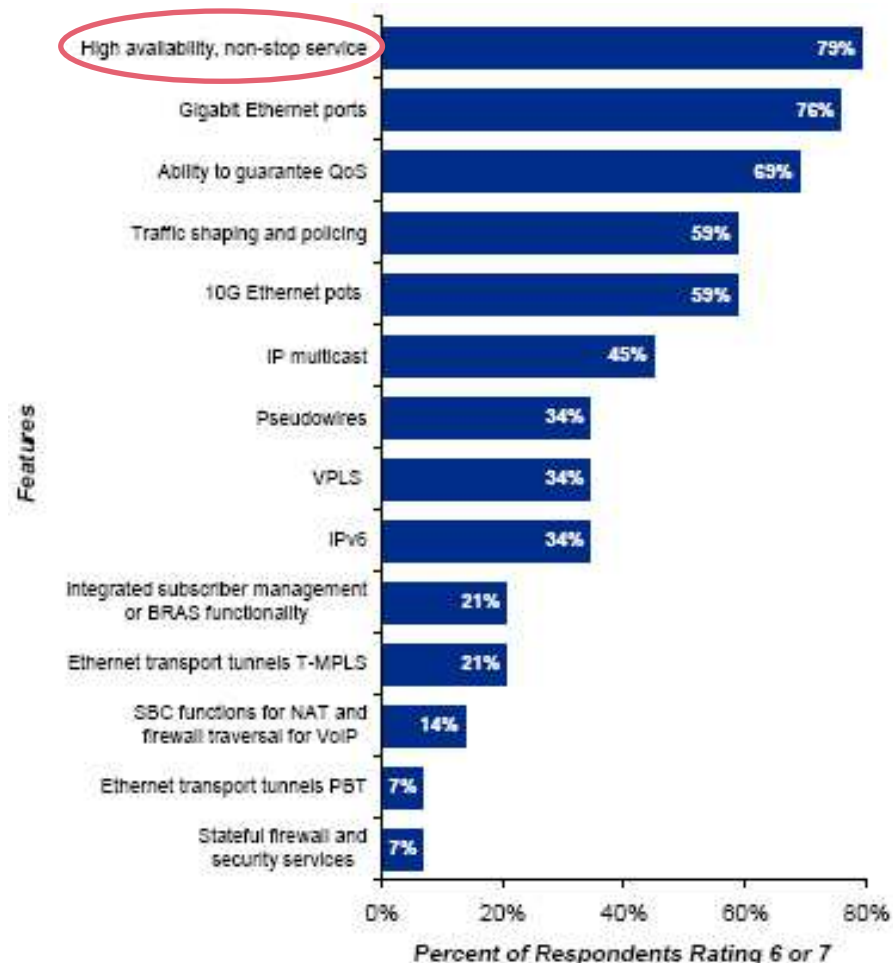
1. Drivers for High Availability
2. Key Components of a Highly Available Network
3. Router Resiliency
4. Network Resiliency
5. Questions ?

# 1

## Drivers for High Availability



## High availability is a top priority for IP operators. Why now? . . .



“The top four criteria operator respondents use in selecting an IP router switch manufacturer are the same as in last year’s study, and all are at higher ratings, with product reliability reaching 79 percent this year.”

Source: Infonetics Research: Service Provider Plans for IP/MPLS: North America, Europe & Asia Pacific 2007, December 2007

Why high availability is a top priority:  
 Customer expectations for always available service

High availability has moved from being a requirement for business services to a requirement for all services because the TV and phone are now part of the bundle.

**BUSINESS COMMUNICATIONS**



**CONSUMER ENTERTAINMENT**



“Always Available”  
 Service  
 Expectation

## Why high availability is a top priority: The impact of downtime

Downtime translates into a major cost to you and to your customers.

### Impact to You:

- Operational and customer service costs to repair
- Warranty costs/Service Level Agreement (SLA) penalties
- Disrupted operations
- Lost business and revenue
- Perceived brand value and quality

**RISK**

### Your Business

#### Business Continuity

Service  
Continuity

Customer  
Satisfaction

Operational  
Continuity

Revenue  
Continuity

#### Customer Experience

According to one study by Network Strategy Partners, the annual cost of unplanned downtime for a typical large carrier ranged from \$960,000 to over \$13 million per year.

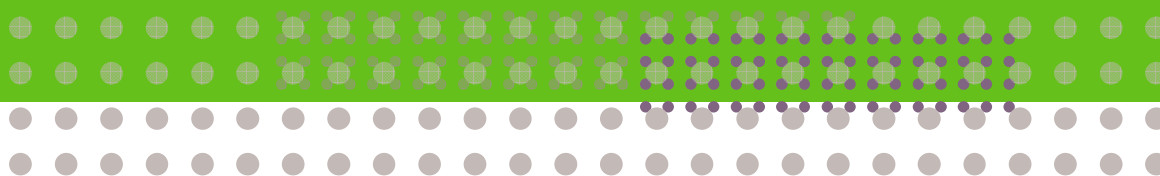
## High availability for next-generation IP networks

The goal of next-generation IP high availability is to deliver continuous IP service availability.

- Next-generation IP high availability provides an achievable IP service availability of five-nine's or greater.
- Next-generation IP high availability can also be measured by:
  - The percentage of time that a carrier can meet Service Level Agreements (SLAs)
  - The impact on brand quality, customer satisfaction, business profitability

