MPLS and Traffic Engineering

Juniper NETWORKS™

Agenda

- MPLS Fundamentals
- Traffic Engineering
- Constraint-Based Routing

MPLS Primer

Juniper NETWORKS™

So……

What's……..

MPLS ??????????????

Why Is MPLS an Important Technology?

- Fully integrates IP routing & L2 switching
- Leverages existing IP infrastructures
- Optimizes IP networks by facilitating traffic engineering
- Enables multi-service networking
- Seamlessly integrates private and public networks
- The natural choice for exploring new and richer IP service offerings
- Dynamic optical bandwidth provisioning

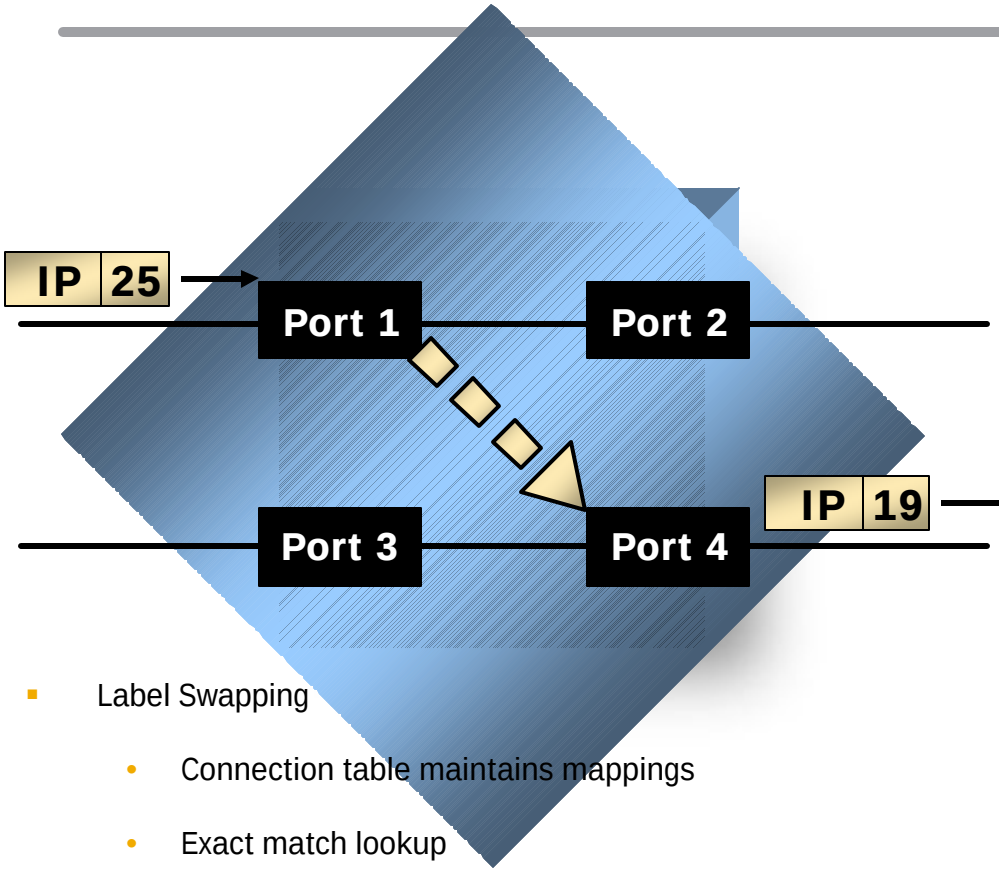
What Is MPLS?

- IETF Working Group chartered in spring 1997
- IETF solution to support multi-layer switching:
  - IP Switching (Ipsilon/Nokia)
  - Tag Switching (Cisco)
  - IP Navigator (Cascade/Ascend/Lucent)
  - ARIS (IBM)
- Objectives
  - Enhance performance and scalability of IP routing
  - Facilitate explicit routing and traffic engineering
  - Separate control (routing) from the forwarding mechanism so each can be modified independently
  - Develop a single forwarding algorithm to support a wide range of routing and switching functionality

MPLS Terminology

- Label
  - Short, fixed-length packet identifier
  - Unstructured
  - Link local significance
- Forwarding Equivalence Class (FEC)
  - Stream/flow of IP packets:
    - Forwarded over the same path
    - Treated in the same manner
    - Mapped to the same label
  - FEC/label binding mechanism
    - Currently based on destination IP address prefix
    - Future mappings based on SP-defined policy

MPLS Terminology

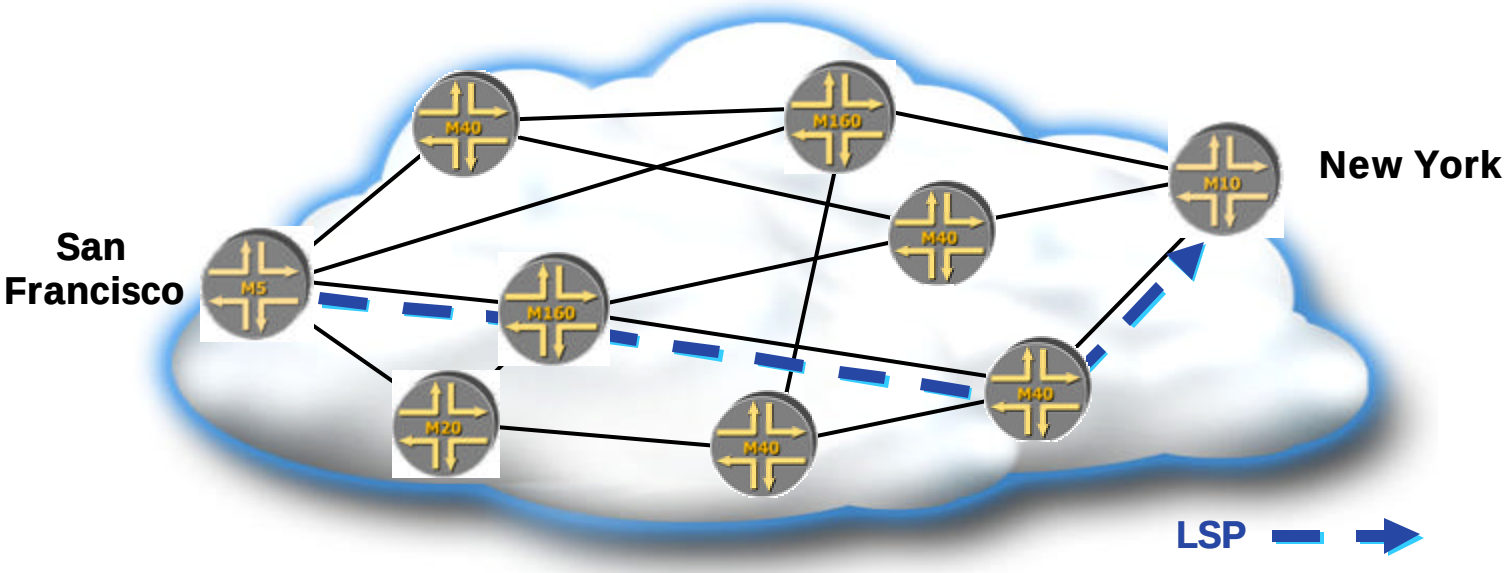


Connection Table

In (port, label)	Out (port, label)	Label Operation
(1, 22)	(2, 17)	Swap
(1, 24)	(3, 17)	Swap
(1, 25)	(4, 19)	Swap
(2, 23)	(3, 12)	Swap

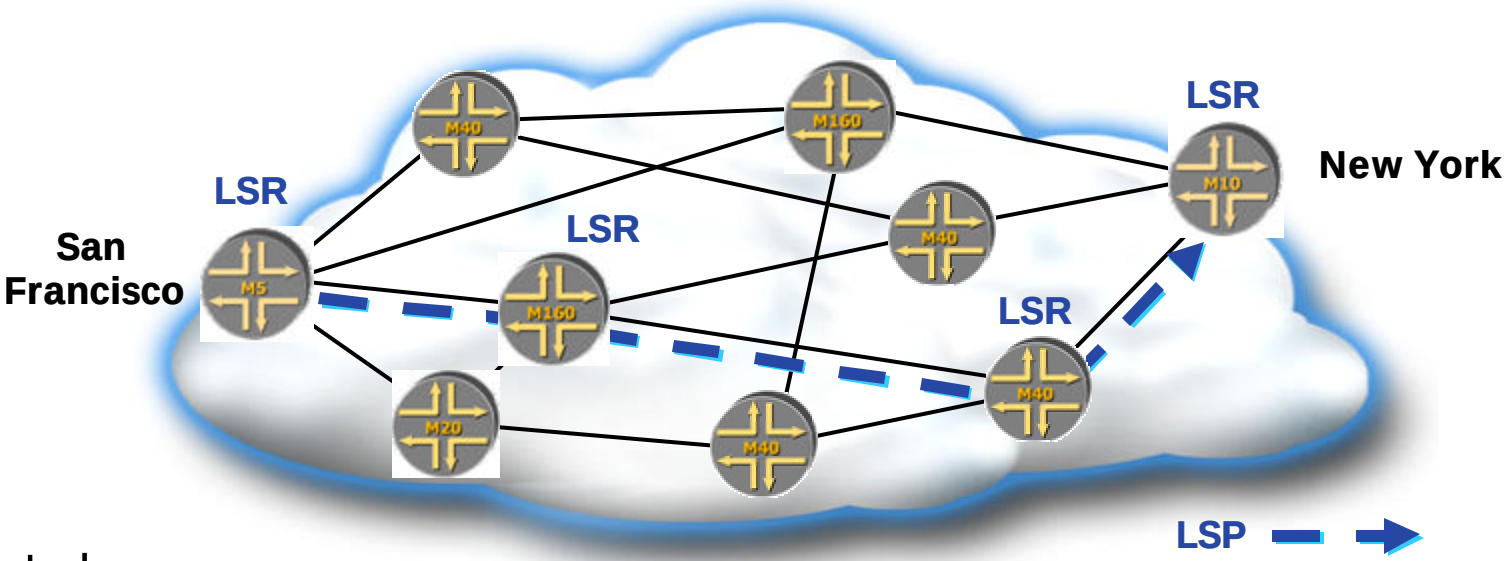
- Label Swapping
  - Connection table maintains mappings
  - Exact match lookup
  - Input (port, label) determines:
    - Label operation
    - Output (port, label)
  - Same forwarding algorithm used in Frame Relay and ATM

MPLS Terminology



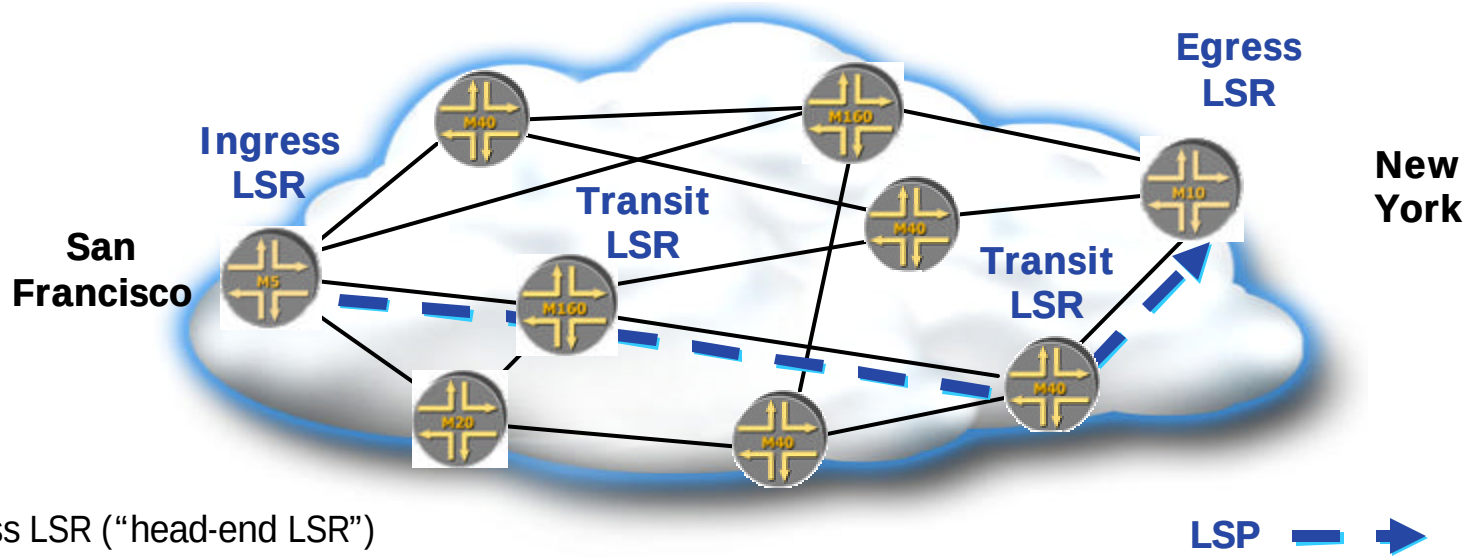
- Label-Switched Path (LSP)
  - Simplex L2 tunnel across a network
  - Concatenation of one or more label switched hops
  - Analogous to an ATM or Frame Relay PVC

MPLS Terminology



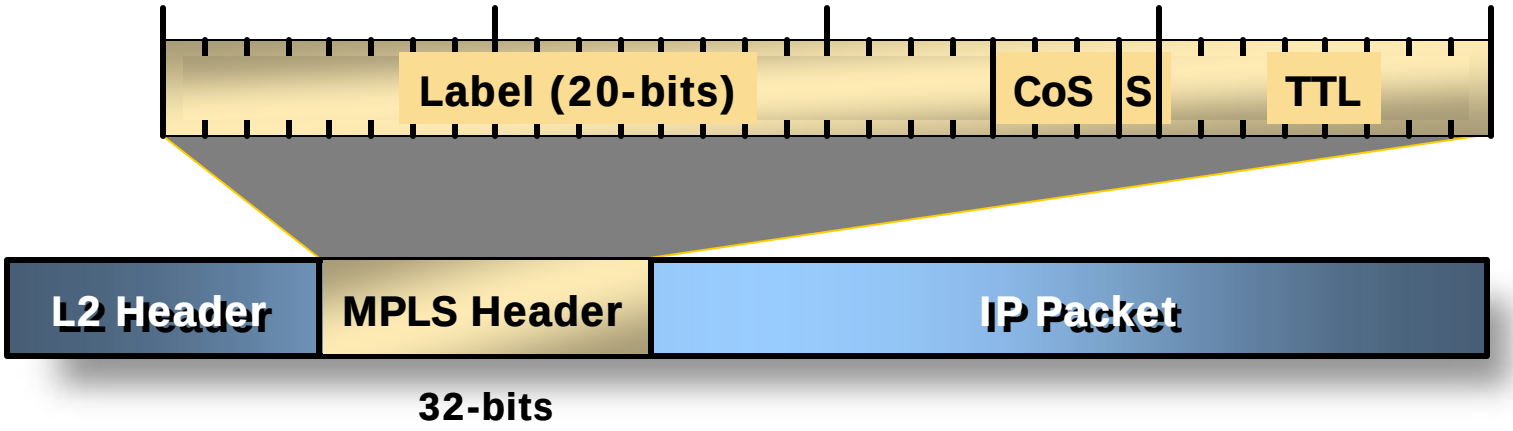
- Labe
  - Forwards MPLS packets using label-switching
  - Capable of forwarding native IP packets
  - Executes one or more IP routing protocols
  - Participates in MPLS control protocols

MPLS Terminology



- Ingress LSR ("head-end LSR")
  - Examines inbound IP packets and assigns them to an FEC
  - Generates MPLS header and assigns initial label
- Transit LSR
  - Forwards MPLS packets using label swapping
- Egress LSR ("tail-end LSR")
  - Removes the MPLS header

MPLS Header

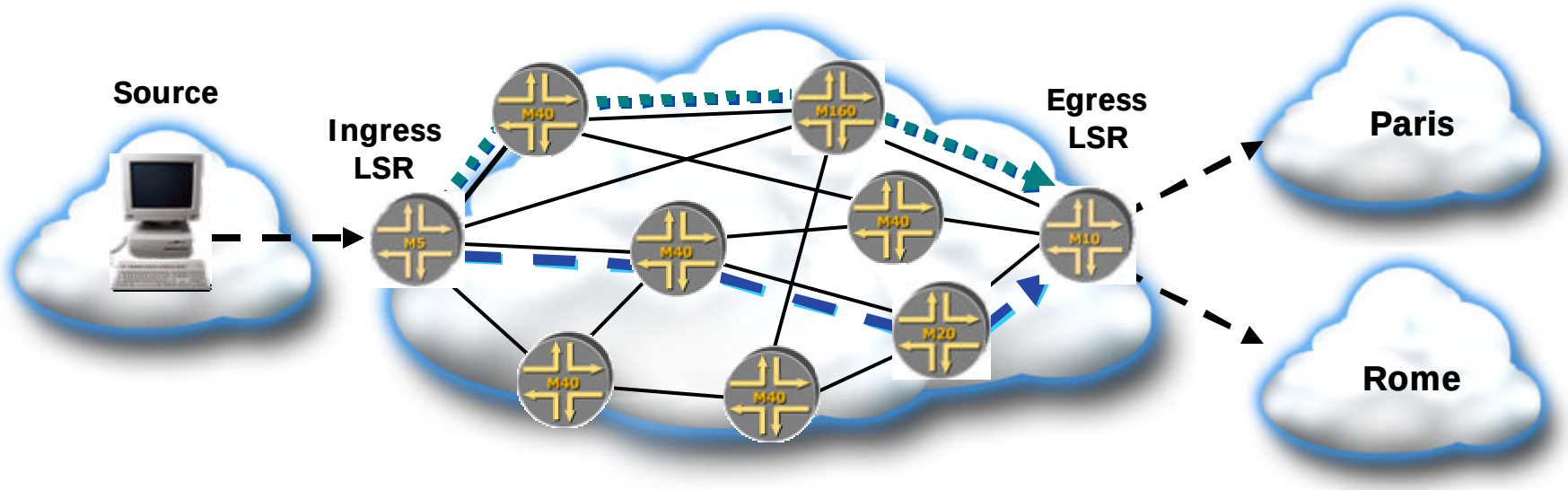


- Fields
  - Label
  - Experimental (CoS)
  - Stacking bit
  - Time to live
- IP packet is encapsulated by ingress LSR
- IP packet is de-encapsulated by egress LSR

Lets Review

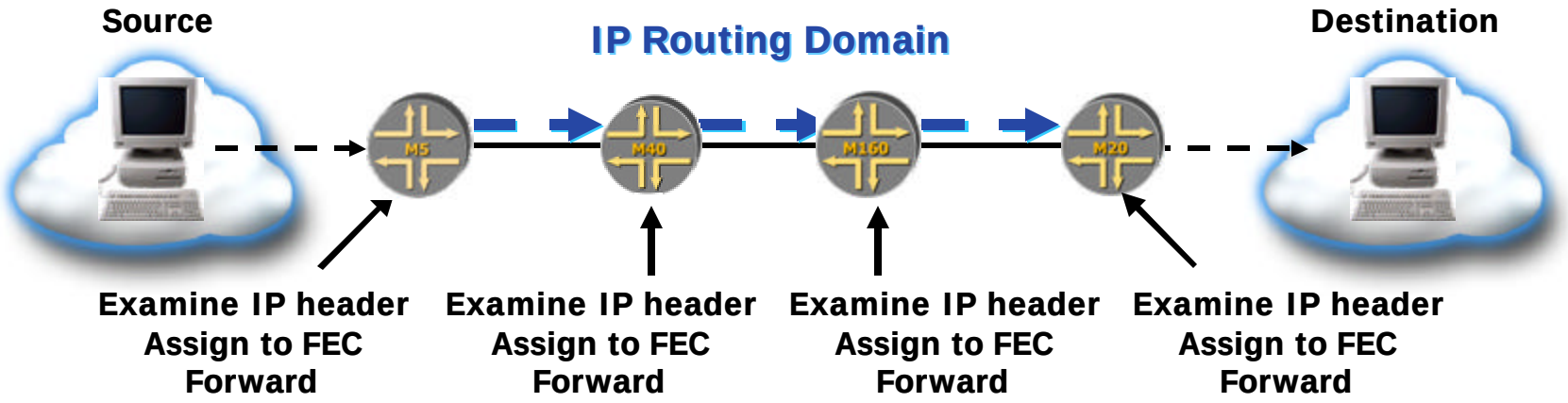
MPLS Packet Forwarding

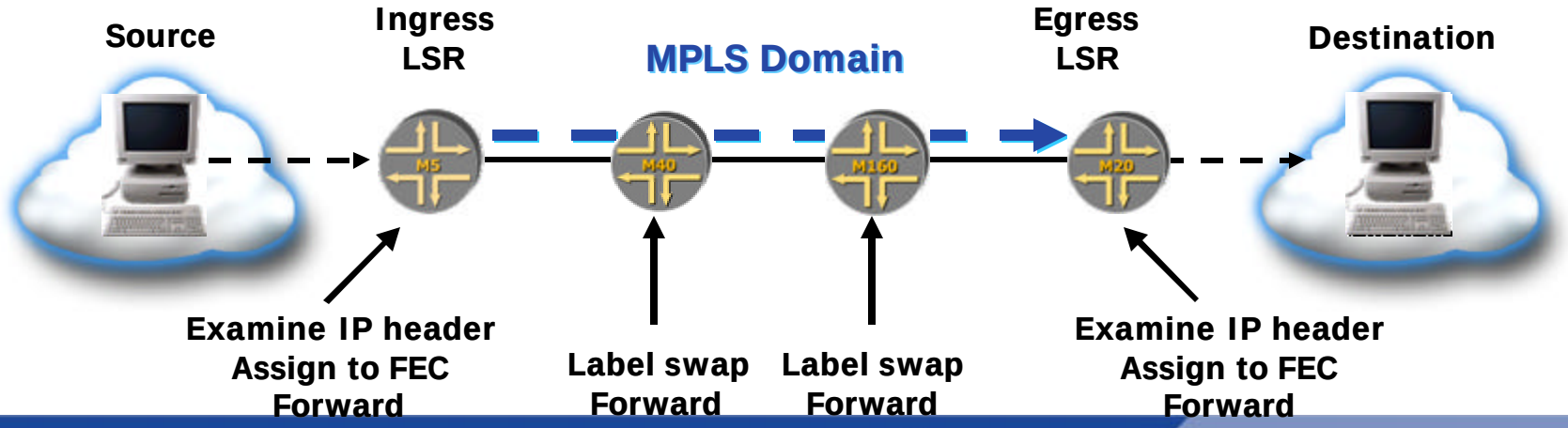
MPLS Forwarding Model



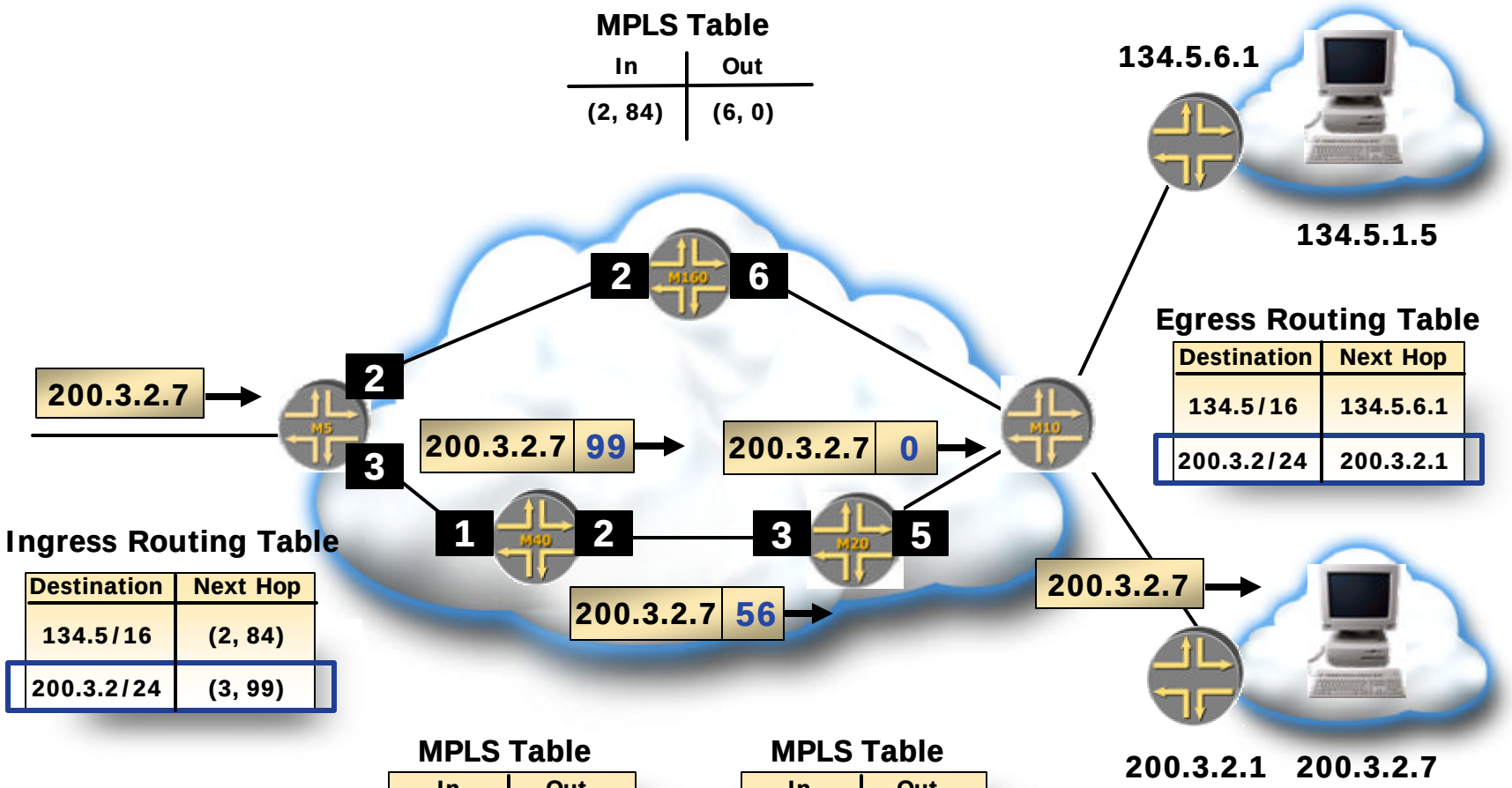
- Ingress LSR determines FEC and assigns a label
  - Forwards Paris traffic on the Green LSP
  - Forwards Rome traffic on the Blue LSP
- Traffic is label swapped at each transit LSR
- Egress LSR
  - Removes MPLS header
  - Forwards packet based on destination address

MPLS Forwarding vs. IP Routing





MPLS Forwarding Example



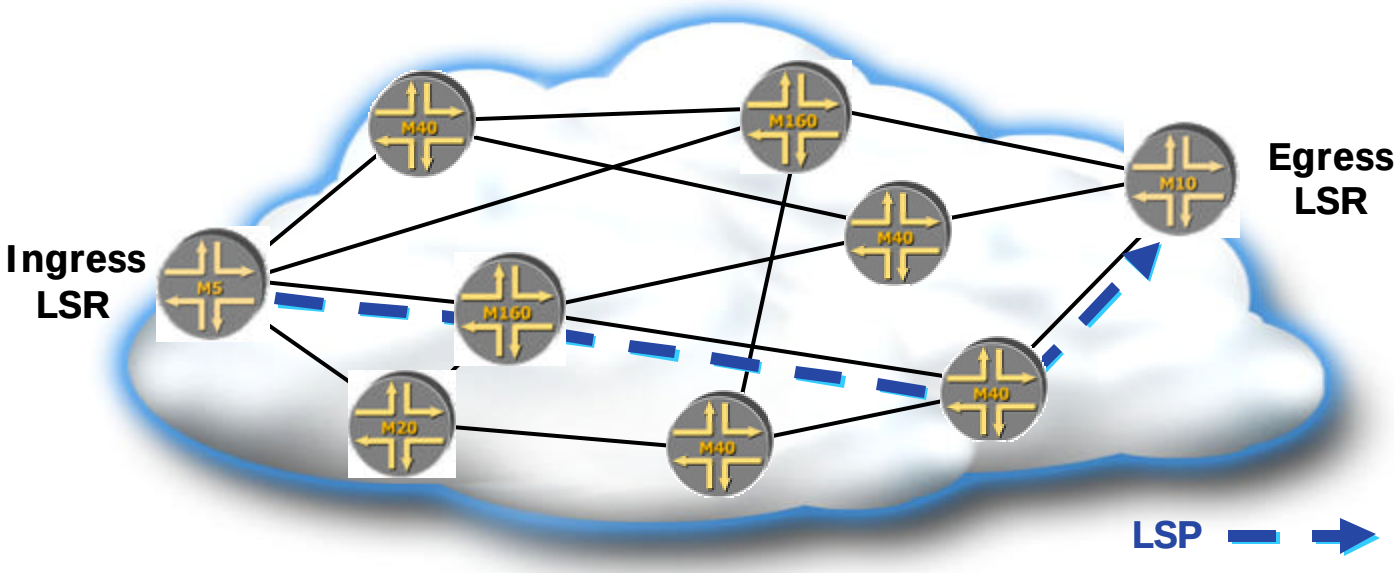
But There's Much More ...

