ADVANCED TOPICS AND FUTURE DIRECTIONS IN MPLS

BRKIPM-3003

Bruce Davie

Cisco Networkers 2007

HOUSEKEEPING

- We value your feedback, don't forget to complete your online session evaluations after each session and complete the Overall Conference Evaluation which will be available online from Friday.
- Visit the World of Solutions on Level -01!
- Please remember this is a 'No Smoking' venue!
- Please switch off your mobile phones!
- Please remember to wear your badge at all times including the Party!
- Do you have a question? Feel free to ask them during the Q&A section or write your question on the Question form given to you and hand it to the Room Monitor when you see them holding up the Q&A sign.

Outline

- IETF Update
- Traffic Engineering
- Layer 3 VPNs
- Quality of Service
- Layer 2 VPNs & Pseudowires

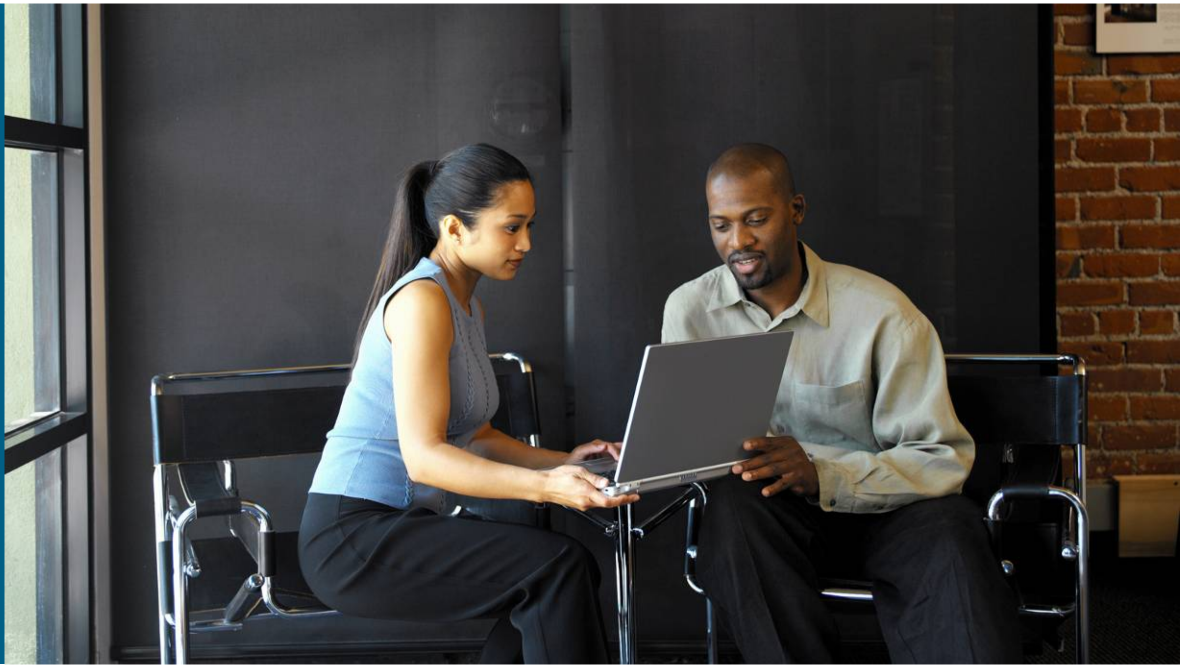
Goals of this session

You should gain

- an understanding of latest developments in the MPLS architecture
- an overview of MPLS standards activities
- a sense of the future trends in MPLS
- an appreciation of what problem MPLS can and cannot address

Not a good place to learn MPLS basics

IETF Update



Internet Engineering Task Force

- Originated MPLS standardization
- Base MPLS technology specifications completed
- Six active working groups
  - MPLS
  - Pseudowire Edge-to-Edge (PWE3)
  - Layer 2 Virtual Private Networks (L2VPN)
  - Layer 3 Virtual Private Networks (L3VPN)
  - Common Control and Measurement Plane (CCAMP)
  - Path Computation Element (PCE)
- Also some relevant work in Transport WG (TSVWG)

Some Recent MPLS RFCs

- RFC 4659—BGP-MPLS VPN Extension for IPv6 VPN (PS)
- RFC 4657—PCE Communication Protocol Generic Requirements (I)
- RFC 4655—A PCE-Based Architecture (I)
- RFC 4577—OSPF as the PE-CE Protocol for BGP/MPLS VPNs (PS)
- RFC 4461—Signaling Requirements for Point-to-Multipoint TE (I)
- RFC 4447—Pseudowire Setup and Maintenance using LDP (PS)
- RFC 4448—Encapsulation Methods for Transport of Ethernet Over MPLS (PS)
- RFC 4379—Detecting MPLS Data Plane Failures (PS)
- RFC 4377—Operations and Management (OAM) Requirements for MPLS
- RFC 4364—BGP/MPLS IP Virtual Private Networks (PS)
- RFC 4221—MPLS Management Overview (I)
- RFC 4216—MPLS Inter-AS TE Requirements (I)
- RFC 4206—LSP Hierarchy with GMPLS (PS)
- RFC 4124—Protocol Extensions for Support of DS-TE (PS)
- RFC 4090—Fast Reroute Extensions to RSVP-TE for LSP Tunnels (PS)

MPLS Working Group

- >35 RFCs published to date
- Current major work areas:
  - OAM (Operations and Management)
  - Point-to-multipoint
    - P2MP TE
    - LDP extensions for point-to-multipoint
    - Label allocation for p2mp
  - Advancing base specs from "Proposed" to "Draft Standard"

Pseudowire Emulation Edge-to-Edge (PWE3) Working Group

- Original charter is near completion
- ATM, Frame Relay, PPP/HDLC, and SONET encaps about to become RFCs
- LDP extensions for signaling & Ethernet encaps are published RFCs
- Current work items:
  - Virtual circuit connection verification (VCCV)—uses MPLS OAM tools over a pseudowire
  - Inter-AS PWs
  - Pseudowire congestion control

L2VPN WG

- Standardizing:
  - Virtual Private LAN Service (VPLS): L2 service that emulates a LAN, allowing standard Ethernet devices to communicate as if connected to a common LAN segment
  - Virtual Private Wire Service (VPWS): L2 service that provides L2 point-to-point connectivity across an IP/MPLS network
  - IP-only L2VPNs (IPLS): L2 service allowing standard IP devices to communicate with each other as if connected to a common LAN segment
- Specs for VPLS (LDP-signaled, BGP-signaled) are on way to RFCs
- IPLS passed last call
- WG still working on L2VPN multicast

L3VPN WG

- Responsible for defining and specifying solutions for supporting Layer 3 provider-provisioned virtual private networks
- RFC 2547 (BGP/MPLS VPNs) now replaced by RFC 4364
- IPv6 VPN extensions published as RFC 4659
- Main current work item: Multicast in BGP/MPLS VPNs

Path Computation Element (PCE) WG

- New WG chartered in 2005

  http://www.ietf.org/html.charters/pce-charter.html

- Responsible for overall PCE architecture, discovery, and signaling, targeted at MPLS and GMPLS

- RFCs:

  PCE Architecture

  Protocol Requirements

- Work items:

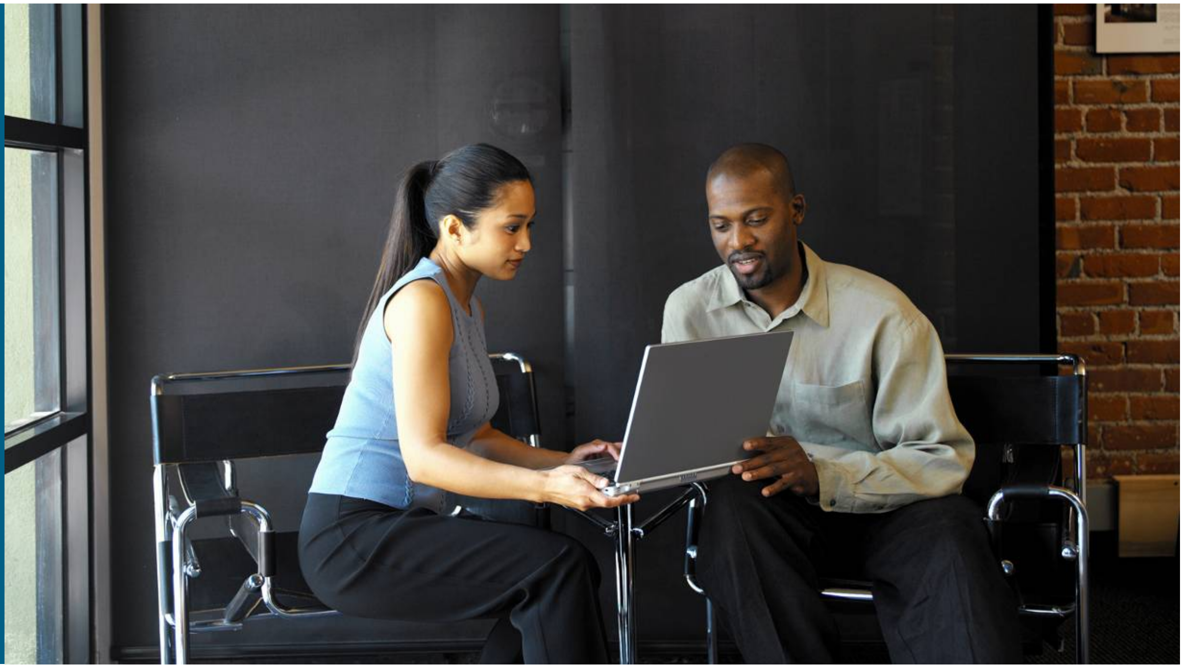
  PCE⇔client communication protocol

  PCE discovery using IGP

IETF Summary

- Base MPLS specifications are complete
- Current main focus areas:
  - Multi-segment pseudowires
  - Layer 3 VPN multicast
  - Inter-AS/Inter-area TE
  - Path Computation Element
  - Point-to-multipoint TE, LDP & OAM

Traffic Engineering



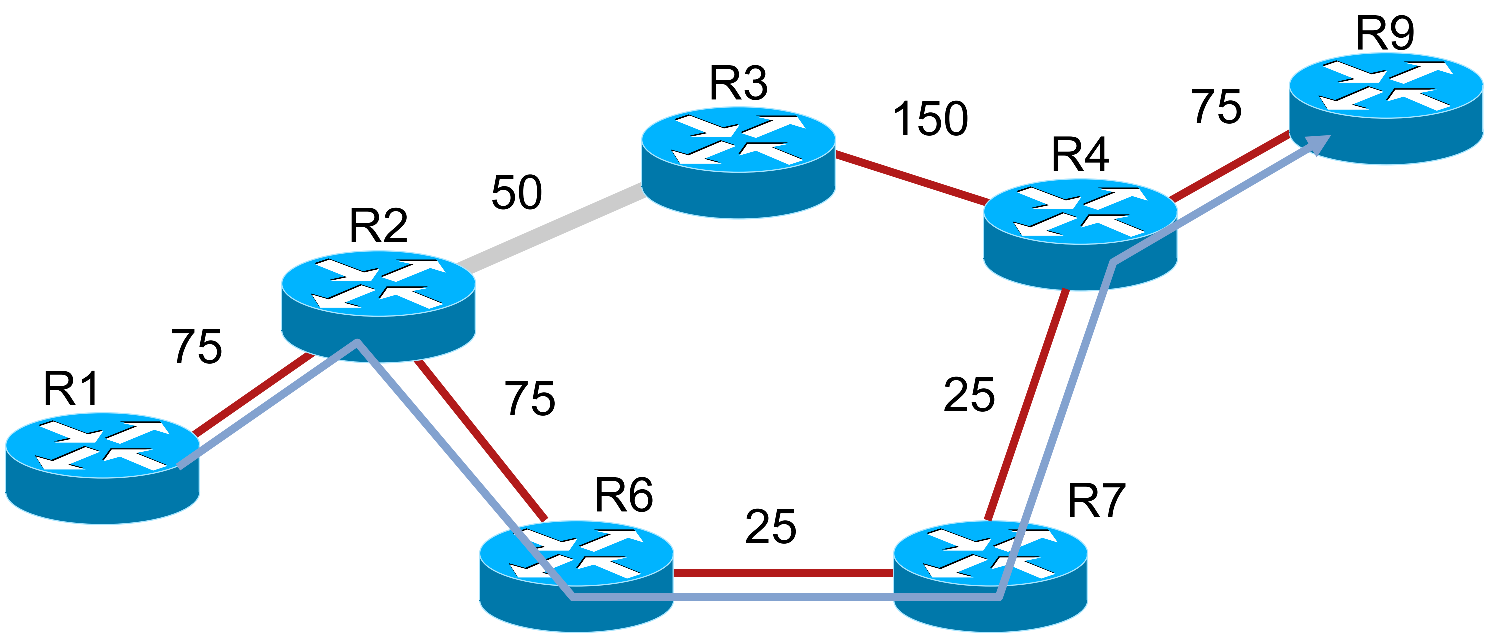
Traffic Engineering Agenda

- Inter-AS and Inter-Area TE
- Path Computation Element (PCE)
- Point-to-Multipoint TE

What's Hard About Inter-AS/Inter-Area TE?

