

# BGP Prefix Independent Convergence

## draft-bashandy-rtgwg-bgp-pic-01

Authors :

Ahmed Bashandy, Cisco Systems

Clarence Filsfils, Cisco Systems

Pradosh Mohapatra, Sproute Networks

Presenter :

Ahmed Bashandy

IETF94, Nov/2015

Yokohama, Japan

# Agenda

- Problem
- Solution
- Handling Limited Hardware Capability
- Recovery from various failure scenarios

# Agenda

- Problem
- Solution
- Handling Limited Hardware Capability
- Recovery from various failure scenarios

# What do we want to Achieve?

- *Serial* nature of reachability propagation
  - ⇒ BGP convergence is *inherently* slow
- Can we adjust the data structure in time complexity that does ***not*** depend on BGP prefixes?
  - If a path is lost or gained, modifications should be independent of the number of prefixes
  - If there is a recovery, recover as soon as possible independent of the number of prefixes

## ***Objective***

**Make re-convergence after topology change independent of the number of BGP prefixes**

# Terminology

- Leaf: A prefix or local label container datastructure
- Path: a recursive or non-recursive path
  - May be primary or backup
- Pathlist: an array of paths
  - Each path carries its own pathindex
- OutLabel-Array: Array of outgoing labels and/or label actions\*
  - A possibly different outlabel-array attached to each leaf
  - Each entry represents an outgoing label and/or label action for a path in the pathlist\*\*
- Adjacency:
  - The L2 information to send a packet to a directly connected router

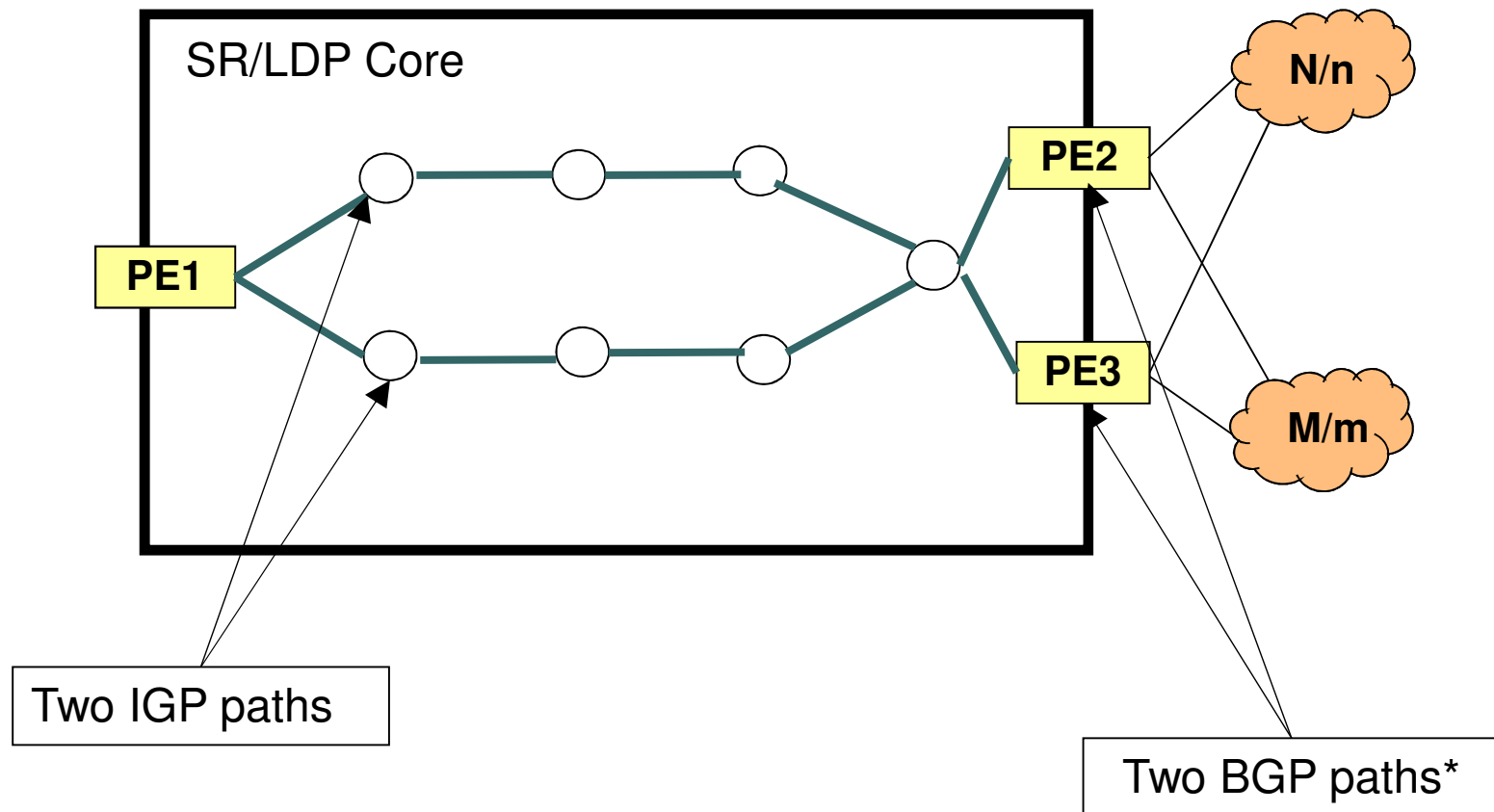
# Agenda

- Problem
- **Solution**
- Handling Limited Hardware Capability
- Recovery from various failure scenarios