... to MPLS than simple packet forwarding!

- How is the physical path for each LSP determined?
- How is an LSP established?
  - Label distribution and coordination
  - Bandwidth reservation
- How does the ingress LSR map traffic to an LSP?
- Does MPLS support a routing hierarchy?
- Can the LSP physical path calculation be performed online?

MPLS Physical Path Determination

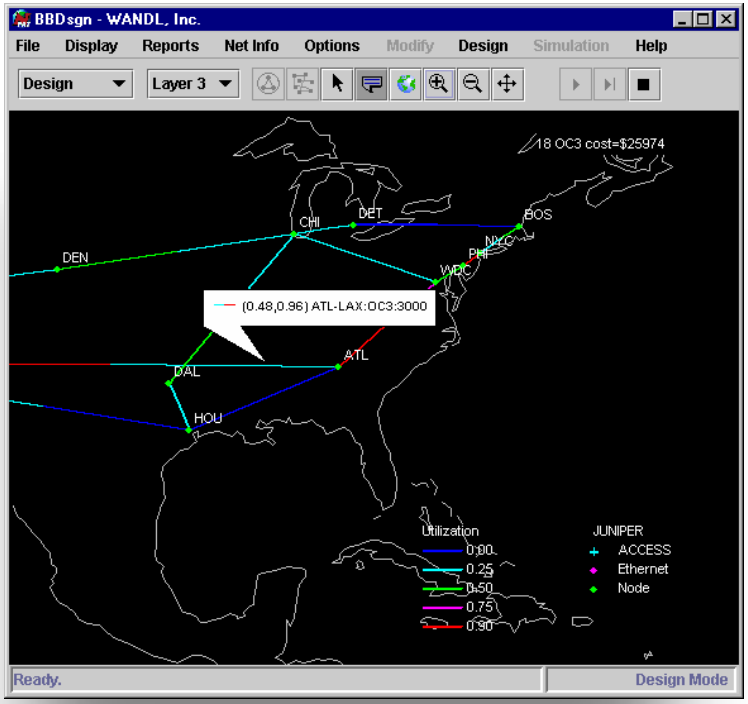
How Is the LSP Physical Path Determined?



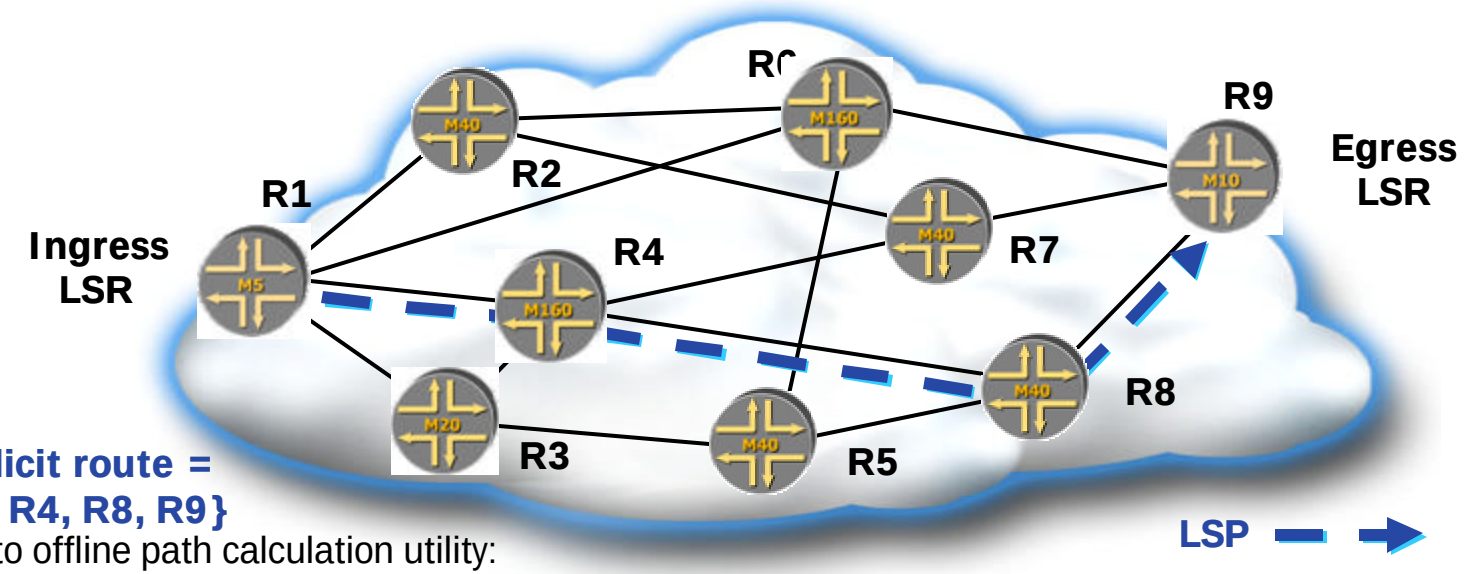
- Two approacnes:
  - Offline path calculation (in house or 3rd party tools)
  - Online path calculation (constraint-based routing)
- A hybrid approach may be used
- Much more about constraint-based routing later!

Offline Path Calculation

- Simultaneously considers
  - All link resource constraints
  - All ingress to egress traffic trunks
- Benefits
  - Similar to mechanisms used in overlay networks
  - Global resource optimization
  - Predictable LSP placement
  - Stability
  - Decision support system
- In-house and third-party tools



Offline Path Calculation



- Input to offline path calculation utility:
  - Ingress and egress points
  - Physical topology
  - Traffic matrix (statistics about city - router pairs)
- Output:
  - Set of physical paths, each expressed as an explicit route

Lets Review

MPLS Signaling Protocols

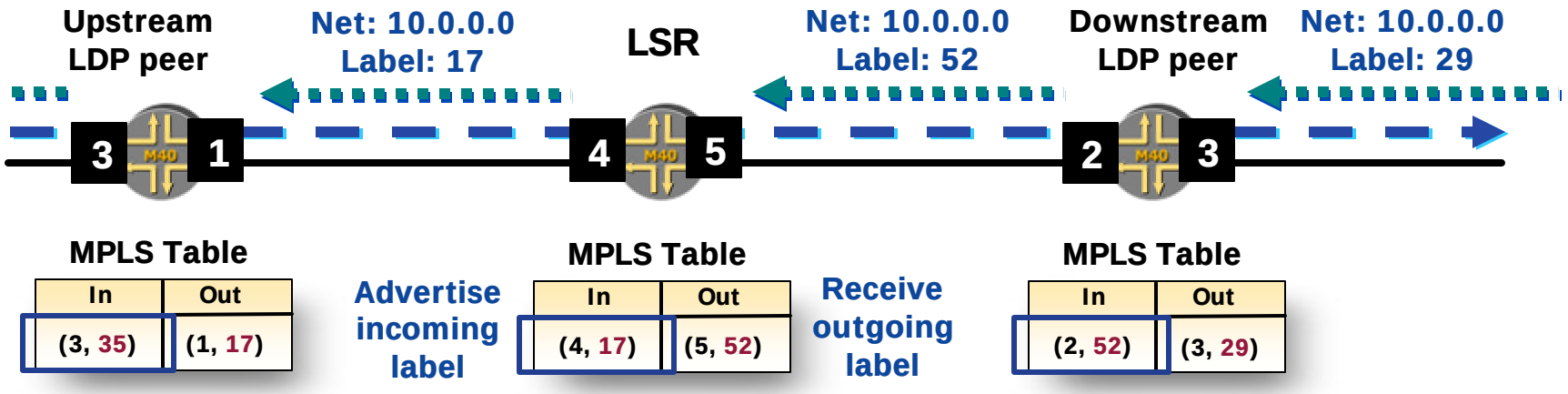
How Is an LSP Established?

- Requires a signaling protocol to:
  - Coordinate label distribution
  - Explicitly route the LSP
  - Bandwidth reservation (optional)
  - Class of Service (DiffServ style)
  - Resource re-assignment
  - Pre-emption of existing LSPs
  - Loop prevention
- MPLS signaling protocols
  - Label Distribution Protocol (LDP)
  - Resource Reservation Protocol (RSVP)
  - Constrained Routing with LDP (CR-LDP)

MPLS Signaling Protocols

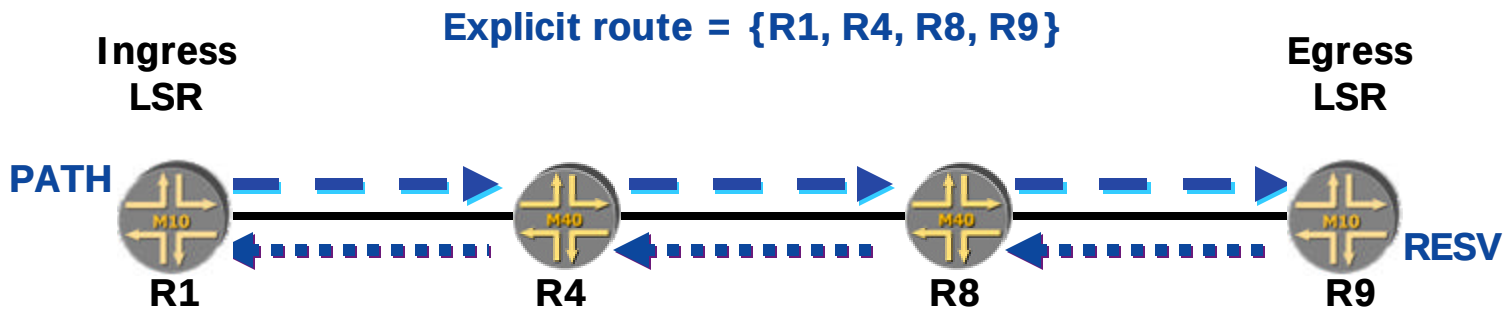
- The IETF MPLS architecture does not assume a single label distribution protocol
- LDP
  - Executes hop-by-hop
  - Selects same physical path as IGP
  - Does not support traffic engineering
- RSVP
  - Easily extensible for explicit routes and label distribution
  - Deployed by providers in production networks
- CR-LDP
  - Extends LDP to support explicit routes
  - Functionally identical to RSVP
  - Not deployed

Label Distribution Protocol (LDP)



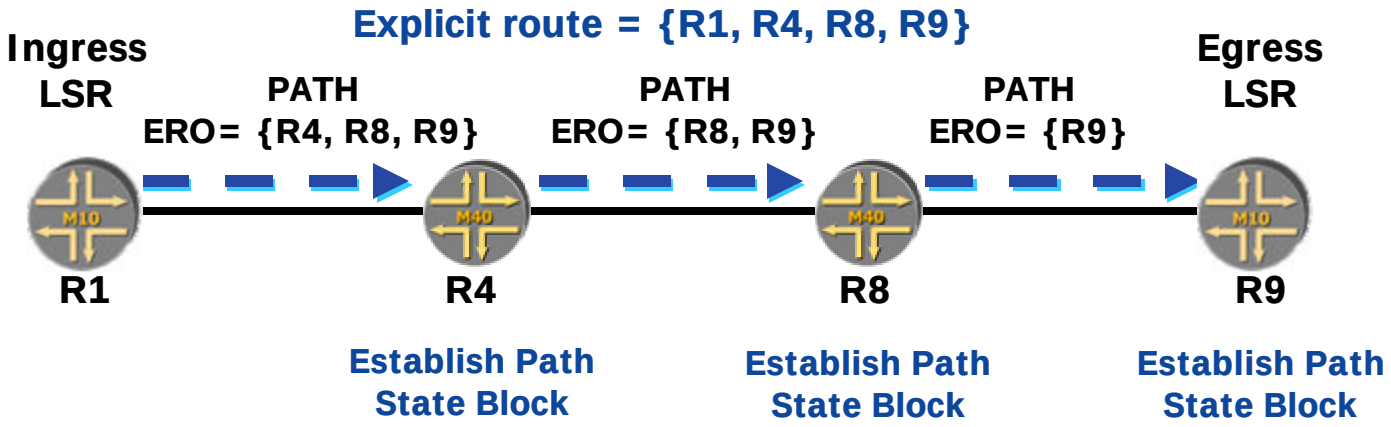
- Labels assigned by downstream peer
- Benefits
  - Labels are not piggybacked on routing protocols
- Limitations
  - LSPs follow the conventional IGP path
  - Does not support explicit routing

Resource Reservation Protocol



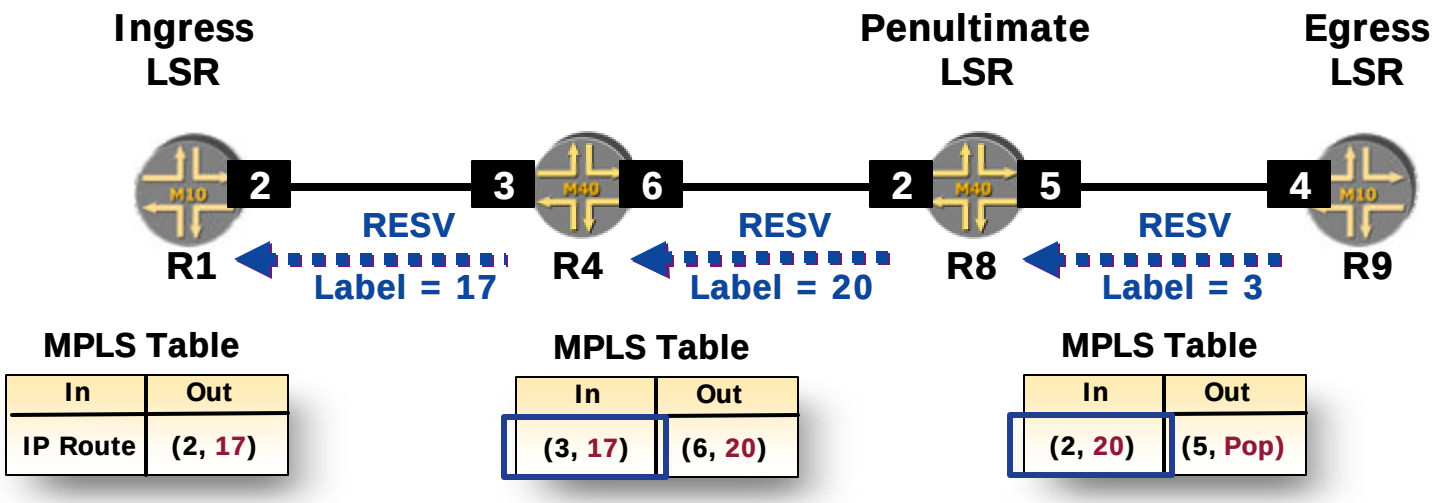
- Internet standard for reserving resources
- RSVP extensions for LSP tunnels
  - Explicit Route Object (ERO)
  - Label Request Object
  - Label Object
  - Session Object
  - Session Attribute Object
  - Record Route Object (RRO)
- RSVP message types
  - PATH: Establish state and request label assignment
  - RESV: Distribute labels & reserve resources
- Runs ingress-to-egress, not end-to-end

Extended RSVP – PATH Message



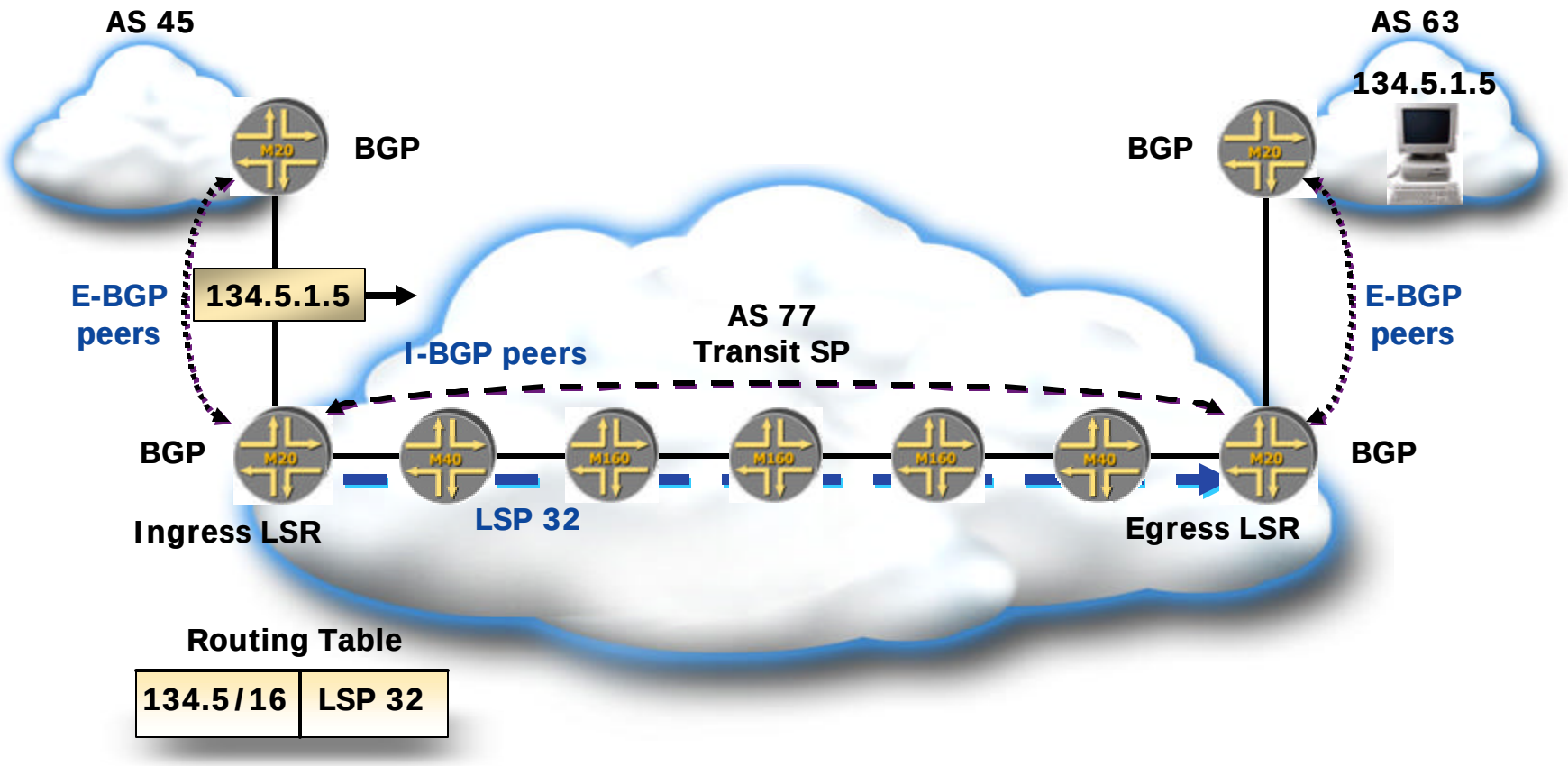
- Explicit route is passed to R1
- R1 transmits a PATH message addressed to R9
  - Label Request Object
  - ERO = {strict R4, strict R8, strict R9}
  - Session object identifies LSP name
  - Session Attributes: Priority, preemption, and fast reroute
  - Sender T_Spec: Request bandwidth reservation

Extended RSVP – RESV Message



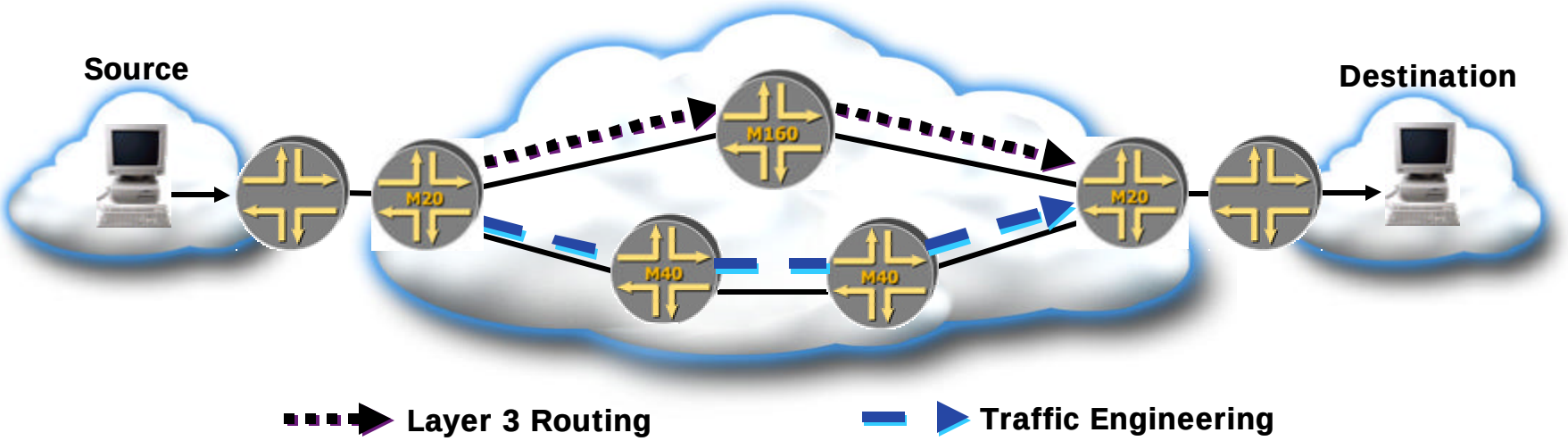
- R9 transmits a RESV message to R8
  - Label = 3 (indicates that penultimate LSR should Pop header)
  - Session object to uniquely identify the LSP
- R8 and R4
  - Stores “outbound” label, allocate an “inbound” label
  - Transmits RESV with inbound label to upstream LSR
- R1 binds label to FEC

How Is Traffic Mapped to an LSP?



- Map LSP to the BGP next hop
- FEC = {all BGP destinations reachable via egress LSR}

What Is Traffic Engineering?