- TE depends on running CSPF at tunnel headend
- This works fine if tunnel headend has complete picture of the network topology
- If tunnel head and tail are not in the same area of a single AS, the head does not know enough about topology to run CSPF
- A classic scale vs. optimality tradeoff:
    - Hierarchy is good for scaling
    - Hierarchy hides information
    - Information hiding makes optimal paths hard to find

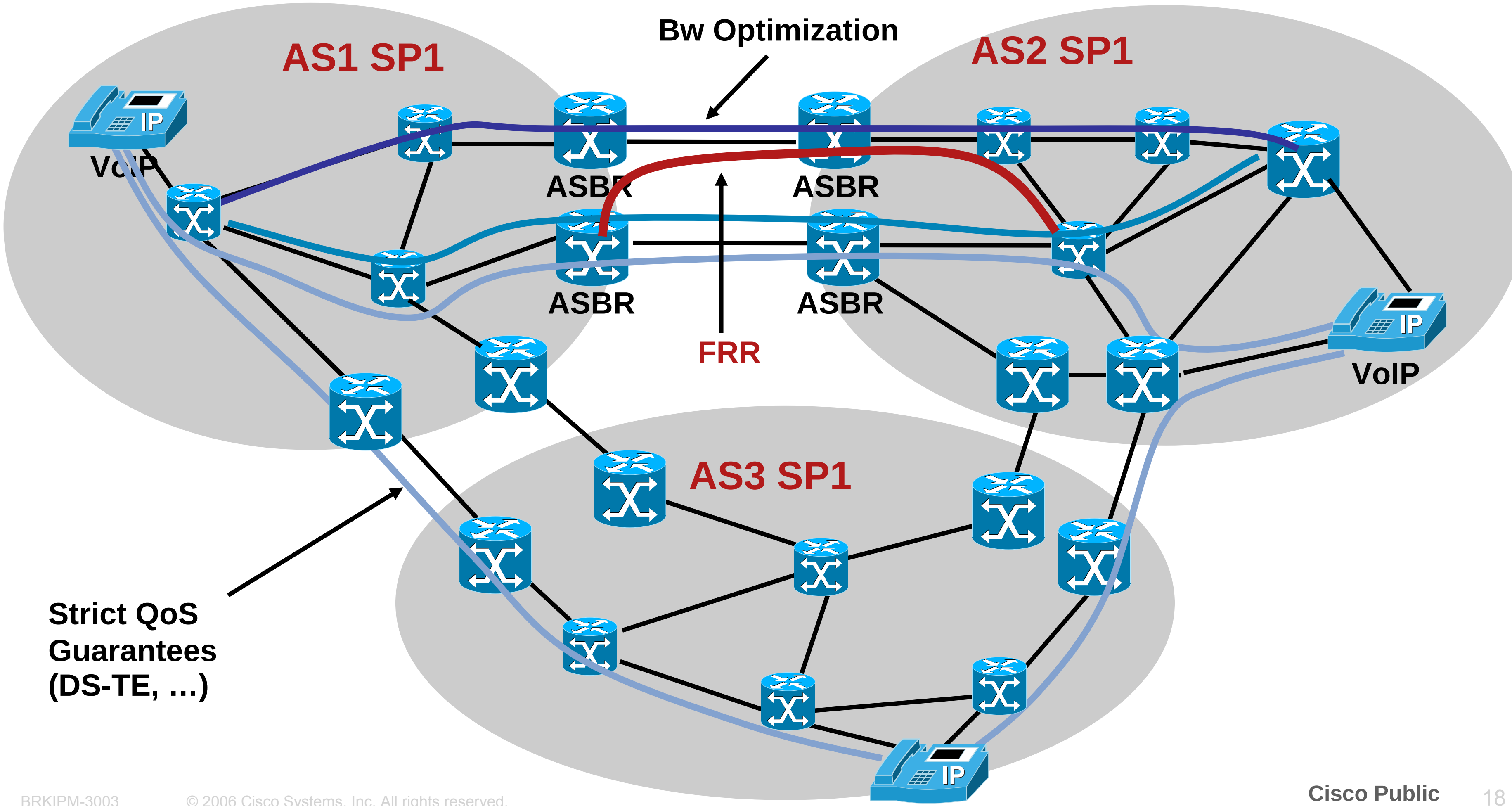
TE Example



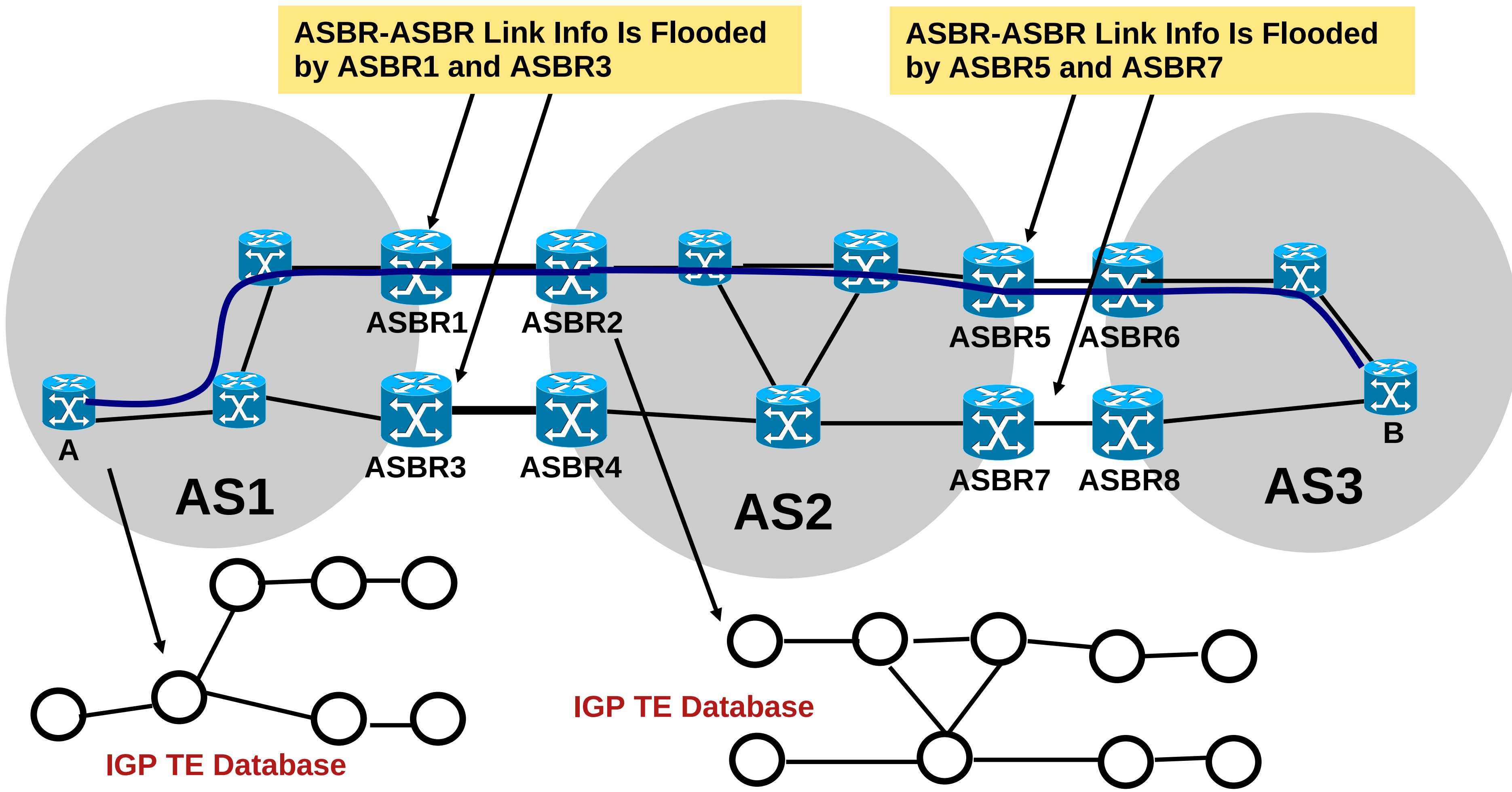
- Trying to route a trunk from R1 to R9 with bandwidth 75 Mbps
- R2-R3 link violates constraint (BW ≥ 75), so delete it
- Pick shortest path on remaining topology
- Update available capacities when path is established

Deployment Scenario: Multi-AS Provider

Seamless TE Plane for Bandwidth Optimization, ASBR FRR Node Protection, Strict QoS Guarantees Across ASes



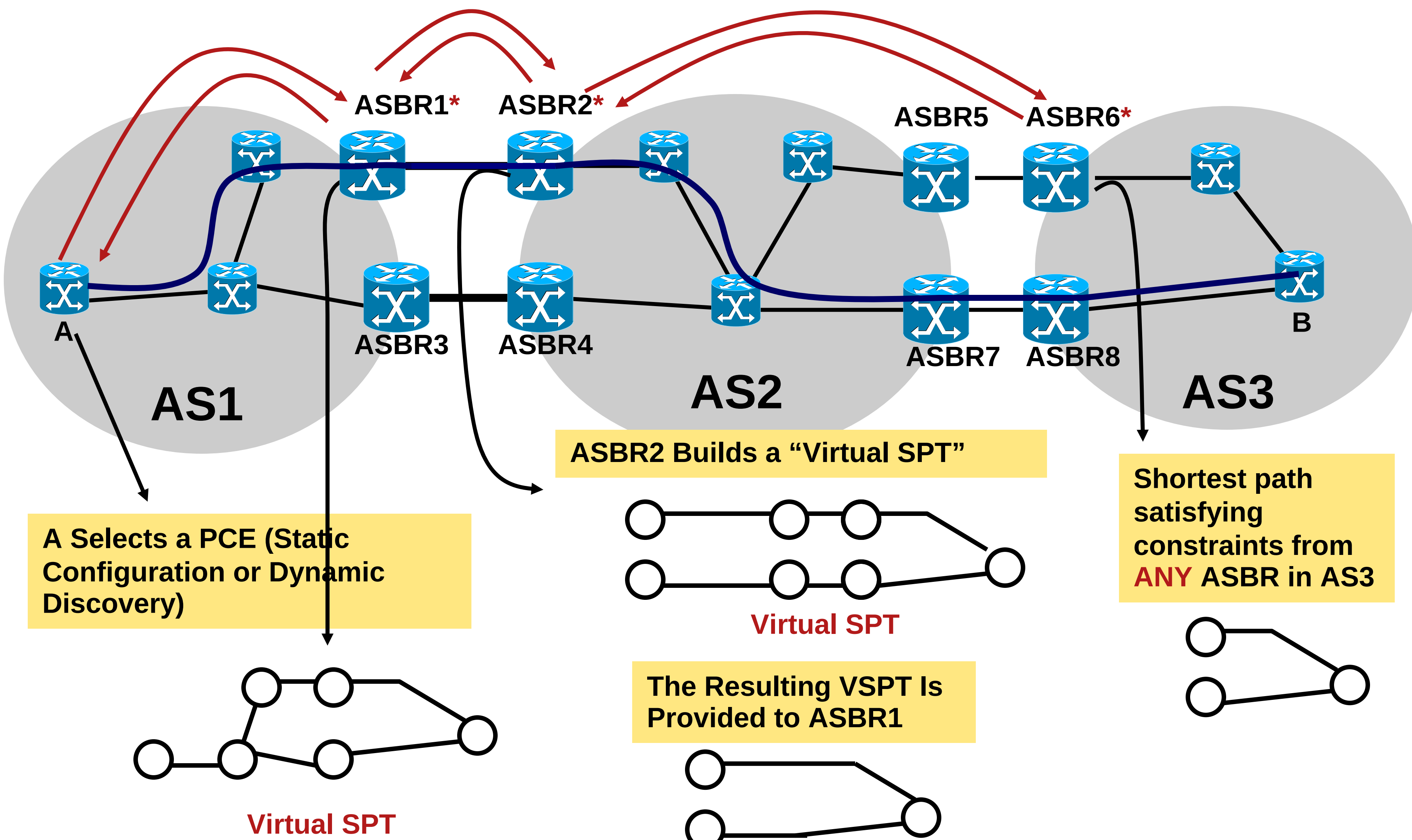
Per-Domain Approach: Loose Hop Expansion