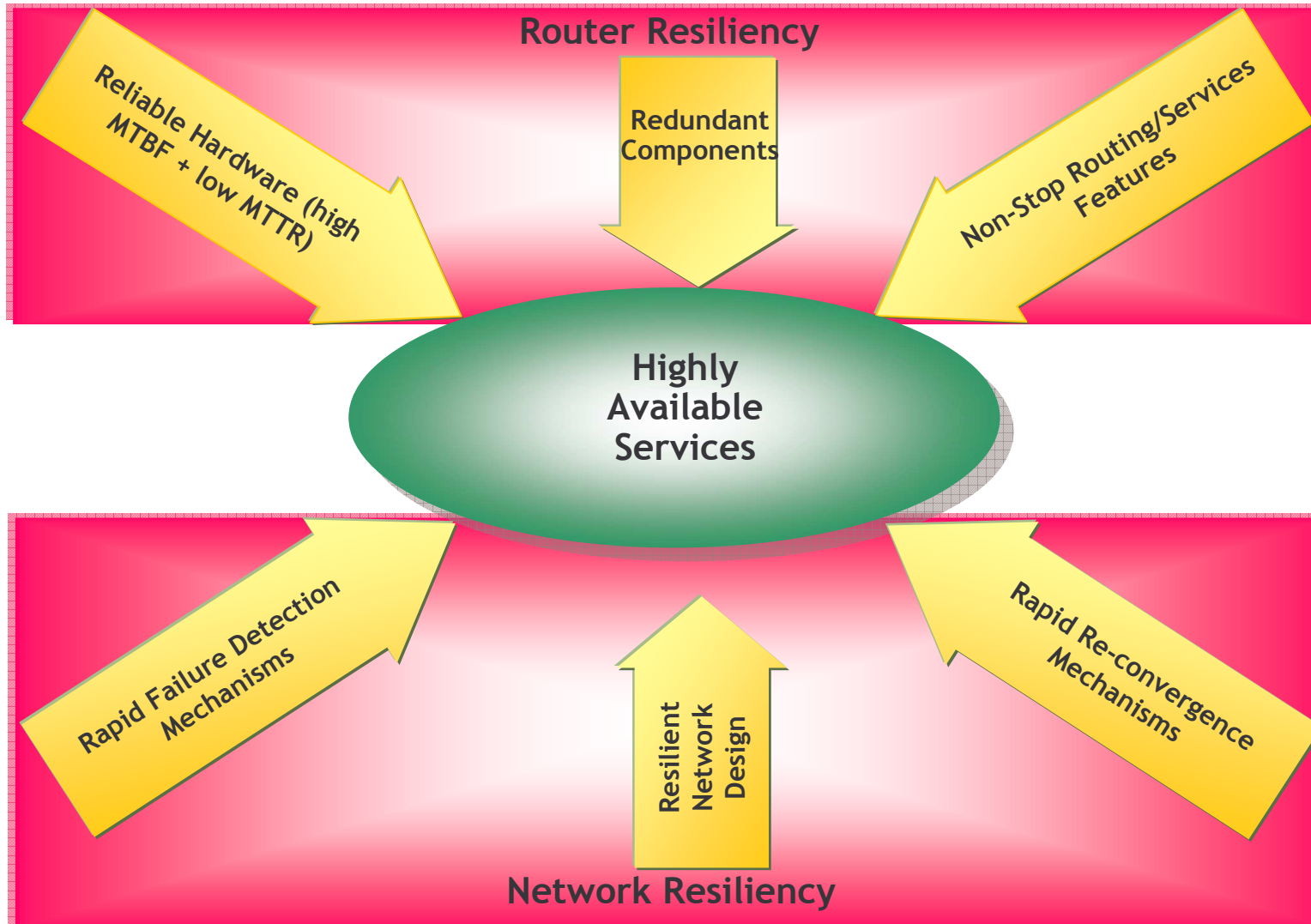
Availability	Service Downtime per Year	
99.9000%	8 hours, 46 minutes	} ← IP High Availability Performance Today
99.9900%	52 minutes, 33 seconds	
99.9990%	5 minutes, 15 seconds	} ← Next-generation IP High Availability Performance
99.9999%	31.5 seconds	

