BGP/MPLS VPN Deployment Experiences and Challenges

Luyuan Fang
luyuanfang@att.com

MPLS 2004

AT&T

Outlines

- Overview of VPN Customer Requirements
- Popular VPN features and implementation challenges
  - Multi-homing load balancing
  - Inter-AS support for iBGP/eiBGP multipath
  - Multi-L3 Services through single physical interface
  - MLPPP for NxT1 VPN support
  - LDP failure detection and recovery
  - Scalability challenges
- Conclusions

Overview of L3 VPN Customer Requirements (1)

- BGP/MPLS VPN is still a very fast growing service - you may see it doubling in size every few months
- BGP/MPLS VPN customer profile observations
  - VPN routes
    - Number of sites per VPN can range from 2 to over 100K with average a few hundreds
    - Routes may be about 2-3 times the number of sites
  - BGP sessions
    - VPN customers use more eBGP connections than Static connections, ~85/15, the opposite for Internet customers
  - QoS policies
    - Over 50% VPN customers requires QoS support

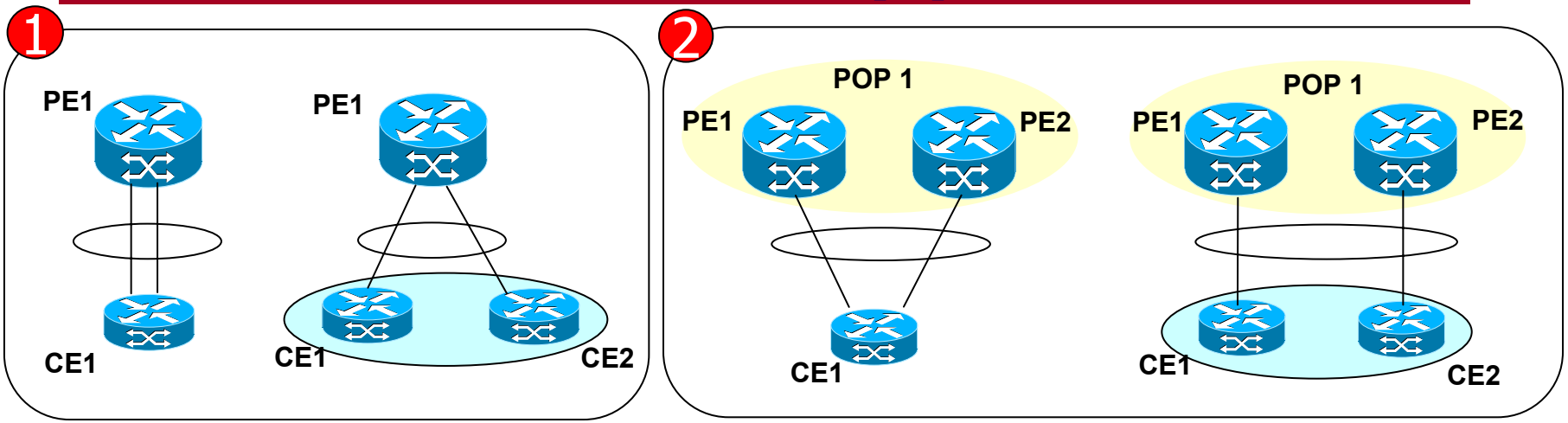
Overview of L3 VPN Customer Requirements (2)

- Customer / peer Access support required
  - Access types:
    - POS, FR, ATM, PPP, MLPPP, IMA, NxT1, FR encapsulation over POS, Ethernet with VLANs, etc.
  - Access speed:
    - From 56K to OC-192
    - All VPN service features need to be supported across large port speed range, on all line cards
  - Access protocol/connection:
    - eBGP, Static for Enterprise VPN
    - eBGP with labels, Static with labels for Carrier's Carrier
    - eBGP with labels, three label stack support for Inter-AS

Overview of L3 VPN Customer Requirements (3)