Distributed Path Computation

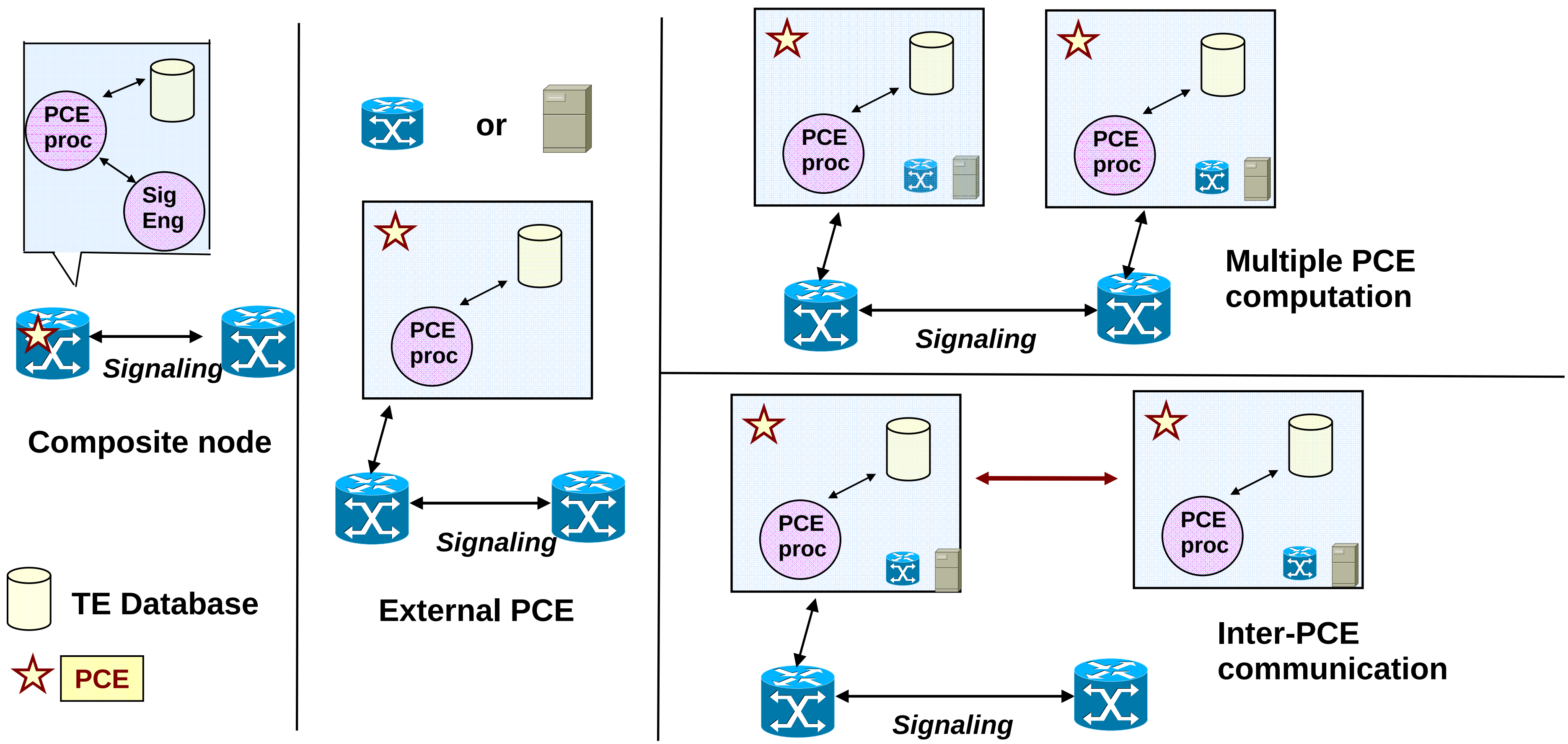
- Key idea: use a "path computation element" (PCE) in each AS
  - PCE is typically an ASBR
- PCEs communicate with each other to gather information about the topology and resources along a sequence of ASes
- PCE for each AS calculates a set of shortest paths from all its ingress ASBRs to the destination
- Each PCE reports only those paths that meet the constraints to the next AS
- Able to calculate shortest path that meets the constraints
  - Caveat: still need to choose the AS-level path
- Able to find a suitable path if path exists

Distributed Path Computation



PCE: Terminology and Architecture

The PCE can be located within an application, on a network node or component, on a standalone server, etc.



Comparison of Approaches

PER-DOMAIN PATH CALCULATION

- No impact on routing or signaling scalability
- Minor protocol extensions
- Doesn't find shortest path in general
- May fail to find paths that exist
- Hard to find diverse paths

DISTRIBUTED PCE APPROACH

- No impact on routing or signaling scalability
- More complex protocol extensions and need for PCEs
- Will find shortest path
- Will find a path if one exists
- Diverse paths possible

Bottom Line: Two Valid Approaches, Complexity vs. Optimality Tradeoff

PCE Discovery

- Clients need to discover PCE(s) within the same domain
  - Accomplished with Link-state flooding of PCE capability
  - Dynamic discovery avoids single point of failure and enables load balancing
- PCEs may need to discover PCEs in adjacent domains
  - In many cases static configuration will suffice - peerings are few and slowly changing

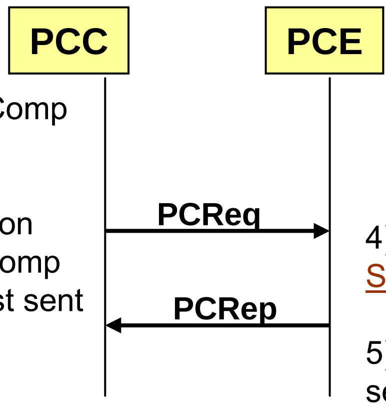
draft-ietf-pce-disco-proto-igp-01.txt

PCE ⇔ PCC Communication

- Based on TCP: reliable transport with flow control
- Open messages provide PCEP session characteristics (Version, Keepalive frequency, Session mode, …)
- Once session is established, path computation requests/replies are exchanged
- Messages types: Request, Reply, Notification, Error
- Request messages include request characteristics (e.g. pre-emption priority) and LSP constraints (bandwidth, delay, etc.)
- Notifications include information on PCE status (e.g. load) — may be used by PCC to select alternate PCE

Protocol Examples

1) Path Comp event
2) PCE selection
3) Path comp request sent



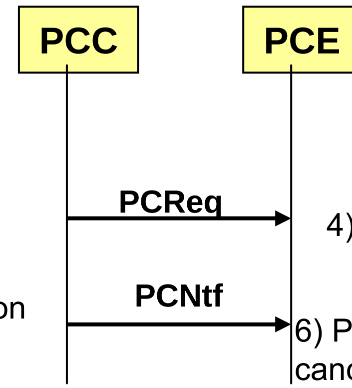
4) Path Comp triggered
Successful computation

5) Computed path(s) sent to the PCC

Path computation request and reply

1) Path Comp event
2) PCE selection
3) Path comp request X sent

5) Path Computation X cancelled

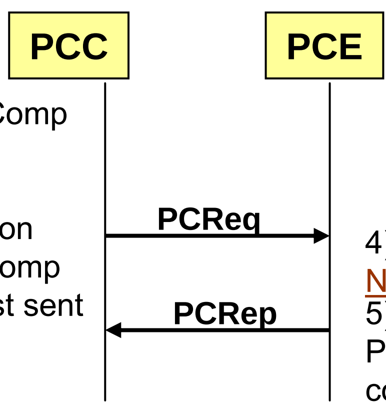


4) Path Comp triggered

6) Path computation X cancelled

Notification (example)

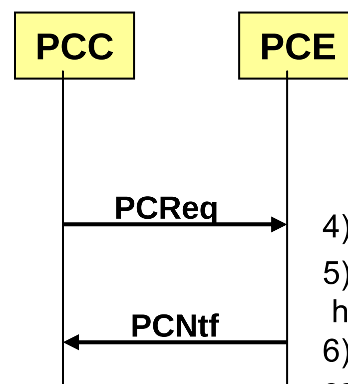
1) Path Comp event
2) PCE selection
3) Path comp request sent



4) Path Comp triggered
No path found
5) Negative reply sent to PCC (optionally with less-constrained path)

Path computation request and reply

1) Path Comp event
2) PCE selection
3) Path comp request X sent



4) Path Comp triggered
5) PCE experiencing high load
6) Path computation X cancelled

Notification (example)

Inter-Domain TE Implementation

- Path computation and signaling:
  - Per-AS/Per-Area Approach—today
  - Distributed path computation (PCE)—prototype
- Inter-AS link flooding
- Reoptimization of Inter-Area/Inter-AS TE LSP
- Policy control at ASBR boundaries (number of TE LSPs, bandwidth, on per-AS basis)
- Integrity object support inter-AS (MD5)
- Fast reroute:
  - ASBR-ASBR link protection
  - ASBR node protection (using nodeID)
  - ABR node protection

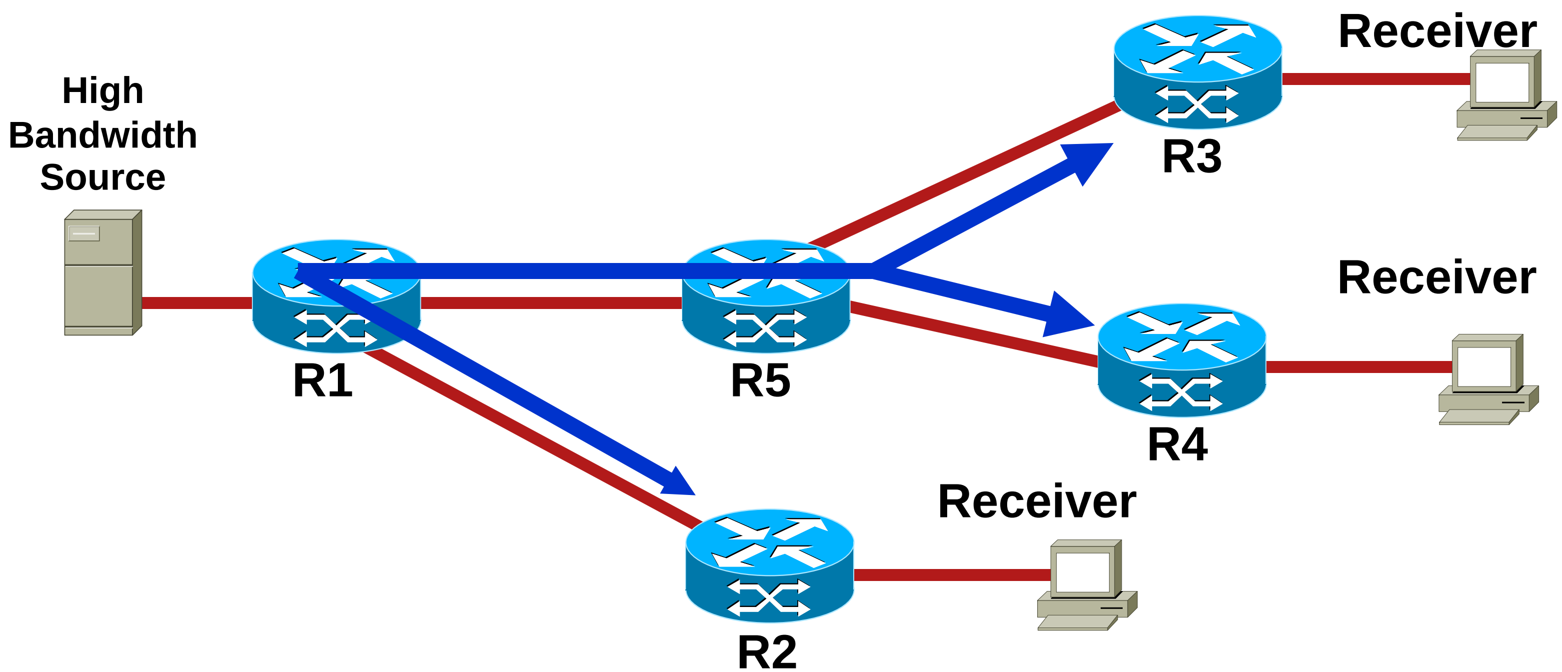
Point-to-Multipoint TE

- Increasing demand to support multicast flows with:
  - High-rate sources (e.g. video/TV distribution)
  - Network optimization (not all traffic on shortest path)
  - QoS guarantees
  - Fast restoration
- This has led to demand for point-to-multipoint TE
- Solutions currently being developed and standardized