# 2 Key Components





## Key Components of a Highly Available Network

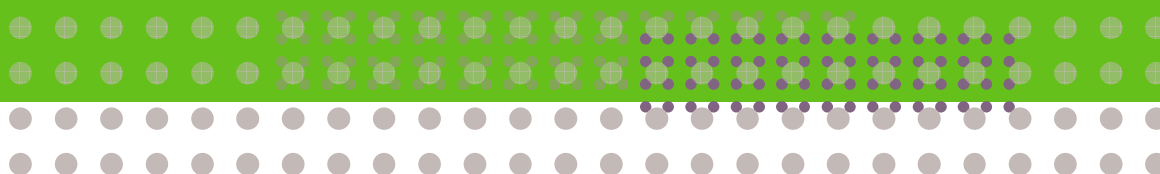


## Key Components of a Highly Available Network

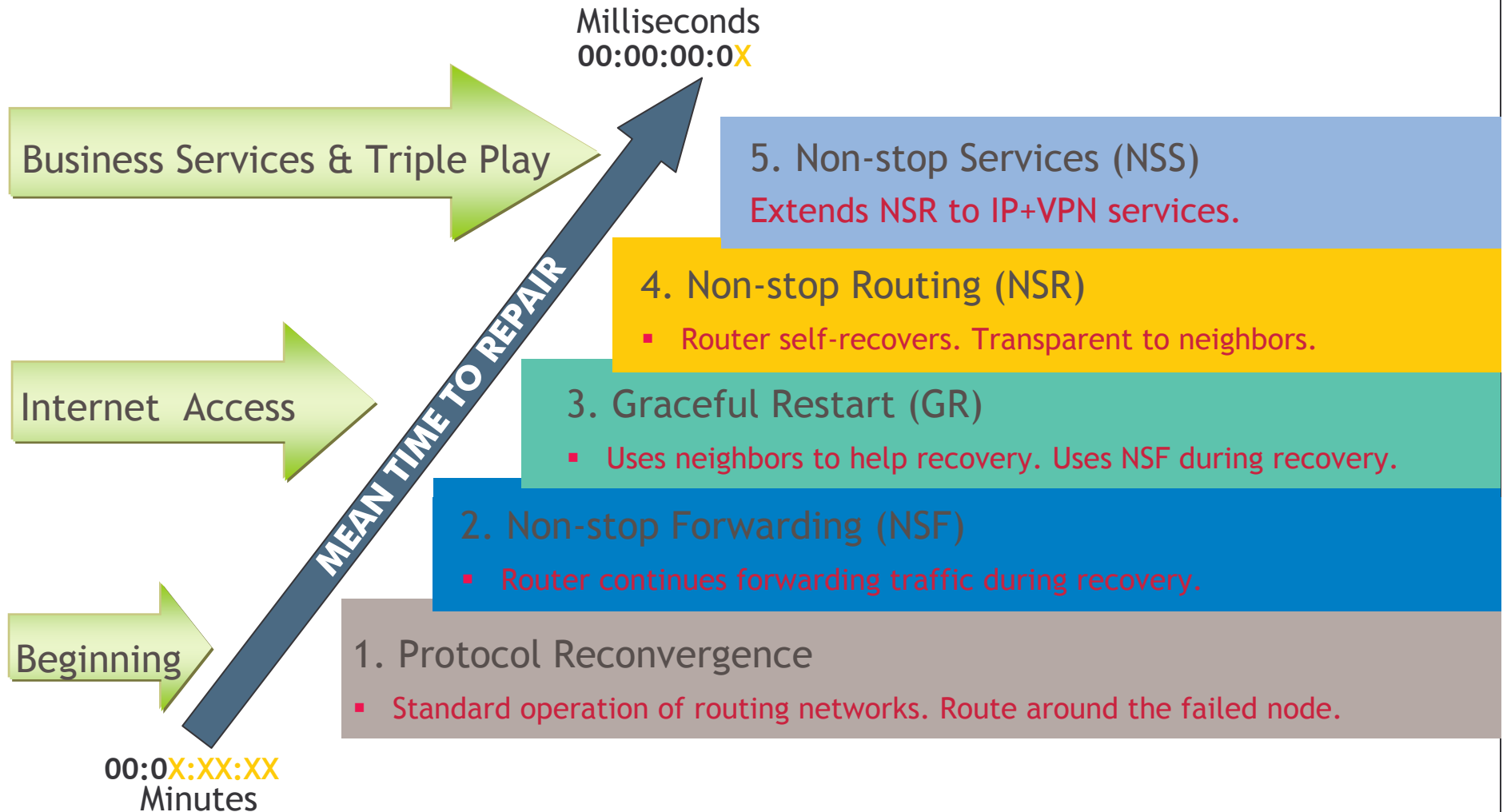
---

**High IP service availability is the result of building and maintaining resiliency at multiple network levels.**

# 3 Router Resiliency



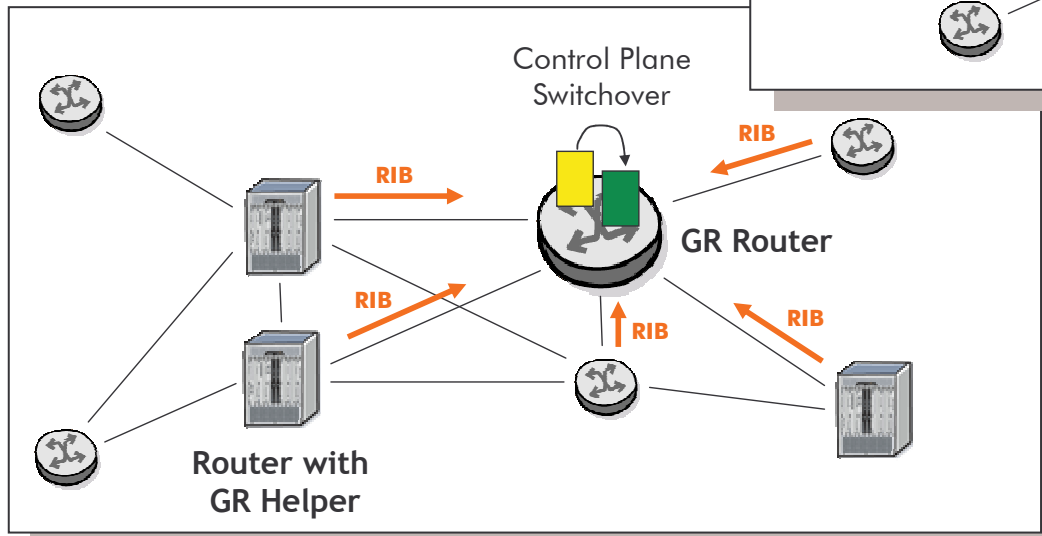
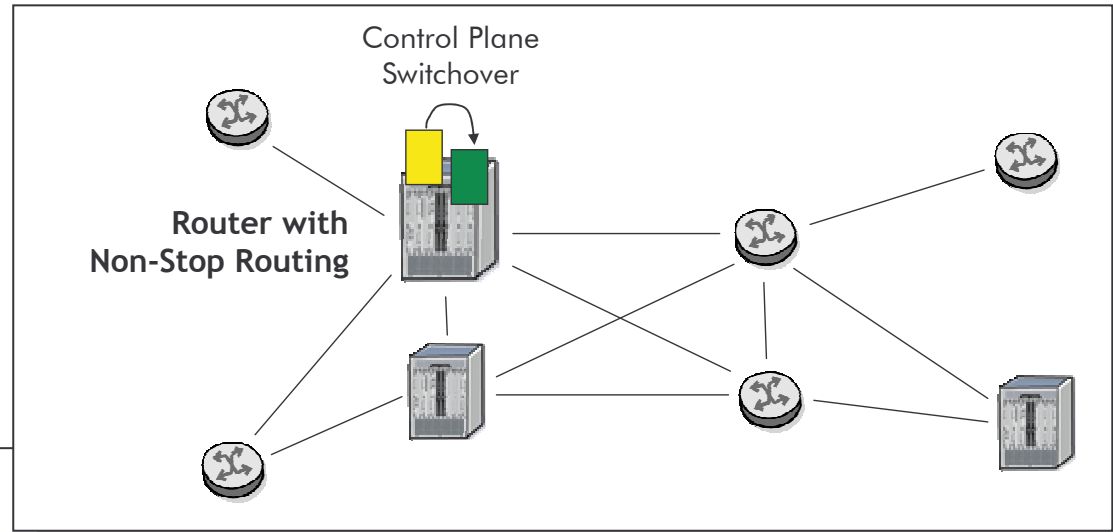
# The Path to Non-Stop Services







# Recovery Comparison - Network Behavior

**Non-Stop Routing** →  
Node-level Recovery

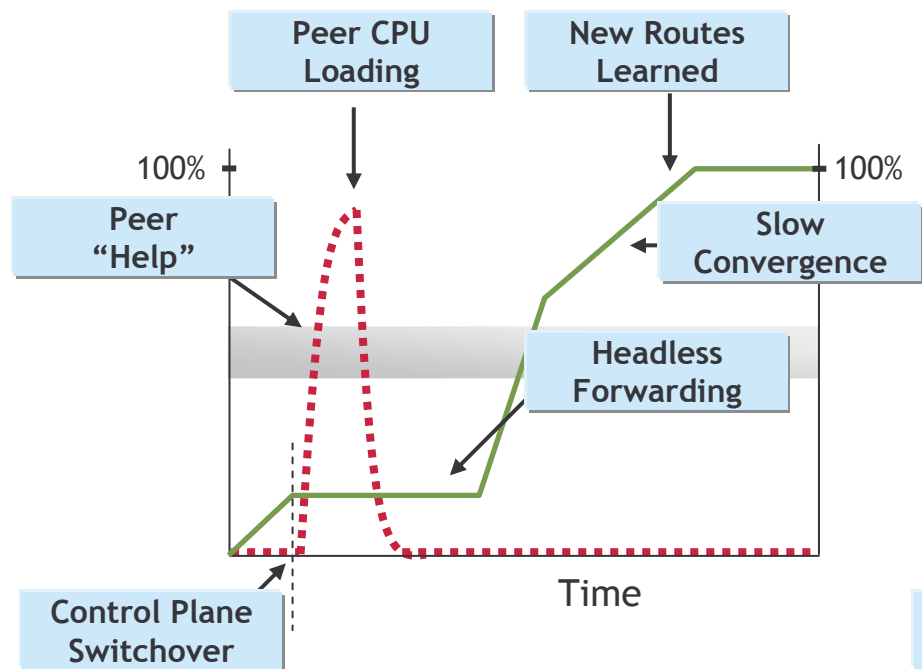
**Graceful Restart** ↙  
Network-level Recovery