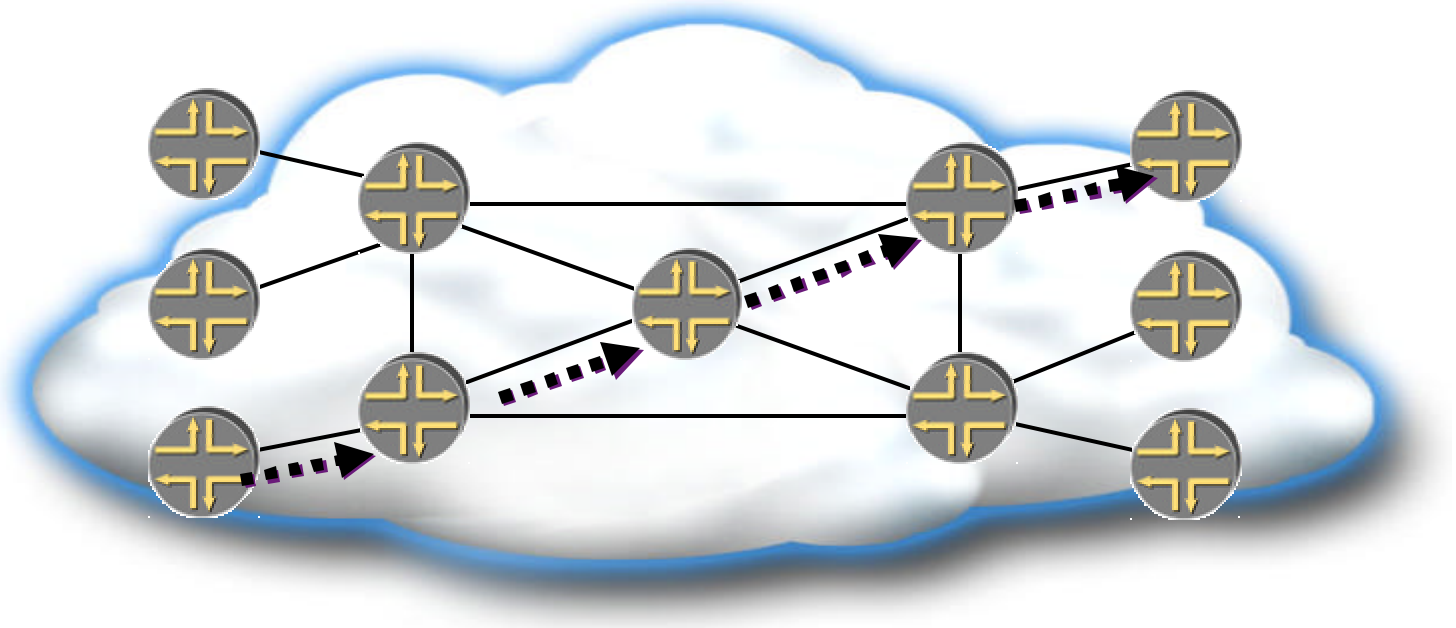


- Ability to control traffic flows in the network
  - Optimize available resources
  - Move traffic from IGP path to less congested path

Lets Review

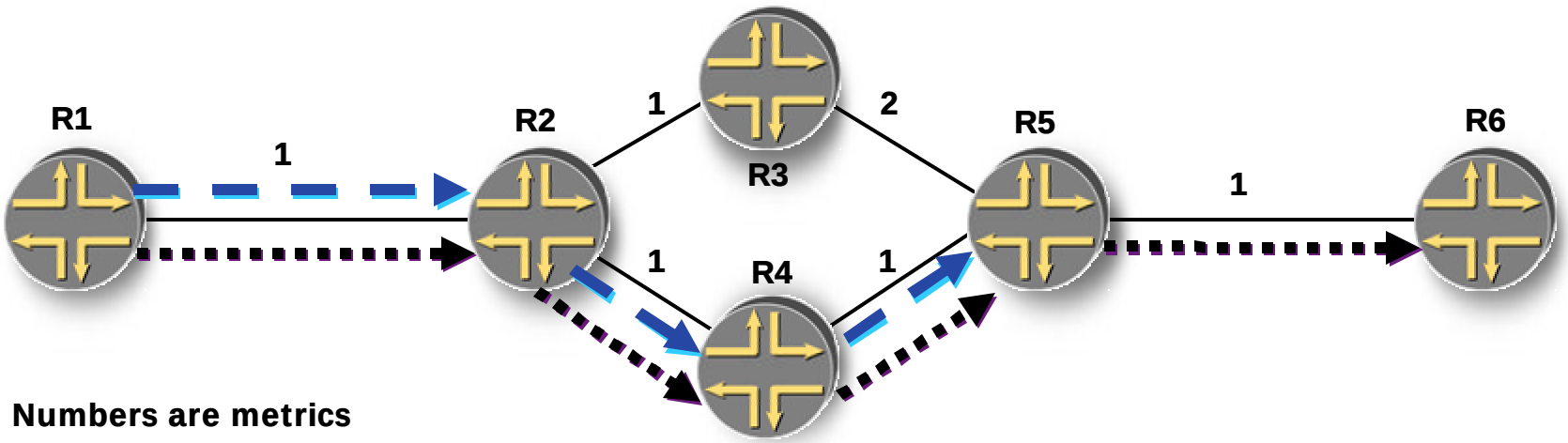
Traffic Engineering

Traffic Engineering Mid 1990s



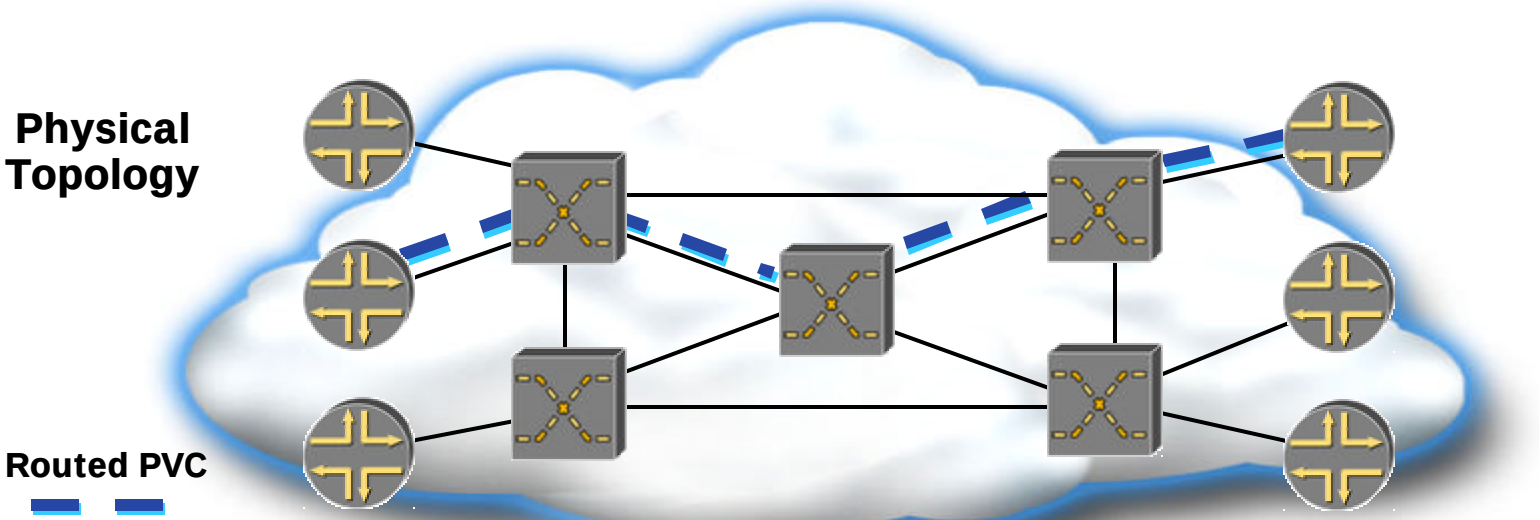
- Infrastructure
  - Routed core
    - Independent L3 decision at each hop
  - DS-1 and DS-3 trunks

Traffic Engineering Mid 1990s



- TE Mechanisms
  - Over provisioning
  - Metric manipulation
- Limitations
  - S/W router became a bottleneck
  - Trial-and-error approach
  - Not scalable

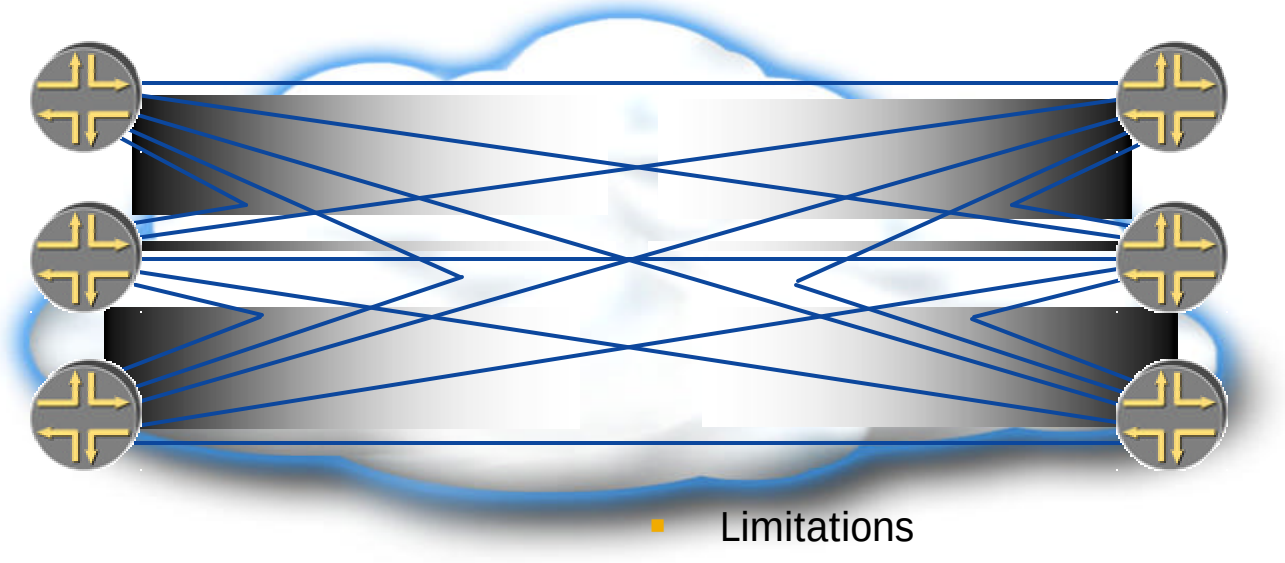
Traffic Engineering Mid to Late 1990s



- Infrastructure
  - Routed edge/ATM core
    - L3 decision at edge router
    - L2 decision at each core switch
  - Dense PVC meshes
  - OC-3, OC-12, and OC-48 trunks

Traffic Engineering Mid to Late 1990s

Logical Topology



- TE Mechanisms
  - PVC routing
  - Overlay network

- Limitations
  - Two networks to manage - IP and ATM
  - Cell tax
  - OC-48+ SAR interfaces
  - “N-squared” PVCs
  - IGP stress

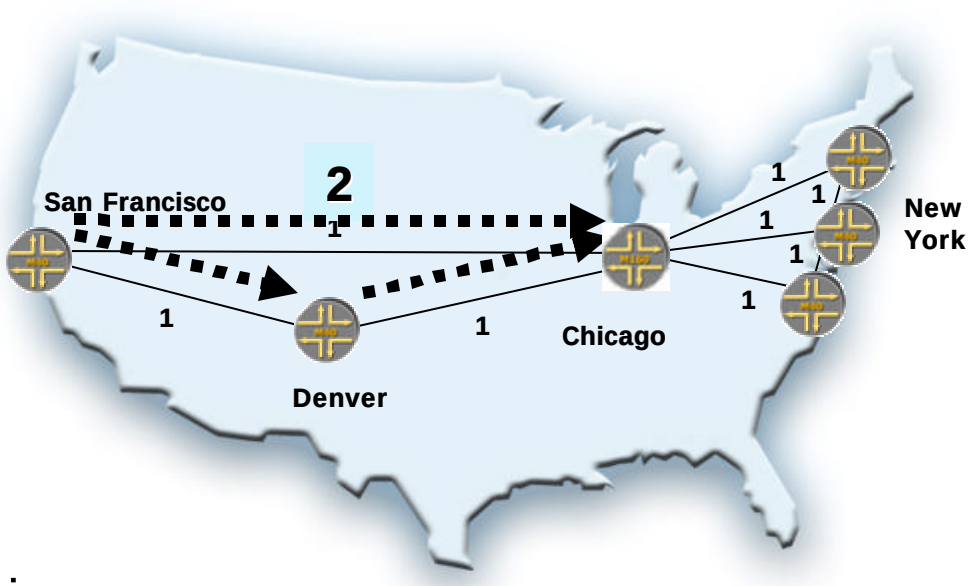
Traffic Engineering in the 21st Century

Question: **Is there a better solution for the 21st century?**

Answer: **Yes … Multiprotocol Label Switching (MPLS)**

- The MPLS Advantage
  - Public and private service integration
  - A fully integrated IP solution
  - Traffic engineering
  - Lower cost
  - A CoS enabler
  - Failover/link protection
  - Multi-service and VPN support

Case Study 1 Deferring a Link Upgrade



- Challenges
  - SF-NY traffic increases
  - Manage expenses by delaying SF-Chicago link upgrade
  - Customer satisfaction

- **IGP metric manipulation**
  - **Manipulation is difficult**
  - **Load balancing is imperfect**
  - **Network destabilization**
  - **Packet misordering**
  - **No fine grained control**

Case Study 1 Deferring a Link Upgrade

SF Routing Table

Destination	Next hop
New York	Blue LSP
Chicago	Chicago
Boston	Chicago
Wash, DC	Chicago

- LSP from SF-to-NY via Denver & Chicago
  - Fine-grained control of SF-NY traffic
  - Network remains stable
  - Packet order maintained

Case Study 2 Utilize Excess Bandwidth

- Challenges
  - Paris to London link is approaching capacity
  - Under-utilized capacity from Frankfurt to London
  - Desire to deliver a "premium" Paris to London service

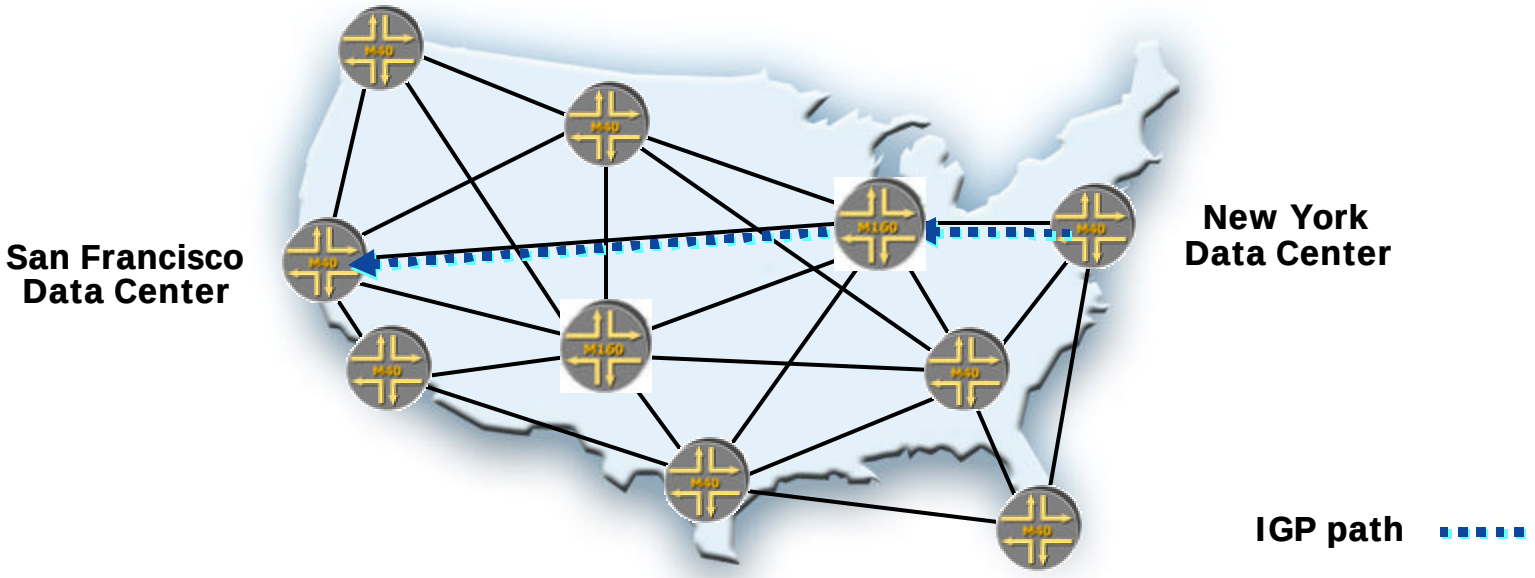
- **Solution**
  - **Premium traffic takes LSP from Paris to London via Frankfurt**

Paris Routing Table

Destination	Next hop
London (premium)	Blue LSP
London (standard)	Direct

Case Study 3

Enhance Service Reliability

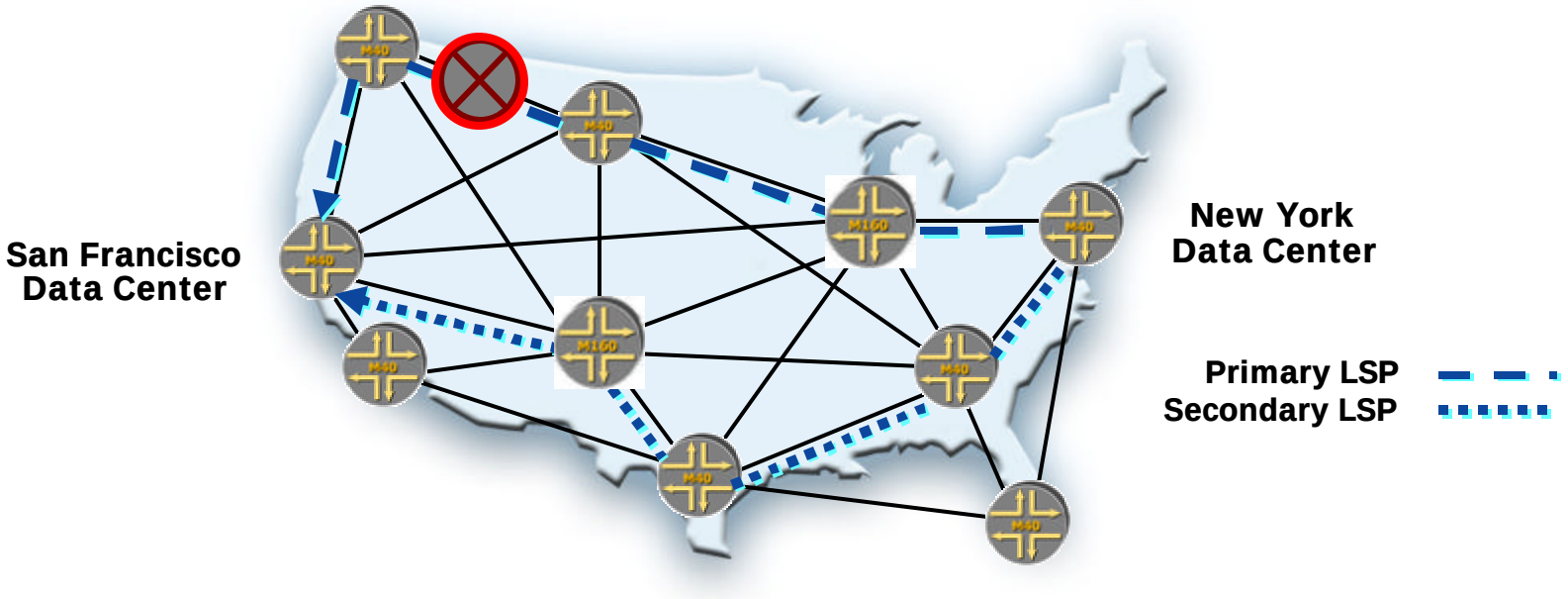


- Challenge
  - Subscriber demands reliable service between SF and NY data centers

- **Motivation**
  - **Avoid the congested IGP path**
  - **Satisfy a highly visible, premium customer**

Case Study 3

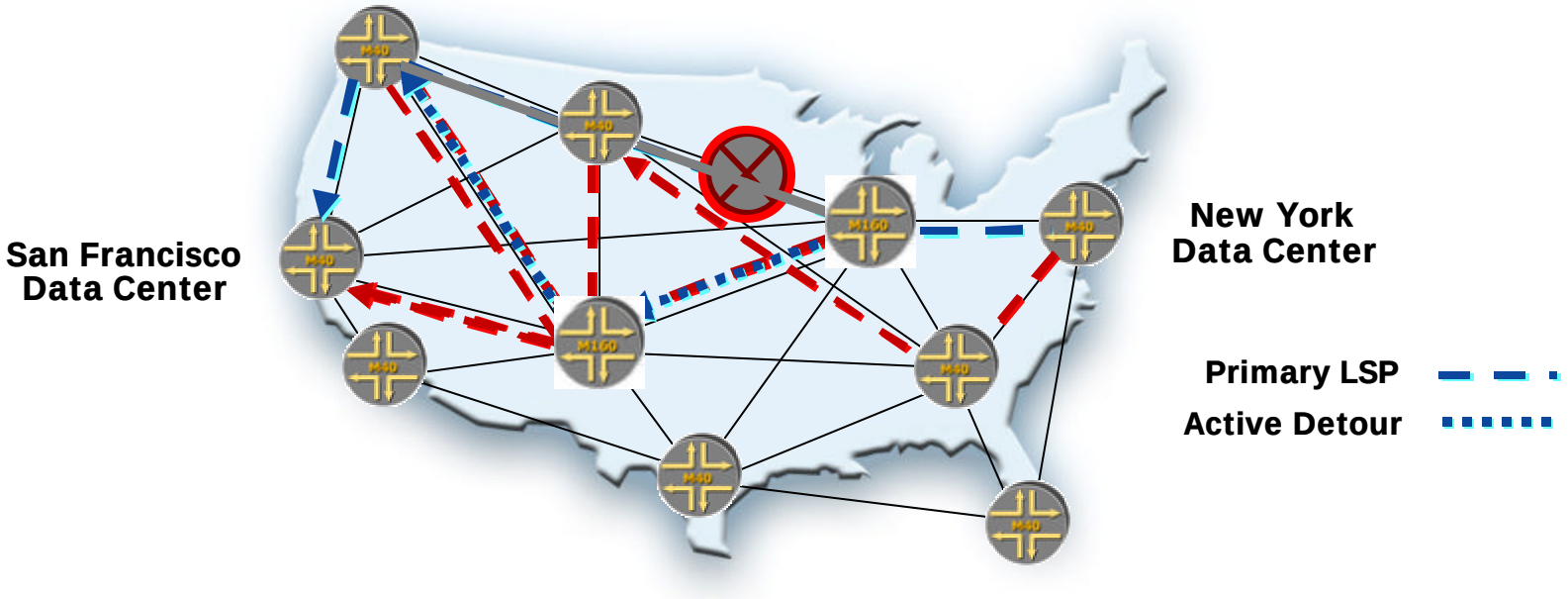
Enhance Reliability - Secondary LSPs



- Standard LSP failover
  - Failure signaled to ingress LSR
  - Calculate & signal new LSP
  - Reroute traffic to new LSP
- Standby Secondary LSP
  - Pre-established LSP
  - Sub-second failover

Case Study 3

Enhance Reliability: Fast Reroute

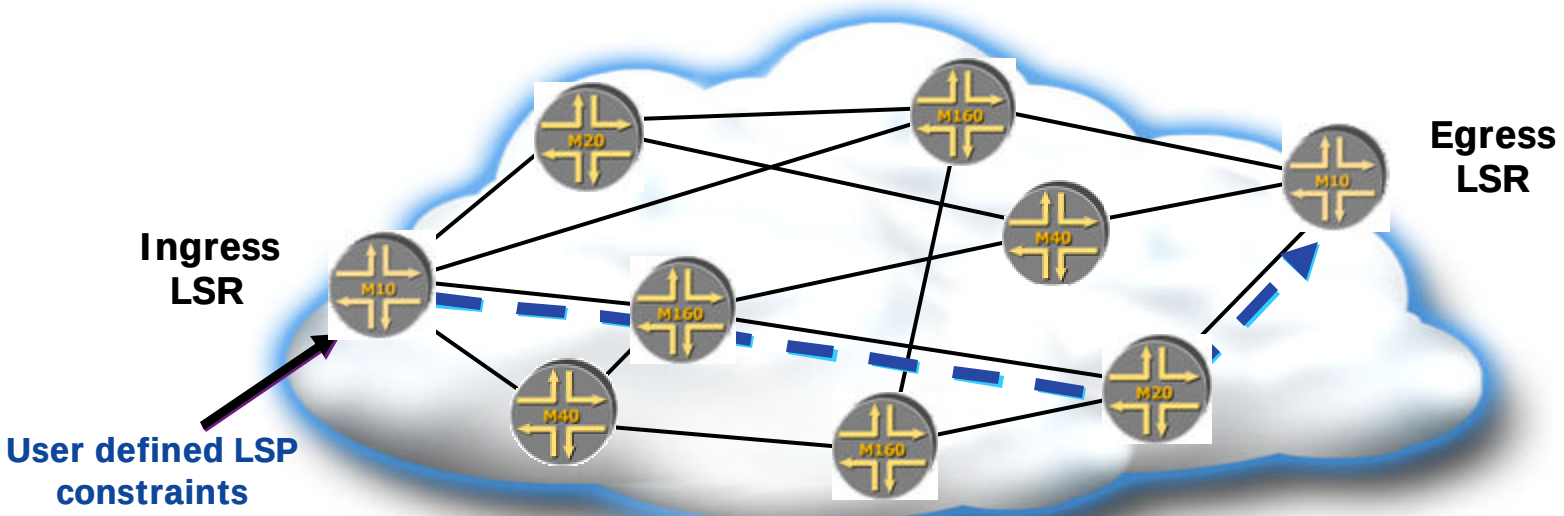


- Ingress signals fast reroute during LSP setup
- Each LSR computes a detour path (with same constraints)
- Supports failover in ~100s of ms

Agenda: Constraint-Based Routing

- Defined
- Operational model
  - Extended IGP
  - Traffic Engineering Database (TED)
  - Operator constraints
  - Constraint Shortest Path First (CSPF) Algorithm
  - RSVP signaling
- Examples
- Online CSPF vs. Offline LSP Calculation

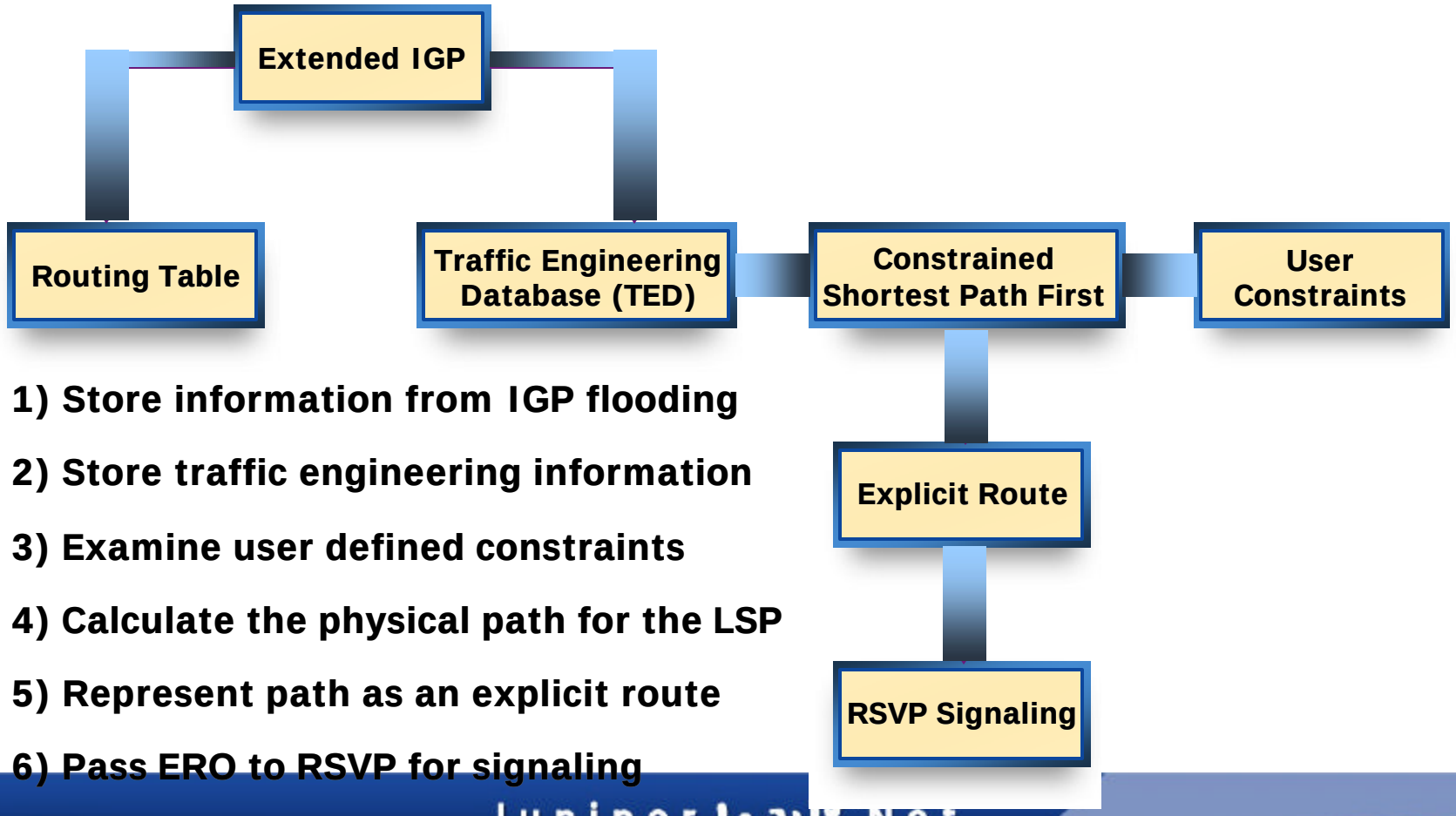
Constraint-Based Routing



- Online LSP path calculation
- Operator configures LSP constraints at ingress LSR
  - Bandwidth reservation
  - Include or exclude a specific link(s)
  - Include specific node traversal(s)
- Network actively participates in selecting an LSP path that meets the constraints