P2MP TE – Basic Concepts

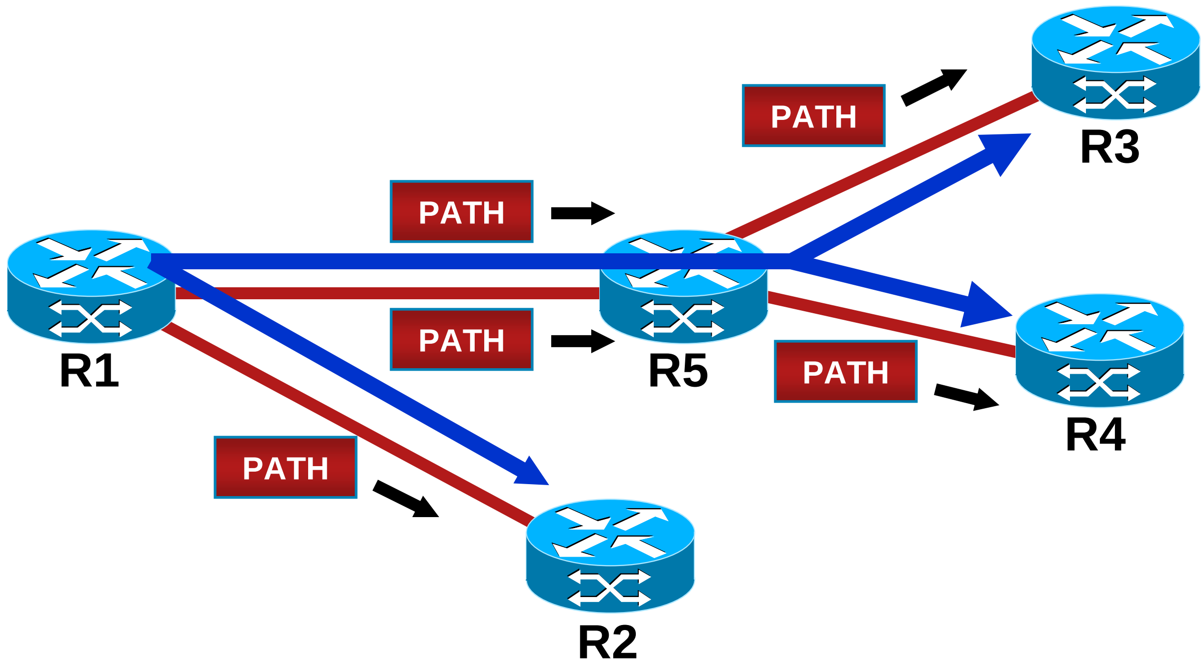
- Each P2MP TE LSP is defined by one head-end and a set of tunnel destinations (or tail-ends)
- Path calculation based on CSPF or explicit path
- P2MP TE LSP segment that runs from source to one leaf forms an S2L sub-LSP
- Each S2L sub-LSP is signaled via a separate RSVP Path message
- TE control plane determines when to perform a “label merge”
- Data-plane builds the label multicast state and merges the S2L sub-LSPs in the forwarding plane

P2MP RSVP-TE



R1 is the head-end
Three tunnel leaves: R2, R3 and R4
R1 sets up and maintains three S2L sub-LSPs via three RSVP Path messages (one per leaf)

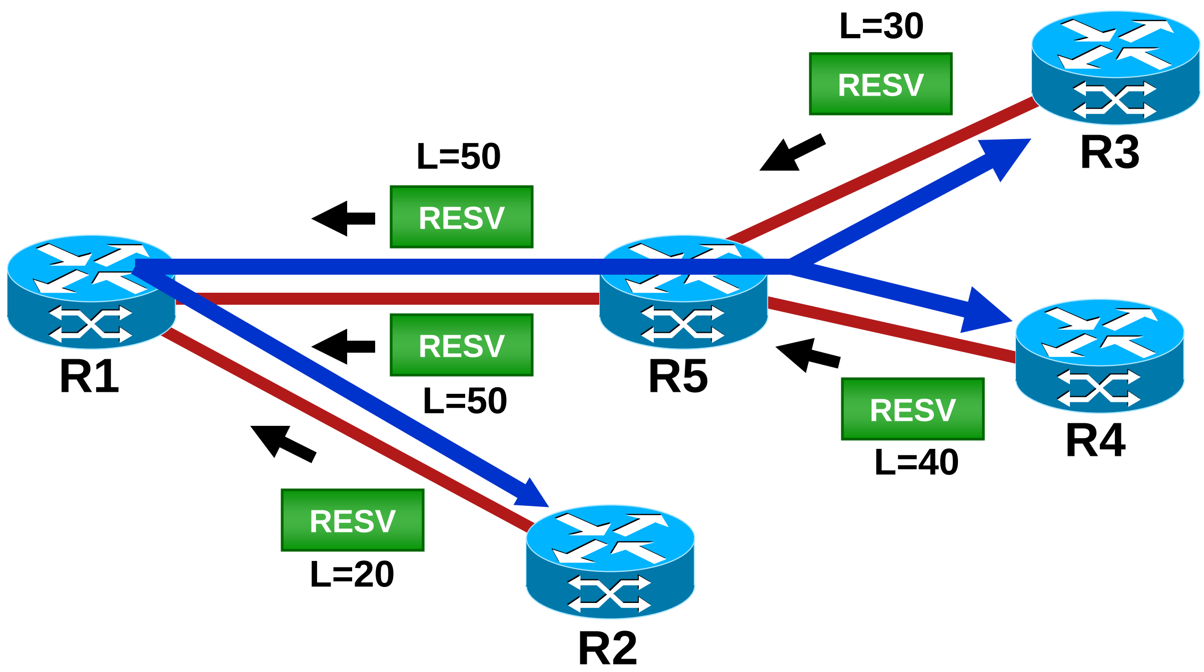
P2MP TE LSP Setup – RSVP PATH Messages



Head-end Router R1 sends three path messages (one per destination)

First PATH message:	**R1 → R5 → R3**
Second PATH message:	**R1 → R5 → R4**
Third PATH message:	**R1 → R2**

P2MP TE LSP Setup – RSVP RESV Message



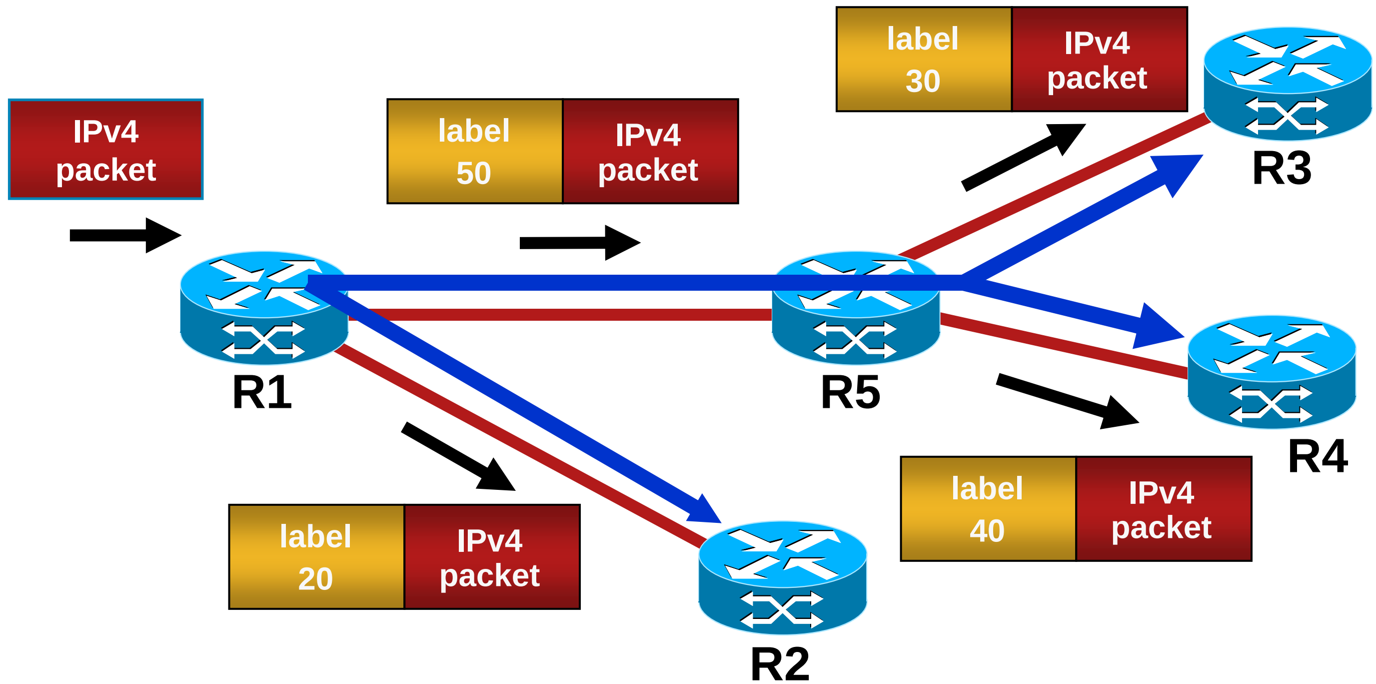
R3 advertises incoming “30”, R4 advertises “40” and R2 advertises “20”

RSVP RESV from R3 and R4 may reach R5 at different times

Upon arrival of RESV from R3, R5 advertises incoming label “50” for the LSP destined for R3

Upon arrival of RESV from R4, R5 realizes that it is a branch point. Hence, R5 also advertises SAME incoming label “50” for LSP destined for R4

P2MP TE LSP Data Plane



Label merging at R5 allows single copy of packet to be sent on R1 → R5 link

TE Standardization

- Inter-AS/inter-area TE
  - RFC 4216: Requirements
  - draft-ietf-ccamp-inter-domain-rsvp-te-03.txt
  - draft-ietf-ccamp-inter-domain-pd-path-comp-03.txt
  - draft-ietf-ccamp-loose-path-reopt-02.txt
- PCE
  - RFC 4655: PCE Architecture
  - RFC 4657: PCE Protocol Requirements
  - draft-ietf-pce-discovery-reqs-06.txt
  - draft-ietf-pce-pcep-02.txt
  - draft-ietf-pce-disco-proto-igp-02.txt
- Point-to-multipoint
  - RFC 4461: Signaling Requirements
  - draft-ietf-mpls-rsvp-te-p2mp-06.txt

Traffic Engineering Summary

- Inter-AS and Inter-Area TE
  - Per-domain and distributed (PCE) approaches
  - Complementary approaches made different tradeoffs
- Point-to-multipoint
  - Simple extensions to point-to-point RSVP-TE
  - Support “provisioned” multicast with TE capabilities

Layer 3 VPNs

L3VPN Agenda

- L3VPN Multicast
  - Recap of Current (deployed) State (draft-rosen[1])
  - Recent Enhancements (L3VPN WG draft[2])
    - Supporting multiple tree types
    - Aggregation
    - Carrying multicast routing in BGP
    - Inter-AS improvements

1. **draft-rosen-vpn-mcast-08.txt**
2. **draft-ietf-l3vpn-2547bis-mcast-02.txt**

L3VPN Multicast - Motivation

- Customers with IP multicast traffic would like to use MPLS VPN services
- RFC 2547/4364 only addresses unicast
- As usual, multicast makes the problem harder
  - Difficult to achieve same scalability as unicast
  - Scalability vs. optimality