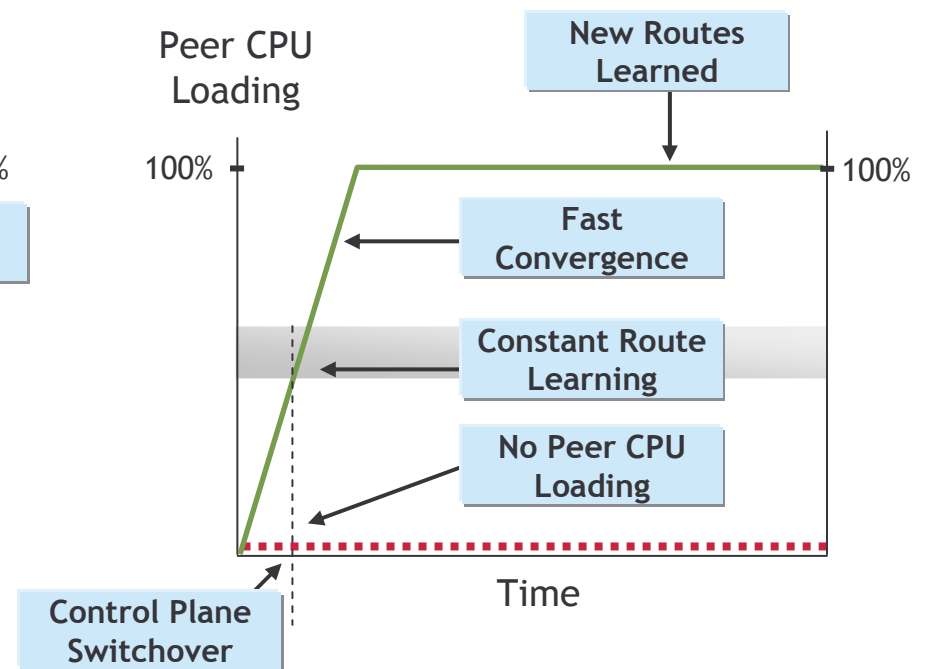
-  GR router
-  Router with GR Helper
-  Newly inactive control plane
-  Newly active control plane

# Recovery comparison: Results

## Graceful Restart



## Non-stop Routing



**“Stop-and-Restart-Routing”**  
Network-impact, Peers Help

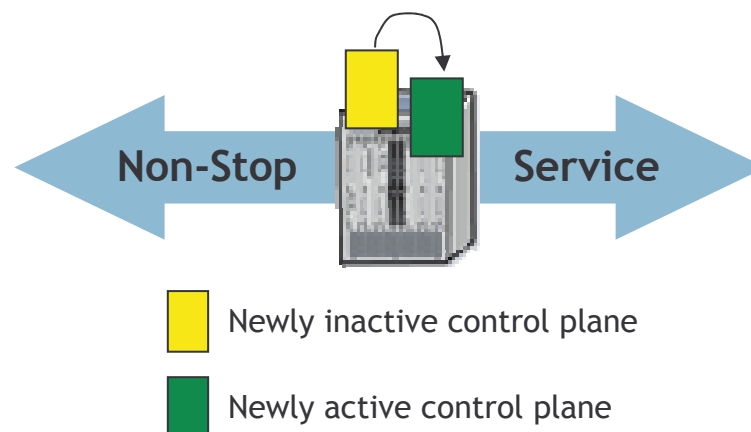
**“Non-stop” Routing**  
Self-contained & Transparent

## Non-Stop Services

### Extending non-stop routing to IP/VPN/3Play services

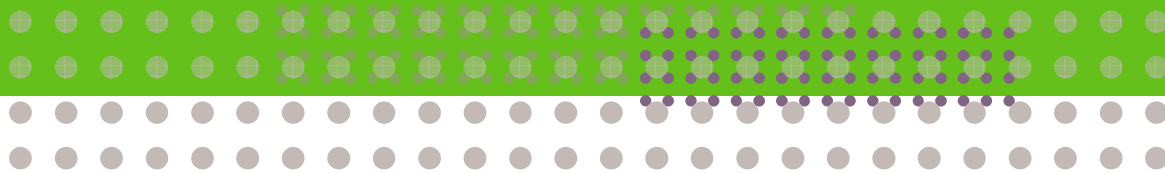
- Non-stop enterprise IP and VPN service
- Non-stop residential voice & video services
  - <math><1\mu\text{s}</math> & <math><4\text{ms}</math> recovery times for h/w & s/w induced switchovers
    - No service outages
    - No SLA violations
  - Continuous dynamic routing
    - No “headless” forwarding
    - No “black holes”

### High Availability Control & Fabric Switchovers



- Self-contained solution in every respect
  - No dependencies on adjacent routers
  - No routing interoperability issues
  - No post-switchover reconvergence

# 4 Network Resiliency



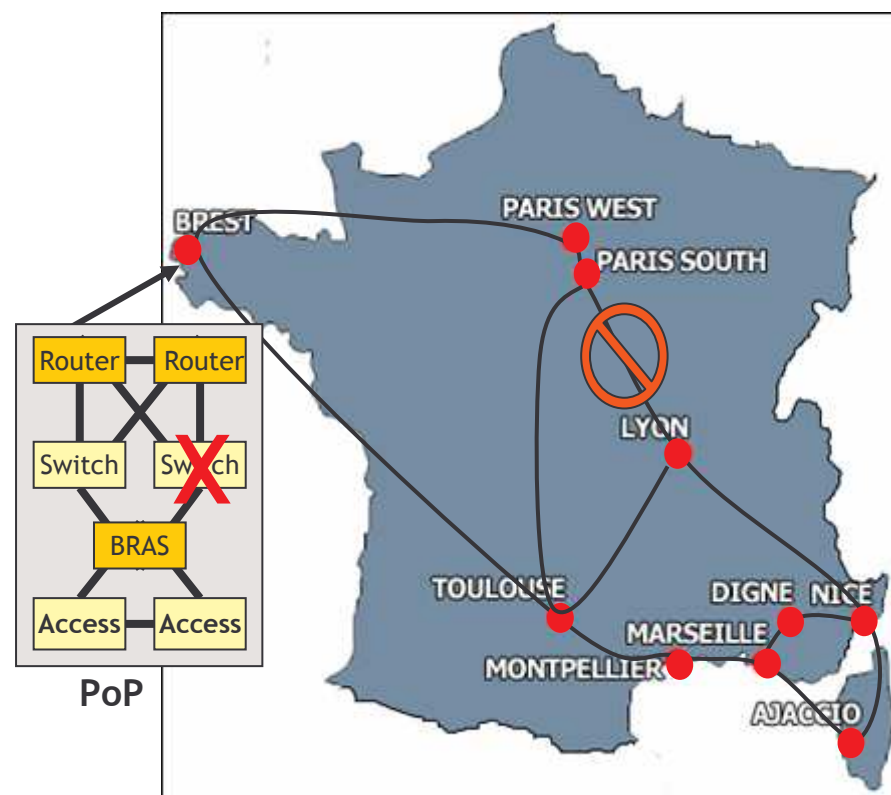


## Network design and architecture

Establishes the baseline network for delivering high-performance services at the lowest possible cost and assuring SLAs can be met.