Constraint-Based Routing: Service Model

Operations Performed by the Ingress LSR



1) Store information from IGP flooding

2) Store traffic engineering information

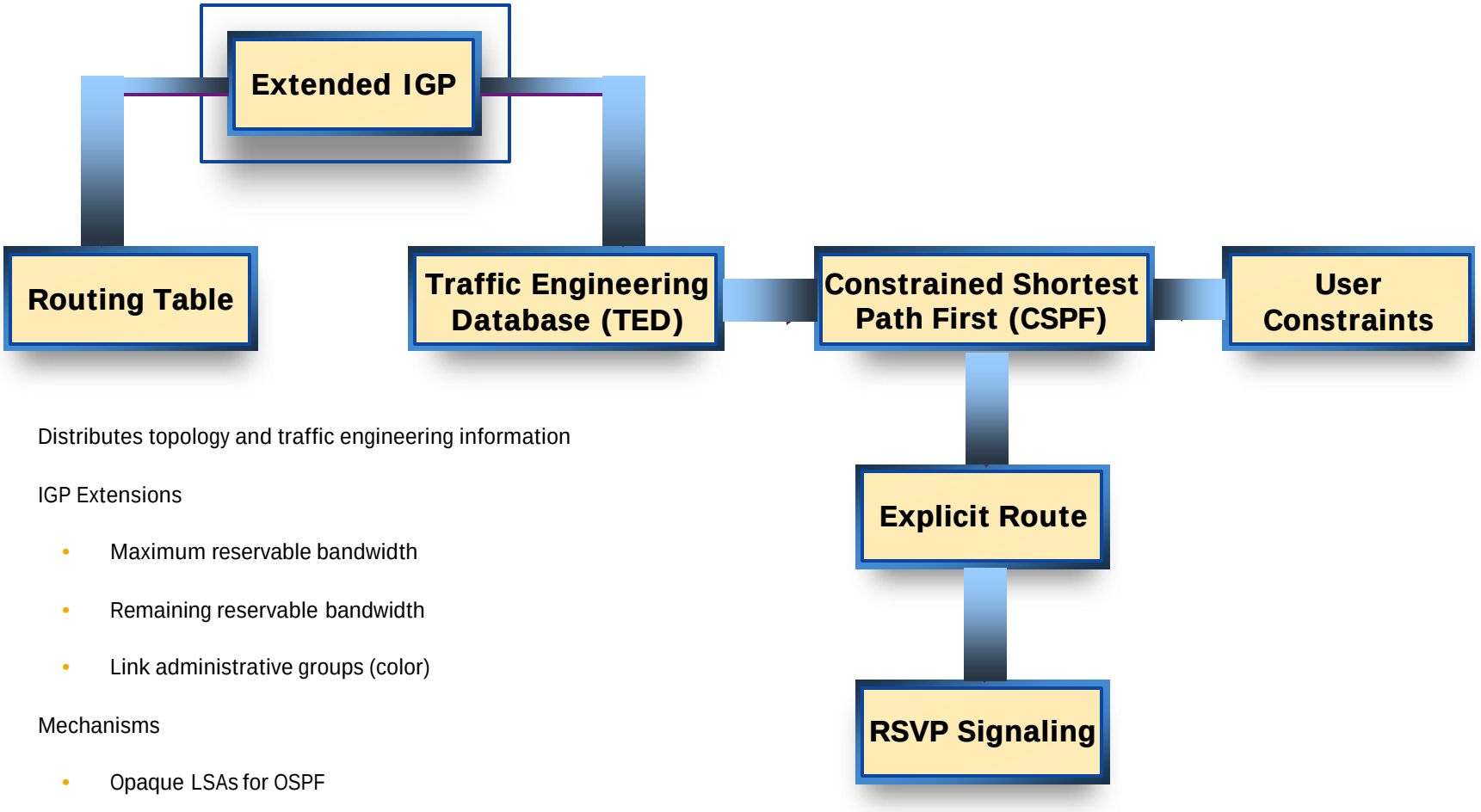
3) Examine user defined constraints

4) Calculate the physical path for the LSP

5) Represent path as an explicit route

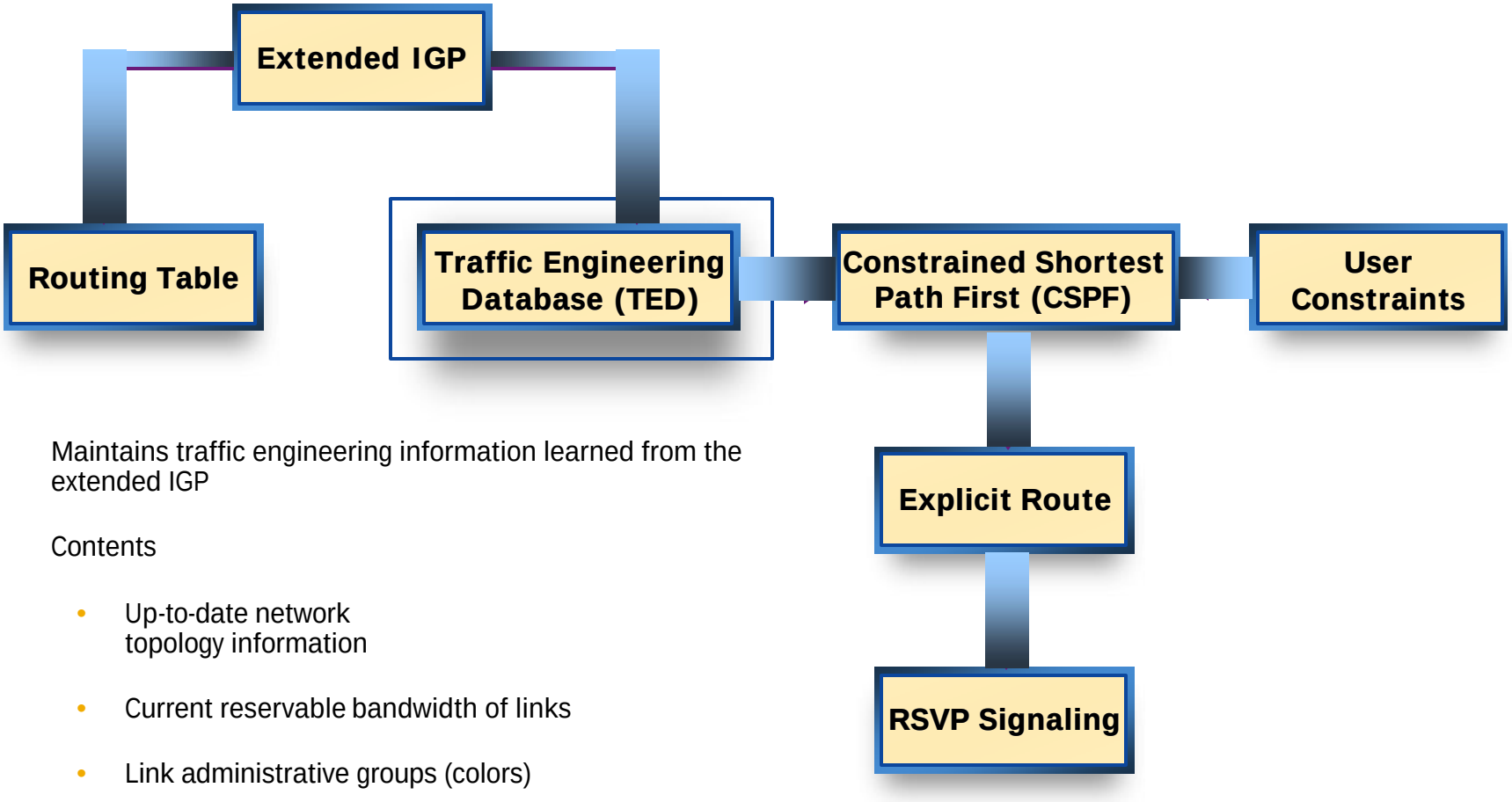
6) Pass ERO to RSVP for signaling

Constraint-Based Routing: Extended IGP



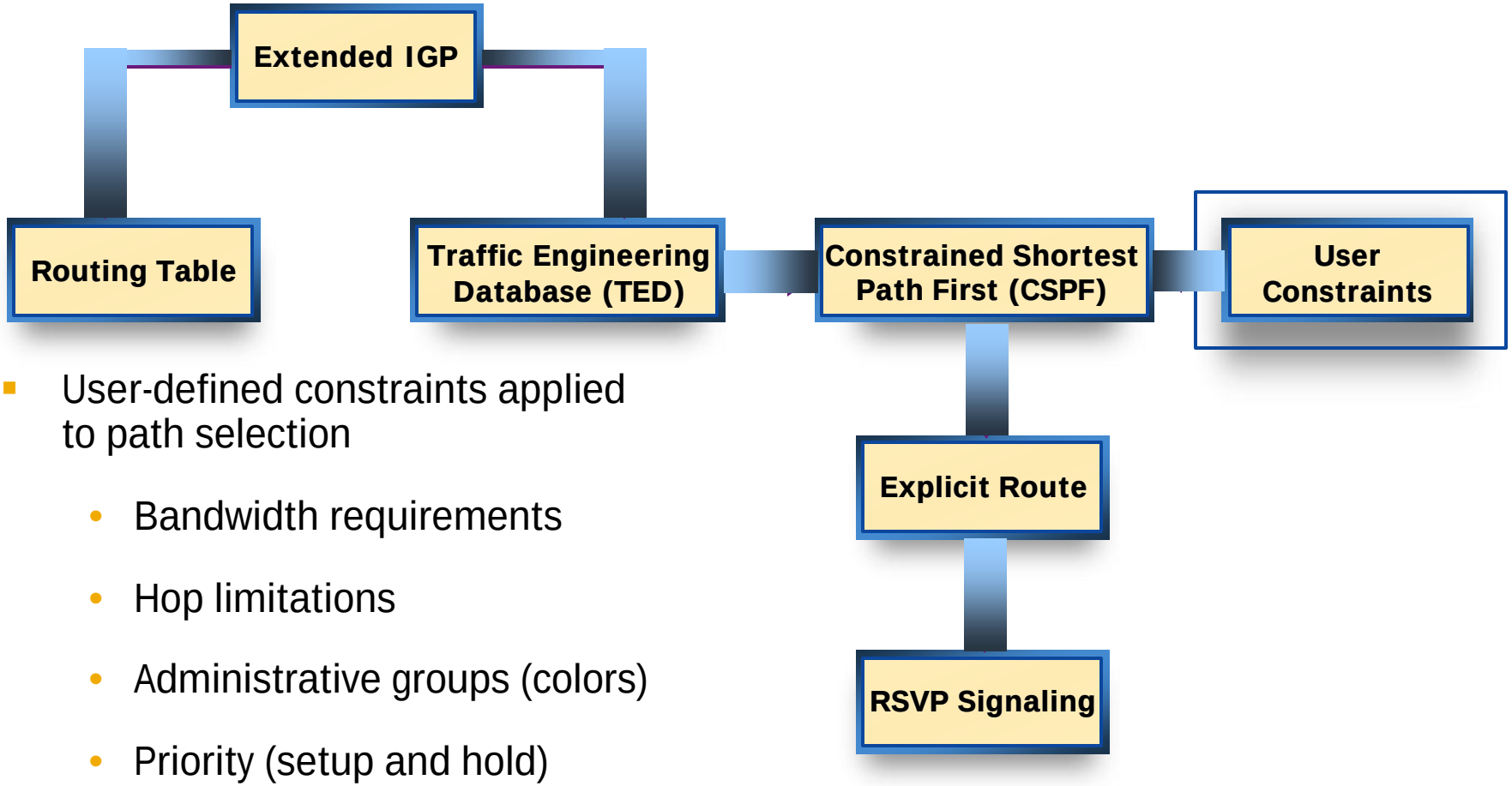
- Distributes topology and traffic engineering information
- IGP Extensions
  - Maximum reservable bandwidth
  - Remaining reservable bandwidth
  - Link administrative groups (color)
- Mechanisms
  - Opaque LSAs for OSPF
  - New TLVs for IS-IS

Constraint-Based Routing: TED



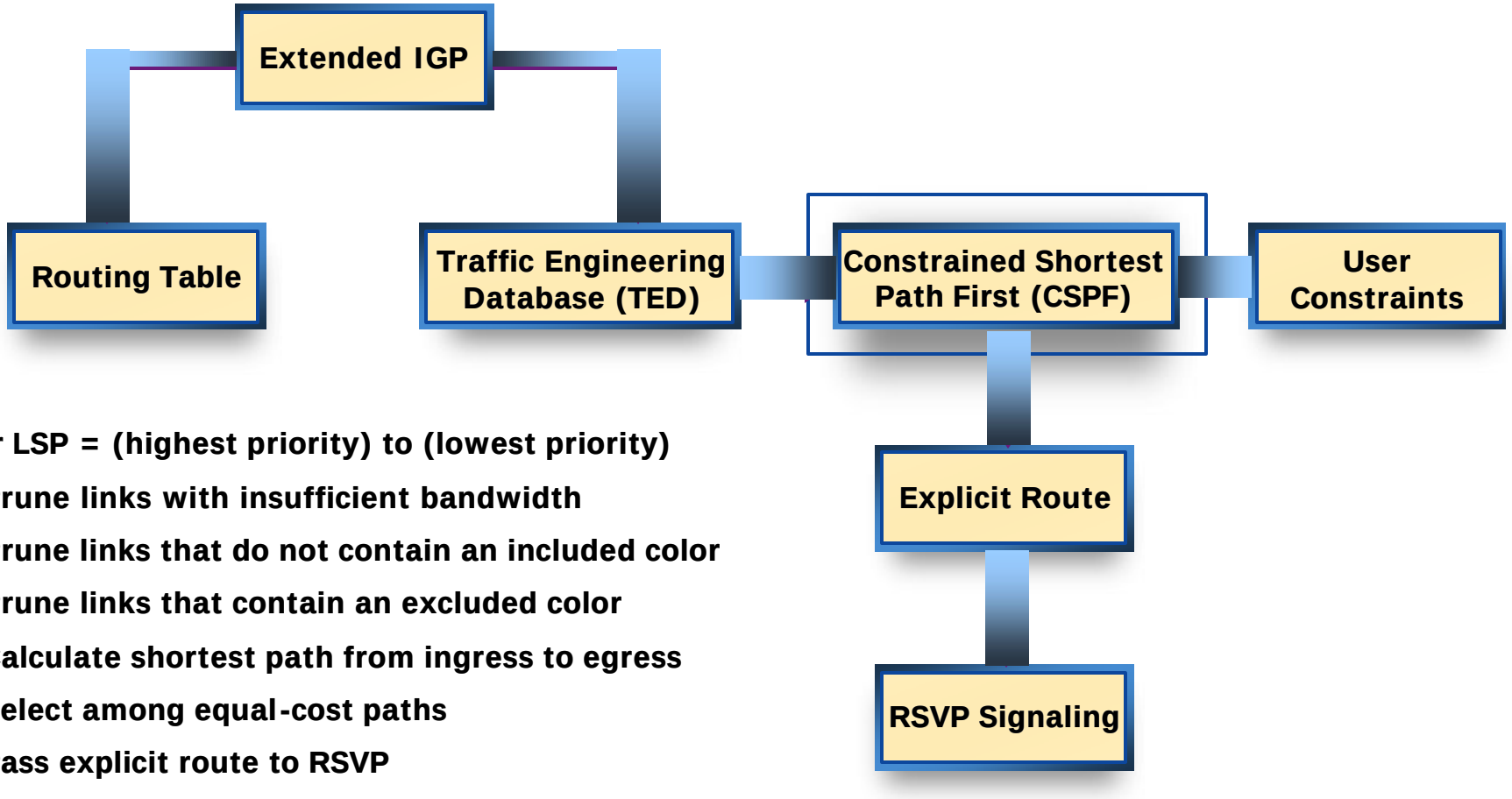
- Maintains traffic engineering information learned from the extended IGP
- Contents
  - Up-to-date network topology information
  - Current reservable bandwidth of links
  - Link administrative groups (colors)

Constraint-Based Routing: User Constraints



- User-defined constraints applied to path selection
  - Bandwidth requirements
  - Hop limitations
  - Administrative groups (colors)
  - Priority (setup and hold)
  - Explicit route (strict or loose)

Constraint-Based Routing: CSPF Algorithm

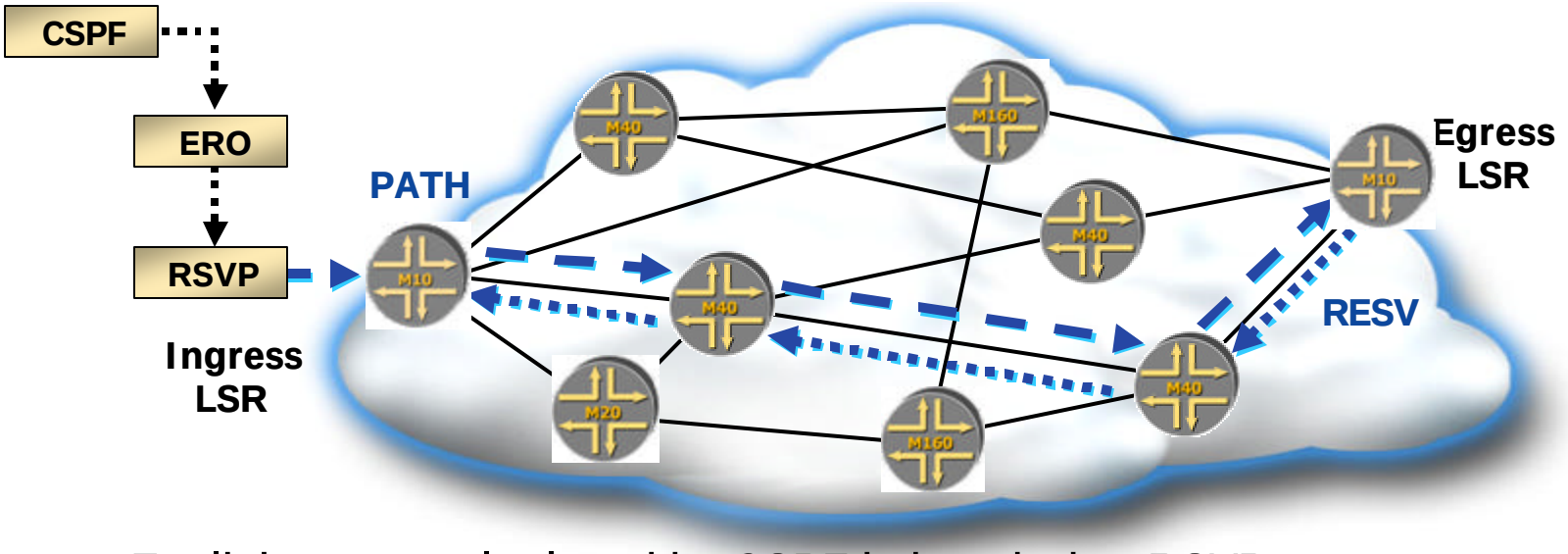


For LSP = (highest priority) to (lowest priority)

- **Prune links with insufficient bandwidth**
- **Prune links that do not contain an included color**
- **Prune links that contain an excluded color**
- **Calculate shortest path from ingress to egress**
- **Select among equal-cost paths**
- **Pass explicit route to RSVP**

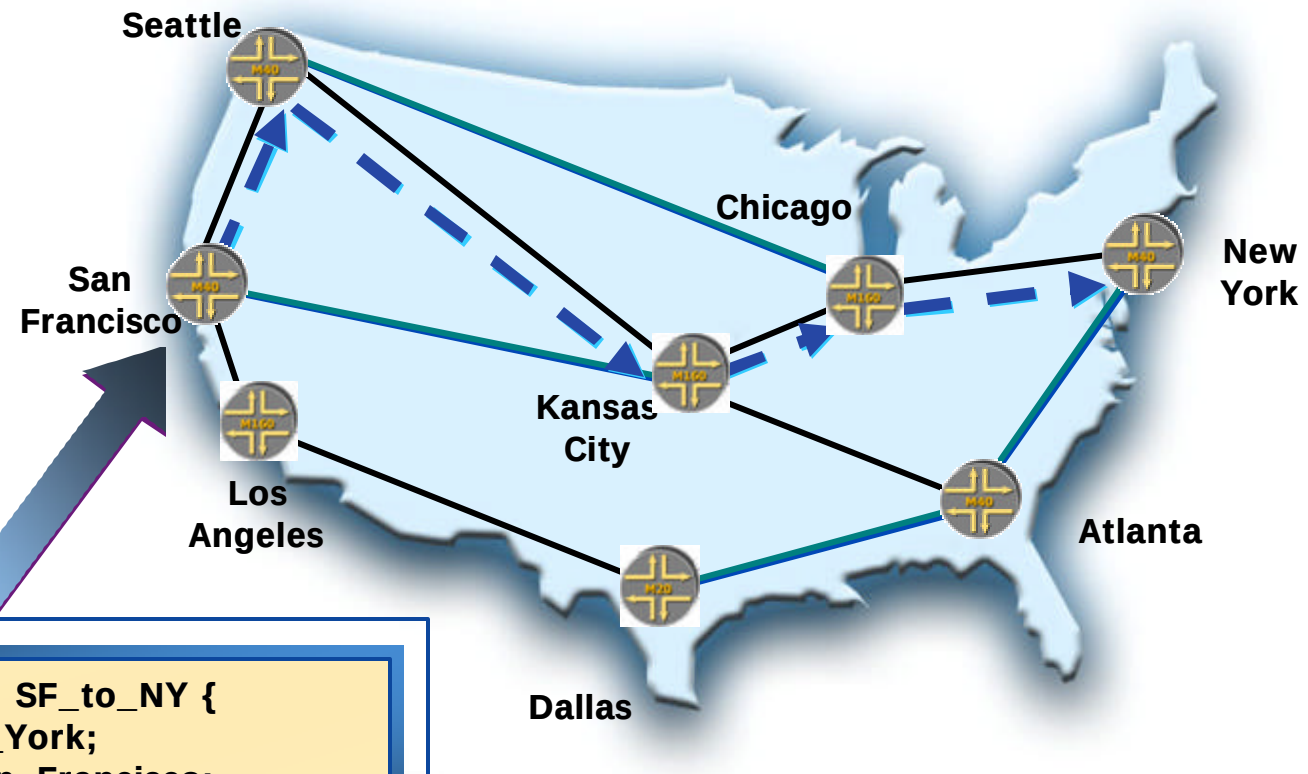
END FOR

Constraint-Based Routing: RSVP Signaling



- Explicit route calculated by CSPF is handed to RSVP
  - RSVP is unaware of how the ERO was calculated
- RSVP establishes LSP
  - PATH: Establish state and request label assignment
  - RESV: Distribute labels & reserve resources

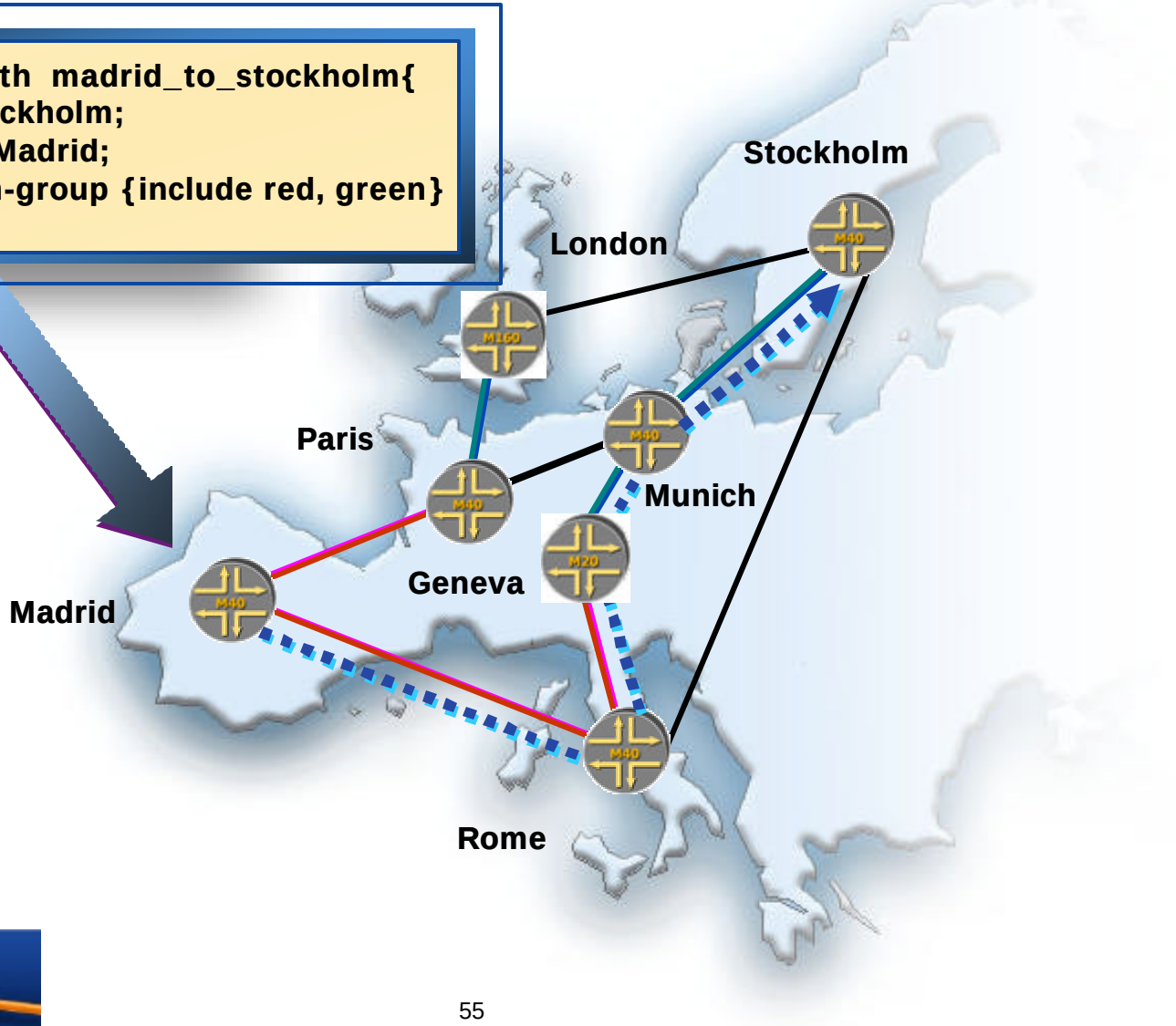
Constraint-Based Routing: Example 1




```
label-switched-path  SF_to_NY {
                to New_York;
                from San_Francisco;
                admin-group {exclude green}
                cspf}
```