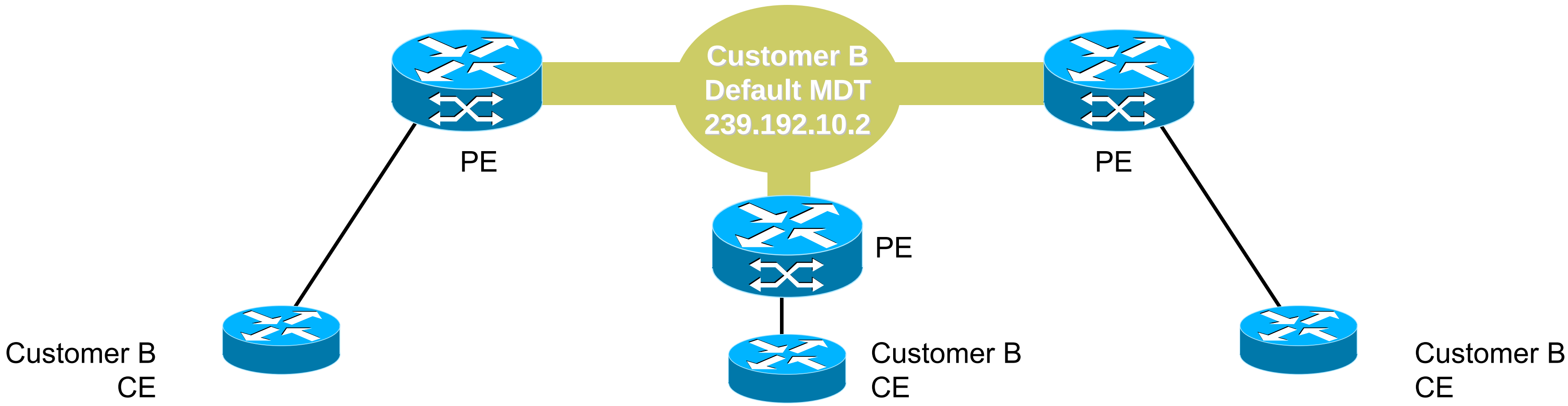
Multicast VPN - Current Deployments

- Based on draft-rosen-vpn-mcast-08.txt
- CE-routers maintain PIM adjacency with PE-router only
  - Similar concept to 2547/4364 VPNs
- P-routers do not hold (S, G) state for individual customers
  - Unlike unicast, there is some per-customer state in P-routers
- PE-routers exchange customer routing information using PIM
  - Contrast to BGP for unicast
- Customer multicast group addresses need not be unique
  - same as unicast addresses

Multicast VPN - Current State (2)

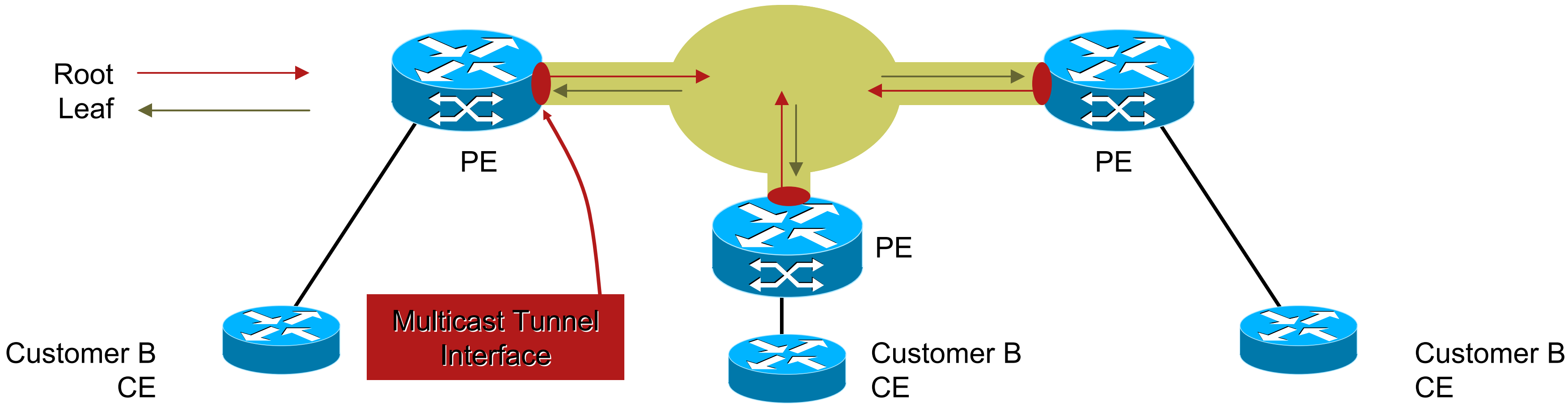
- Multicast domain is a set of multicast enabled VRF’s (mVRF’s) that can send multicast traffic to each other
  - e.g. VRFs associated with a single customer
- Maps all (S, G) that can exist in a particular VPN to a single (S, G) group in the P-network
  - This is the Multicast Distribution Tree (MDT)
  - Amount of P-state is a function of # of VPN’s rather than # of (S, G)’s of all customers
  - This is not as good as unicast, but better than the alternative
- Mapping is achieved by encapsulating C-packet into P-packet using GRE

Default Multicast Distribution Tree



- PE routers build a default MDT in the global table for each mVRF using standard PIM procedures
- All PEs participating in the same mVPN join the same Default MDT
- Every mVRF must have a Default MDT
- MDT group addresses are defined by the provider
    - Unrelated to the groups used by the customer

Default Multicast Distribution Tree



- Default MDT is used as a permanent channel for PIM control messages and low bandwidth streams
- Access to the Default MDT is via a Multicast Tunnel Interface
- A PE is always a root (source) of the MDT
- A PE is also a leaf (receiver) to the MDT rooted on remote PEs

Limitations of Current Model

- At least one multicast tree per customer in core
    - No option to aggregate multicast customers on one tree
- Multicast traffic is GRE (not MPLS) encapsulated
    - Bandwidth and encaps/decaps cost
    - Operational cost - different mcast and unicast data planes
- PIM the only way to build core trees
    - Can’t leverage p2mp RSVP-TE, mLDP
- PE-PE routing exchange using per-customer PIM instances
- Inter-AS challenges

PMSI: P-Multicast Service Interface

- New terms introduced to decouple tree from service
- Three types of PMSI
  - MI-PMSI: Multipoint Inclusive, all→all
    - all PEs can transmit to all PEs
  - UI-PMSI: Unidirectional Inclusive, some→all
    - Unidirectional, selected PEs can transmit to all PEs
  - Selective: S-PMSI, some→some
    - Unidirectional, selected PEs can transmit to selected PEs

draft-ietf-l3vpn-2547bis-mcast

Supporting Multiple Tree Types

- A PMSI is scoped to a single mVPN
- PMSI is instantiated using a tunnel (or set of tunnels)
- Tunnels may be:
  - PIM (any flavor)
  - MPLS (mLDP or p2mp RSVP-TE)
  - Unicast tunnels with ingress PE replication
- Can map multiple PMSIs onto one tunnel (aggregation)
- Encaps a function of tunnel, not service

Mappings to Old Terminology

- Default MDT
  - MI-PMSI, instantiated by PIM Shared Tree or set of PIM Source Trees
- Data MDT
  - S-PMSI, instantiated by PIM Source Tree
- New terminology helpful in:
  - Describing the complete set of options
  - Allowing multiple instantiations of same service, without changing service specification

Aggregation

- Conflicting goals:

  Scale: Minimize P-router state ⇒ Use as few trees as possible

  Optimality: Send traffic at most once on each link, and only to PEs that need it ⇒ Use a tree for each customer multicast group

- Solution: lots of options

  Draft-rosen has one MDT per VPN, and optional data MDT for high BW or sparse customer groups

  New draft also allows a tunnel to be shared among multiple mVPNs

    Better aggregation, less optimality

    Requires a de-multiplexing field (e.g. MPLS label)

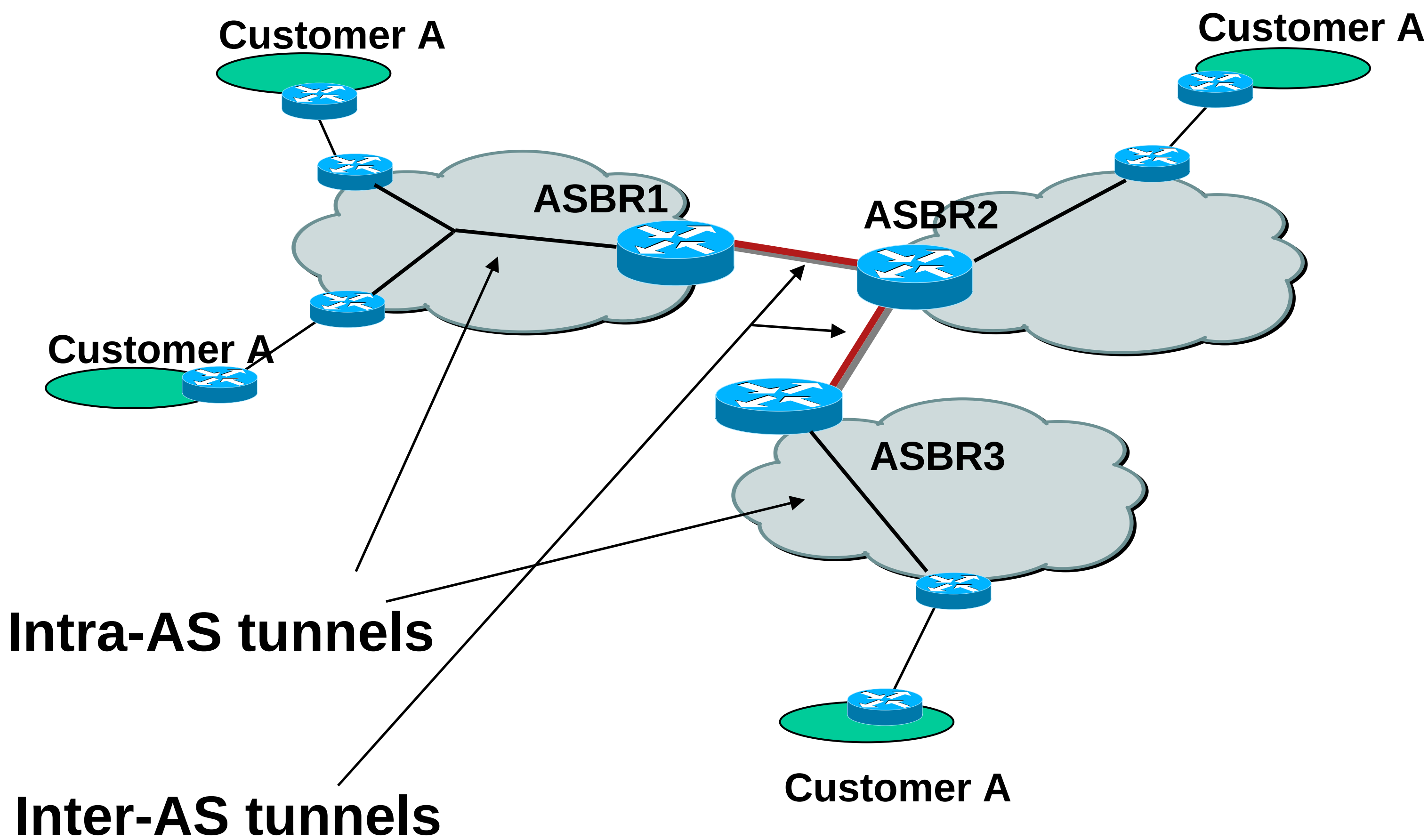
PE-PE routing exchange

- In draft-rosen, C-PIM instances exchange PIM messages over the MDT as if it were a LAN
  - Per-customer PIM peering among the PEs
  - By contrast, one BGP instance carries all customer unicast routes among PEs
  - PIM Hellos can be eliminated, but Join/Prunes remain
- In new draft, BGP is proposed, as in unicast
  - New AFI/SAFI
  - Advertisement contains essentially the same info as a PIM join or prune (source, group, PE sending the message)
  - RDs are used to disambiguate customer multicast group and source addresses
  - BGP route reflectors may be used

Inter-AS

- Current (draft-rosen) approach: tunnel spans multiple ASes
  - Undesirable fate-sharing, must agree on tunnel type
- New (draft-ietf) approach: may “splice” tunnels from different ASes
  - Allows each AS to build its tunnels independently of other ASes
  - Scaling now independent of number of PEs in other ASes
  - Group membership announced using BGP

Inter-AS tunnels



L3 VPN standardization

- RFC4364! No more “2547bis”

  Also L3VPN MIB (RFC4382), applicability statement (RFC4365), OSPF for PE-CE (RFC 4577)

- draft-rosen-vpn-mcast-08.txt

  pre-standard but deployed

- draft-ietf-l3vpn-ppvpn-mcast-reqts-08.txt
- draft-ietf-l3vpn-2547bis-mcast-02.txt
- draft-ietf-l3vpn-2547bis-mcast-bgp-01.txt

L3VPN Summary

- Multicast VPN: improving the solution
  - Support different multicast tree types, including p2mp MPLS-TE and mLDP
  - More flexible aggregation
  - Use of BGP to carry C-routes
  - Better scaling and provider independence for inter-AS

Quality of Service

QoS Agenda

- Tunnel-Based Admission Control (TBAC)
- Interprovider QoS
- Routing Protocol Support for QoS

Is Admission Control Needed?