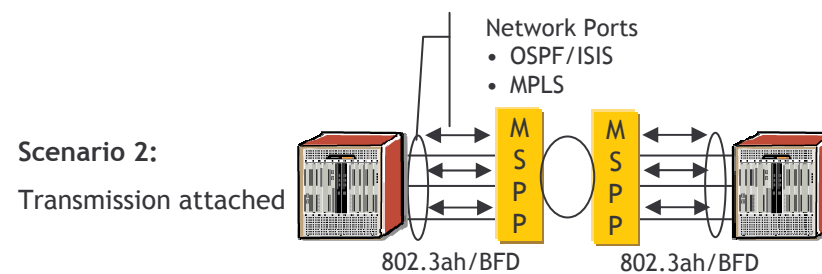
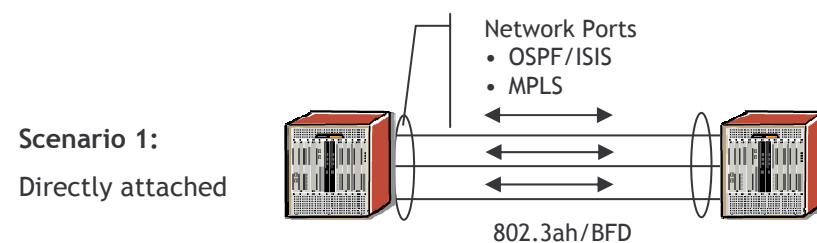
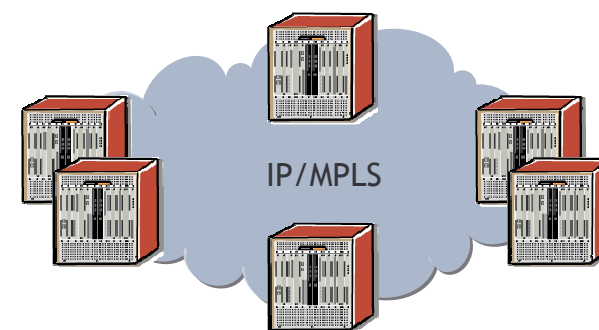
### Key considerations

- Network topology, connectivity, and routing
- SLA requirements
- Bandwidth requirements
- PoP design
- Single points of failure
- Redundancy strategy and cost
- Security vulnerabilities
- Disaster recovery
- Available budget
- Service and network migration requirements