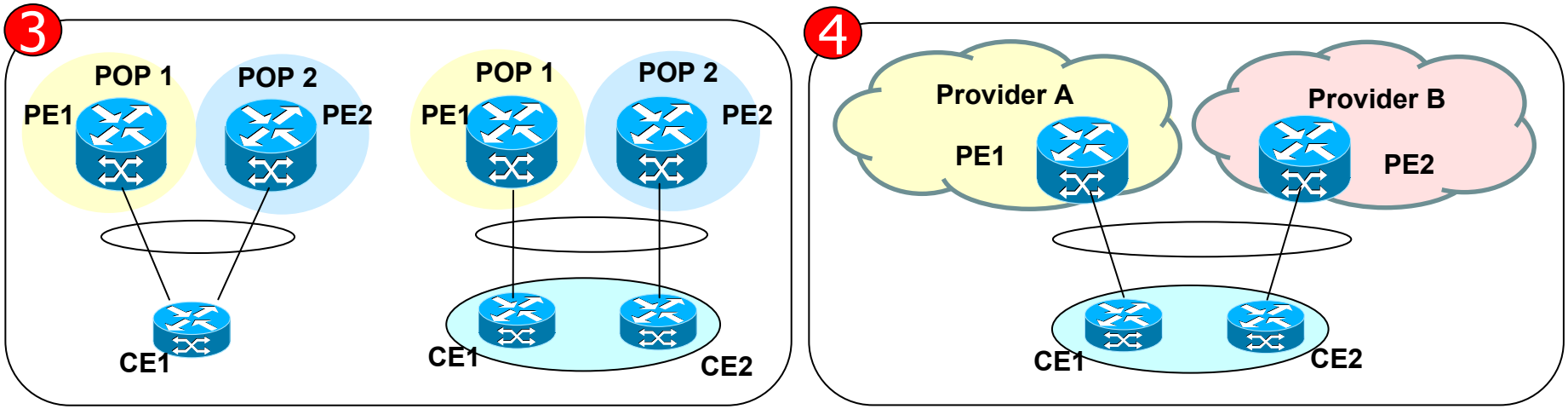
- Customer requirements for MPLS VPNs
  - QoS support for all types of VPNs
  - MLPPP for NxT1 VPNs with QoS and MVPN
  - Multi-homing load-balancing and redundancy
  - Multi-service port: Internet and VPNs through the same customer interface
  - Carrier's Carrier
  - Global reachability - Inter-AS and Inter-provider
  - Multicast VPN, Multicast Inter-AS
  - IPv6 VPN

Multi-homing load balancing / redundancy scenarios (1)



1. Single PE / Multi-link load balancing/redundancy
   - using eBGP multihop eBGP multipath as in normal IP services.
   - Not very beneficial for customer: no PE redundancy
   - Low network impact
2. Multi-PE / Single POP load balancing/redundancy
   - Customer address splitting, or using iBGP and eBGP auto load balancing
   - Most popular customer config
   - PE memory impact when using iBGP or eiBGP load balancing

Multi-homing load balancing / redundancy scenarios (2)