It depends on the environment & goals

- QoS degradation acceptable
  - e.g. free voice on the Internet
- Overprovisioning
  - e.g. corporate voice on switched gigabit campus
- Overprovisioning + Diffserv
  - e.g. corporate voice/video on switched gigabit campus
- “Right-sizing” of links + Diffserv for Voice/Video
  - Reject sessions which “don’t fit” (e.g. during failure) to preserve QoS of other sessions
  - Pre-empt “less important” traffic during unexpected overload
  - e.g. corporate voice/video on WAN links, Mobile Phone Trunking, PSTN Class 5 replacement, military networks, emergency calls

This is the Call Admission scenario

Is Admission Control Needed?

It depends on the environment & goals

Abundant Bandwidth

- Q[illegible] e[illegible]table
  - [illegible]nt[illegible]net
- Overprovisioning
  - e.g. corporate voice on s[illegible]itched giga[illegible] campus
- Overprovisioning + Diff[illegible]erv
  - e.g. corporate voice/vide[illegible] on switche[illegible] gigabit campus
- “Right-sizing” of links + [illegible]iffserv fo[illegible] Voice/Video
  - Reject sessions which “c[illegible]n’t fit” (e.g. [illegible]uri[illegible] [illegible] of other sessions
  - Pre-empt “less important[illegible] traffic durin[illegible] [illegible]un[illegible]
  - e.g. corporate voice/video on WAN lin[illegible]s, [illegible] [illegible]TN Class 5 replacement, military networks, emerg[illegible]ency calls

Greater QoS Assurance

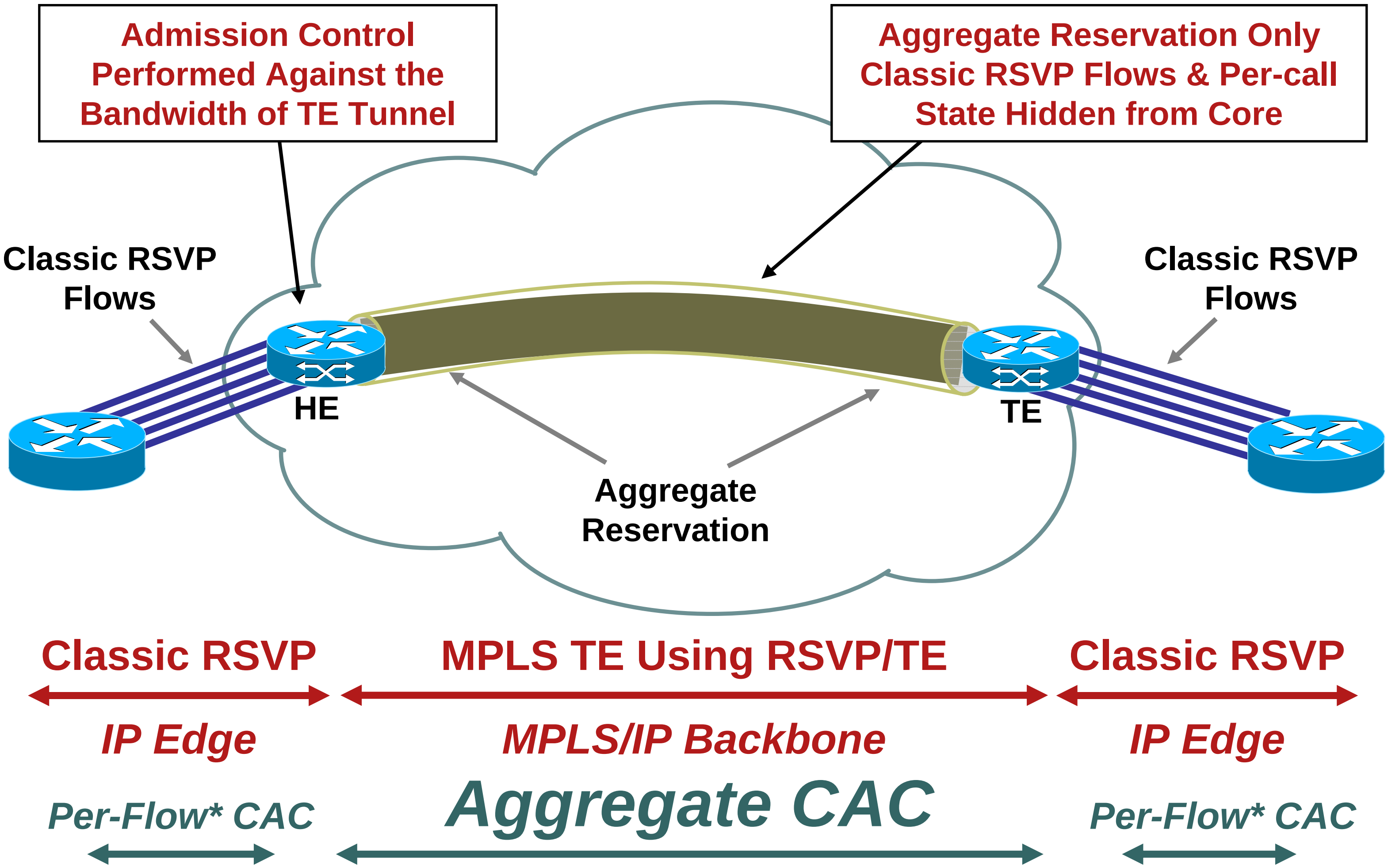
This is the Call Admission scenario

IETF Components for Scalable CAC solution

- Clean separation between Bearer Control and Call Control
  - Call Control (e.g. SIP, H.323) completely leaves it to Bearer Control to make the right QoS/CAC decisions and report
- On-Path ("in-band") CAC using RSVP
  - Explicit CAC decision on actual path followed by sessions
- Use of TE/DS-TE tunnel for Aggregate Bearer Control in Core
- Use of RSVP for finer Bearer Control on the edge
- RSVP Aggregation over MPLS TE Tunnels
- Synchronization between RSVP and Call Control (e.g. SIP) on end-device

draft-ietf-tsvwg-rsvp-dste-03.txt

CAC Scalability: RSVP Aggregation



* Hierarchical Aggregation allows Aggregate CAC at edge as well

Scalable Bearer Control in Core

- No per-session bearer in core
- Aggregate Bearer Control (e.g. one reservation per PoP pair)
- MPLS TE (or DS-TE) tunnel is ideal Aggregate Bearer:
  - Bandwidth Reservation
  - Aggregate CAC
  - Operational experience at large scale
  - Constraint Based Routing
  - Path engineered against many parameters (delay metric, max voice utilization, …)
  - Protection by MPLS Fast ReRoute
  - Dynamic Resizing
  - Support for different classes of service via DS-TE

RSVP for QoS?

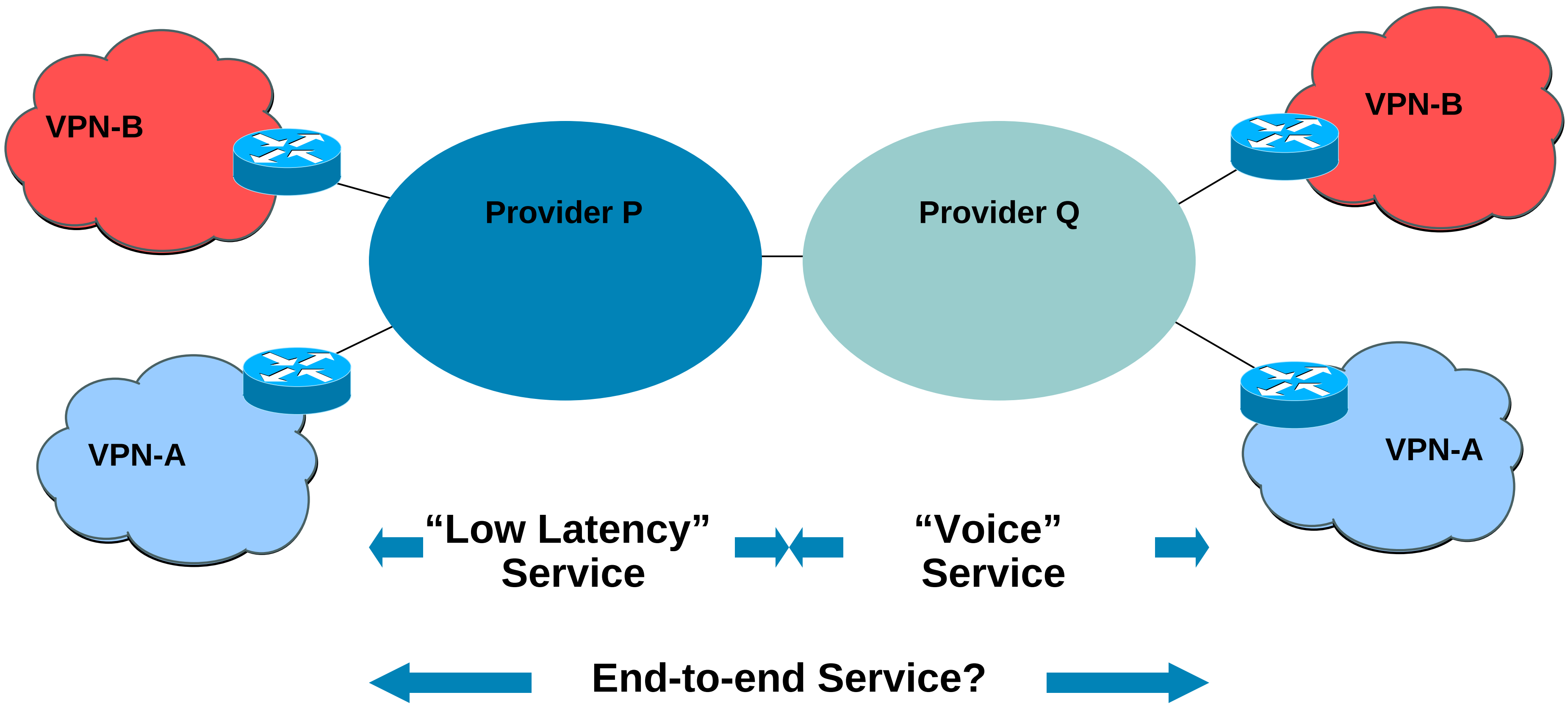
- "I thought RSVP was…
  - Dead
  - Unscalable
  - Only for TE
- Scalability issues are all around per-flow reservations
  - We avoid those or push them to edges
- RSVP is undergoing resurgence due to
  - Greater deployment of QoS-dependent applications (e.g. video)
  - Need for policy-aware admission control (e.g. preemption of less important traffic during overload)

Pace of Application Deployment

RSVP Aggregation: Key Features

- Flexibility to perform CAC on one or more segments:
  - GW→PE, PE→PE, PE→GW
- No assumption of symmetric bandwidth
- Range of TE Tunnel deployment models:
  - PE→PE mesh, P→P mesh (GW not directly connected to TE Headend), any combination
- Flexible granularity of GW-GW RSVP reservations
  - May be per-call, per-gateway pair, or between these extremes
  - Hierarchical Aggregation Support
- No restrictions on scope of RSVP signaling
  - End-to-end RSVP signaling, RSVP signaling localised onto GW→Headend segment (while retaining CAC over TE Tunnel)
- Dynamic adaptation to topology change
  - If a GW is suddenly reachable through a different tunnel, CAC adjusts immediately (and reservation is maintained if it fits)

The Challenge of Interprovider QoS



- Other issues: measurement, monitoring, troubleshooting, impairment allocation

Interprovider QoS Axioms

- Leverage what works today

  e.g. Diffserv deployed in majority of single-provider VPNs (at edges)

- Don't constrain providers more than necessary

  e.g. Leave them free to overprovision the core, or apply more complex mechanisms like DS-TE

- Mechanisms should support wide range of services/applications

  e.g. VOIP, MPLS-VPNs, Internet,…

- Troubleshooting/monitoring must be addressed
- Don't neglect business/economic issues

Service Classes

- Key goal: ability to build an end-to-end service from the concatenated services of multiple providers
- Achieving this goal requires:
    - A small set of common services supported by all providers
    - Agreement on the metrics (loss, delay, jitter, etc.) by which services are defined
    - Agreed methodology for allocating impairment budgets
- A provider can offer many services at the edge mapping to a few classes in the core
    - One way to avoid the "commoditization" concern of service class standardization

Routing for Interprovider QoS

- Problem:
  - Provider may wish to send traffic with QoS assurances via one provider and best effort via another
  - BGP has no means to identify the QoS capabilities supported along a path
  - BGP (usually) selects only one path to a destination

Basics of BGP functionality

- What can BGP do?
  - Find routes which (claim to) support a given QoS end-to-end
- What can't BGP do?
  - Treat QoS as anything other than opaque
  - Signal dynamic path characteristics (e.g., instantaneous loss or delay)