Constraint-Based Routing: Example 2

```
label-switched-path  madrid_to_stockholm{
                to Stockholm;
                from Madrid;
                admin-group {include red, green}
                cspf}
```



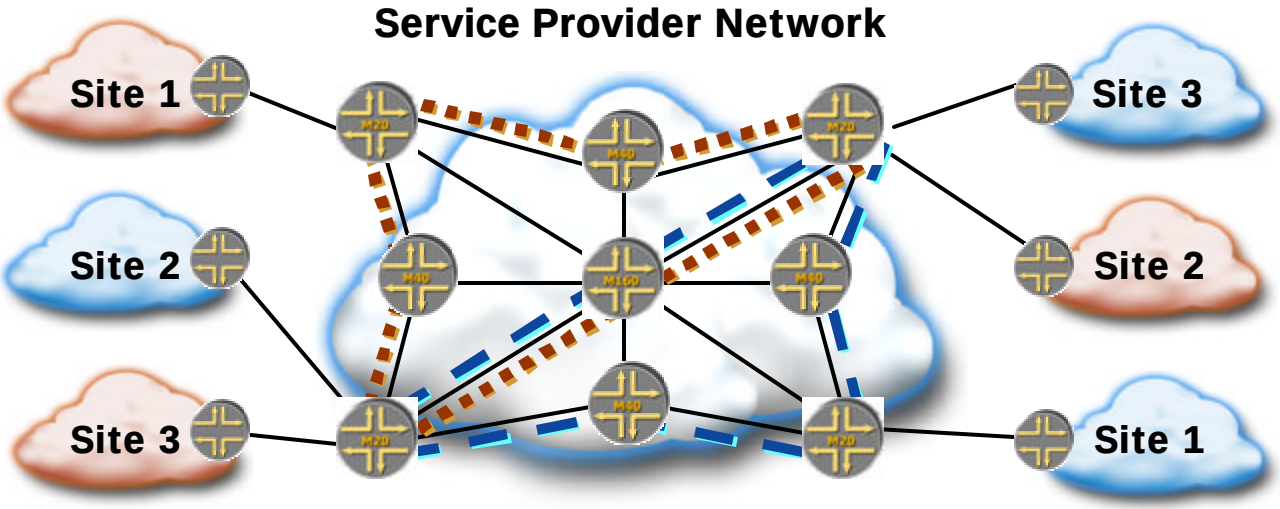
Summary

- MPLS
  - Label Switching
  - Alternate to IP Routing
  - Traffic Engineering – Optional
- Signalling Protocols
  - RSVP
  - LDP

Lets Review

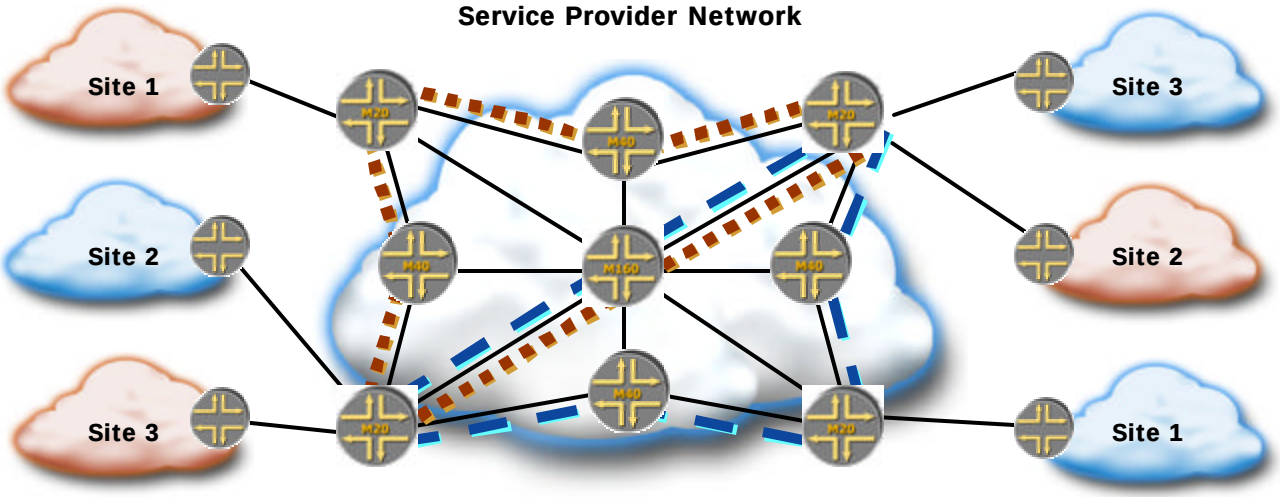
MPLS VPNs

MPLS: A VPN Enabling Technology



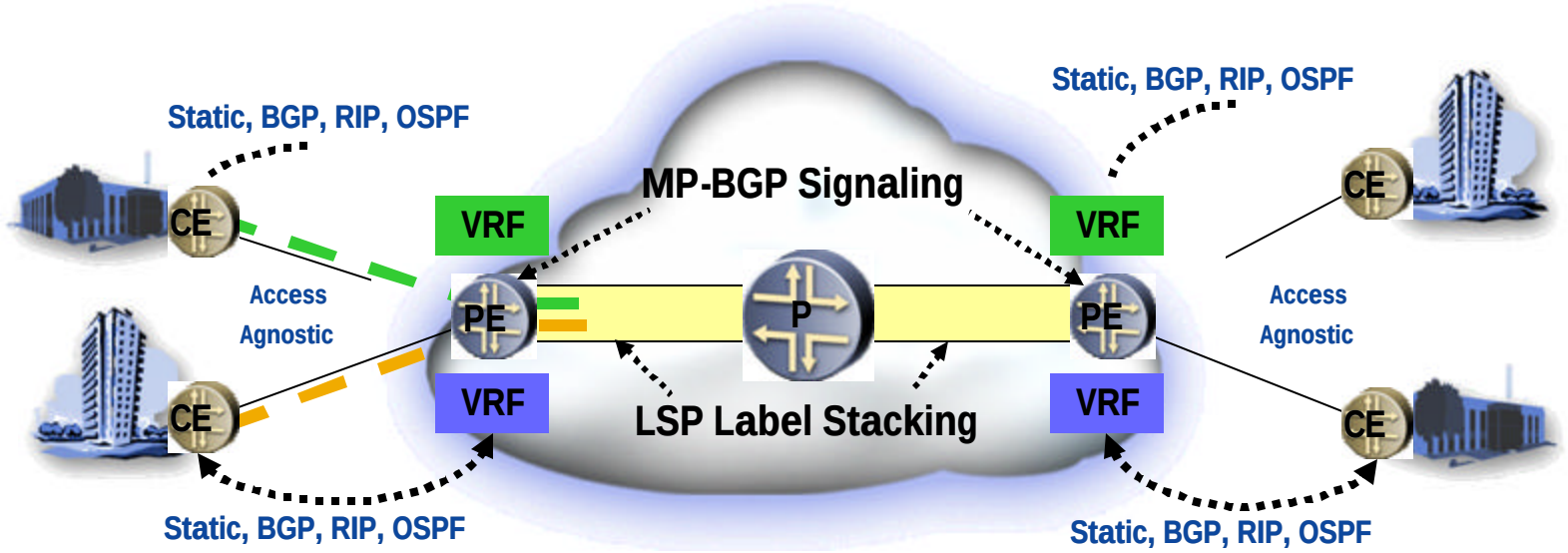
- For subscribers
  - Seamlessly integrates public and private networking
  - Permits a single connection to the service provider
  - Supports rapid delivery of new services
  - Minimizes operational expenses
  - Provides higher network reliability and availability (SLAs)

MPLS: A VPN Enabling Technology



- For service providers
  - Standards based, IP-centric solution
  - Traffic engineering
  - Overcomes limitations of overlay models
  - Supports multiple service-delivery models
  - Delivers core flexibility to support multiple services
  - By combining IP and layer 2 in a convenient way, it is the natural choice for exploring richer VPN models

Layer 3 VPNs - RFC 2547bis



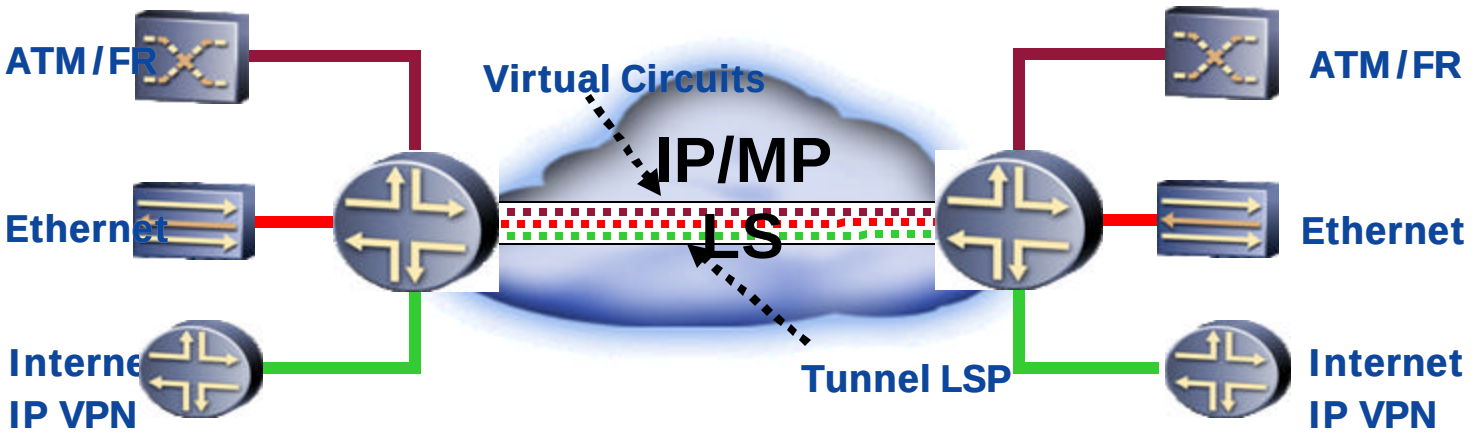
How it works?

- MPLS label stacking optimizes LSPs in the core
- Each PE router has a routing instance per VPN - VRF
- Learns/distributes routes via either BGP, OSPF, RIP or static routes from/to CE
- Routing & VPN membership information distributed automatically via MP-BGP
- Can substitute IPSec & GRE tunnels for LSPs

Benefits:

- Standards based/interoperable
- Ease of provisioning
- Uses scalable BGP/MPLS in the core
- Supports overlapping address space
- Flexible and scalable IP QOS
- Automatic full mesh or hub & spoke
- Supports wide rage of access types

Layer 2 VPNs



- Consolidate multiple service networks onto a single core network
- Focus of two IETF working groups
  - Provider Provisioned VPN (PPVPN)
    - Layer 2 VPNs over tunnels - Draft-kompella-ppvpn-l2vpn
    - Virtual Private LAN service - Draft-kompella-ppvpn-vpls
  - Pseudo Wire Emulation Edge to Edge (PWE3)
    - Various IETF drafts supporting encapsulation and service emulation of pseudo wires.
    - Also known as Draft-Martini

Module Objectives

- After completion of this chapter, you will be able to:
  - Define the roles of P, PE, and CE routers
  - Describe the format of VPN-IPv4 addresses
  - Explain the role of the route distinguisher (RD)
  - Describe the flow of RFC 2547bis control information
  - Explain the operation of the RFC 2547bis forwarding plane

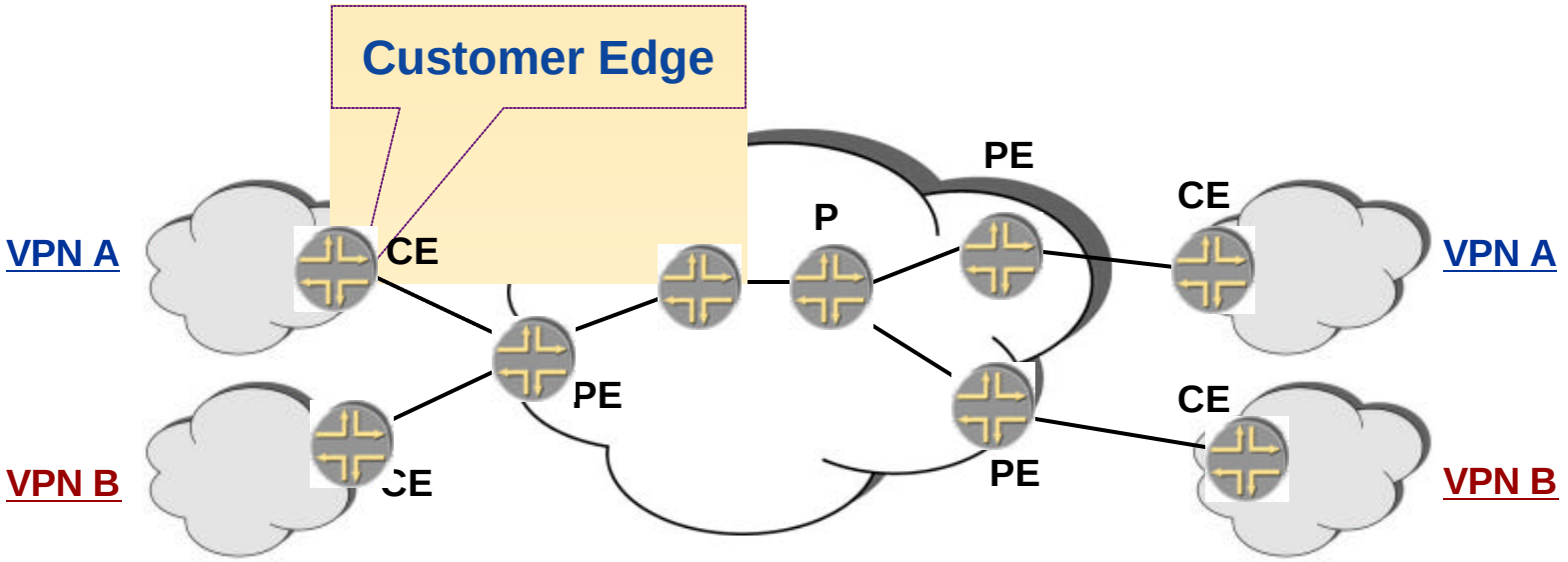
Agenda: Layer 3 MPLS VPNs

- RFC 2547bis terminology
- VPN-IPv4 address structure
- Operational characteristics
  - Policy-based routing information exchange
  - Traffic forwarding

Agenda: Layer 3 MPLS VPNs

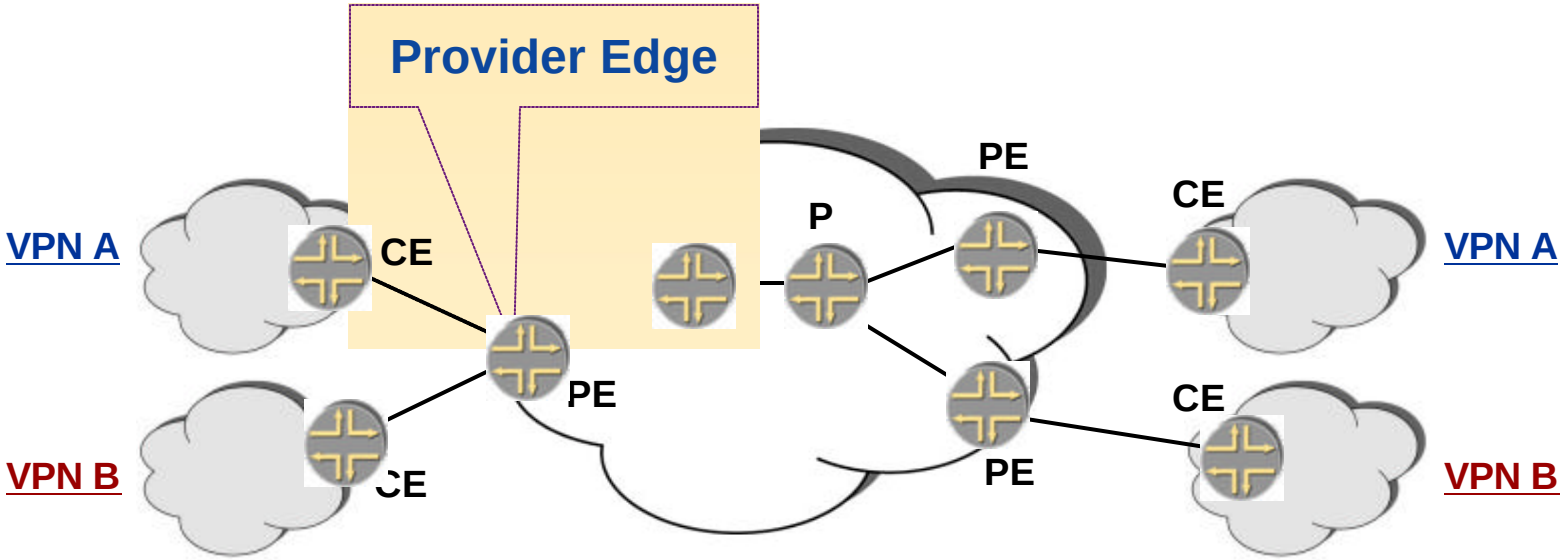
- RFC 2547bis terminology
- VPN-IPv4 address structure
- Operational characteristics
  - Policy-based routing information exchange
  - Traffic forwarding

Customer Edge Routers



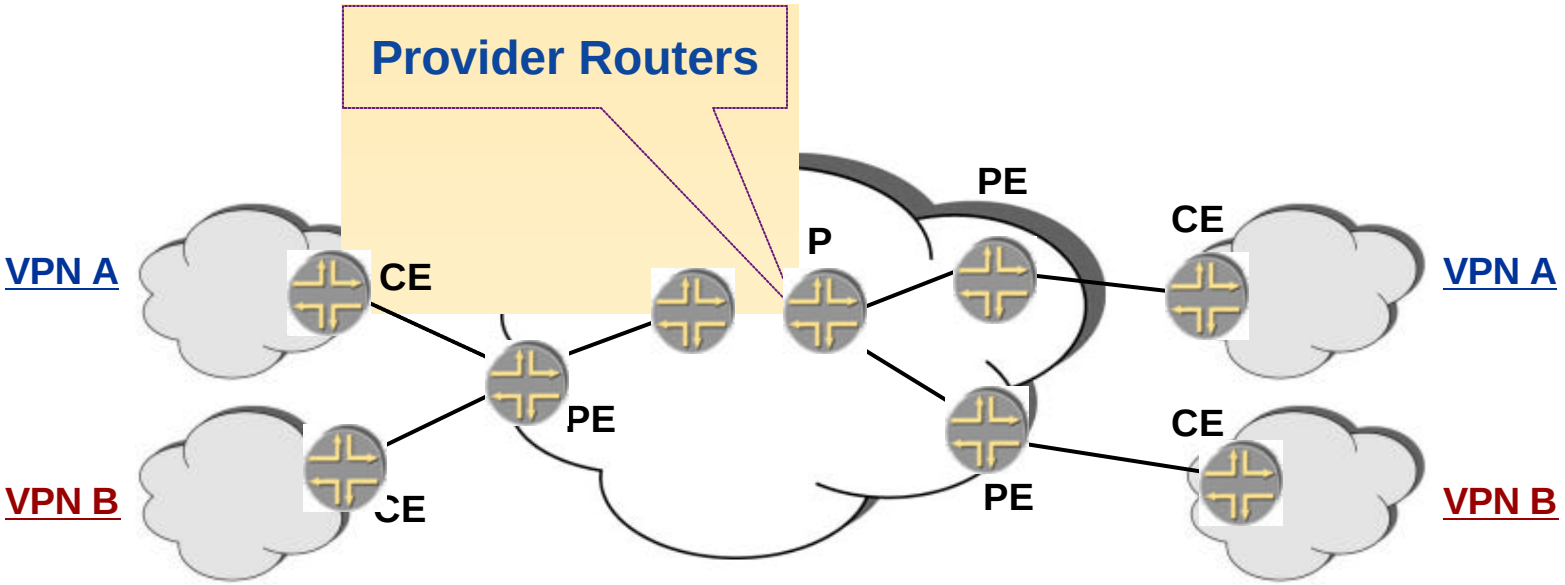
- Customer edge (CE) routers
  - Located at customer premises
  - Provide access to the service provider network
  - Can use any access technology or routing protocol for the CE/PE connection

Provider Edge Routers



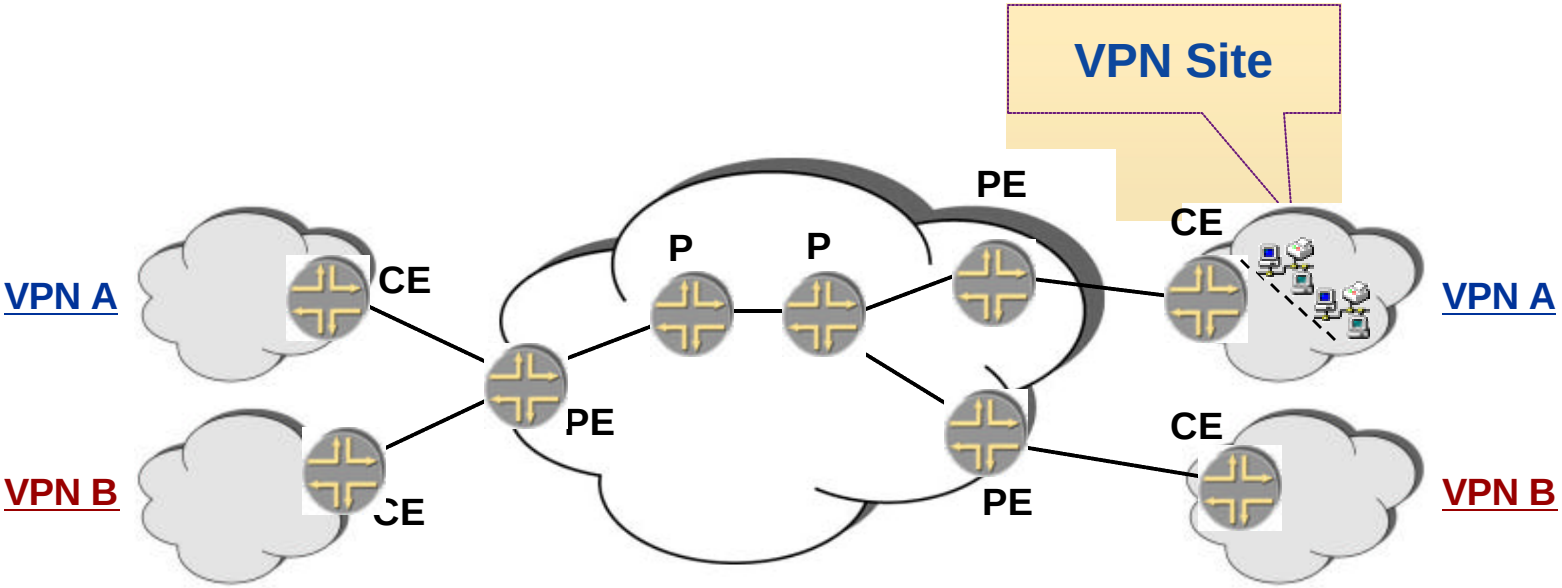
- Provider edge (PE) routers
  - Maintain VPN-specific forwarding tables
  - Exchange VPN routing information with other PE routers using BGP
  - Use MPLS LSPs to forward VPN traffic

Provider Routers



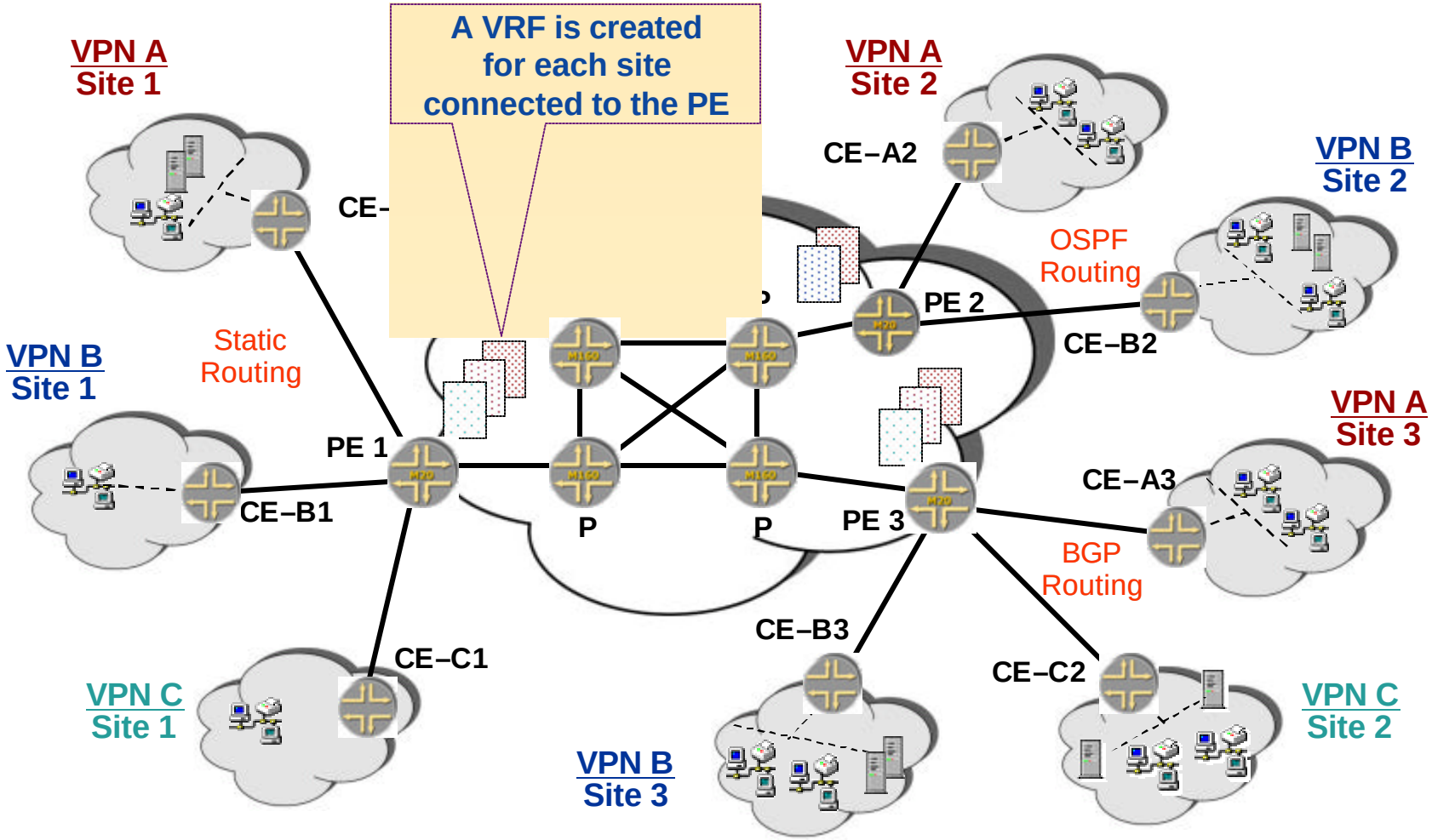
- Provider (P) routers
  - Forward VPN data transparently over established LSPs
  - Do not maintain VPN-specific routing information