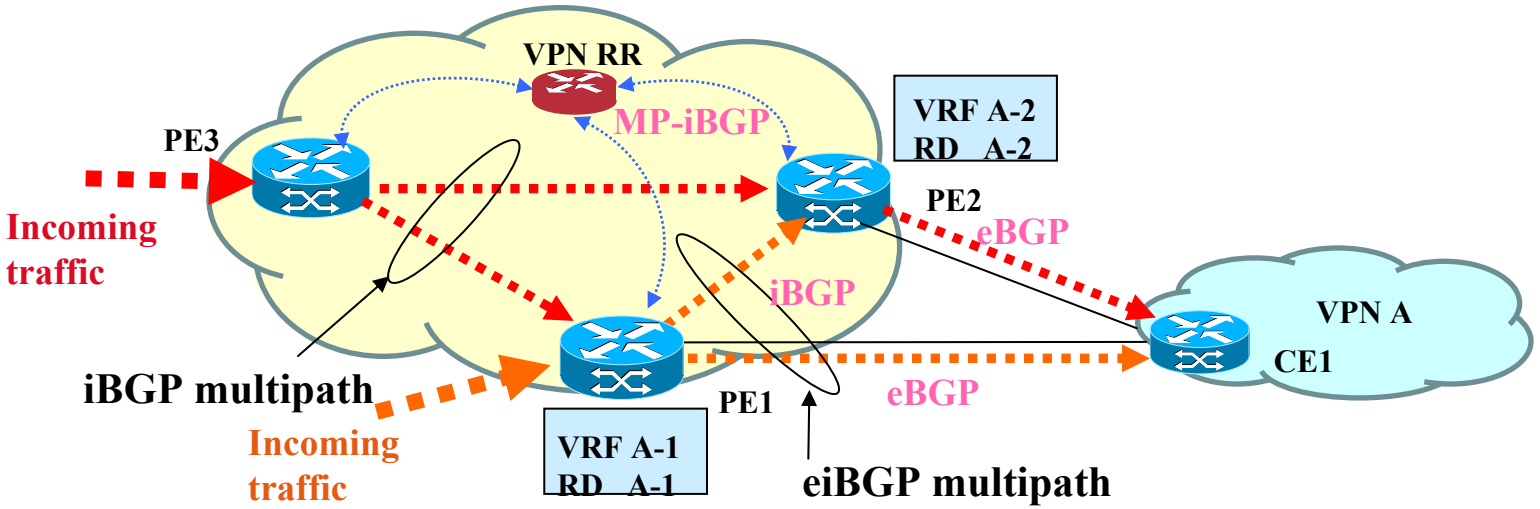
3 Multi-PoP / Single Provider load balancing / redundancy

- Customer address splitting, or using iBGP and eBGP auto load balancing
- Popular customer config, especially by large customers
- PE memory impact, network capacity impact when using iBGP or eiBGP load balancing

4 Multi-Provider load balancing / redundancy

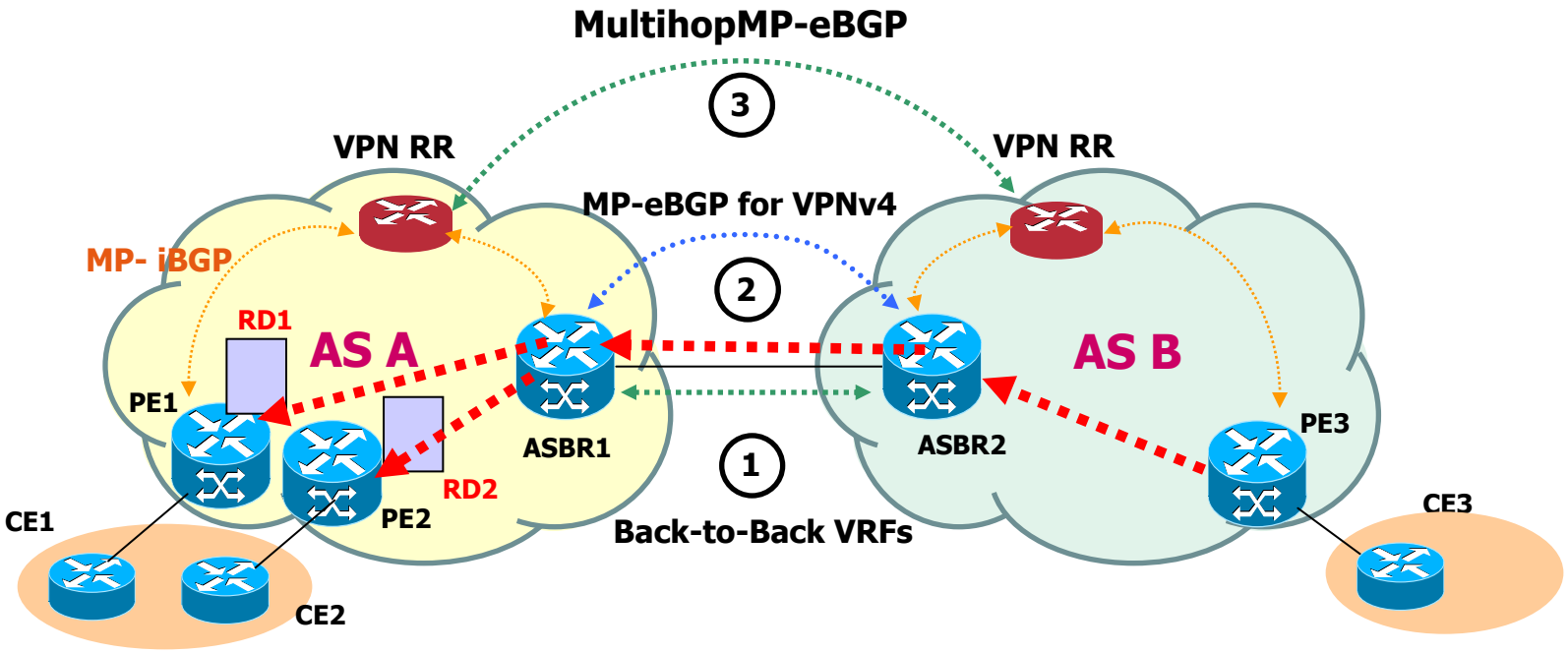
- Not very common, but required by customers who need provider diversity
- Fail-over control from customer side, using GLBP or other techniques
- Challenges: two SP networks often have different delay / jitter, CoS, VPN feature offerings, and network capacity. Require SPs coordination