BGP for Service (QoS) Routing

- BGP well-suited to carrying multiple classes of routing information (MP-BGP)
- Could consider QoS as a distinct class of routes
    - Service classes, metrics, etc. are opaque — BGP simply signals reachability
- Small number of classes = tractable problem
- Solution approach: Minimal extensions to BGP to:
    - allow a set of routes (NLRI) to be associated with a given service class
    - advertise up to one path per class to given prefix

QoS Standardization

- RSVP items in the Transport Area
  - draft-ietf-tsvwg-rsvp-dste-01.txt
  - draft-ietf-tsvwg-rsvp-bw-reduction-02.txt
  - draft-ietf-tsvwg-rsvp-ipsec-00.txt (actually RSVP aggregation)
- Interprovider QoS
  - http://cfp.mit.edu/groups/internet/qos.html
  - draft-ietf-idr-bgp-multisession-02.txt
  - draft-djernaes-simple-context-update-00.txt

QoS Summary

- Tunnel-Based Admission Control (TBAC)
  - Part of the scalable admission control solution
  - Leverages the use of RSVP by end systems/gateways
- Interprovider QoS
  - Important next step beyond today's Diffserv deployments
- Routing for QoS
  - Simple increment to BGP to advertise "per class" NLRI

Layer 2 VPNs

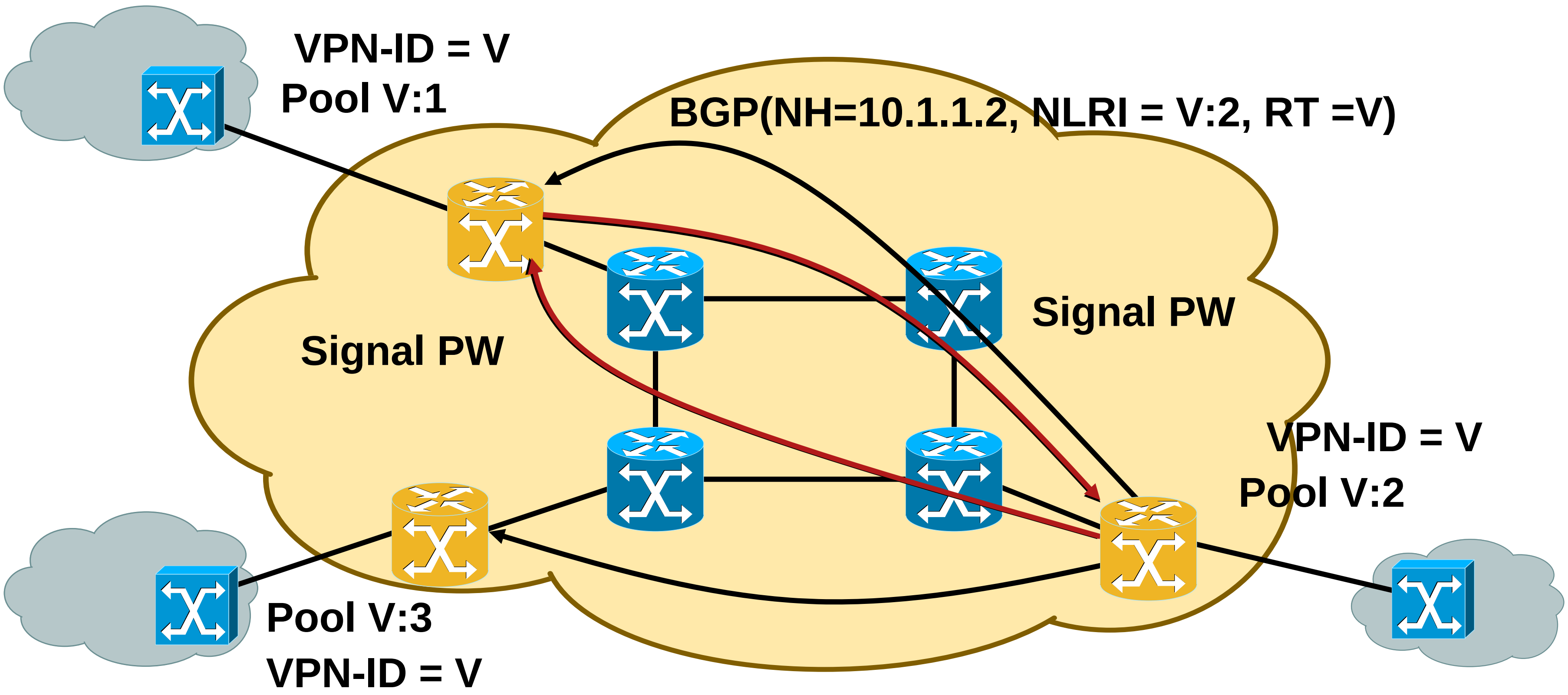
L2VPN Agenda

- L2VPN Autodiscovery
- Inter-AS L2VPNs

Separation of Discovery and Signaling

- Signaling and discovery are separable parts of L2VPN establishment
- Discovery (finding members of an L2VPN) is a point-multipoint task
- Signaling (establishing the pseudowires) is a point-point task
- By separating the tasks, you can choose a suitable protocol for each:
    - LDP, L2TPv3 for PW signaling
    - BGP, RADIUS, etc. for discovery

L2VPN Auto-Discovery
"Colored Pools"



1. Assign pool names and VPN-IDs
2. BGP advertises pools

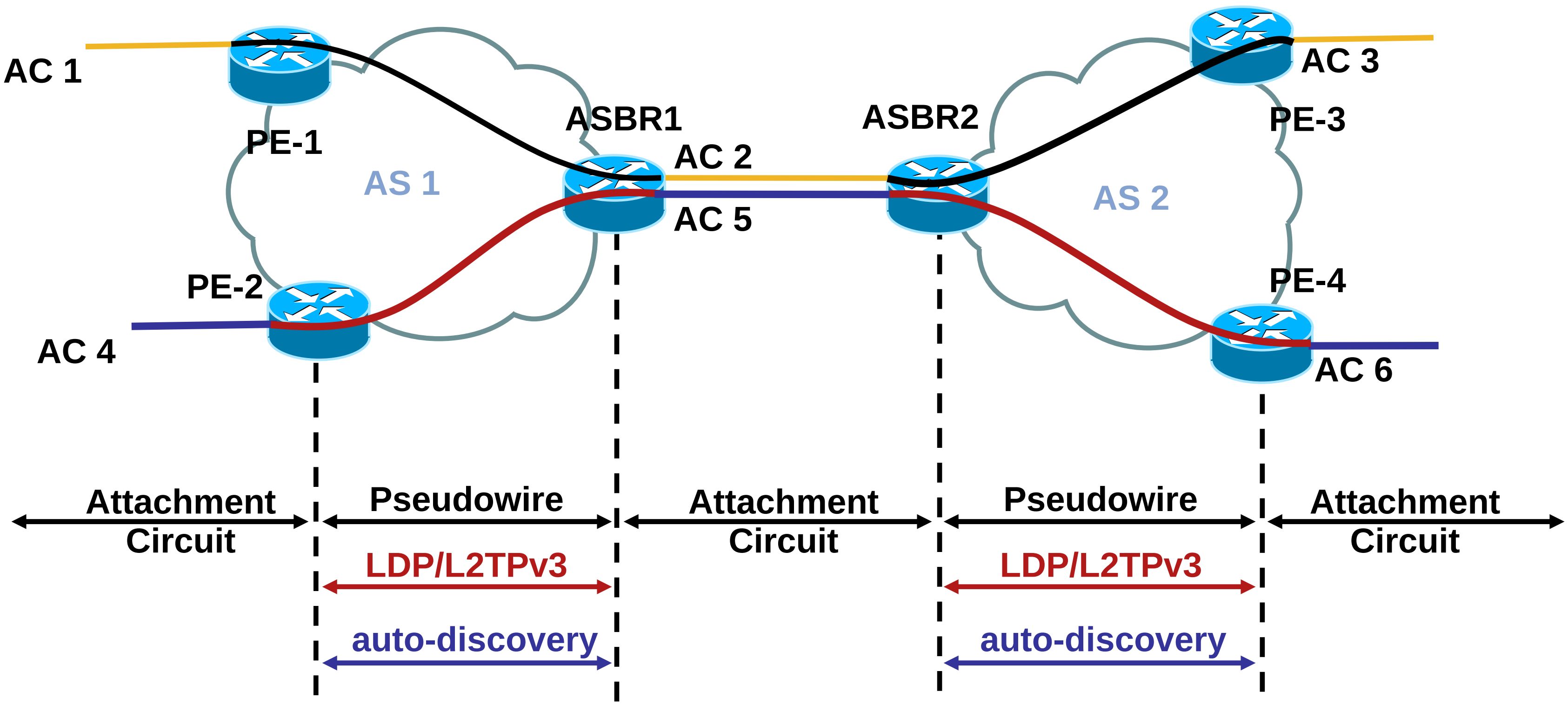
draft-ietf-l2vpn-signaling-08.txt

3. PE automatically signals PW between 2 members of pool
4. Far PE signals reverse direction

Summary of Inter-AS L2VPN Options

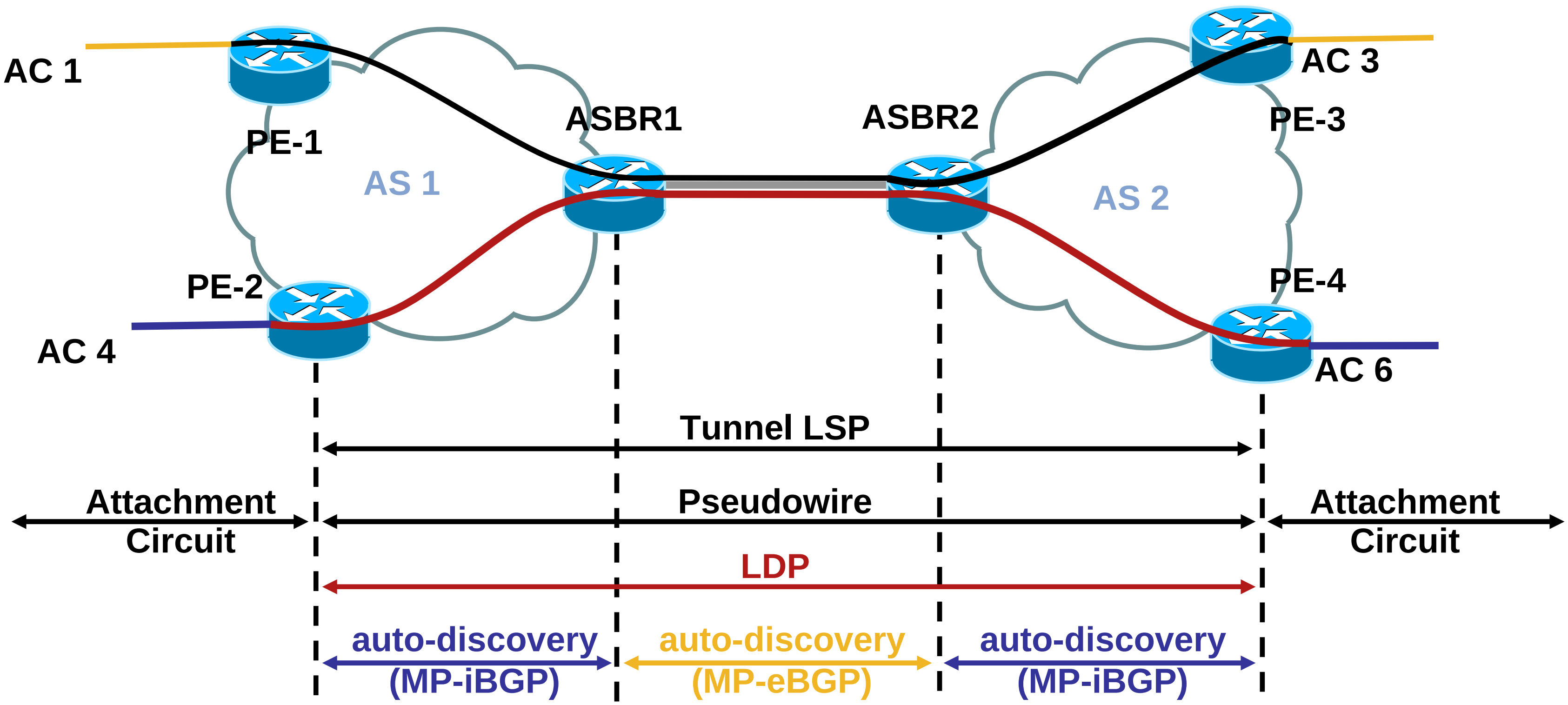
- Interconnected attachment circuits
  - Like RFC 4364 option (a)
  - Good isolation between providers, more provisioning effort
- Multi-AS tunnel
  - Like option (c)
  - Requires more trust between providers
  - PE-PE IP tunnels also an option
- Multi-Segment PW
  - Like option (b)
  - Provides more control, good scaling of signaling