VPN Sites



- A site is a collection of machines that can communicate without traversing the SP backbone
- Each VPN site is mapped to a PE router interface
  - Routing information is stored in different tables for each site

VPN Routing and Forwarding Tables (VRFs)



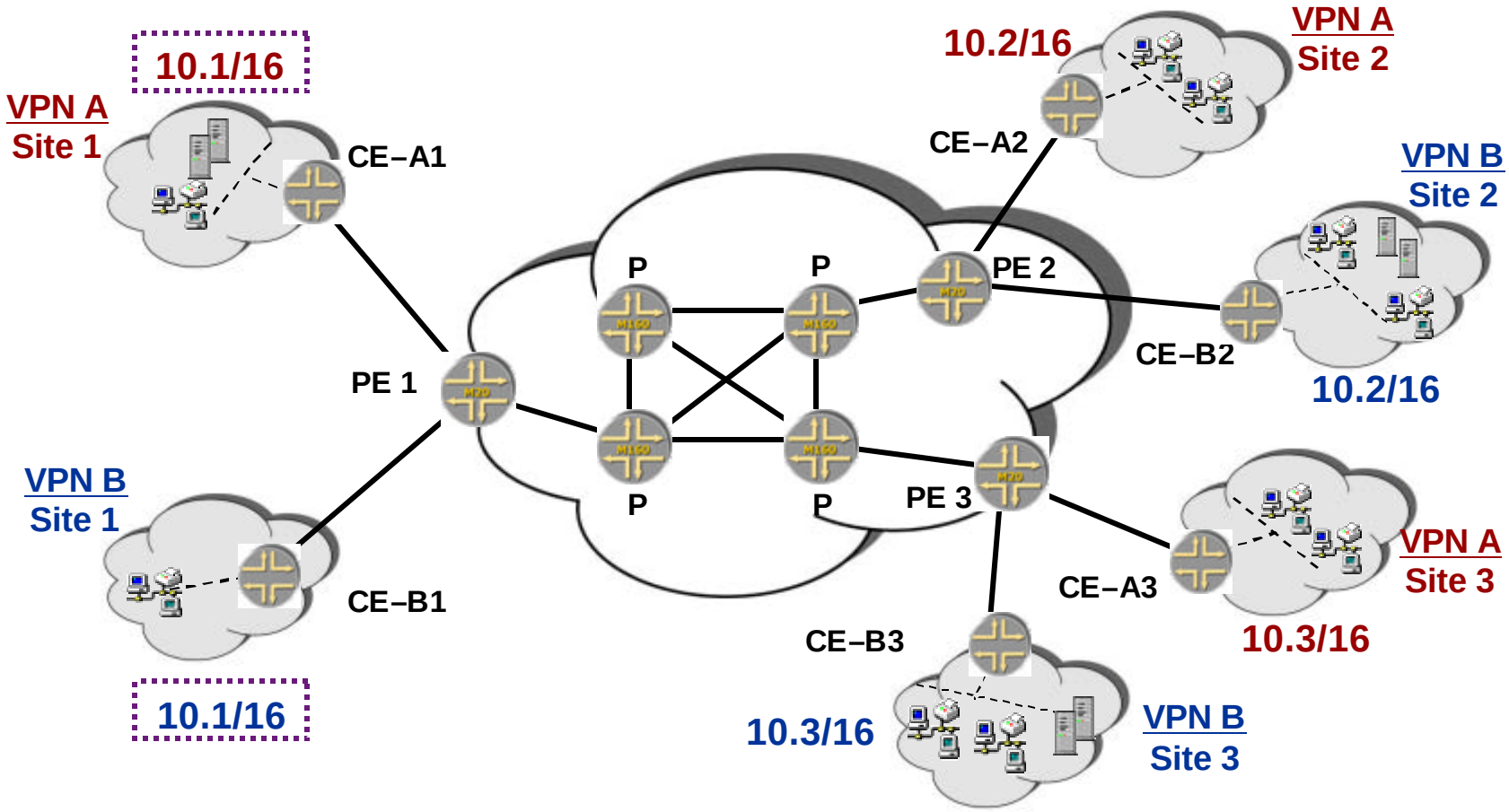
VRFs

- Each VRF is populated with:
  - Routes received from directly connected CE sites associated with the VRF
  - Routes received from other PE routers with acceptable MP-BGP attributes
- Packets from a given site are only matched against the site's corresponding VRF
  - Provides isolation between VPNs

Agenda: Layer 3 MPLS VPNs

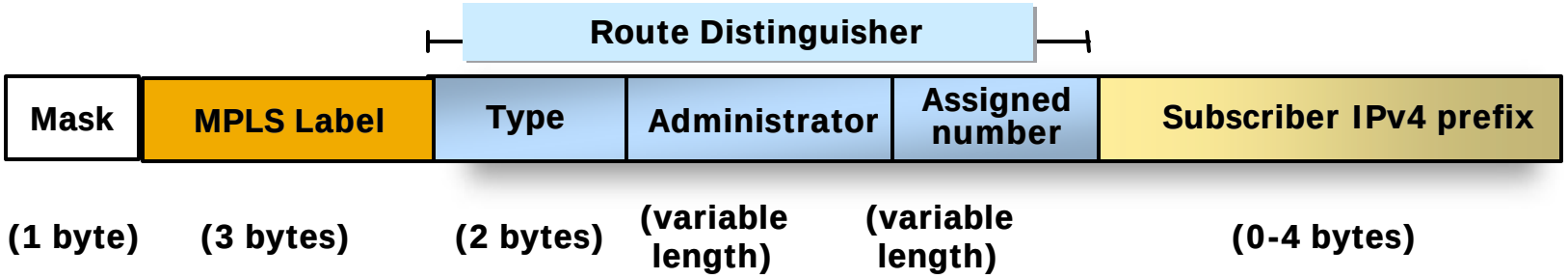
- RFC 2547bis terminology
- ➔ VPN-IPv4 address structure
- Operational characteristics
  - Policy-based routing information exchange
  - Traffic forwarding

Overlapping Address Spaces



VPNs A and B use the same address space

VPN-IPv4 NLRI Format

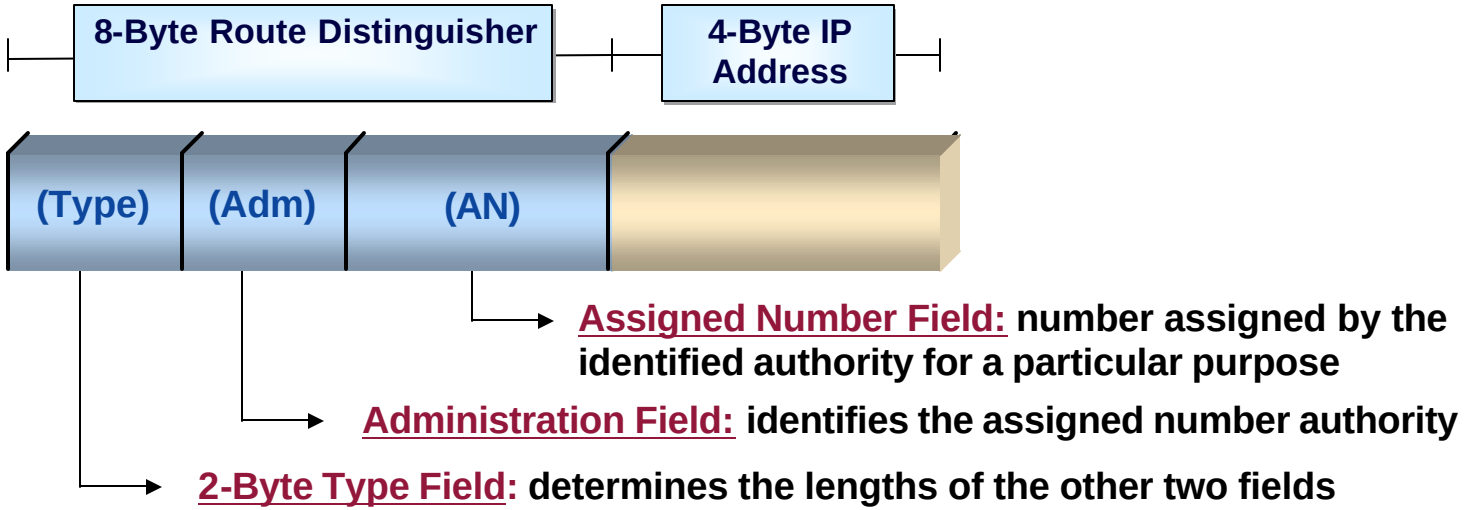


- VPN-IPv4 address family
  - New BGP-4 Sub Address Family Identifier (SAFI 128)
    - Consists of MPLS label + RD + subscriber IPv4 prefix
  - Route distinguisher disambiguates IPv4 addresses
    - allows SP to administer its own “numbering space”
- VPN-IPv4 addresses are distributed by MP-BGP
  - Uses multiprotocol extensions for BGP4 (RFC 2283)
- A /32 IPv4 prefix produces a mask of /120 (15 octets)
  - JUNOS software CLI displays (and the examples in this class) only show IPv4 prefix length (that is, /32)

The VPN-IPv4 Address Family

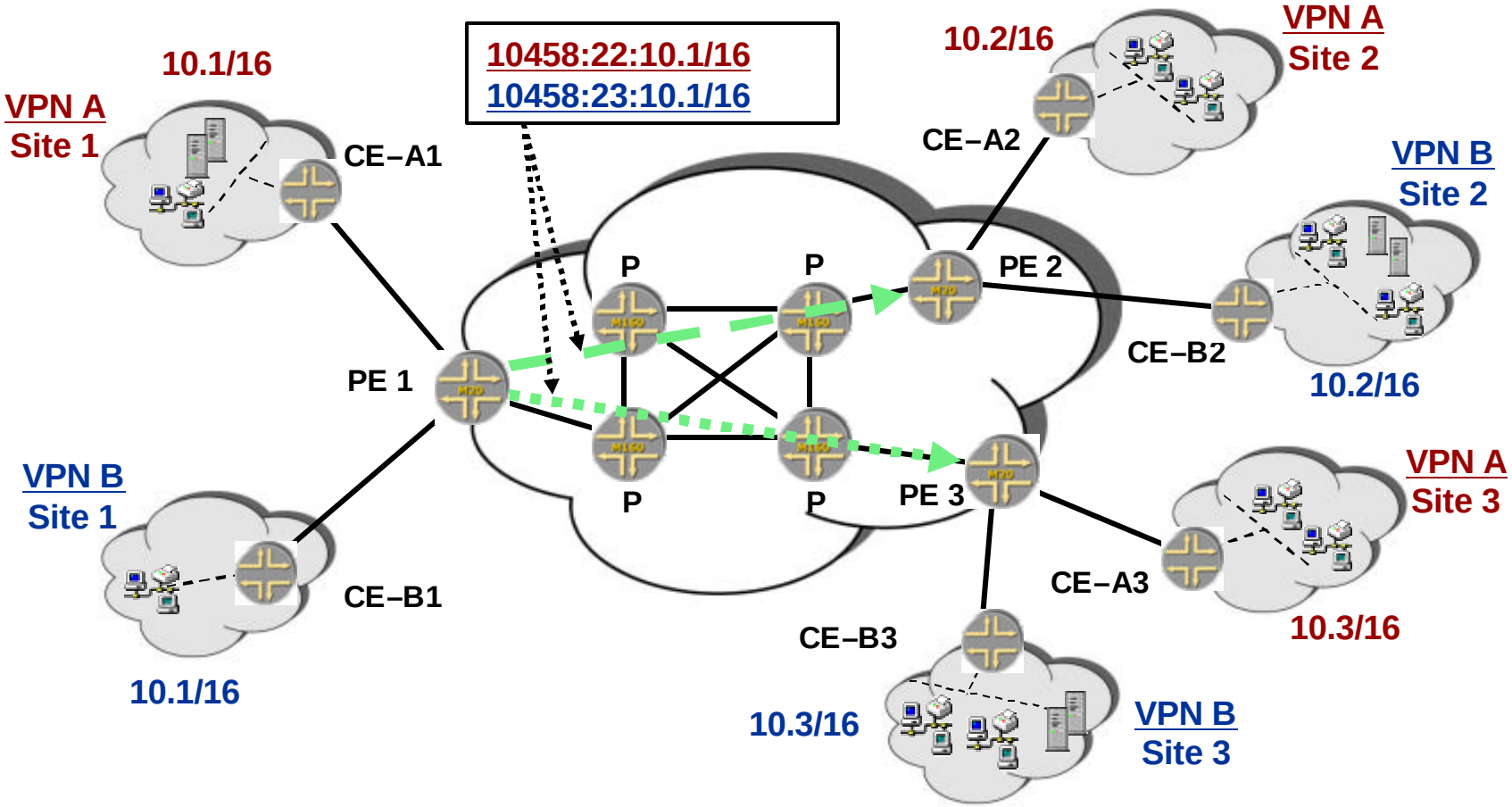
- RD disambiguates IPv4 addresses
- VPN-IPv4 routes
  - Ingress PE router prepends RD to IPv4 prefix of routes received from each CE device
  - VPN-IPv4 routes are exchanged between PEs using MP-BGP
  - Egress PE router converts VPN-IPv4 routes into IPv4 routes before inserting into site's routing table
- VPN-IPv4 is used only in the control plane
  - Data plane uses MPLS encapsulated IPv4 packets

Route Distinguisher Formats



- Two values are defined for Type Field: 0 and 1
  - Type 0: Adm Field = 2 bytes, AN Field = 4 bytes
    - Adm field should contain an autonomous system number (ASN) from IANA
    - AN field is a number assigned by SP
  - Type 1: Adm Field = 4 bytes, AN field = 2 bytes
    - Administration field should contain an IP address assigned by IANA
    - Assigned Number field is a number assigned by SP
  - Examples: 10458:22:10.1.0.0/16 or 1.1.1.1:33:10.1.0.0/16

Using RDs to Disambiguate Addresses



The overlapping routes from A and B cannot be compared as they have unique RDs

Agenda: Layer 3 MPLS VPNs

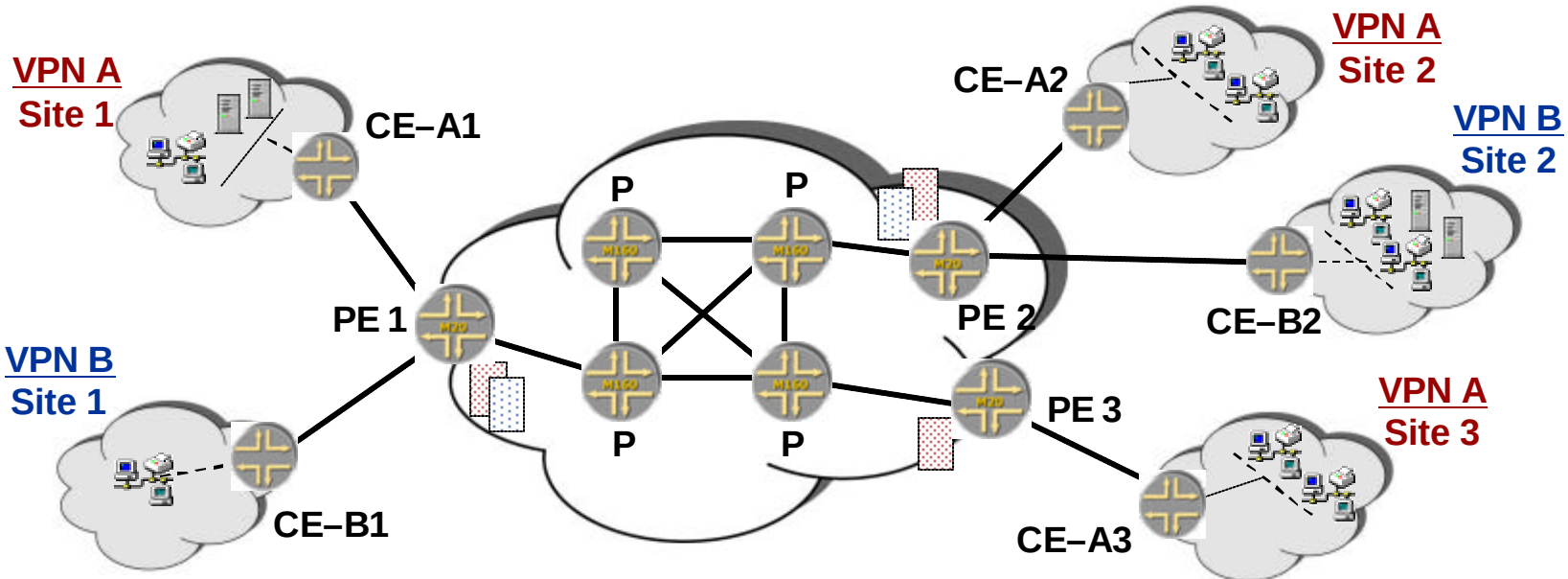
- RFC 2547bis terminology
- VPN-IPv4 address structure

➔ Operational characteristics

  ➔ Policy-based routing information exchange

  - Traffic forwarding

2547bis: Operational Overview



- Control flow (signaling plane)
  - Routing information exchange between CE and PE routers
    - Independent at both ends
  - Routing information exchange between PEs
  - LSP establishment between PEs (RSVP or LDP signaling)
- Data flow (forwarding plane)
  - Forwarding user traffic

RFC 2547bis Policies

- VPNs defined by administrative policies
  - Used for connectivity and QoS guarantees
  - Defined by customers
  - Implemented by service providers
- Full mesh or hub-and-spoke connectivity
  - Logical VPN topology results from the application of export and import Route Target policies

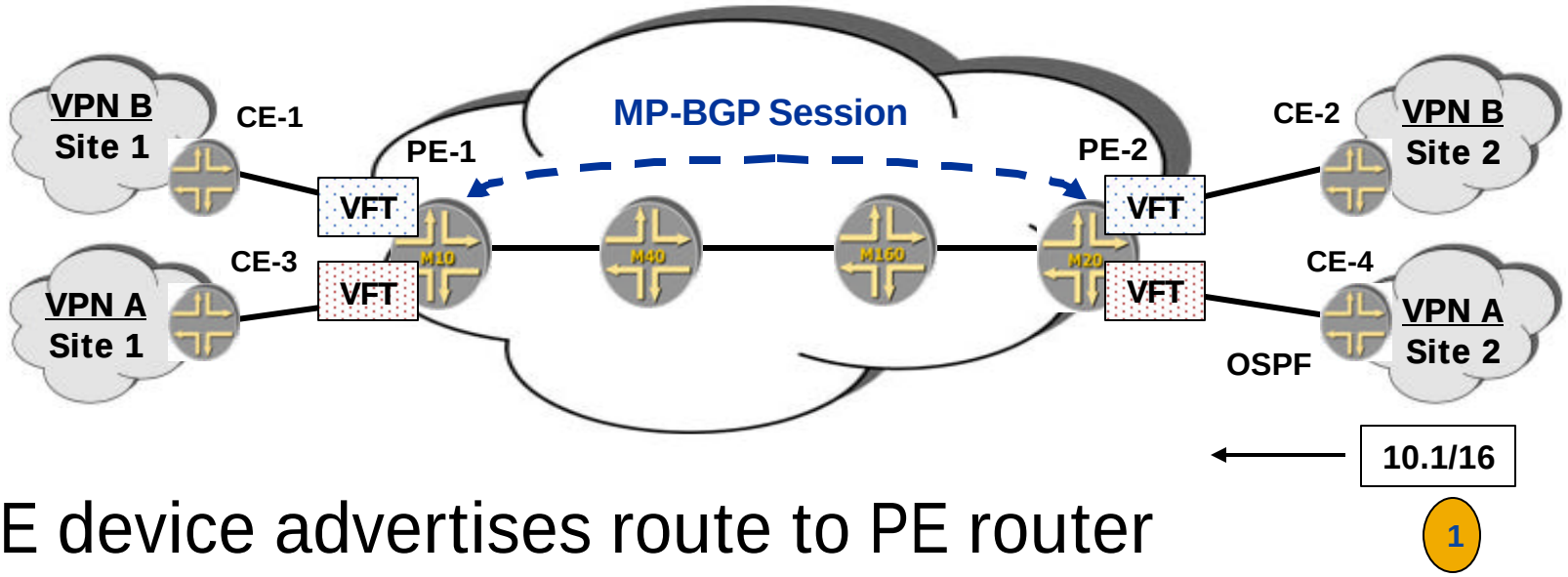
PE-PE Route Distribution

- Distribution of routes is controlled by BGP extended community attributes and VRF policy
  - Route Target
    - Identifies a set of VRFs to which a PE router distributes routes
  - Site of Origin/Route Origin
    - Identifies the specific site from which a PE router learns a route
- Structured similarly to the RD
  - 8 bytes in length
    - 2-byte Type field, 6-byte Value field
  - Type 0
    - 2-byte Global Administrator subfield (ASN)
    - 4-byte Local Administrator subfield
  - Type 1
    - 4-byte Global Administrator subfield (IANA-assigned IP Address)
    - 2-byte Local Administrator subfield

Route Targets

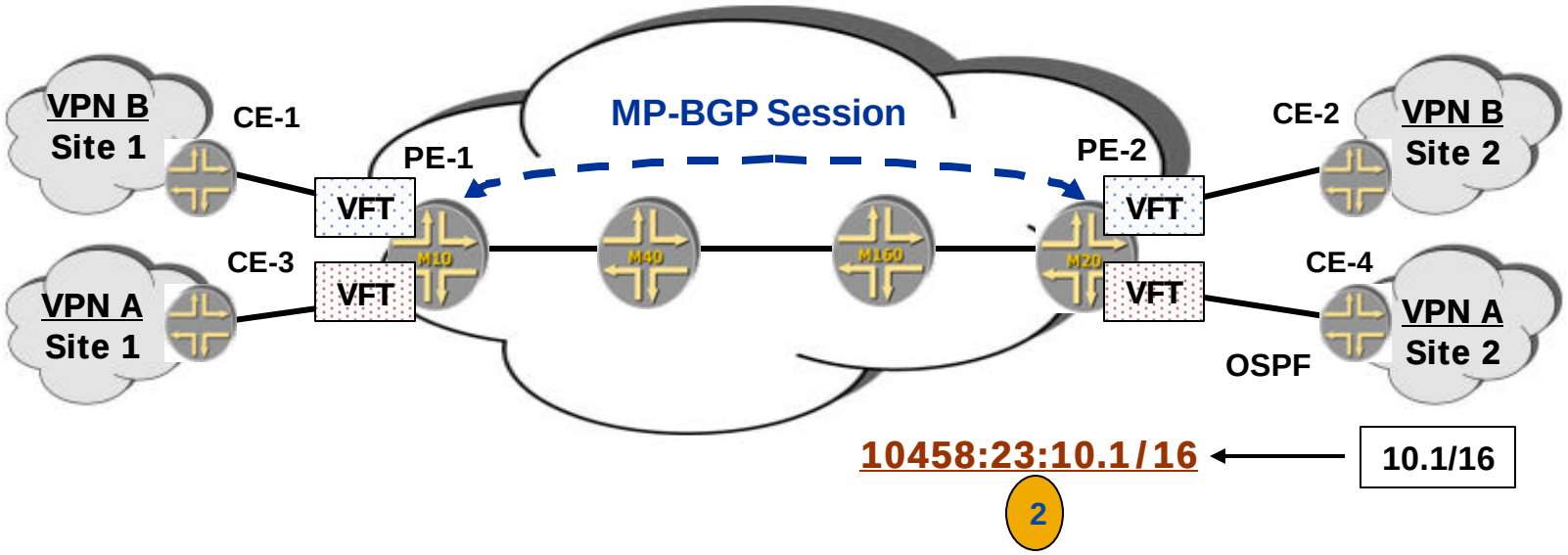
- Each VPN-IPv4 route advertised through MP-BGP is associated with a Route Target attribute
  - Export policies define the targets associated with routes a PE router sends
- Upon receipt of a VPN-IPv4 route, a PE router decides whether to add that route to a VRF
  - Import policies define which routes to add to a given VRF
- Route isolation between VRFs is accomplished through careful policy administration
  - SP provisioning tools can determine the appropriate export and import targets automatically

Exchange of Routing Information (1 of 7)



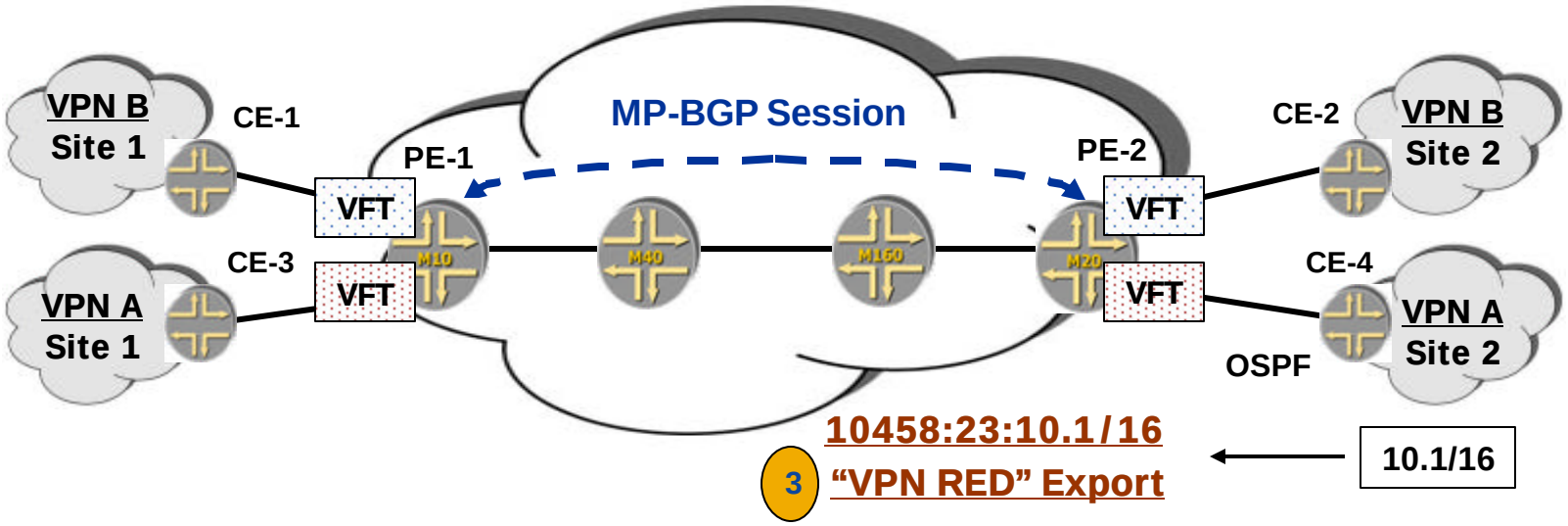
- CE device advertises route to PE router
  - Using traditional routing techniques (for example, OSPF, IS-IS, RIP, BGP, and static routes)

Exchange of Routing Information (2 of 7)