iBGP and eiBGP BGP Multipath Load Sharing



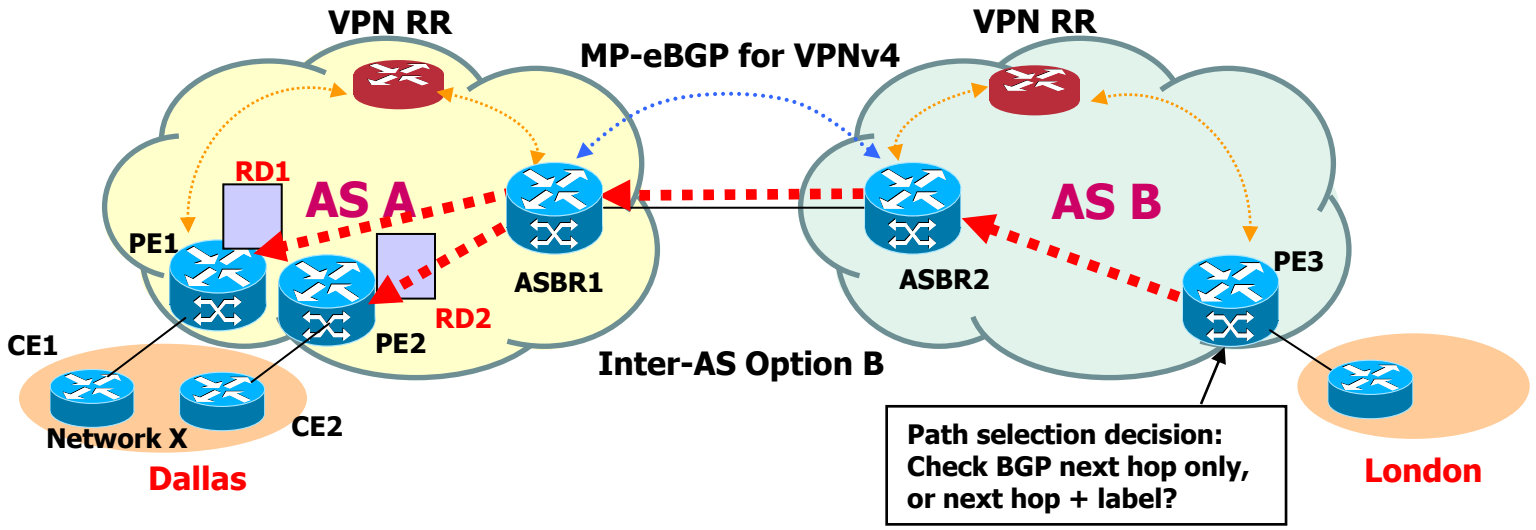
- iBGP multipath - Network based load sharing among multi iBGP paths
- eiBGP multipath - Network based load sharing between both iBGP and eBGP paths
- When VPN RR is used, need separate VRFs with different RDs on the PEs
- Scaling issue: 1) additional memory consumption on PEs. 2) Network capacity impact