# Basic Idea

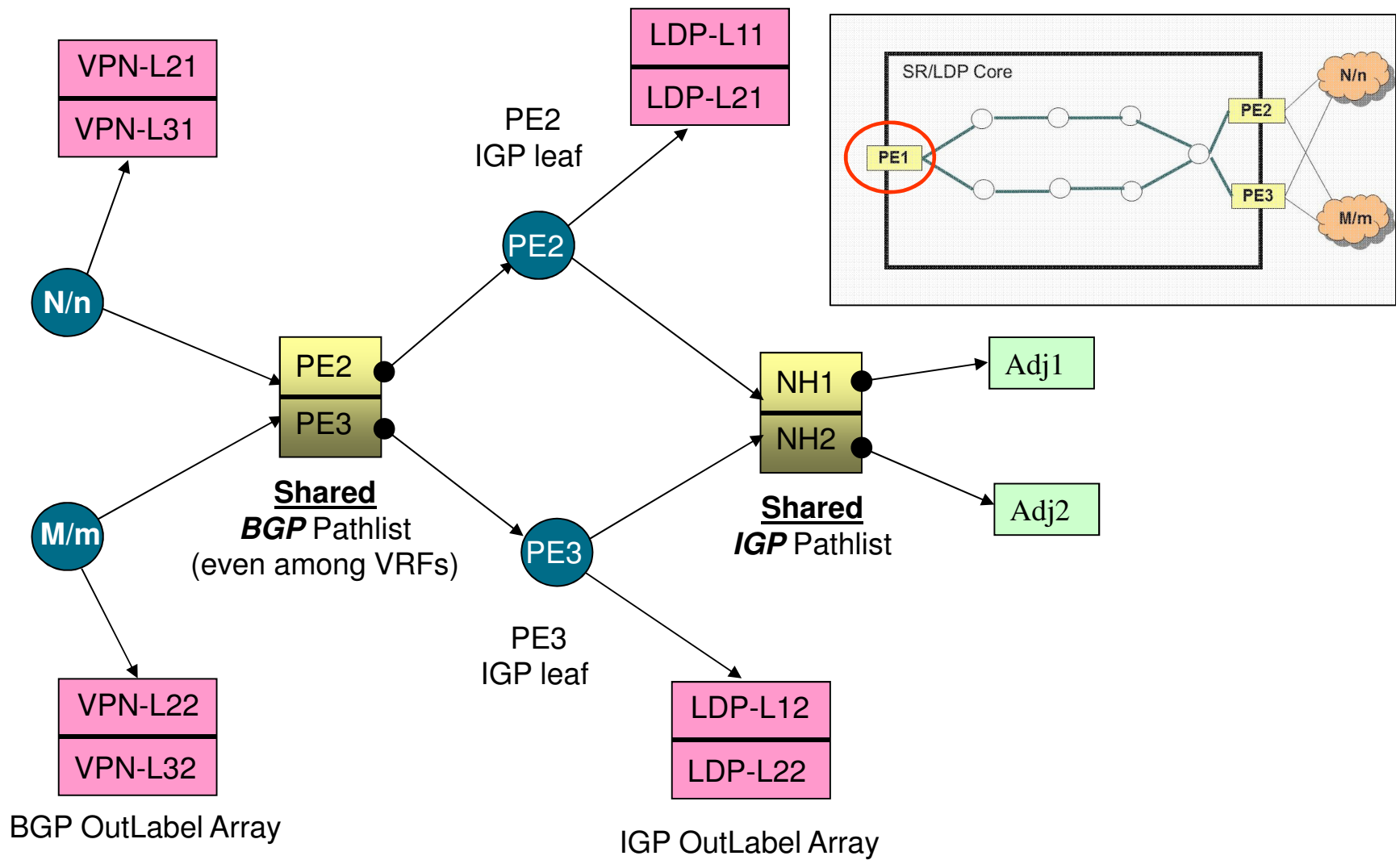
- BGP-PIC is really a **FIB** feature
- BGP-PIC need not involve BGP at all: e.g. works for pseudo-wires
- *Hierarchical* and *shared* forwarding chains
- On topology changes
  - FIB modifies pathlists immediately without modifying leaves
  - Restore traffic very quickly
  - BGP modifies leaves later at a slower pace
- Behavior is internal to the router
  - Completely *transparent* to operator
  - *Incrementally* deployable

# Sample Topology: VPN in SR/LDP Core





# Forwarding Chain on *Ingress* PE

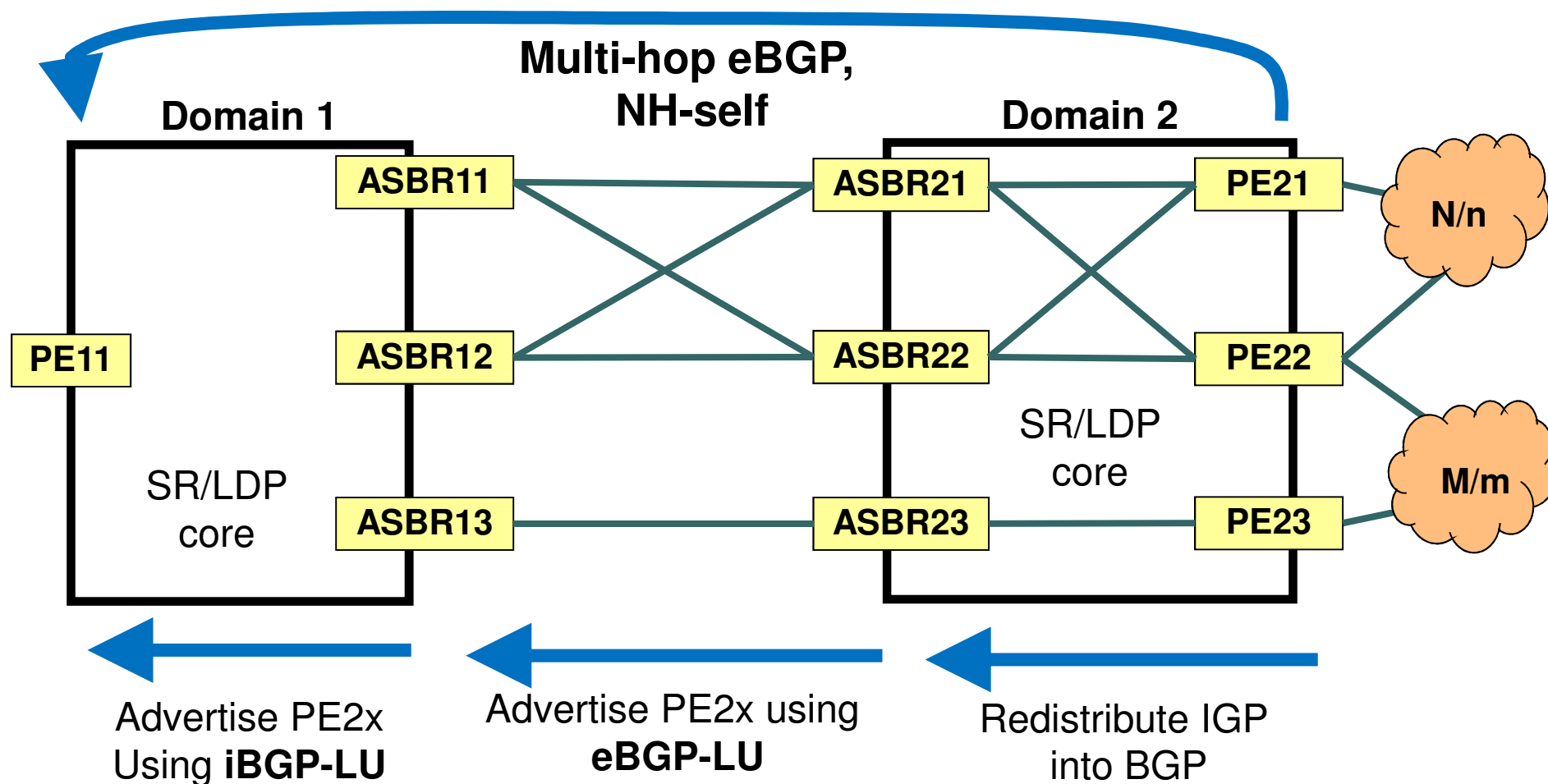


# Agenda

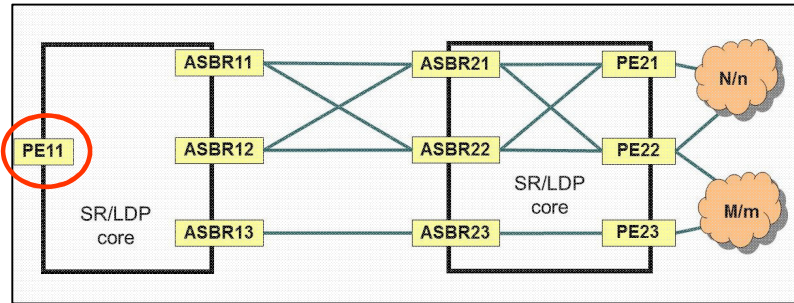
- Problem
- Solution
- Handling Limited Hardware capability
- Recovery from various failure scenarios

# Flattening a Forwarding Chain: Sample Topology

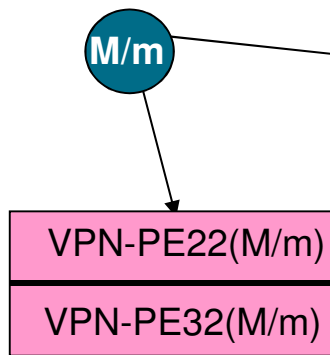
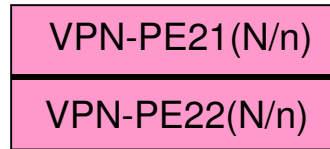
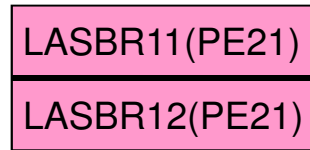
- Inter-AS Option C, but
  - ABRs advertise remote domain core using **BGP** instead of IGP