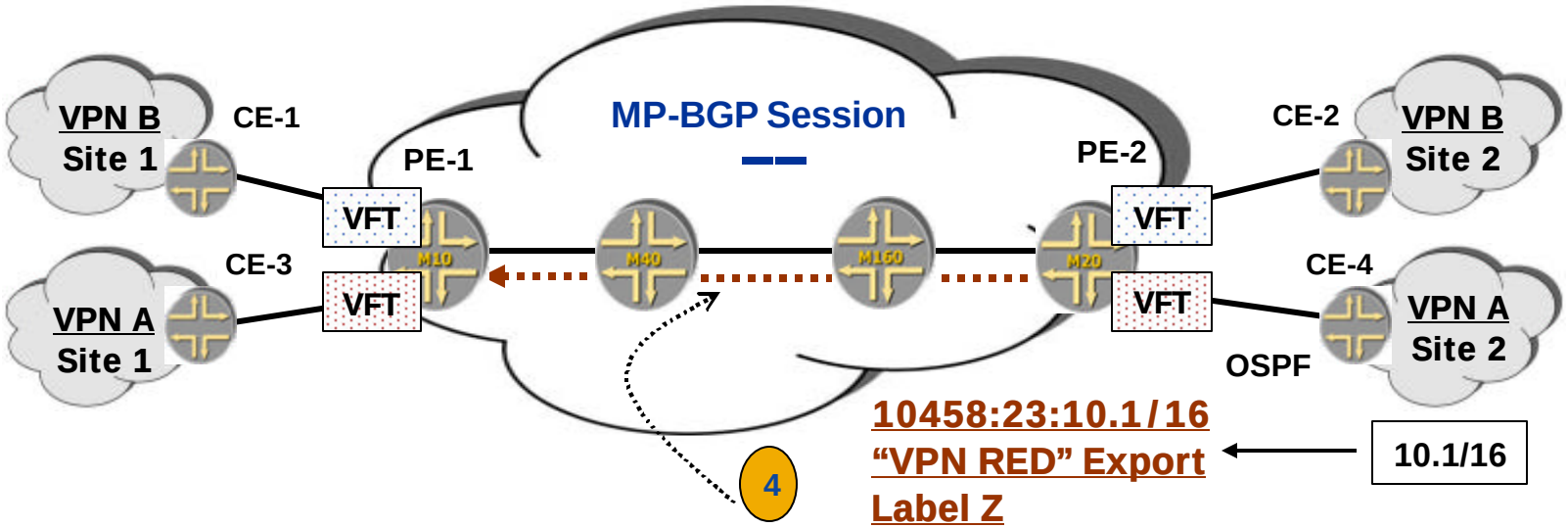
IPv4 address is added to the appropriate VRF

Exchange of Routing Information (3 of 7)



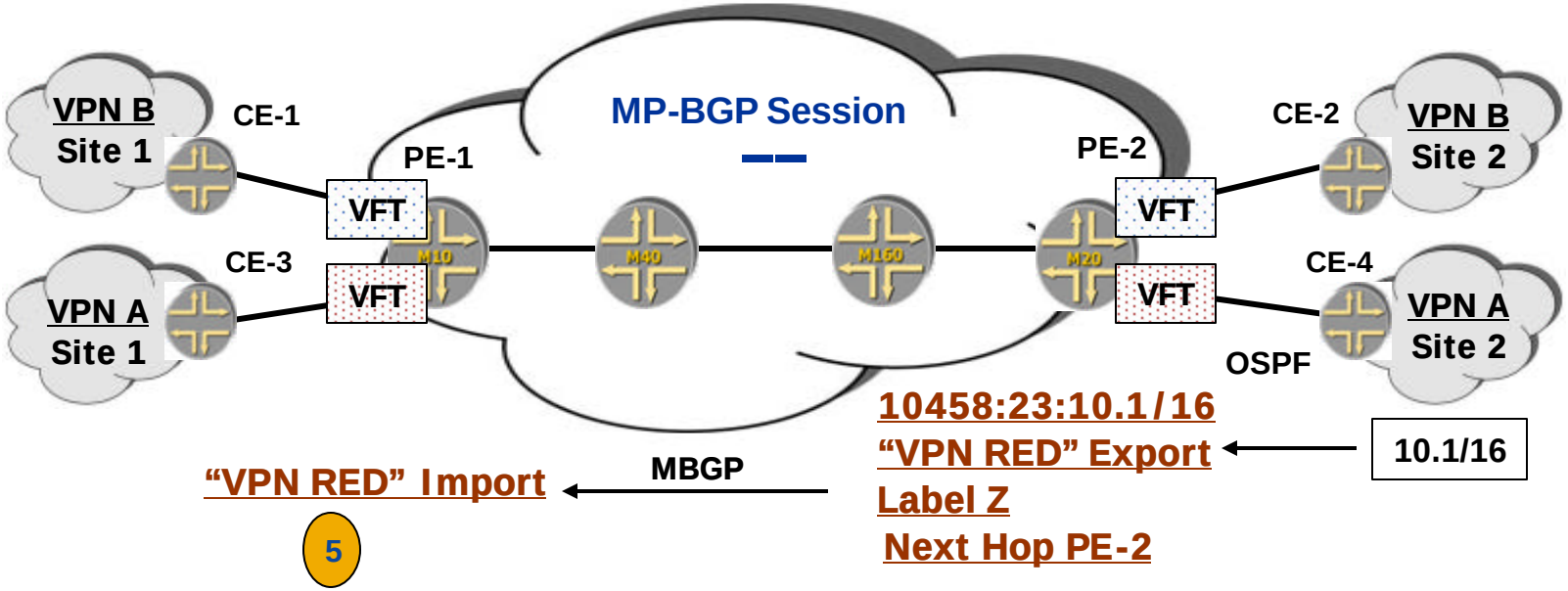
- VRF is associated with an export policy
  - VRF export adds "VPN RED" Route Target

Exchange of Routing Information (4 of 7)



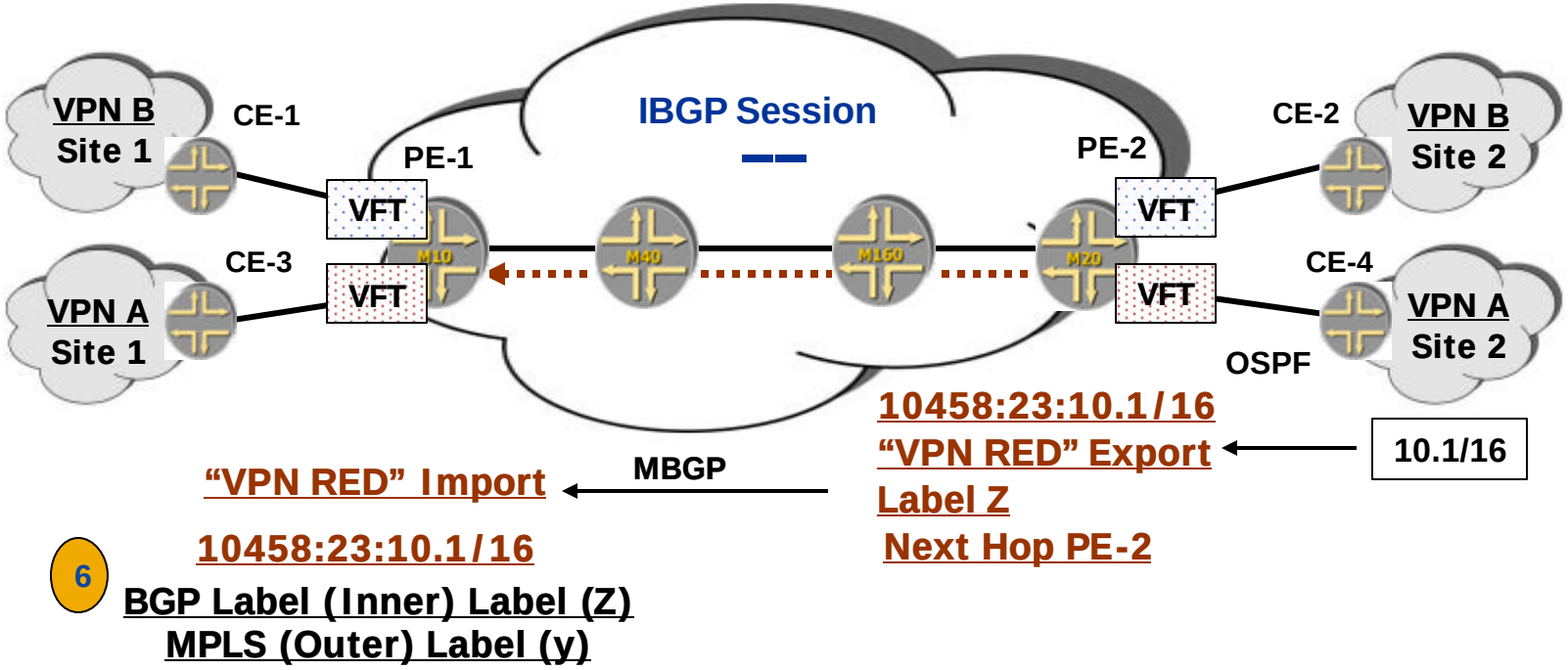
- VPN-IPv4 NLRI is advertised to other PEs
  - Inner label (a.k.a "VRF Label", "BGP Label")
  - Extended communities
    - Route Target
    - Site of Origin
  - BGP next hop (RID of advertising PE router)

Exchange of Routing Information (5 of 7)



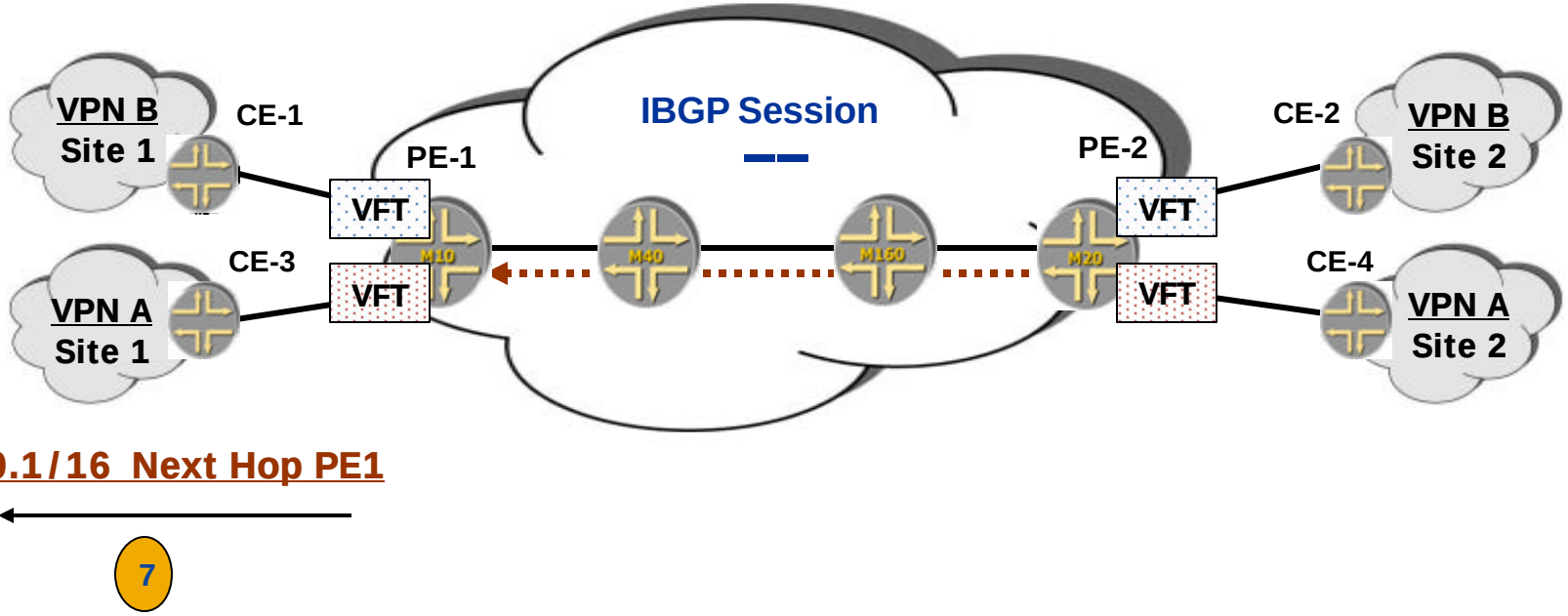
- Each PE router is configured with import Route Targets
  - Import Route Target is used to incorporate VPN-IPv4 routes into VRFs selectively
    - If import Route Target matches Route Target attribute in BGP route, the route is installed into the `bgp.l3vpn` table and copied into appropriate VRF(s)
    - Based on configured import policies, 10458:23:10.1/16 is copied into the red VRF but not the blue VRF

Exchange of Routing Information (6 of 7)



- Each VPN-IPv4 route in a VRF is associated with:
  - Inner (VRF) label to reach the advertised NLRI (carried in BGP update)
  - Outer label to reach the PE router
- All routes associated with the same VRF interface can share a common label

Exchange of Routing Information (7 of 7)



- Each IPv4 route installed in a VRF can be advertised to the CEs associated with that VRF
  - For example, RIP, OSPF, and BGP
  - Routing policy can be used on the PE-CE link to control the exchange of routing information further

Agenda: Layer 3 MPLS VPNs

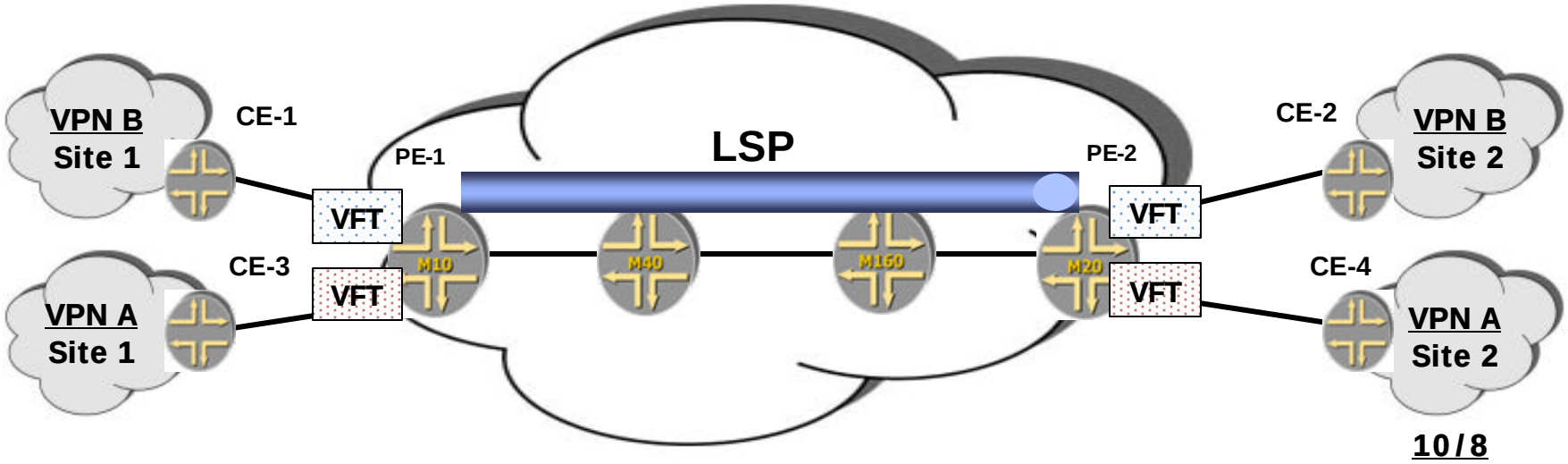
- RFC 2547bis terminology
- VPN-IPv4 address structure

➔ Operational characteristics

  - Policy-based routing information exchange

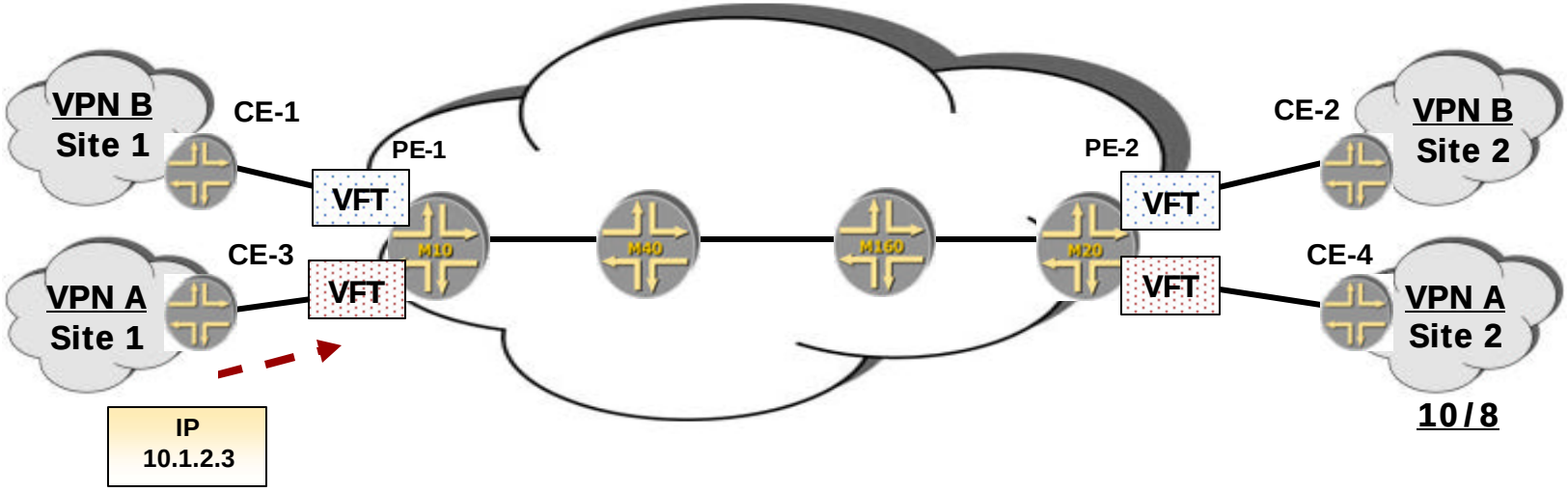
  ➔ Traffic forwarding

Data Flow (1 of 7)



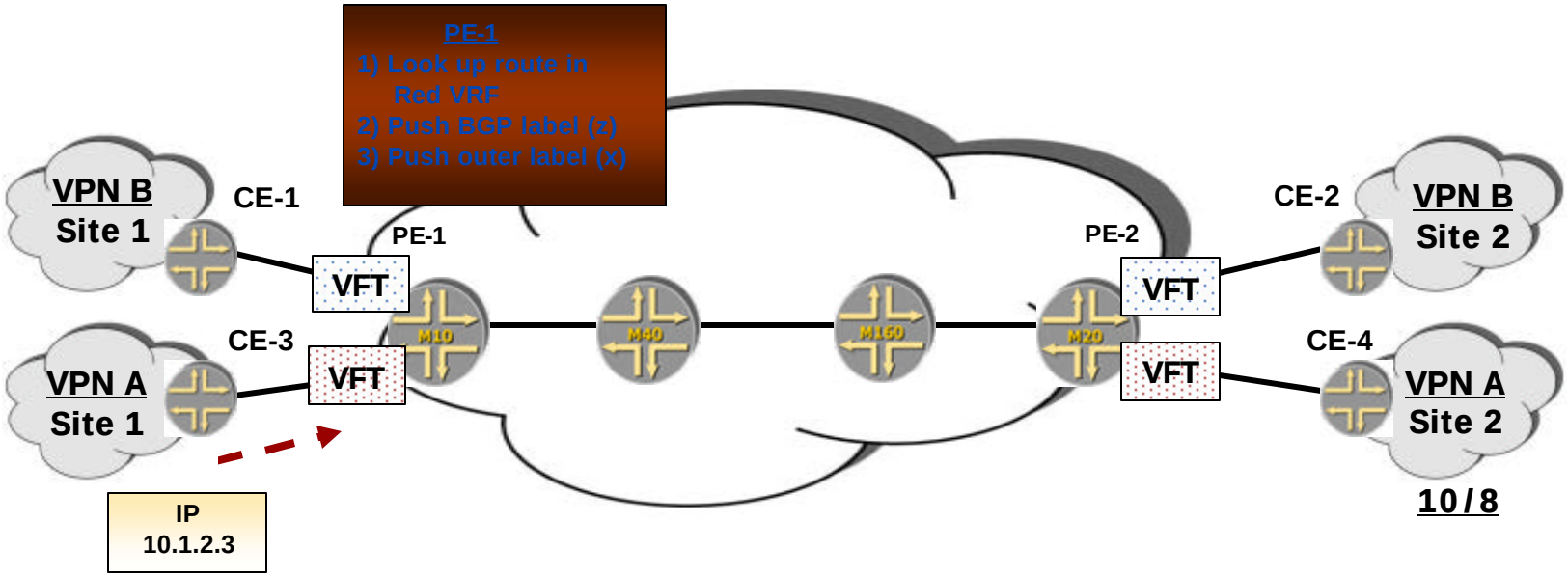
- The PE-to-PE LSP must be in place before forwarding data across the MPLS backbone
  - LSPs are signaled through LDP or RSVP

Data Flow (2 of 7)



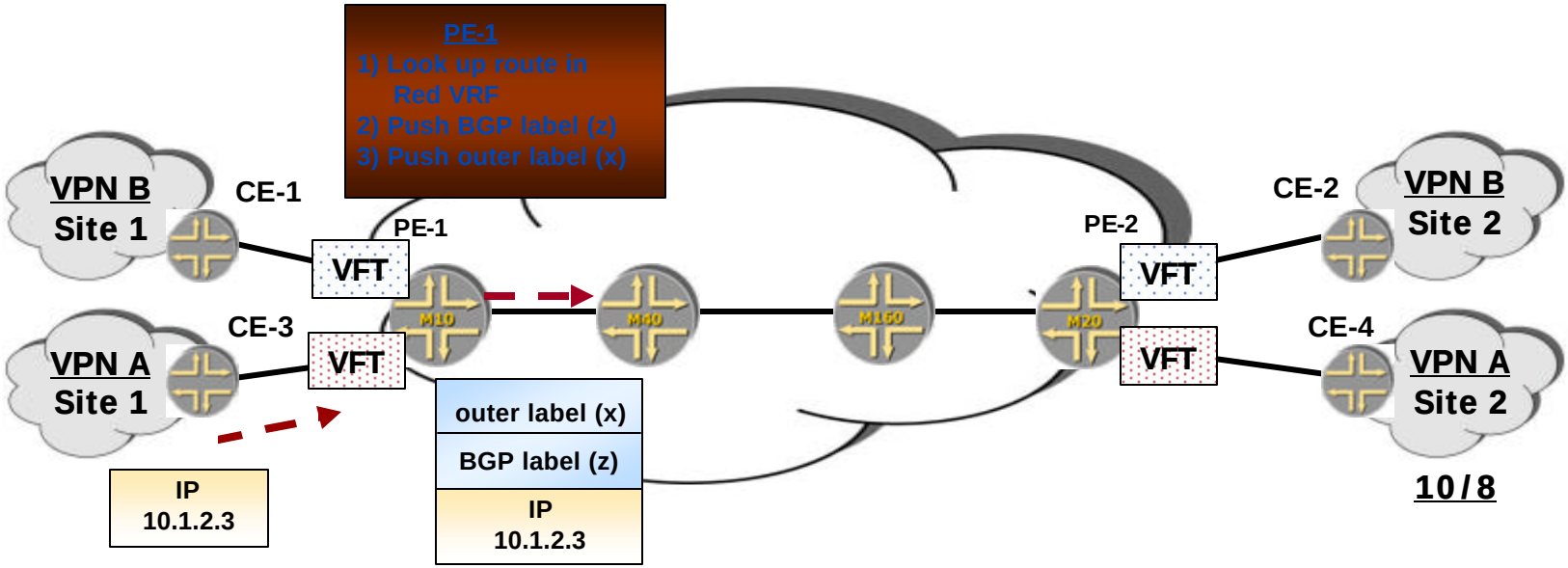
The CE device performs a traditional IPv4 lookup and sends packets to the PE router

Data Flow (3 of 7)



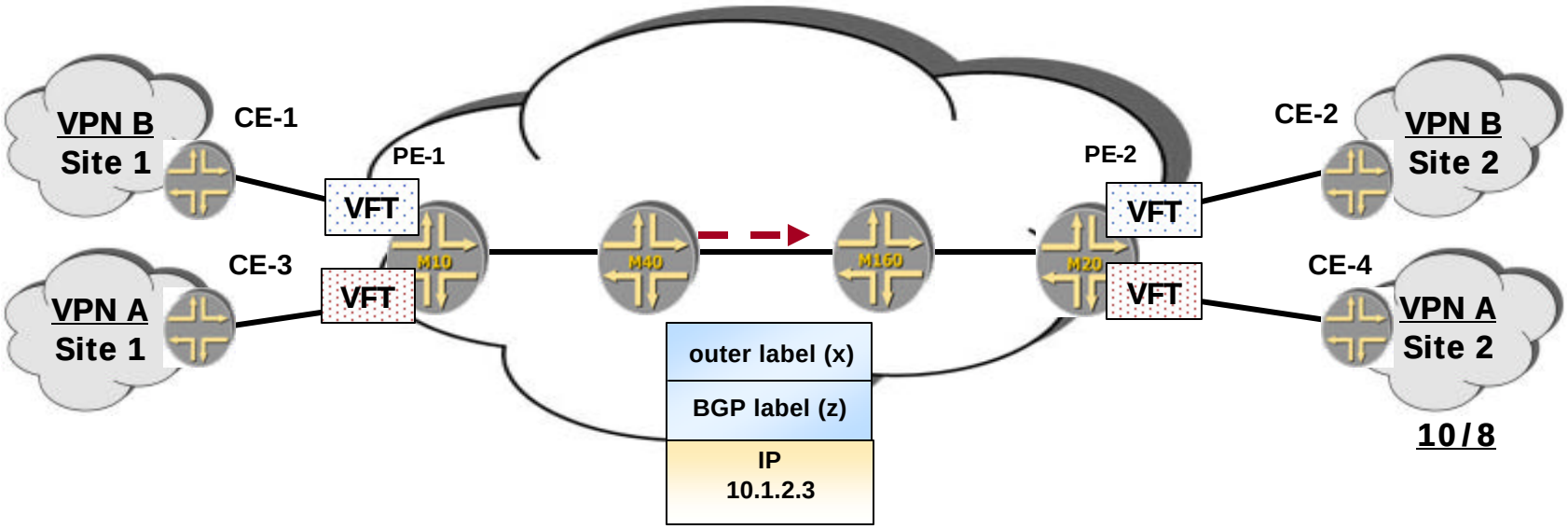
- The PE router consults the appropriate VRF for the inbound interface
- Two labels are derived from the VRF route lookup and are *pushed* onto the packet

Data Flow (4 of 7)