Inter-AS supporting iBGP and eiBGP BGP Multipath (1)



- iBGP / eiBGP multipath should not be impacted with Inter-AS implementation
- Inter-AS Option A and C can support, Option B - depending on the correct implementation

Inter-AS supporting iBGP and eiBGP BGP Multipath (2)



- iBGP and eiBGP are ingress features - the multipath selection decision is made at the ingress PE.
- Every LSP need to be considered in the path selection process
  - PE3 in London receives both paths advertised for network X through PE1 and PE2 with different RDs, labels, and same next hop – there are two distinct LSPs
  - Only one path will be used if PE3 compare BGP next hop alone – it is ASBR2 for both paths
  - Both paths (via PE1, and PE2) will be used if PE3 compare BGP next hop and labels
  - Same for Intra-AS case if 2 or more VRFs for a single VPN on the same PE

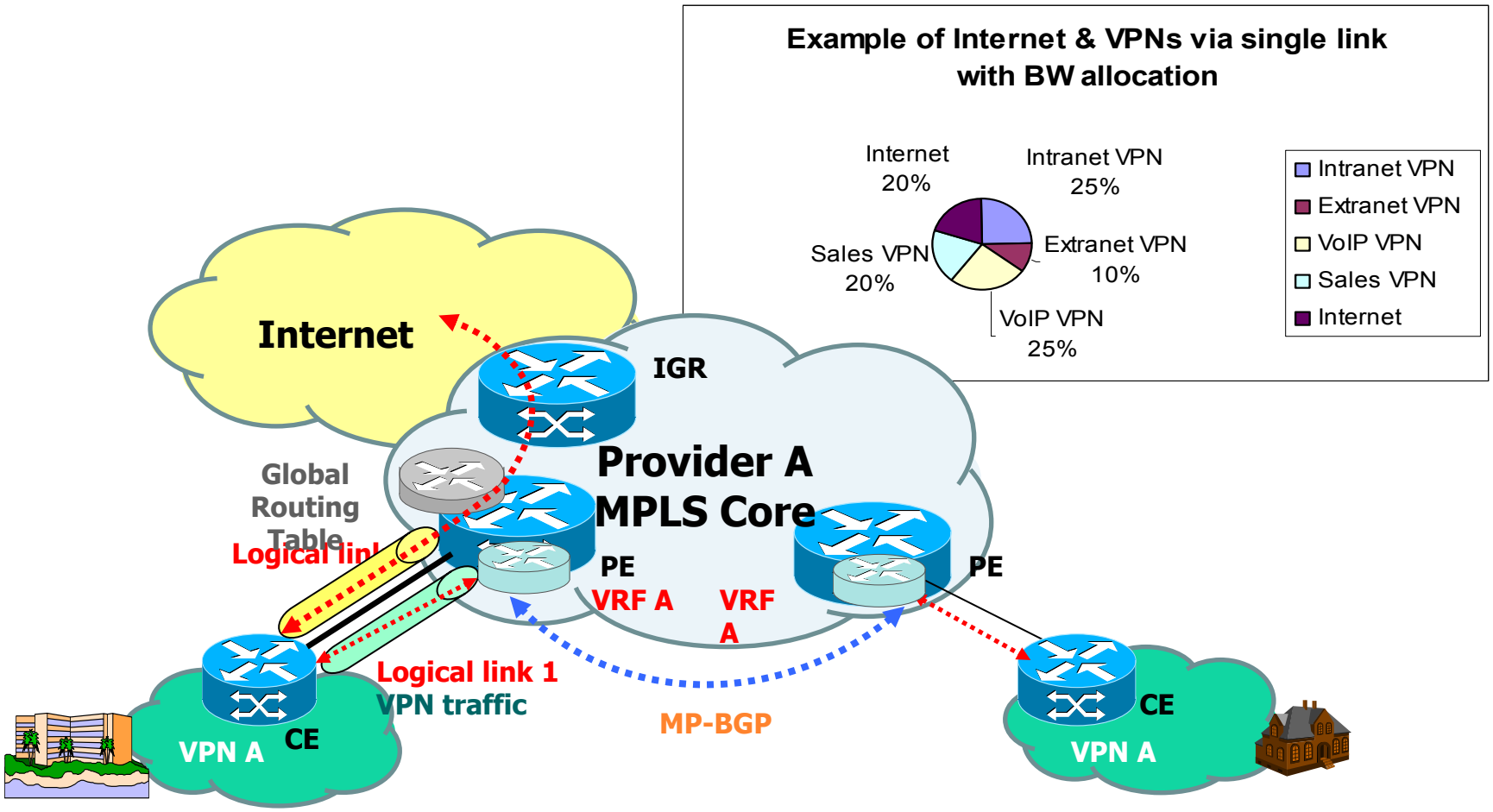
Multi-services via single physical link (1)