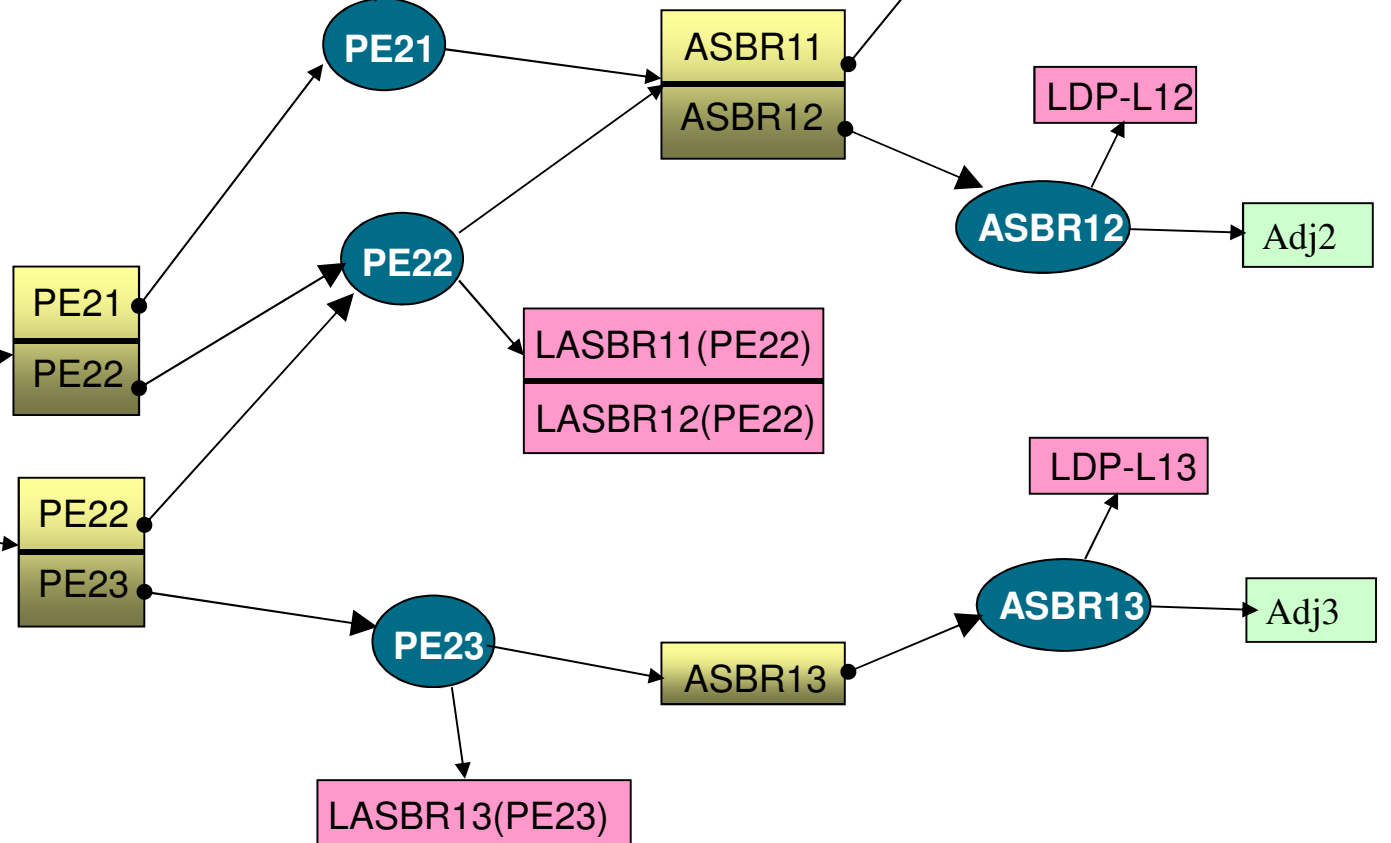
# Forwarding Chain on PE11



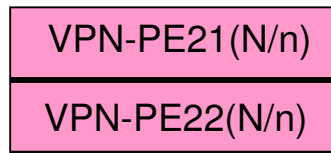
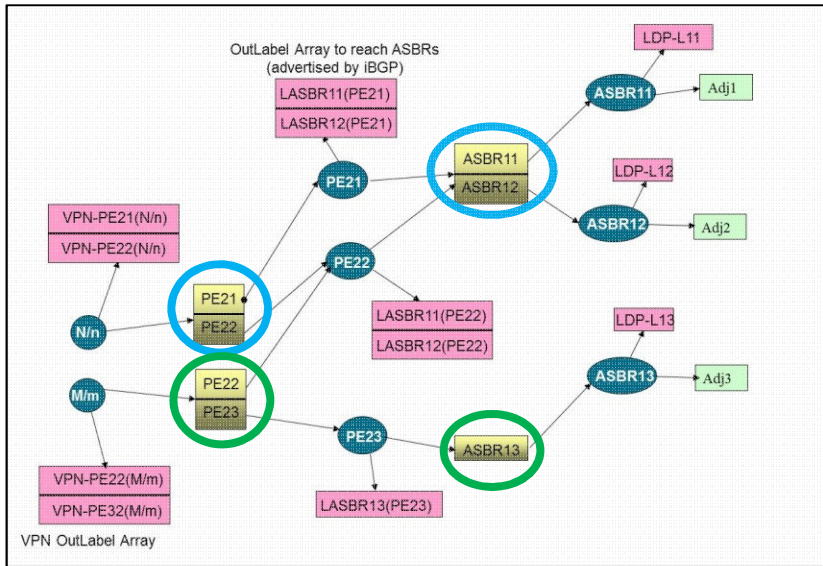
OutLabel Array to reach ASBRs  
(advertised by iBGP)