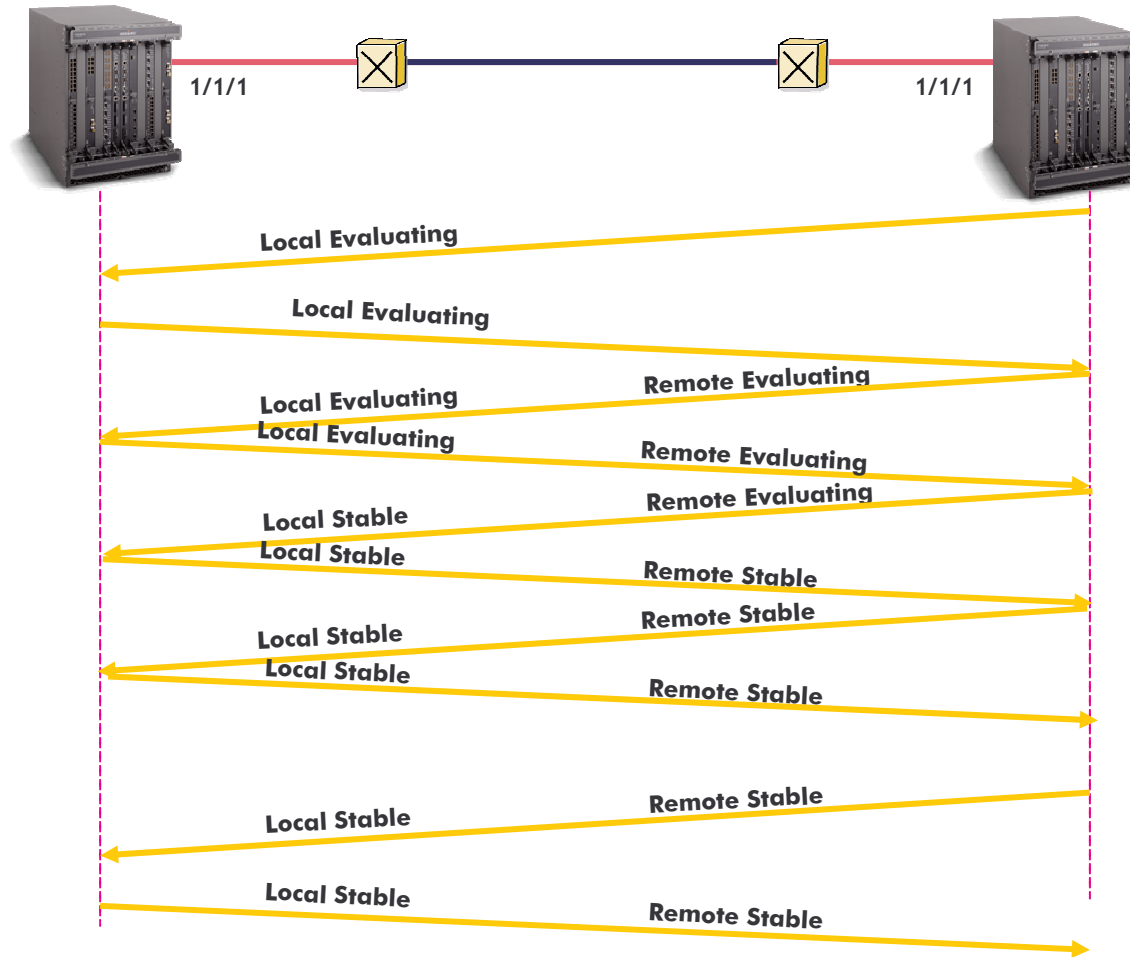


## Network layer fault recovery: Fault detection

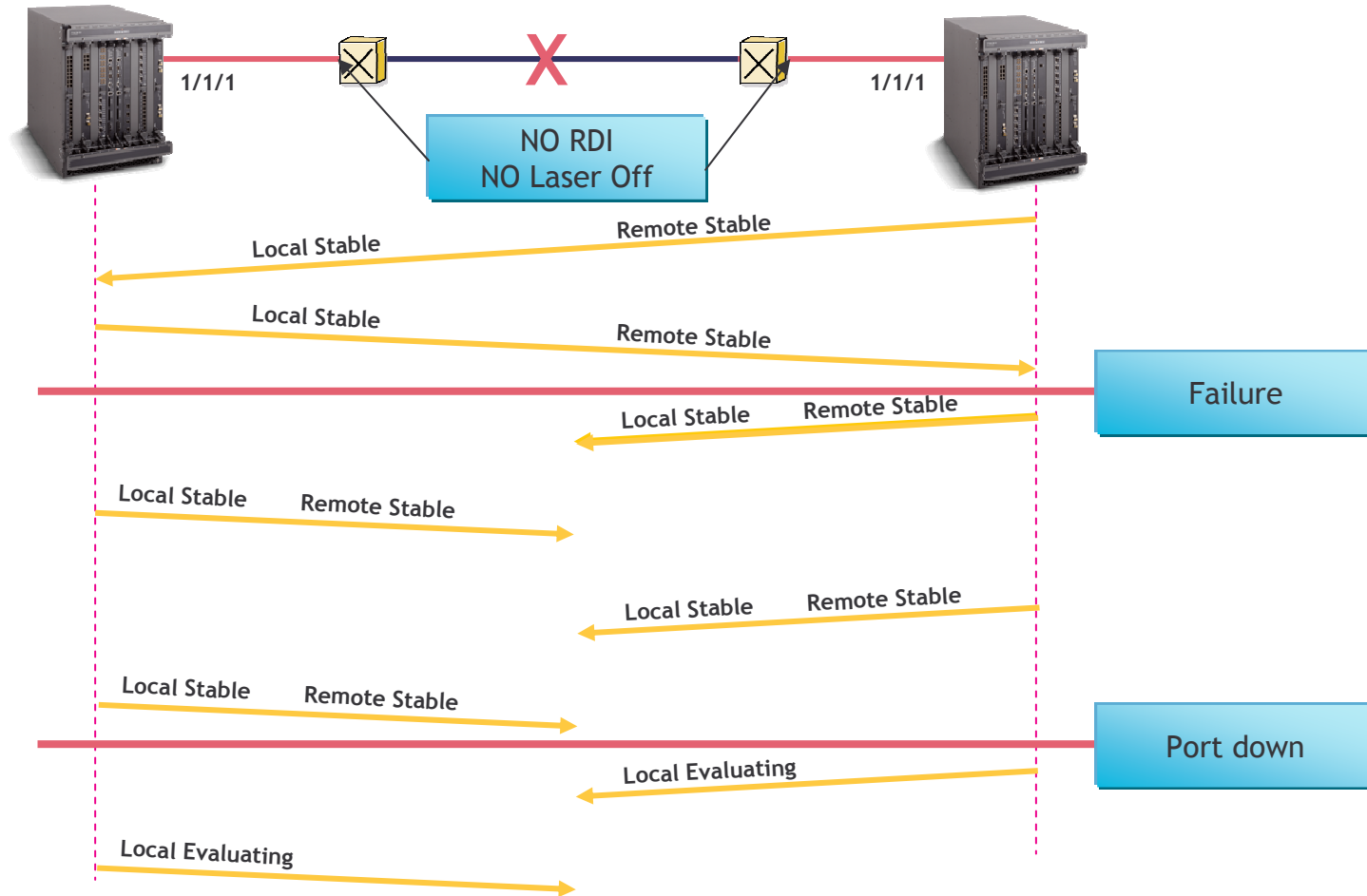
- Fault recovery time: detection + recovery
- A trigger is required to activate the re-convergence protocols
  - OAM or LOS for Sonet/SDH
  - LOS or RDI for Ethernet
- Direct connection or indirect connection may change the available triggers
  - Local link failure results in loss-of-light (LOS) with rapid local detection
  - Failures within a transmission network require propagation of failure or a higher-layer trigger
- When no other trigger is available consider BFD or 802.3ah
  - Can also act as a last resort trigger even with other mechanisms available
  - 802.3ah for Fast Reroute trigger
  - BFD for routing protocols
  - BFD over LAG
  - BFD for RSVP (BFD triggers equivalent actions as interface down)



# 802.3ah Example - OAM Capability Discovery



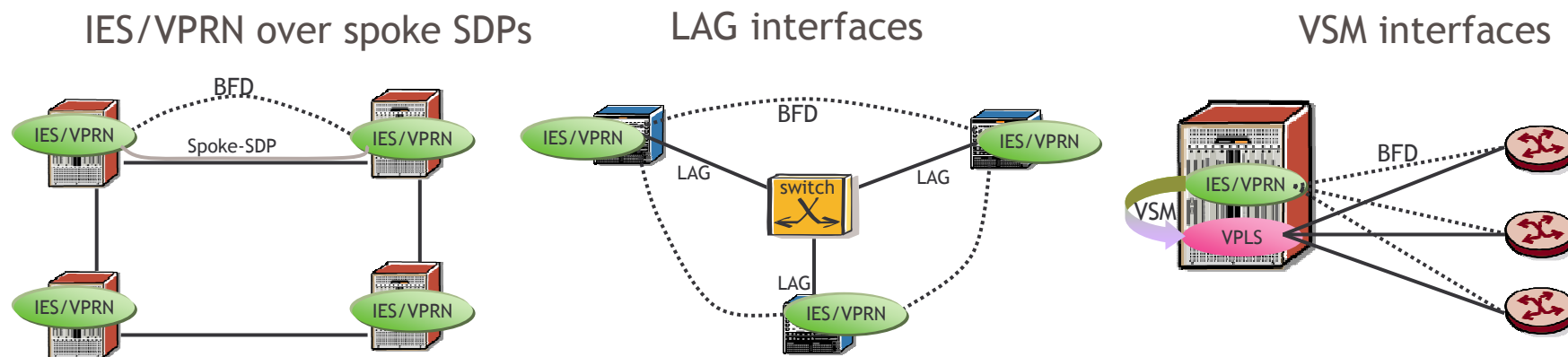
# 802.3ah Example - Using OAM bidirectional link failure indications



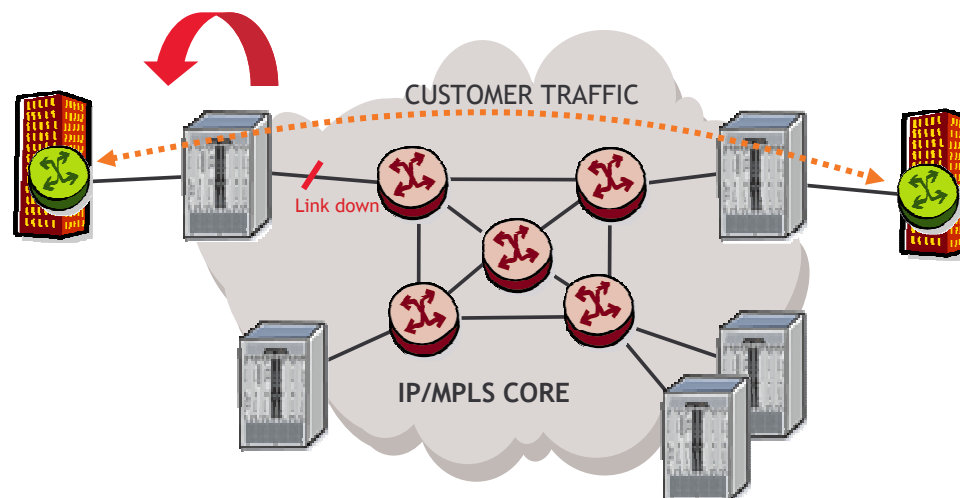


## BFD usage options

- BFD support for: OSPF, IS-IS, PIM, BGP and Static Routes
- BFD on RSVP Interface
- BFD can run over Router Base NW, VPRN interfaces and on IES interfaces
- Centralized and distributed options
- Centralized BFD for:



## OAM enhancements: Link-level Loss Forwarding

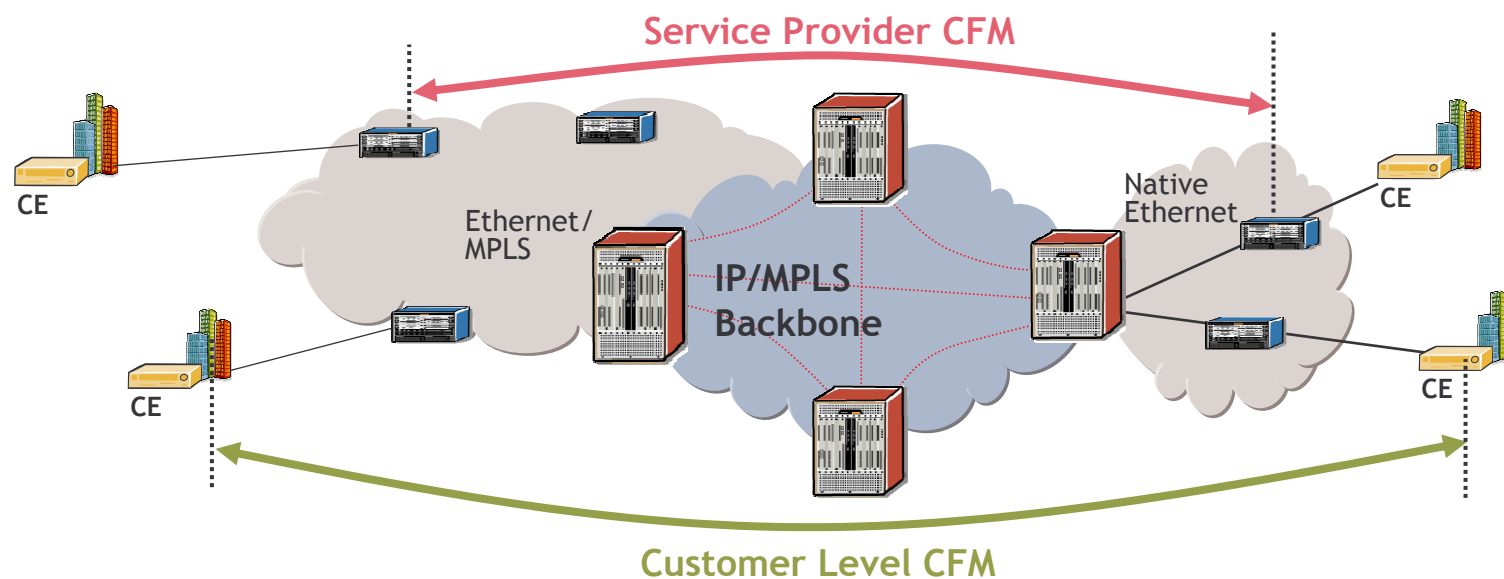


- Need e2e OAM fault notification for Ethernet VLL (pseudowire) service
  - End user wants to activate a backup connectivity over another service provider
- This feature shuts down the laser on the interface to the CE under one of the following conditions:
  - Local fault on the PW or service
  - Remote fault on the Attachment Circuit or Pseudowire (signaled with label withdrawal or T-LDP status bits)

## Ethernet OAM for services and network fault detection

### Ethernet level OAM - 802.1ag

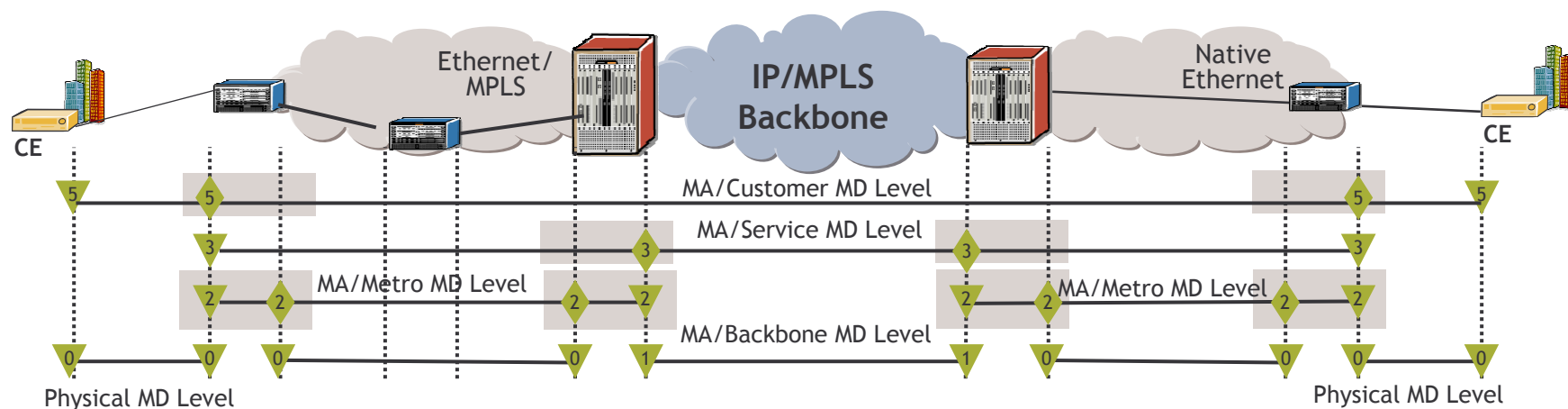
- All the way to the customer CPE
- When a network fault is detected, it generates alarms indicating the location of the fault
- Not suitable for fast protection of the network (use 802.1ah EFM, BFD)
- Support for Management End Points (MEP), Intermediate Points (MIP) & Management Domains (MD)
- Fault management tools
  - Loopback message/reply
  - Linktrace message/reply
  - Periodic Continuity Checks





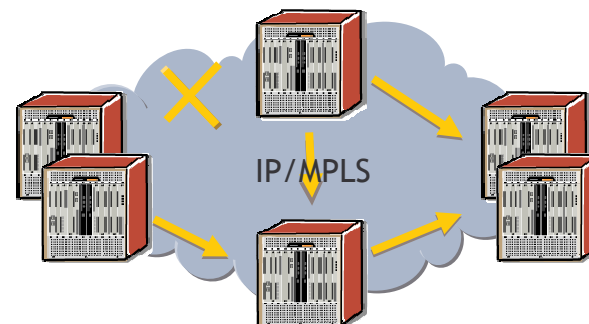
## IEEE 802.1ag connectivity fault management

- MEP: Maintenance end points
- MIP: Maintenance intermediate points
- MD: Maintenance domain - defines a level (0-7) and a span of control
  - Higher level = Wider domain
  - The levels also define a hierarchy. MDs can nest but cannot overlap.
  - All CFM messages exchanges flow within the domain. Lower level MEPs are forwarded transparently.
- MA: Maintenance Association - defines a set MEPs and MIPs that share a common MD.