- A single physical port supports multiple L3 Services via multiple Logical Links – Cost efficient for Customers
  - Internet Service
  - Multiple VPNs
    - Intranet VPN, Extranet VPN, Content delivery, VoIP VPN, etc.
- Feature support requirements
  - Each LL must support all features just as a physical port
  - Access speed support LL implementation: T1 to OC-192
  - Large number of LL supported
  - No degradation of performance on the PEs
  - Scalable: large number BGP sessions and routes support

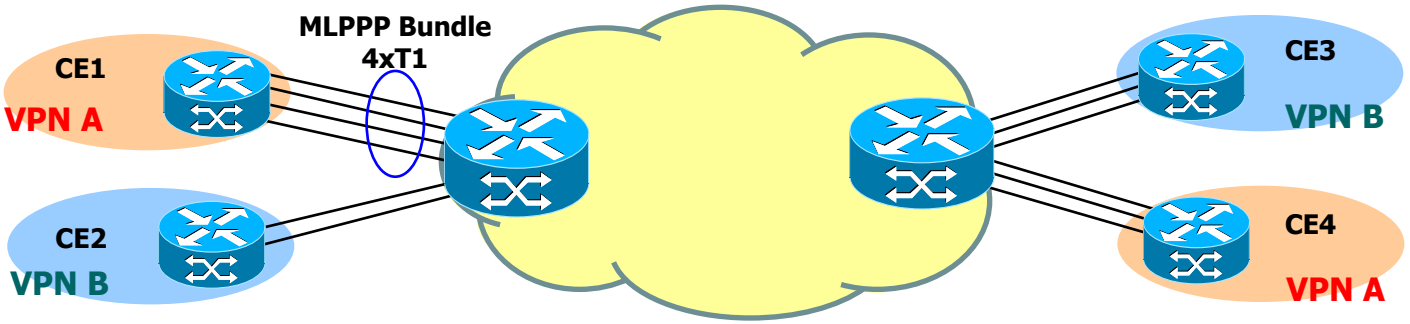
Multi-services via single physical link (2)

- Logical Links implementation
  - One implementation option is to config. Frame Relay Encapsulation over leased lines
    - PE may support both Internet routing tables and VRFs for VPNs
    - PE may only support VPNs - Internet PVC tunnel through PEs to IGR or others Internet service PEs
    - CE may implement "VRF lite"
- Bandwidth allocation and QoS per LL
  - Rate limiting using CAR
    - Max BW for per LL, cannot burst over to other LL, no CoS
  - BW control with QoS policies
    - Guaranteed minimum BW, allow burst over unused BW, providing CoS
  - QoS requirements
    - Per LL queuing
    - Hierarchical CoS, priority across LLs and per-LL queuing