VPN OutLabel Array

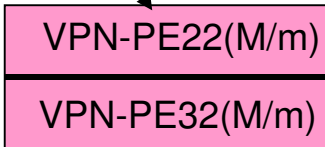


# Flattened\* Forwarding Chain on PE11

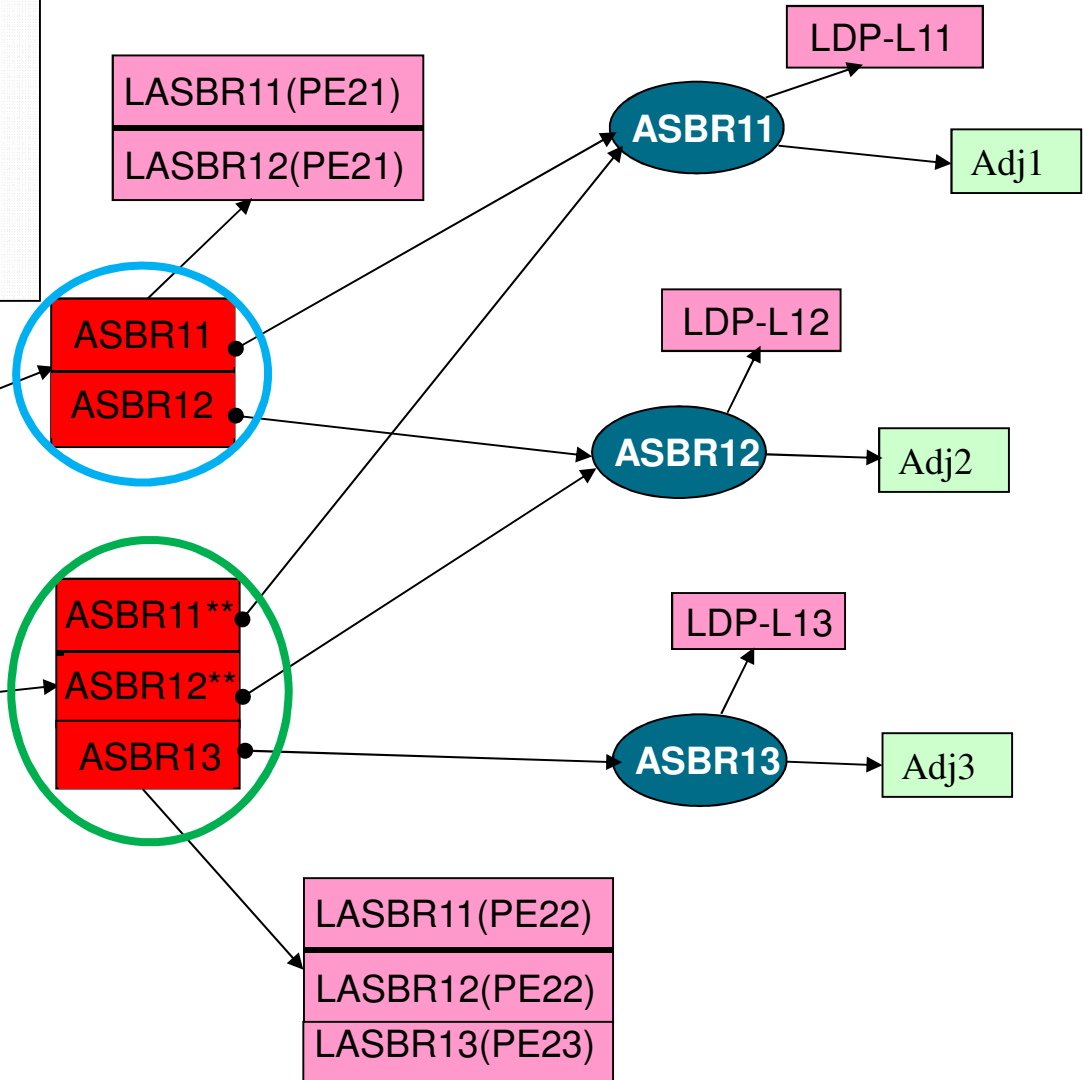


N/n

M/m



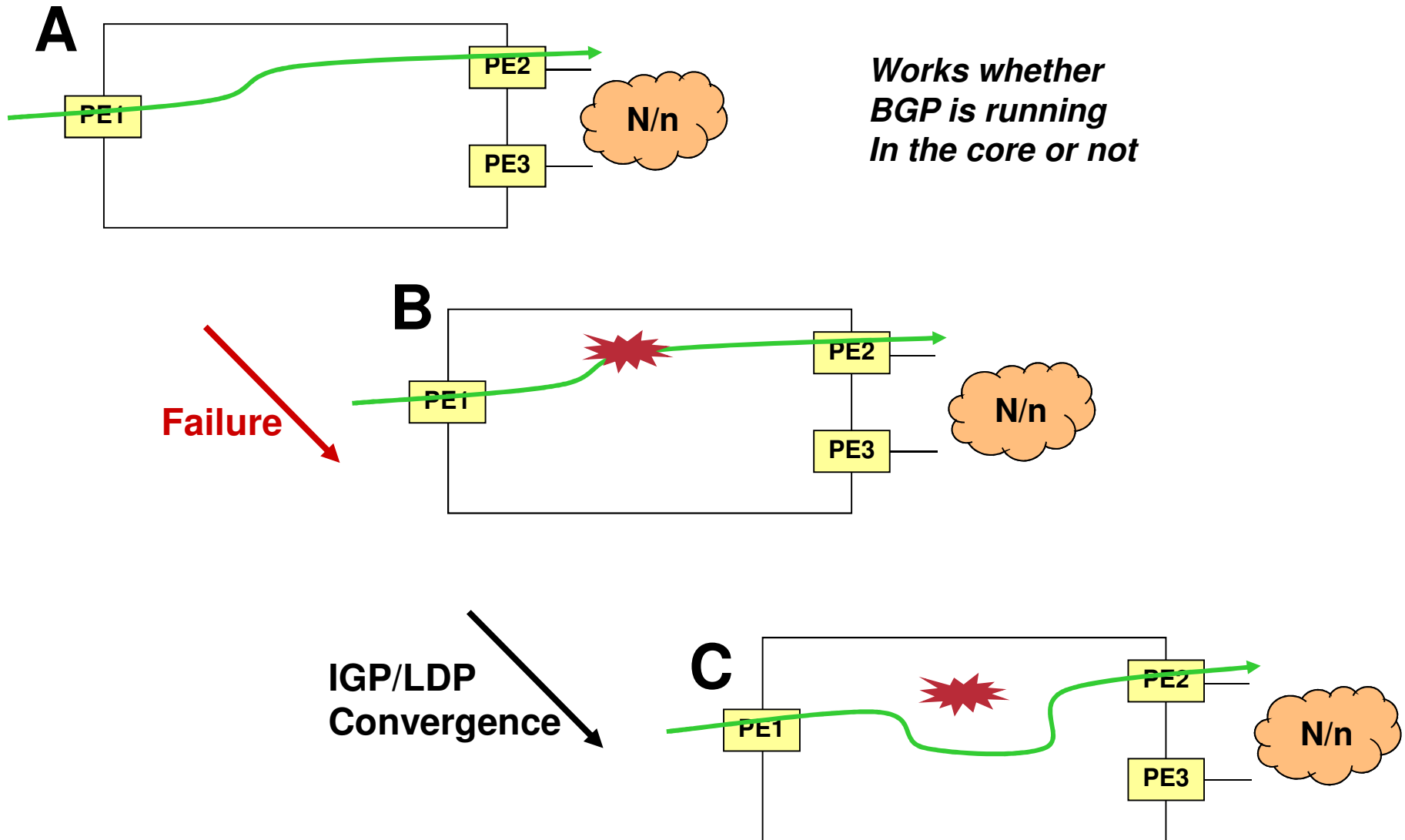
VPN OutLabel Array



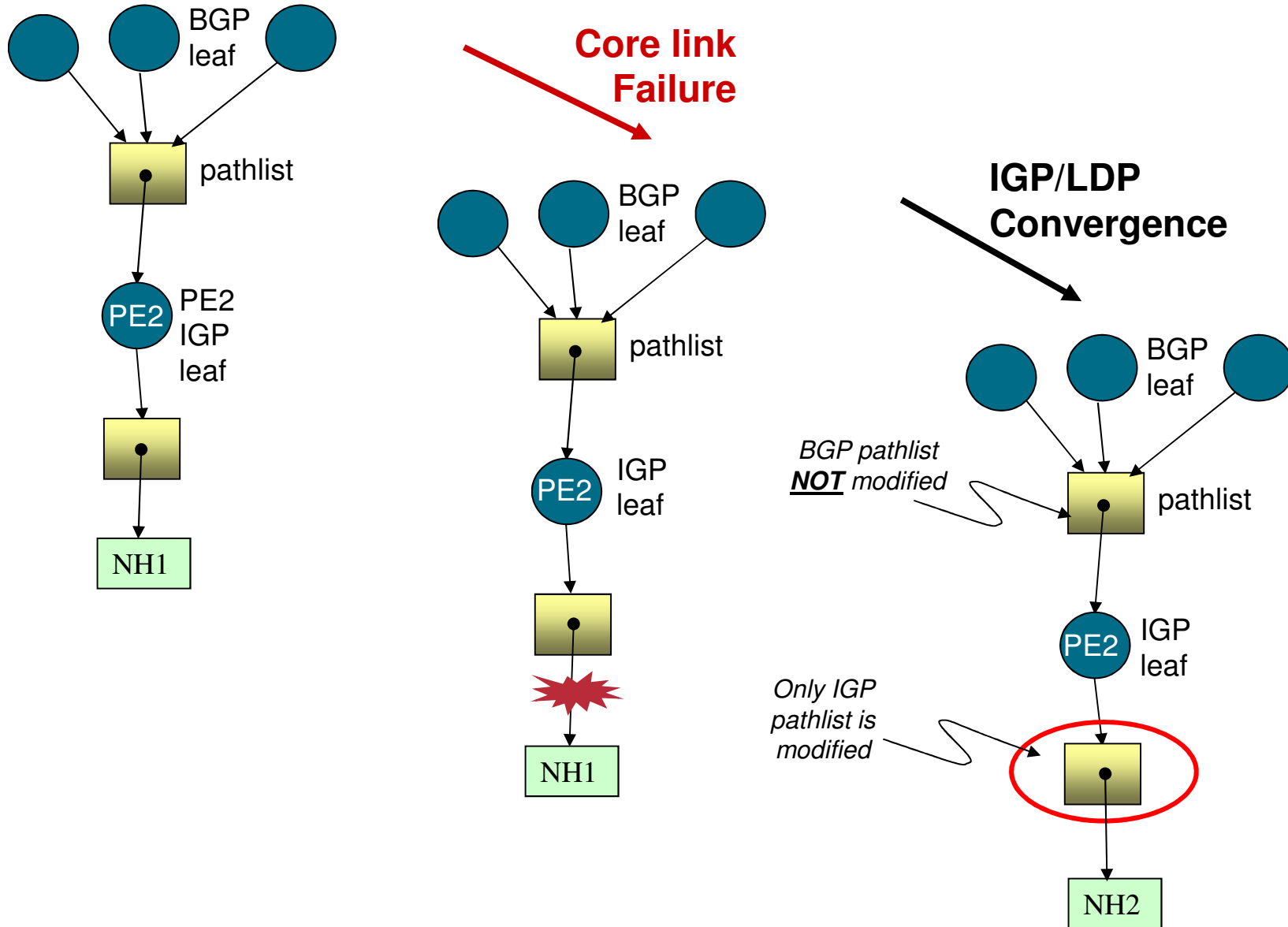
# Agenda

- Problem
- Solution
- Handling Limited Hardware Capability