## Network layer convergence change propagation: IGP

- Upon detection of a link failure, the local node must generate a new IS-IS LSP/OSPF LSA to reflect the current state of its local interfaces
- The time for a network to fully converge following a link-state change is essentially derived from the following inputs:
  - Time taken for the source system to generate and flood the LSP/LSA to adjacent neighbors
  - Time taken for the LSP/LSA to propagate to adjacent neighbors
  - Time taken for the adjacent neighbors to re-flood the LSP/LSA and subsequently execute an SPF to re-compute the SPT topology. It is worthy of note that the LSP/LSA must be re-flooded BEFORE an SPF is executed.



- OSPF
  - spf-wait <max-spf-wait (msec)> <spf-initial-wait (msec)> <spf-second-wait (msec)>
  - lsa-arrival <msec>

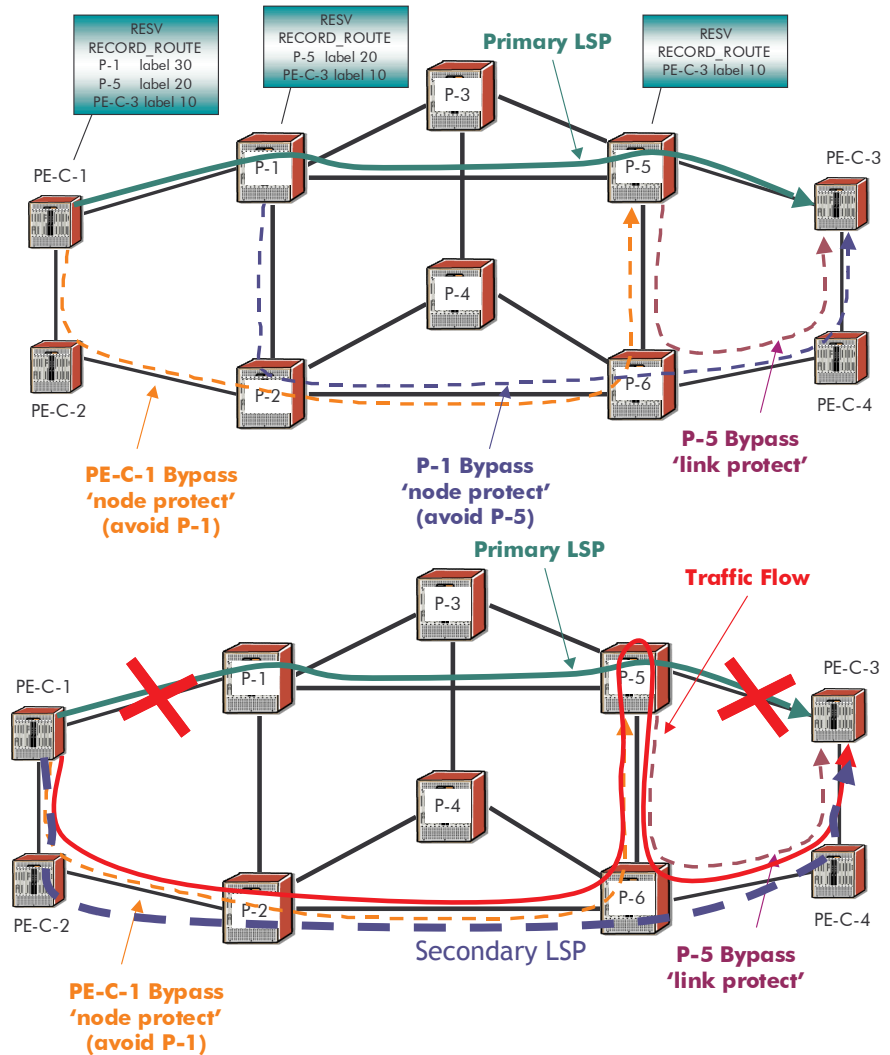
### Lsa-generate <msec>

- ISIS
  - spf-wait <spf-wait (s)> <initial-wait (msec)> <second-wait (msec)>
  - lsp-wait <lsp-wait (s)> <initial-wait (s)> <second-wait (s)>

# Network layer convergence

## Next best path calculation and path restoration: MPLS/IGP

- MPLS Fast Reroute (RFC4090) provides a standard restoration mechanism across the network
  - Facility: 1:n protection or Detour: 1:1 protection
  - Link and node protection
  - Restoration time is sub-50 ms
- MPLS primary/secondary
  - Fault propagated to the LER for primary/secondary decision
  - Head-end decision on primary/secondary
  - More control over the routing of the protected path
- LDP
  - Simple/scalable protocol to operate (multipoint to point approach)
  - Relies on the IGP convergence (200-300 ms depending on the network span)



- MPLS Fast Reroute usually refers to several aspects related to RSVP-TE failure recovery
- Failover sequence of events (link/node)
  1. Failure detected by adjacent LSR (for example, link failure)
  2. Adjacent LSR moves traffic onto pre-established bypass LSP (sub-50 ms)
  3. LSR propagates error message back to LER
  4. LER moves traffic onto secondary LSP (order of magnitude is 100s of ms depending on configuration)
  5. LER attempts to re-establish new primary path (global revertive)
- When primary LSP recovers, traffic re-establishes on the primary LSP
  - Make-before-break to re-established primary path
- Typically, for critical applications or strict constraints on path routing, the recommendation is to build a primary LSP with a secondary backup LSP
  - Detour/Bypass for sub-50ms link/node protection (may be sub-optimal)
  - A secondary backup LSP can be pre-engineered to provide a 2nd best path during the primary failure

# Network layer convergence



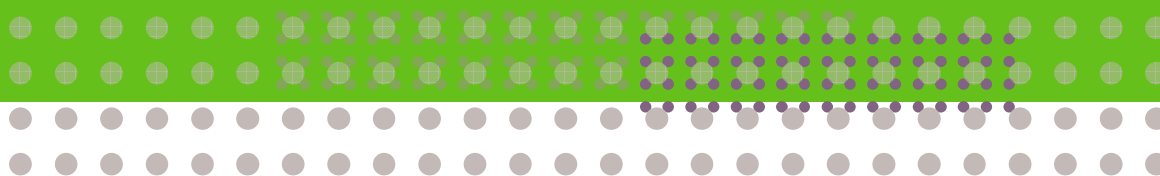
Next best path calculation and path restoration: LDP or RSVP

	LDP	RSVP
<b>Protocol</b>	Multipoint to point protocol	Point to point protocol
<b>Scalability</b>	Scales more simply, One LSP per PE	Mesh issues in large networks Complexity to mitigate
<b>Simplicity</b>	Simple operation	More control, more complexity
<b>Convergence</b>	200 Msec - 2 sec depending on the network topology	+/- 50 msec depending on protection design
<b>Traffic engineering</b>	Not available	Available
<b>Protection mechanisms</b>	LDP ECMP	FRR Multiple LSPs per SDP

# 5

## Questions

[paresh.khatri@alcatel-lucent.com.au](mailto:paresh.khatri@alcatel-lucent.com.au)



[www.alcatel-lucent.com](http://www.alcatel-lucent.com)