Multi-services via single physical link (3)



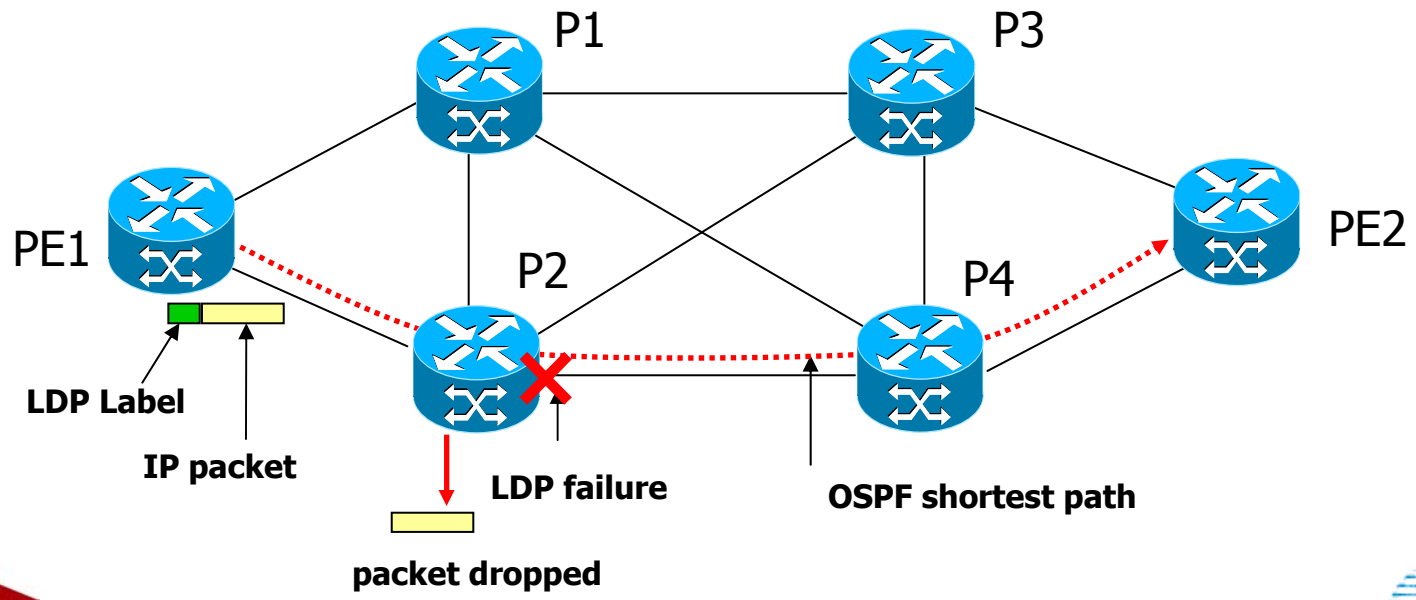
MLPPP for NxT1 VPN Support



- High demands by customers who need VPN with BW between 2T1 – 8T1
- Requirements
  - Basic MLPPP NxT1 bundles for VPN with QoS support
  - Support Multicast VPN with QoS
  - Stability – feature must not cause instability of the port adopter or entire router on both PE and CE
  - No packet loss or reordering issue
  - Scaling – support large number of bundles and MLPPP interfaces on the given line card or PE which supports QoS, MVPN, etc. for all customers

LDP Failure Detection and Recovery (1)

- Failure condition: IGP is operational but LDP fails
- Result: Blackholing LDP label switched packets. E.g. VPN traffic or all traffic if Internet Route Free Core
  - *"LDP Failure Detection and Recovery", IEEE Communications Vol.42 No.10,* October 2004. Fang, L., Atlas, A., Chiussi, F., Kompella, K. and G. Swallow.