- Recovery from various failure scenarios

# Core Link/Node Failure

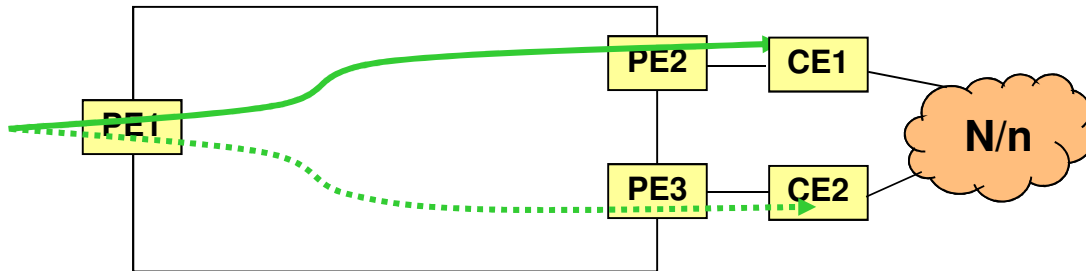


# Core Link/Node Failure: FIB on PE1





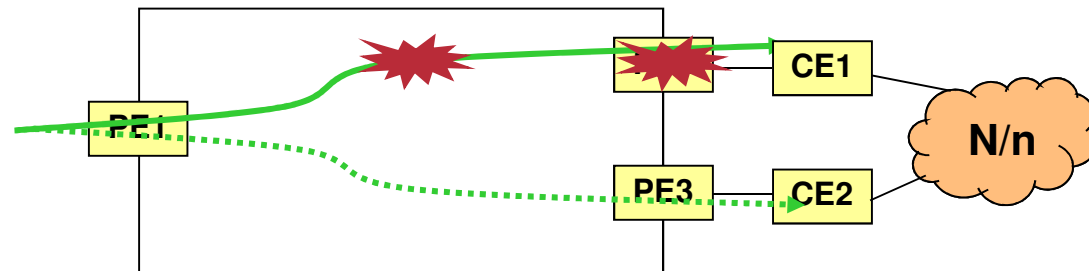
# Unipath PE Node Failure: Protection at *Ingress* PE



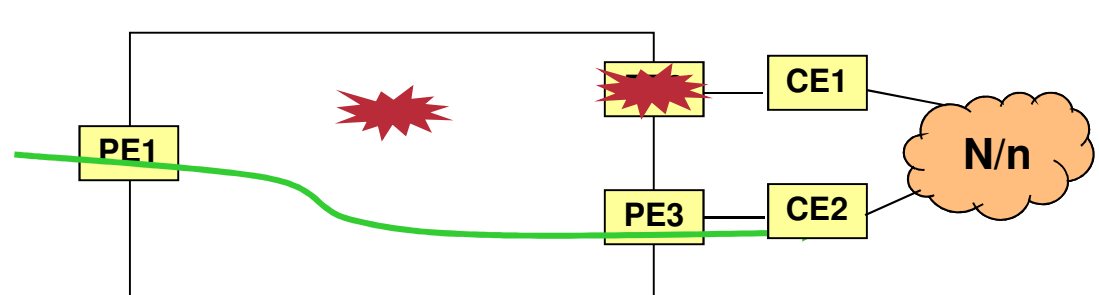
*Works whether BGP is running in the core or not\**