Connected Attachment Circuits



Analogous to L3VPN Model (a) — Back-to-Back VRF

Multi-AS Tunnel LSP Model

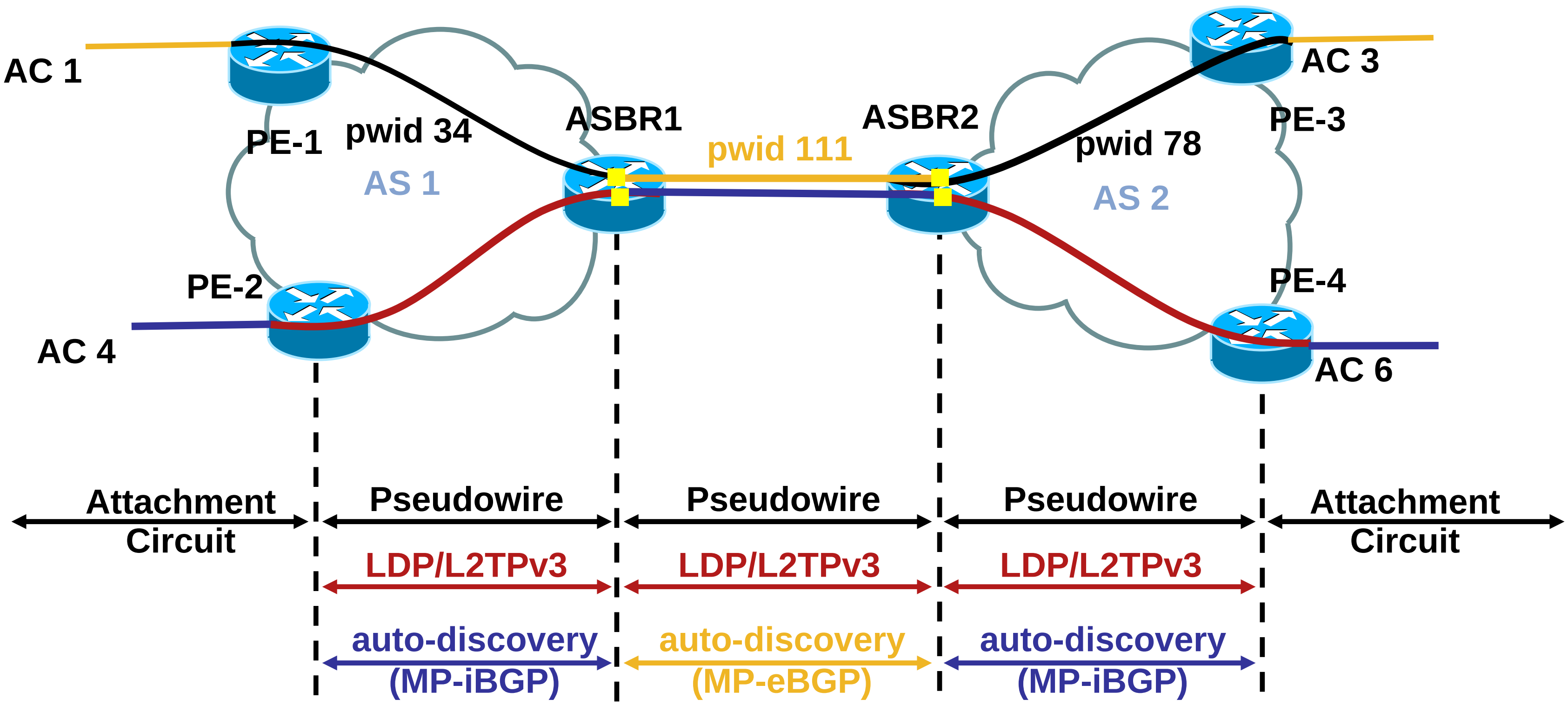


- PE-PE LSP is built as per L3VPN option (c)

    Addresses of PEs + labels carried in BGP

- PW signaling from PE-PE

Multi-Segment Pseudowire Model



- Can be manually configured as per attachment circuit model
- Can support auto-discovery analogous to L3VPN Model (b)—eBGP used between ASBRs
- Limiting PW signaling to ASBRs gives control over policy and avoids mesh of PE-PE signaling

draft-ietf-pwe3-segmented-pw-02.txt

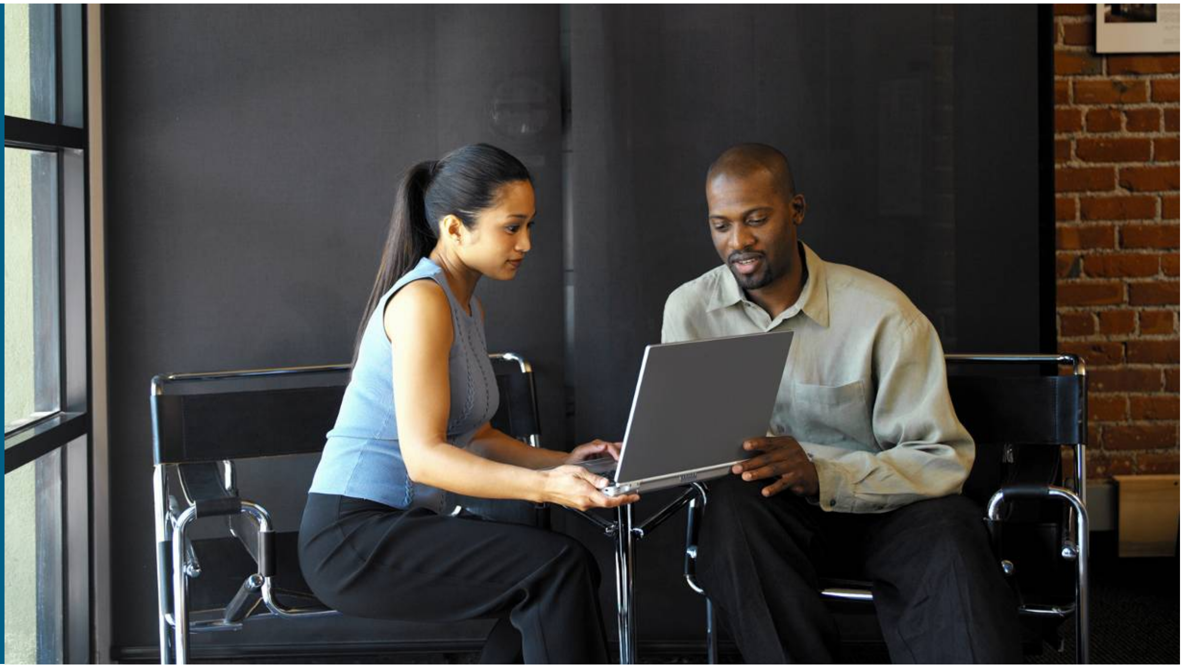
L2VPN Standardization

- Main VPLS drafts: to RFC
  - draft-ietf-l2vpn-vpls-ldp-09.txt
  - draft-ietf-l2vpn-vpls-bgp-08.txt
  - draft-ietf-l2vpn-signaling-08.txt
- Multi-Segment PW
  - draft-ietf-pwe3-ms-pw-requirements-02.txt
  - draft-ietf-pwe3-segmented-pw-02.txt
  - draft-ietf-pwe3-dynamic-ms-pw-01.txt

L2VPN Conclusions

- Inter-AS support emerging as requirement for L2VPNs
  - Both multiprovider and single-provider applications
- Range of inter-AS models are possible, similar to those for L3VPNs
  - Tradeoffs among trust, control, configuration cost
- Separation of discovery from signaling provides flexibility and modularity
- Scalability appears no worse than single-AS case

Concluding Remarks



MPLS: New Developments

- Traffic engineering
  - Moving beyond single-area, single-AS deployments
  - Path Computation Elements
  - Point-to-multipoint
- L3VPN
  - Multicast - improving scalability & flexibility
- QoS
  - Scalable admission control using TBAC
  - Inter-provider QoS gathering momentum
- L2VPNs
  - Signaling with LDP, auto-discovery with BGP
  - Inter-AS operation the next step—options similar to L3VPNs

Meet the Experts
IP and MPLS Infrastructure Evolution

- Andy Kessler
  Technical Leader
- Beau Williamson
  Consulting Engineer
- Benoit Lourdelet
  IP services Product manager
- Bertrand Duvivier
  Consulting Systems Engineer
- Bruce Davie
  Cisco Fellow
- Bruce Pinsky
  Distinguished Support Engineer



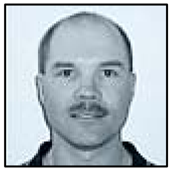



Meet the Experts
IP and MPLS Infrastructure Evolution

- Gunter Van de Velde
  Technical Leader
- John Evans
  Distinguished Systems Engineer
- Oliver Boehmer
  Network Consulting Engineer
- Patrice Bellagamba
  Consulting Engineer
- Shannon McFarland
  Technical Leader







Meet the Experts
IP and MPLS Infrastructure Evolution

- Andres Gasson
  Consulting Systems Engineer

- Steve Simlo
  Consulting Engineer

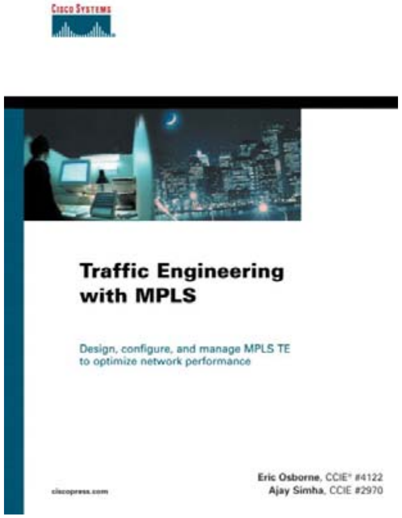
- Toerless Eckert
  Technical Leader

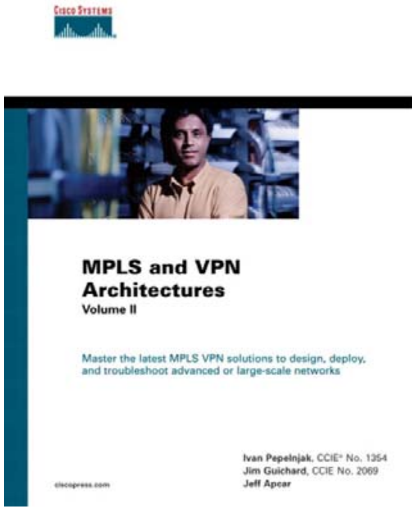
- Dino Farinacci
  Cisco Fellow & Senior Software Engineer

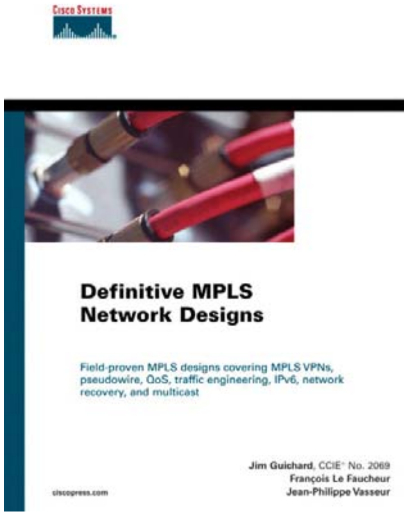
Recommended Reading

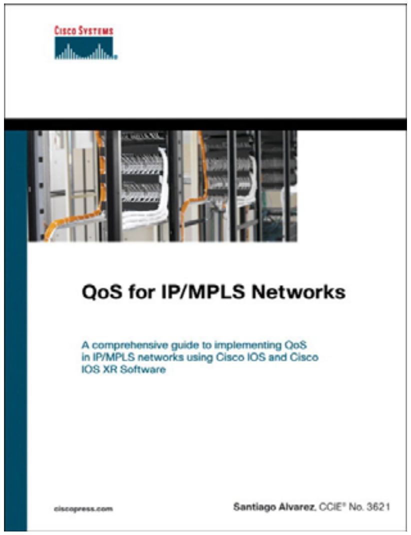
BRKIPM -3003

- Traffic Engineering with MPLS
- MPLS and VPN Architectures, Volume II
- Definitive MPLS Network Designs
- QoS for IP/MPLS Networks









Available in the Cisco Company Store