LDP Failure Detection and Recovery (2)

- Cause of traffic loss due to LDP failure
  - LDP and IDP are separate protocols
  - IGP makes routing decisions without knowledge of LDP status
- Packets dropping location
  - Independent Control: at the point of failure
  - Ordered Control: at the ingress of LSP
- Causes of LDP failure
  - LDP session failed
  - Operator errors
  - Protocol implementation error
  - Stale label or label corruption
  - Race condition with IGP.

LDP Failure Detection and Recovery (3)

- LDP failure detection and isolation
  - LSP Ping: Verify end to end connectivity
  - LSP traceroute: Identify the failed node/link
    - *"Detecting MPLS Data Plane Failures," draft-ietf-mpls-lsp-ping-06.txt*, July, 2004. K. Kompella, and G. Swallow.
  - BFD (Bidirectiojnal Forwarding Detection) with LSP Ping
    - sub-second detection time
      - *"BFD for MPLS LSPs", draft-ietf-bfd-mpls-00.txt*, July 2004, R. Aggarwal, K. Kompella, T. Nadeau, G. Swallow.
  - LSR self test: Test one hop of an LSP
    - verify both its control plane and its data plane
      - The control plane test uses LSP Ping
      - the data plane test is a simple extension of LSP Ping
    - sub-second detection time
    - *"Label Switching Router Self-Test," draft-ietf-mpls-lsr-self-test-02.txt*, Feb.2004. G. Swallow, K. Kompella, and D. Tappan.

LDP Failure Detection and Recovery (4)

- Failure recovery
  - Manual correction by operator
    - Correction if it is config error
    - Reset may help if protocol error
    - Down the failed link to force reroute at the last resource
  - Automatic correction
    - Router self detect and self healing if possible
    - Raising IGP metrics to move traffic off the broken interface, or keep IGP metrics high before IGP is converged when bring links up
      - *"LDP IGP Synchronization", draft-jork-ldp-igp-sync-00*, Oct. 2004. M. Jork, A. Atlas, L. Fang.
    - Possible protocol changes: a node can inform another that LDP on a specific link is faulty.
    - LDP convergence
      - *"Loop-free alternates for IP/LDP Local Protection", draft-atlas-ip-local-protect-00.txt*, February 2004. A. Atlas et al.

Scalability is still a challenge

- BGP/MPLS VPN's healthy growth keeps lifting the bar on scaling requirements
- Customers love rich features, SPs must have scalable solutions
  - iBGP/eiBGP multipath – may be X times additional memory consumption
  - Multi LL per physical links multiply the numbers needed for BGP sessions and routes on the PE
  - MLPPP can add significant stress to CPU, and may impact PE stability
- VPN scaling number?
  - # of BGP sessions – needs to be X times the total interfaces, especially considering multi LL scenarios, X cannot be less than 1
  - # of VRF – should be > # of interfaces
  - # of routes – consider iBGP and eiBGP multipaths, multi LL scenarios, and some VPNs with extra large # of rts, a million is not too much
  - # of QoS policies – consider per LL queuing
  - # of MLPPP bundles – consider enough processing power to support large number of bundles while PE is loaded with MVPN, multipath, QoS, etc.

Conclusions

- BGP/MPLS VPN is in high demand by customers from very small to very large
  - Scaling improvement is still urgently needed for supporting VPN services
    - e.g. number of routes, number of BGP sessions, number of logical interfaces, number of QoS policies, number of MLPPP bundles
  - All features should be supported across all platforms, all line cards
- Features requirements
  - QoS for all types of VPNs, including logical interface support and multi-level queues
  - Multi-homing load-balancing and redundancy in all scenarios
  - Multi-service port: Internet and VPNs through the same customer interface
  - MLPPP for NxT1 with QoS for basic VPN or MVPN
  - Multicast VPN, and Multicast Inter-AS VPN
  - Global reachability - Inter-AS and Inter-provider
  - Carrier's Carrier
  - IPv6 VPN - e.g. 6PE solution with min impact to the core
- Challenges
  - Improvements in Reliability, Security and Scalability in both vendor implementation and provider operation