*PE3 must be pre-programmed as a backup path*

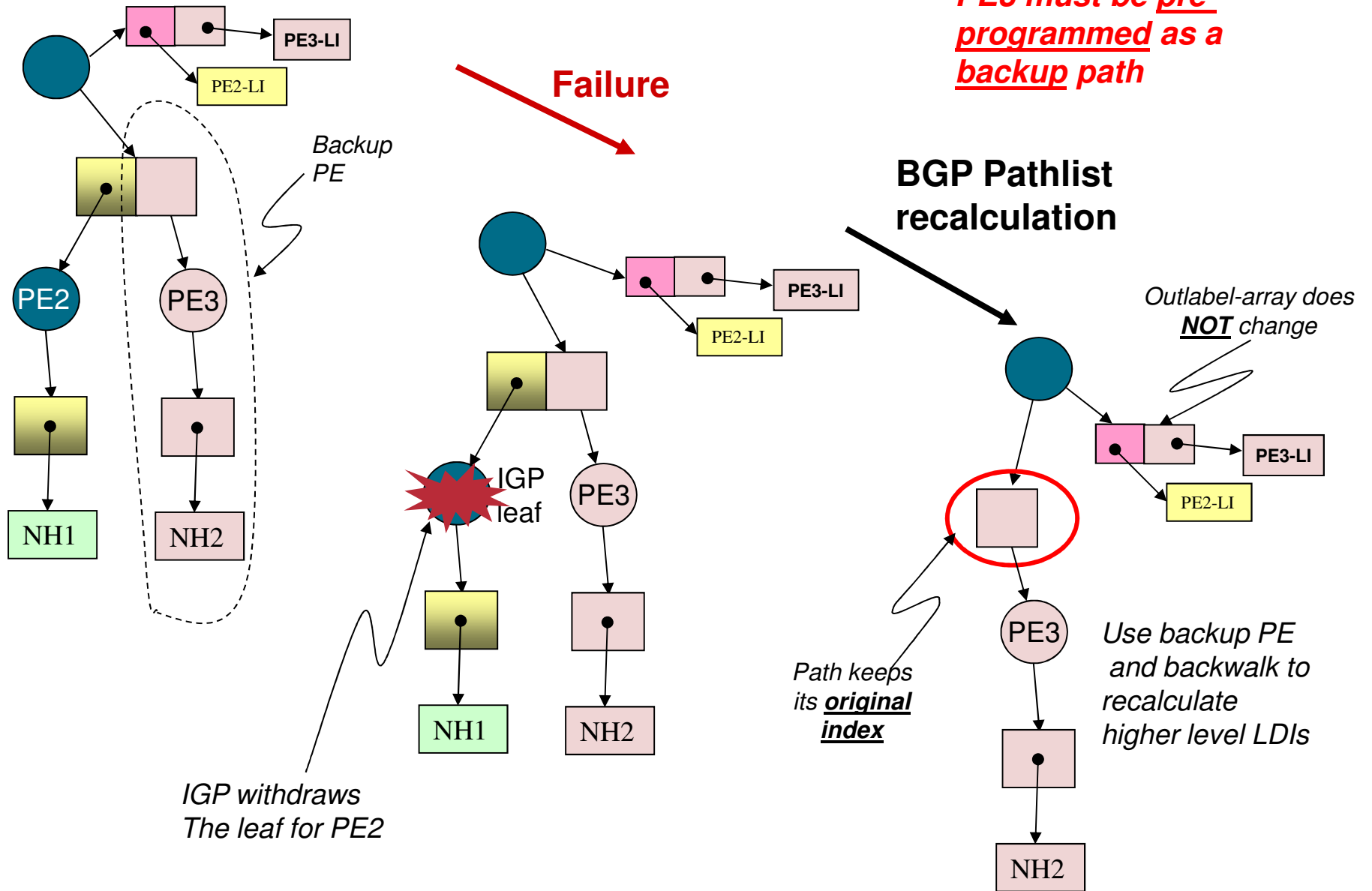
**Core or Edge node Failure**



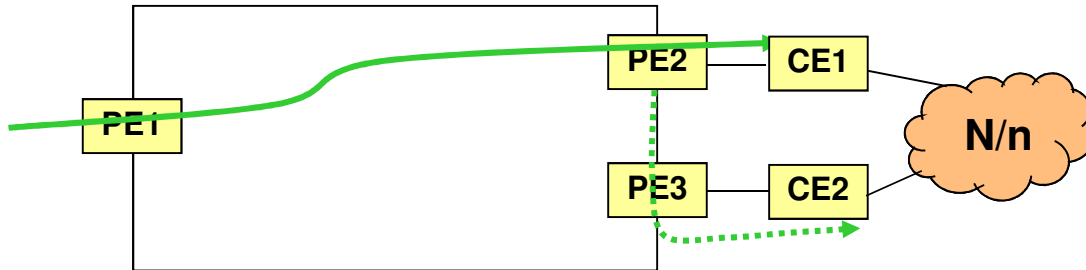
**FIB Correction**



# Unipath PE Node Failure: Protection at *Ingress* PE

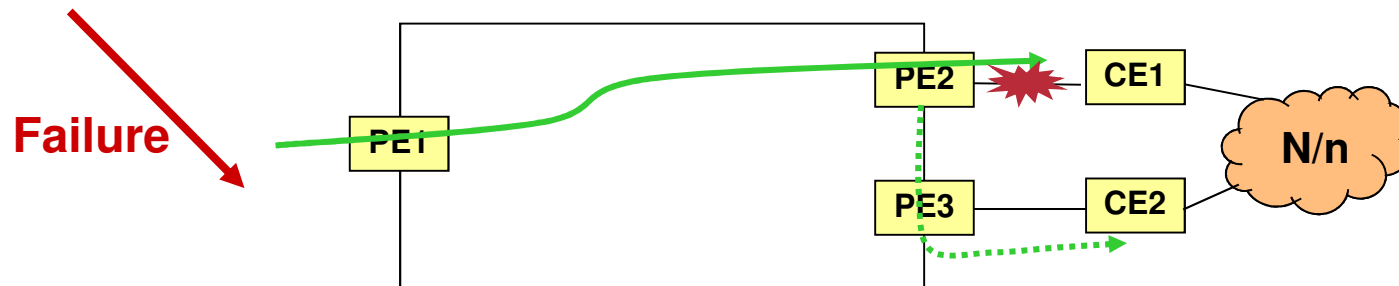


# Unipath PE-CE Link Failure: Protection at Egress PE

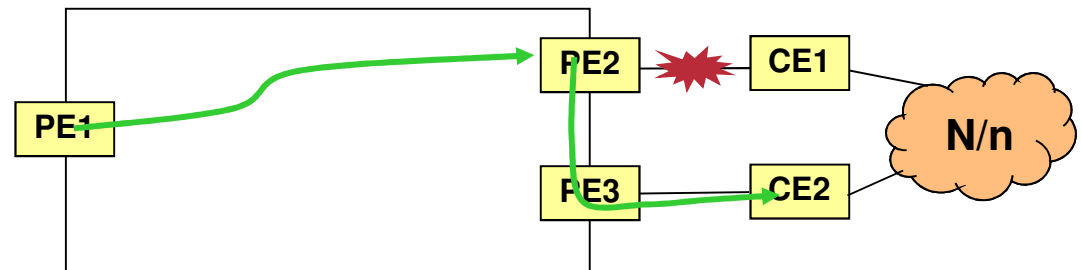


*More useful for  
BGP-free core*

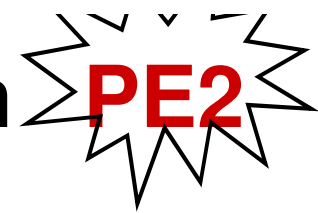
*PE3 must be pre-programmed as a  
backup path*



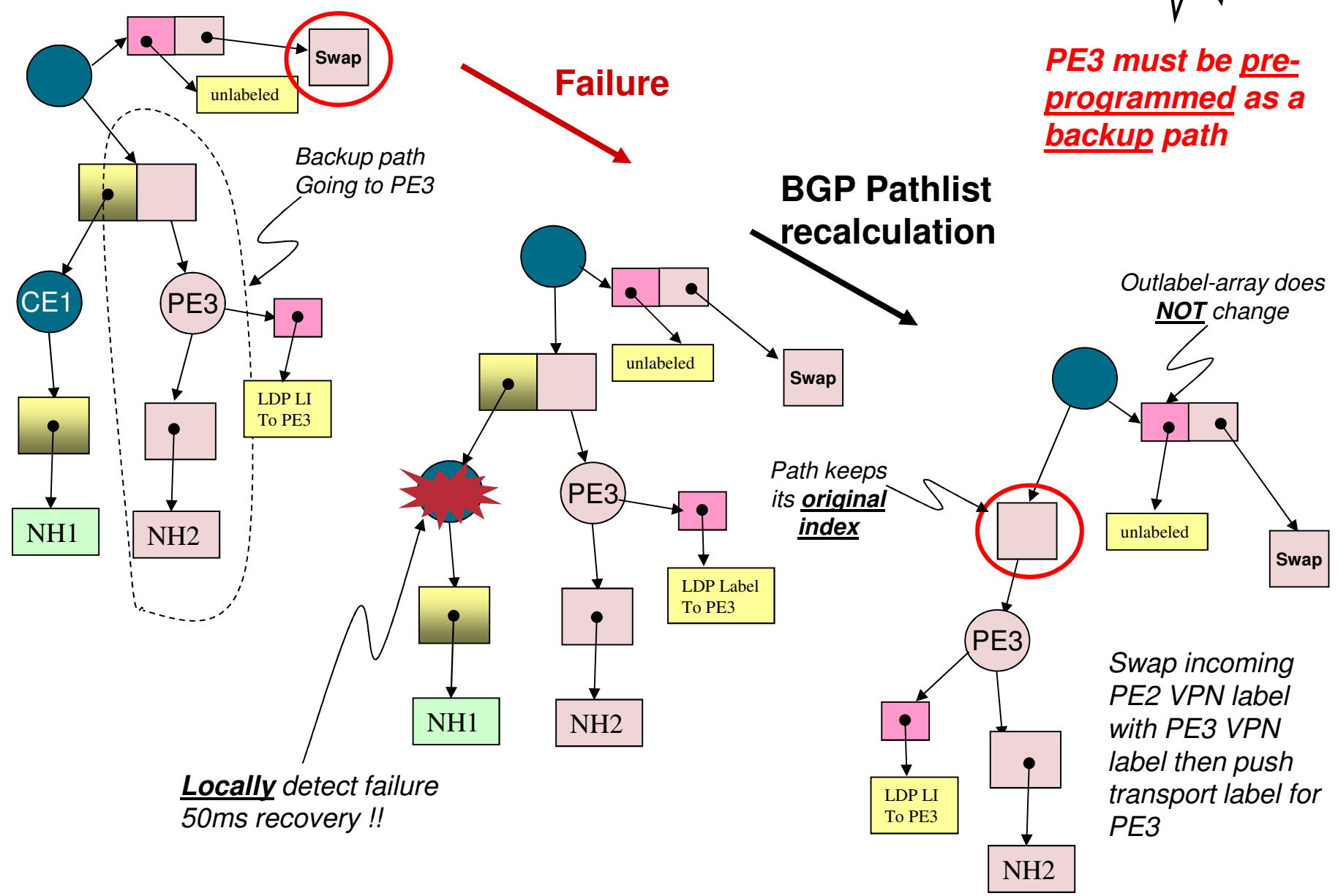
**FIB  
Correction**



# Unipath PE-CE Link failure: FIB on **PE2**

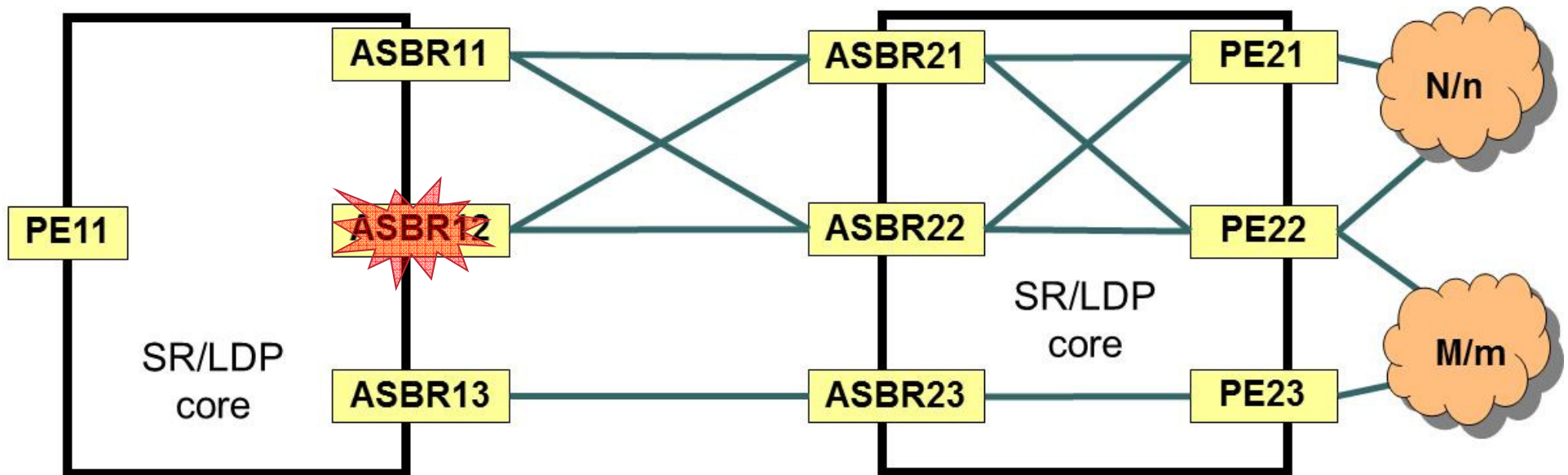


*PE3 must be pre-programmed as a backup path*

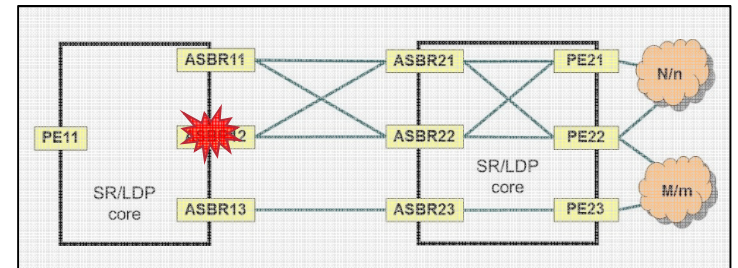
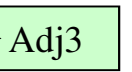
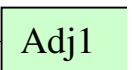
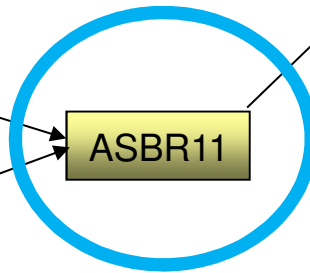
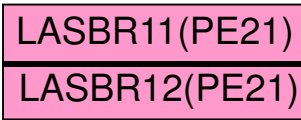
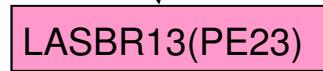
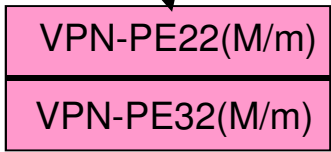
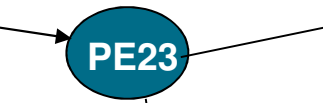
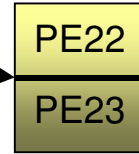
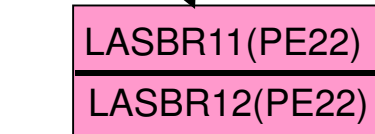
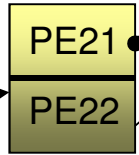
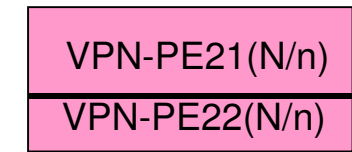
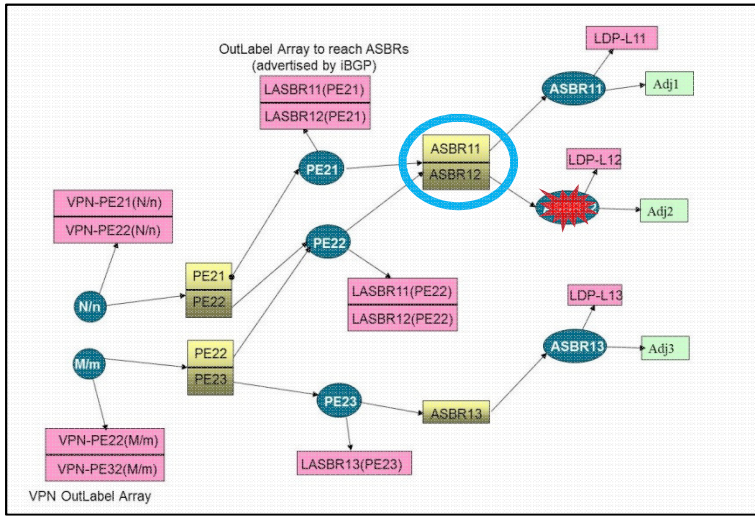


# Effect of Flattening: ASBR12 fails

- Suppose ASBR12 is no longer reachable in domain 1



# Unflattened Chain After Failure



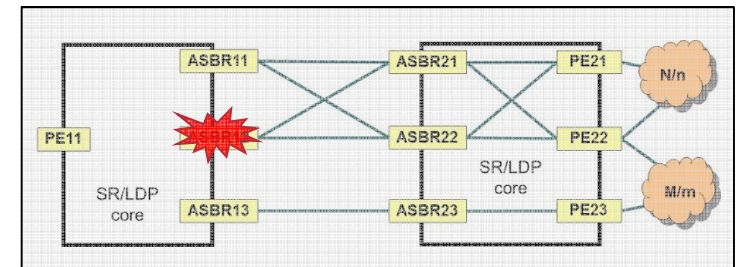
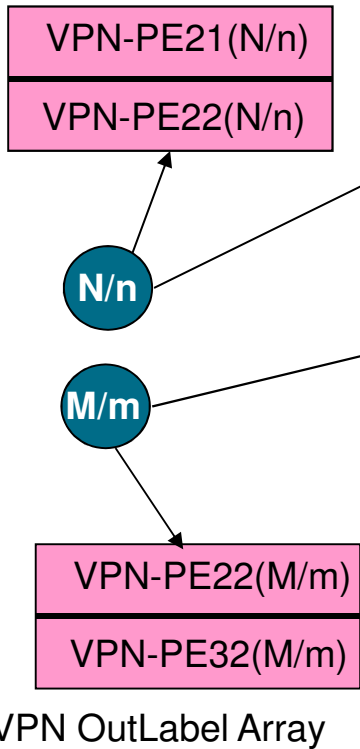
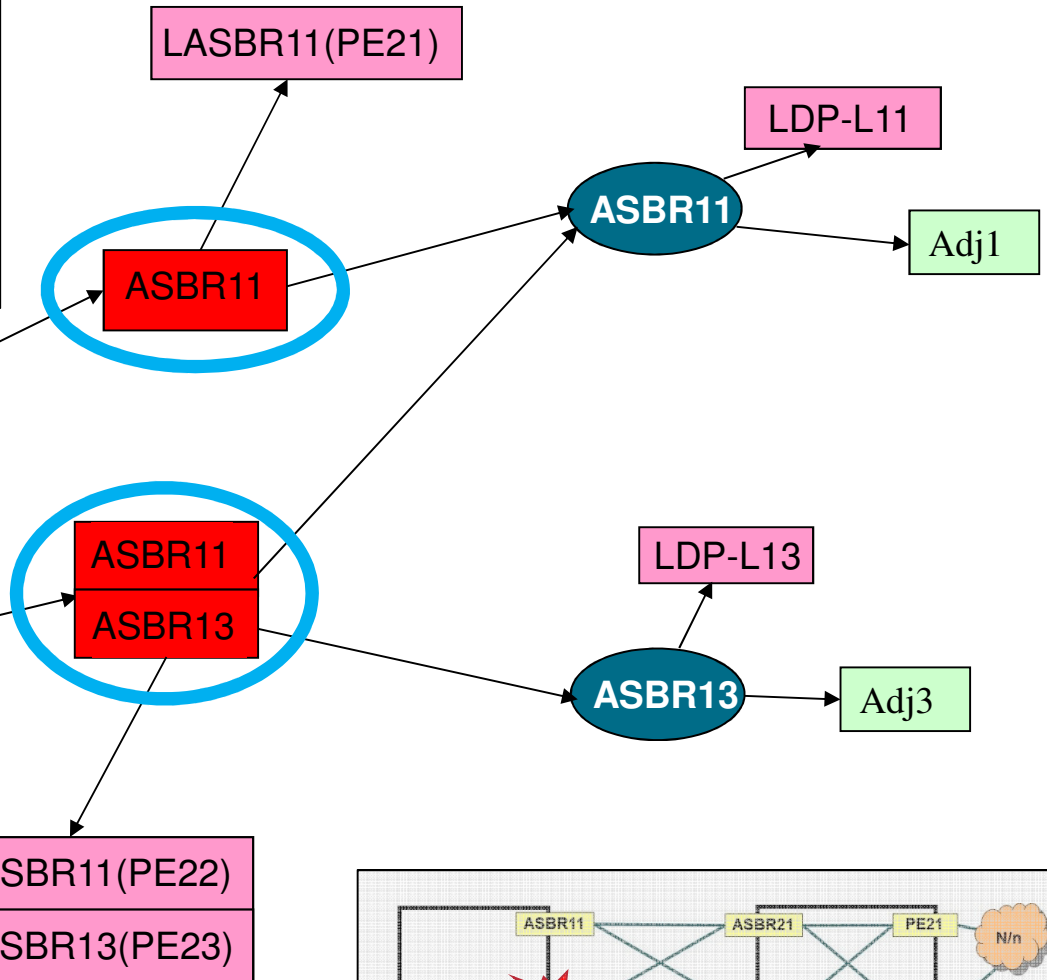
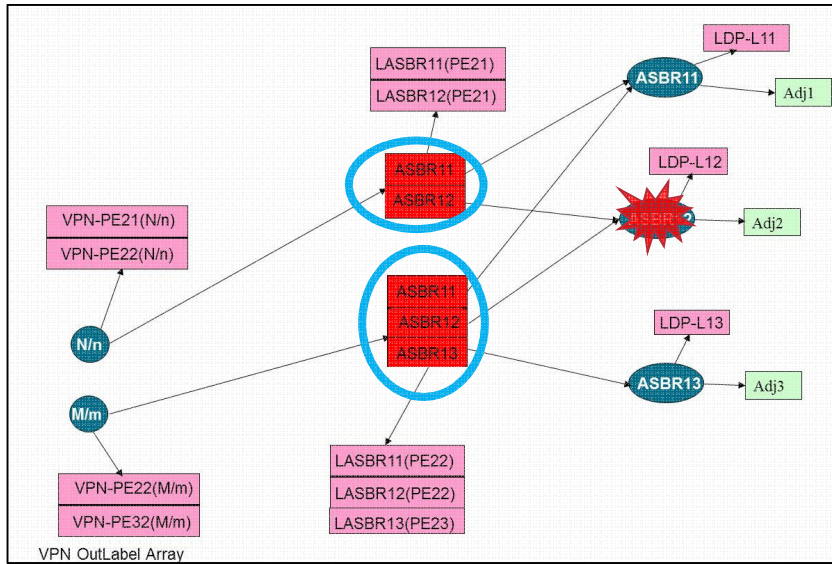
VPN OutLabel Array

BGP PIC

draft-bashandy-rtgwg-bgp-pic-01

IETF 94. Nov/2015, Yokohama

# Flattened Chain After Failure



# Conclusion

- Single elegant design to handle many convergence/protection cases
- BGP prefix (as well as other) independent convergence
- Support for any hierarchy depth with graceful PIC degradation for shallower hierarchy



# Request to become a Working Group Document

!!