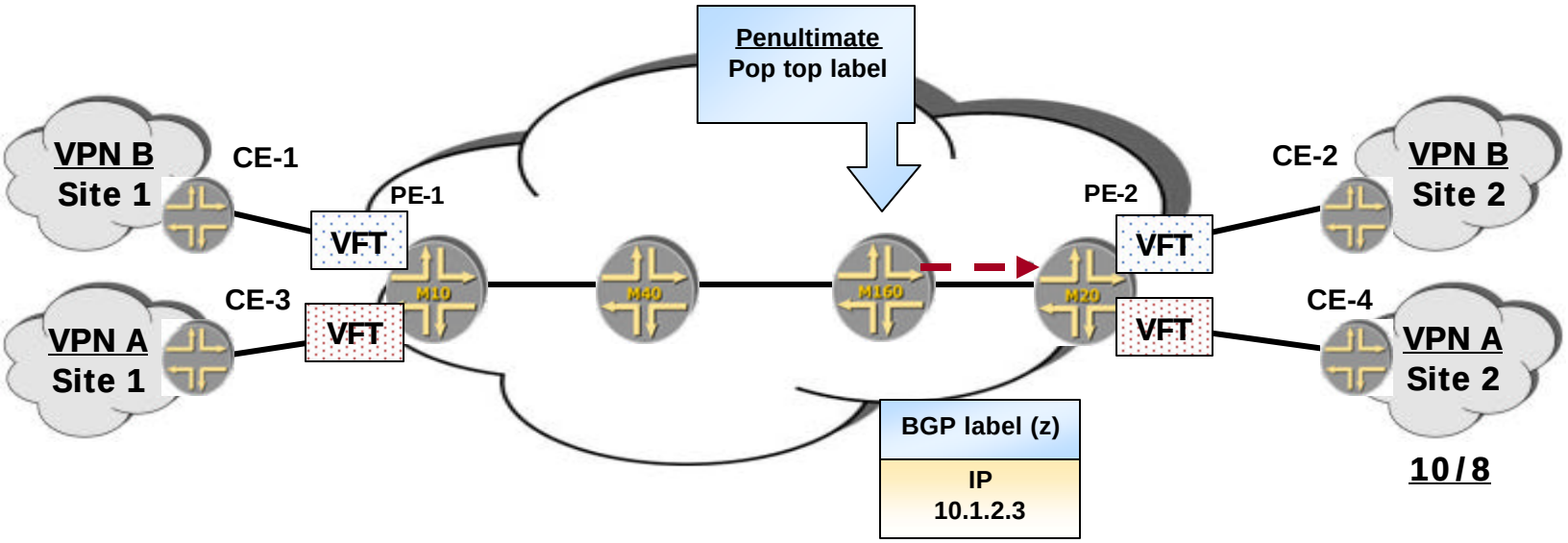
- Packets are forwarded using two-level label stack
  - Outer (MPLS) label
    - Identifies the LSP to egress PE router
    - Resolves BGP next hop through inet.3
    - Distributed by RSVP or LDP
  - Inner BGP label
    - Identifies outgoing interface from egress PE to CE
    - Communicated in BGP updates (control plane)

Data Flow (5 of 7)



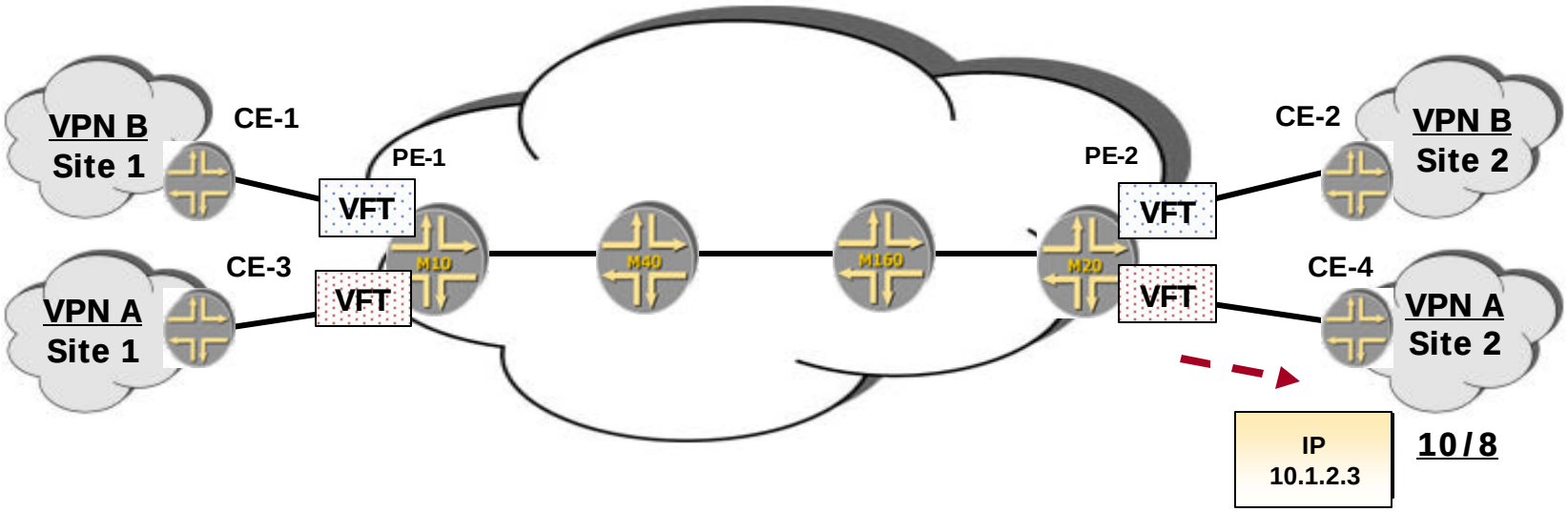
- After packets exit the ingress PE router, the outer label is used to traverse the service provider
  - P routers are not VPN-aware

Data Flow (6 of 7)



Penultimate hop popping (before reaching the egress PE router) removes the outer label

Data Flow (7 of 7)



- The inner label is removed at the egress PE router
- The native IPv4 packet is sent to the outbound interface associated with the label

Module Review

- Can you now:
  - Define the roles of P, PE, and CE routers?
  - Describe the format of VPN-IPv4 addresses?
  - Explain the role of the route distinguisher (RD)?
  - Describe the flow of 2547bis control information?
  - Explain the operation of the 2547bis forwarding plane?

L3 VPN Configuration

Module Objectives

- After completing this module, you will be able to perform the following:
  - Create VRFs
  - Write and apply VRF policy
  - Configure BGP extended communities
  - Configure a point-to-point Layer 3 VPN topology using RSVP

Agenda: Configuring Layer 3 VPNs

- Preliminary steps
- PE configuration
  - VRF instance
    - Assign route distinguisher
    - Associate VRF interfaces
  - VRF policy
    - Create and apply BGP extended communities
  - PE-CE routing protocol
    - AS-override
    - Site of Origin community
    - OSPF Domain Identifier community

Agenda: Configuring Layer 3 VPNs

- Preliminary steps
- PE configuration
  - VRF instance
    - Assign route distinguisher
    - Associate VRF interfaces
  - VRF policy
    - Create and apply BGP extended communities
  - PE-CE routing protocol
    - AS-override
    - Site of Origin community
    - OSPF Domain Identifier community

2547bis Preliminary Configuration

- Preliminary steps:
  1. Choose and configure the IGP for PE and P routers
  2. Configure MP-IBGP peering among PE routers
     - Must include VPN-IPv4 NLRI capability
  3. Enable the LSP signaling protocol(s)
  4. Establish LSPs between PE routers
- The PE routers perform VPN-specific configuration

PE-PE MP-IBGP Peering

- PE-to-PE MP-IBGP sessions require VPN-IPv4 NLRI
- JUNOS software automatically negotiates BGP route refresh

```
[edit]
lab@Amsterdam# show protocols bgp
group int {
    type internal;
    local-address 192.168.24.1;
    family inet {
        unicast;
    }
    family inet-vpn {
        unicast;
    }
    neighbor 192.168.16.1;
}
```

MP-IBGP Peering: PE-PE

```
lab@Amsterdam> show bgp neighbor
Peer: 192.168.16.1+179 AS 65412 Local: 192.168.24.1+1048 AS 65412
  Type: Internal    State: Established    Flags: <>
  Last State: OpenConfirm   Last Event: RecvKeepAlive
  Last Error: None
  Options: <Preference LocalAddress HoldTime AddressFamily Rib-group Refresh>
  Address families configured: inet-unicast inet-vpn-unicast
  Local Address: 192.168.24.1 Holdtime: 90 Preference: 170
  Number of flaps: 0
  Peer ID: 192.168.16.1     Local ID: 192.168.24.1     Active Holdtime: 90
  Keepalive Interval: 30
  NLRI advertised by peer: inet-unicast inet-vpn-unicast
  NLRI for this session: inet-unicast inet-vpn-unicast
  Peer supports Refresh capability (2)
  Table inet.0 Bit: 10000
    Send state: in sync
    Active prefixes: 0
    Received prefixes: 0
    Suppressed due to damping: 0
  Table bgp.l3vpn.0 Bit: 30000
    Send state: in sync
    Active prefixes: 8
    Received prefixes: 8
    Suppressed due to damping: 0
  Table vpna.inet.0 Bit: 40000
    Send state: in sync
    Active prefixes: 7
    Received prefixes: 8
```

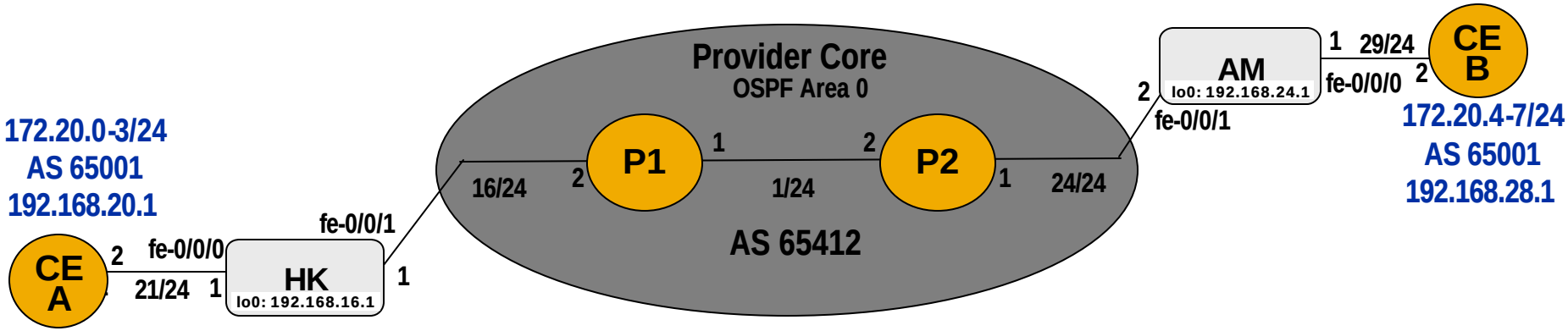
Agenda: Configuring Layer 3 VPNs

- Preliminary steps
- ➔ PE configuration
  - ➔ VRF instance
    - ➔ Assign route distinguisher
    - ➔ Associated VRF interfaces
  - VRF policy
    - Create and apply BGP extended communities
  - PE-CE routing protocol
    - AS-override
    - Site of Origin community
    - OSPF Domain Identifier community

PE Configuration

- PE routers do all VPN-specific configuration
- PE routing instance
  - Create routing instance and list associated VRF interfaces
  - Assign a route distinguisher
  - Link the VRF to import and export policies
  - Configure PE-CE routing protocol properties
- VPN policy
  - Create and apply BGP extended communities (for example, Route Target/Site of Origin)
  - Create VRF import and export policies

Sample Layer 3 VPN Topology



- Network characteristics
  - Interface addressing is 10.0.x/24 (except loopbacks)
  - IGP is single area OSPF
  - RSVP signaling between PE devices, LSPs established between PEs (CSPF not required)
  - Full MP-IBGP mesh between PEs, lo0 peering, VPN-IPv4 NLRI
  - CE-PE link running eBGP
  - Full mesh Layer 3-VPN between CE-A and CE-B
- Actual lab topology will differ–this is a *sample* network

VRF Routing Instances

VRFs are created at the [edit routing-instances] configuration hierarchy

```
[edit routing-instances vpna]
lab@HK# set ?
Possible completions:
+ apply-groups          Groups from which to inherit
                     configuration data
  instance-type         Type of routing instance
> interface             Interface name for this routing instance
> protocols             Routing protocol configuration
> route-distinguisher   Route Distinguisher for this instance
> routing-options       Protocol-independent routing option
                     configuration
+ vrf-export            Export Policy for vrf instance RIBs
+ vrf-import            Import Policy for vrf instance RIBs
```

A Sample VRF Configuration

Creating a VRF called *vpn-a* with BGP running between the PE and CE

```
[edit routing-instances vpn-a]
lab@HK# show
instance-type vrf;
interface fe-0/0/0.0;
route-distinguisher 192.168.16.1:1;
vrf-import vpna-import;
vrf-export vpna-export;
protocols {
    bgp {
        group ce-a {
            type external;
            peer-as 6501;
            neighbor 10.0.6.2;
        }
    }
}
```

Agenda: Configuring Layer 3 VPNs

- Preliminary steps
- PE configuration
  - VRF instance
    - Assign route distinguisher
    - Associated VRF interfaces
  - ➔ VRF policy
    - ➔ Create and apply BGP extended communities
  - PE-CE routing protocol
    - AS-override
    - Site of Origin community
    - OSPF Domain Identifier community

Sample VRF Import Policy

- Installs routes learned from other PEs via MP-IBGP
  - Routes with the specified community are installed in the associated VRF

```
[edit policy-options]
lab@HK# show policy-statement vpn-a-import
term 1 {
        from {
            protocol bgp;
            community vpn-a-target;
        }
        then accept;
    }
    term 2 {
        then reject;
    }
}
```

Sample VRF Export Policy

```
lab@HK# show policy-statement vpn-a-export
term 1 {
    from protocol bgp;
    then {
        community add vpn-a-target;
        community add ce-name-origin;
        accept;
    }
}
term 2 {
    then reject;
}
```

- This policy advertises routes learned via BGP from the CE, while adding the Route Target and Origin communities
  - Matching routes are sent to MP-IBGP peers that have advertised VPN-IPv4 NLRI capabilities

Extended BGP Communities

```
community ce-name-origin members origin:192.168.16.1:100;

community vpn-a-target members target:65412:100;
```

- The `origin` tag allows the specification of Site of Origin community
  - SoO can be used to prevent routing loops when a user has multiple AS numbers
- The `target` tag specifies the Route Target
  - Policy matches on the Route Target control which routes are imported into a given VRF
- Boolean operations possible

PE-CE Policy

- JUNOS software import/export policies can be applied to VRF instances
  - BGP and RIP allow both import and export
  - Link-state protocols allow only export
- Affects routes being sent and received over the PE-CE link

PE-CE BGP Routing/Policy Example

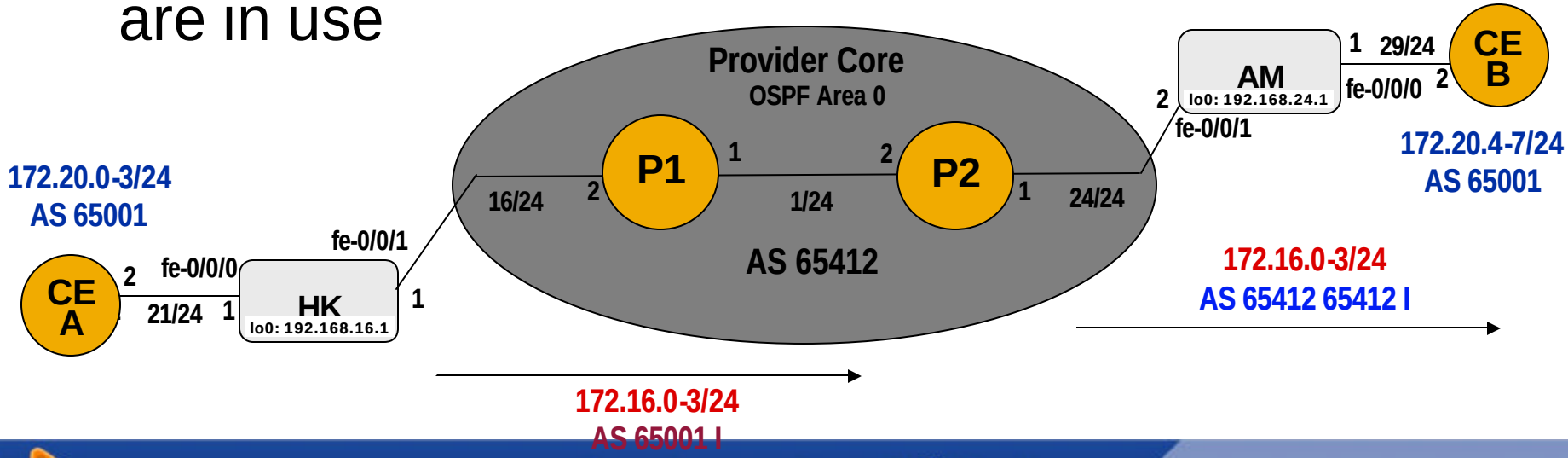
```
lab@Hong-Kong# show routing-instances
vpna {
    .  .  .
    }
    protocols {
        bgp {
            import site-a;
            group ext {
                type external;
                peer-as 65001;
                as-override;
                neighbor 10.0.21.2;
            }
        }
    }
[edit]
lab@ Hong-Kong # show policy-options policy-statement site-a
from protocol bgp;
then {
    as-path-prepend “64512 64512”;
    community add cust-a;
    accept;
}
```

Agenda: Configuring Layer 3 VPNs

- Preliminary steps
- PE configuration
  - VRF instance
    - Assign route distinguisher
    - Associated VRF interfaces
  - VRF policy
    - Create and apply BGP extended communities
  - ➔ PE-CE routing protocol
    - ➔ AS-override
    - ➔ Site of Origin community
    - ➔ OSPF Domain Identifier community

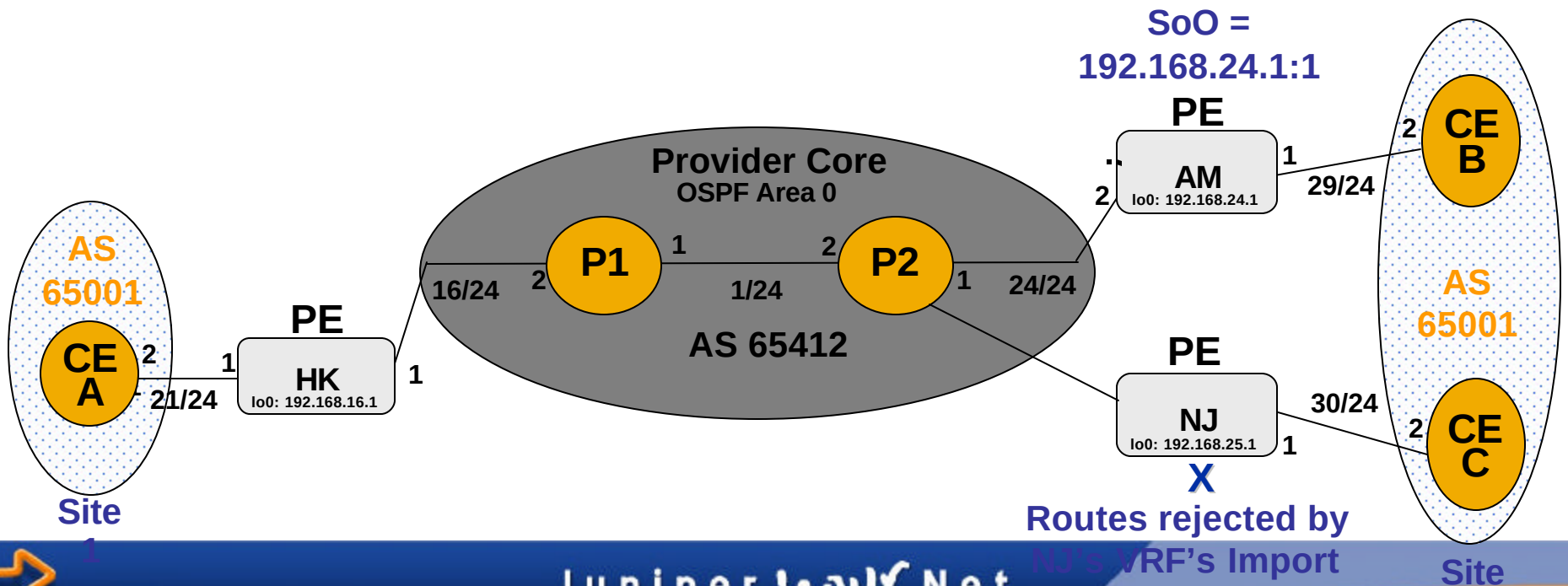
AS-Override

- Use this knob when CE routers belong to the same AS
- Causes the PE to overwrite CE-A's AS # with the provider's AS # (two provider AS #s in AS-path)
- The "autonomous-system loops n" knob can also be used
- Remove-private can also work if private AS numbers are in use



Site of Origin (SoO)

- Use this knob when CE router is dual-homed and AS-override is required (Corner case)
  - `as-override` required to allow route exchange between CE-A and CE-B/C
- SoO (and policy) prevents advertising routes back to the source
  - Advertising these routes back to the CE can cause forwarding loops with equipment that prefers eBGP over IGP-learned routes



PE-CE OSPF Routing

- Requires a separate OSPF process for each VRF
- Carries OSPF routes across backbone as BGP routes
- Routes can appear in CE as external LSAs (type 5|7) or summary LSAs (type 3)
  - Cannot support stub/totally-stubby areas
  - Summary LSA support requires domain ID
    - JUNOS software ≥ 5.0 supports Domain ID community
- PE VRF exports from OSPF, imports from BGP

Basic OSPF VRF Example

```
lab@Hong-Kong# show routing-instances vpna
instance-type vrf;
interface fe-0/0/0.0;
route-distinguisher 192.168.16.1:1;
vrf-import vpna-import;
vrf-export vpna-export;
protocols {
    ospf {
        export bgp-to-ospf;
        area 0.0.0.0 {
            interface fe-0/0/0.0;
        }
    }
}

lab@Hong-Kong# show policy-options
. . .
policy-statement bgp-to-ospf{
    from protocol bgp;
    then accept;
}
```

Policy needed! OSPF does not redistribute BGP routes by Default

OSPF VRF Policy (Basic)

```
lab@Hong-Kong# show policy-options
policy-statement vpna-import {
    term 1 {
        from {
            protocol bgp;
            community vpna-target;
        }
        then accept;
    }
    term 2 {
        then reject;
    }
}
policy-statement vpna-export {
    term 1 {
        from protocol ospf;
        then {
            community add vpna-target;
            accept;
        }
    }
    term 2 {
        then reject;
```

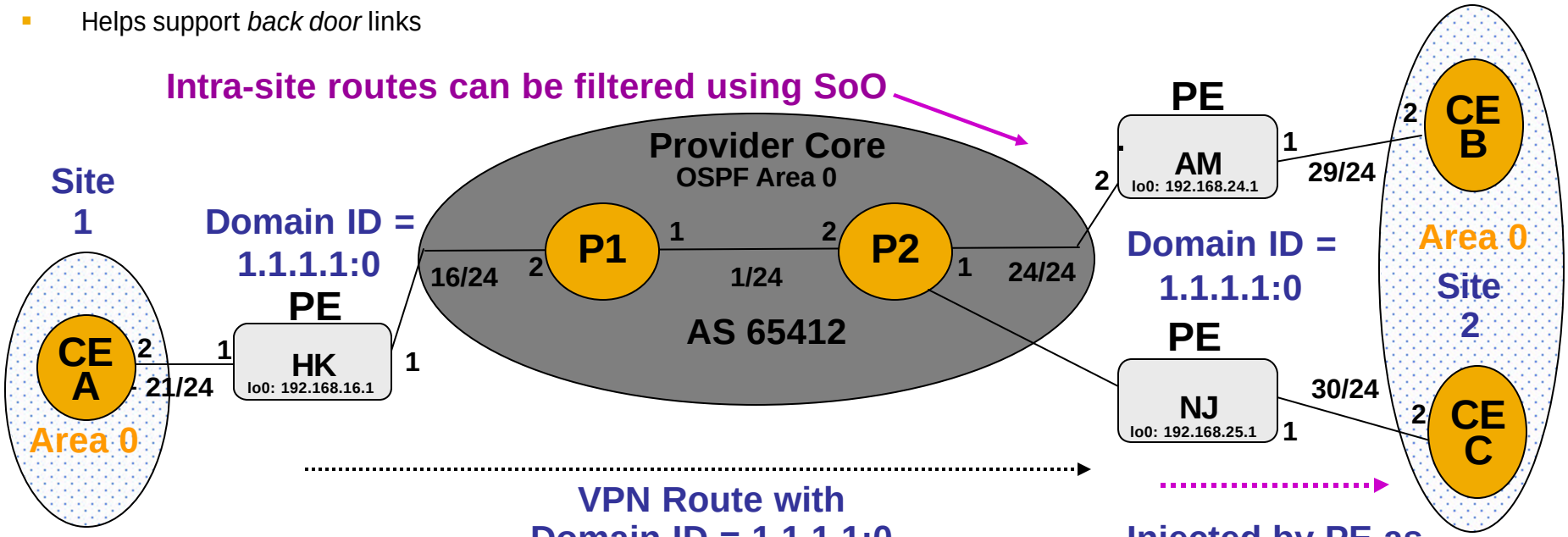
Basic OSPF Configuration Results

- Routes appear in CE as AS-external and summary LSAs
  - Lack of Domain ID causes implicit match and summary LSA generation

```
lab@ce-a> show ospf database

    OSPF link state database, area 0.0.0
 Type       ID               Adv Rtr           Seq      Age  Opt  Cksum  Len
Router   10.0.21.1        10.0.21.1       0x8000000f    62  0x2  0xf8c7  36
Router  *192.168.20.1     192.168.20.1    0x80000025    61  0x2  0xafaf  48
Network *10.0.21.2       192.168.20.1    0x8000000d    61  0x2  0x24eb  32
Summary  192.168.28.1     10.0.21.1       0x80000003    62  0x82 0x52e   28
Summary  200.0.0.0         10.0.21.1       0x80000003    62  0x82 0xcd22  28
    OSPF external link state database
 Type       ID               Adv Rtr           Seq      Age  Opt  Cksum  Len
Extern  *10.0.21.0       192.168.20.1    0x80000015    61  0x2  0x9f84  36
Extern   10.0.29.0       10.0.21.1       0x80000005    62  0x2  0x9f95  36
Extern  *172.20.0.0      192.168.20.1    0x80000013    61  0x2  0x6a17  36
Extern   172.20.4.0      10.0.21.1       0x80000005    62  0x2  0x9202  36
Extern   192.168.28.0    10.0.21.1       0x80000002    62  0x2  0x9343  36
```

The OSPF Domain ID

- Allows OSPF routes to appear as type 3 LSAs (intra-area summary)
  - Up/Down bit and VPN route tag to prevent looping
- Uses three BGP extended communities:
  - OSPF Route Type (Type : 0x8000)
  - OSPF Domain ID (VPN of Origin) (Type : 0x8005)
  - OSPF Router ID (Type : 0x8001)
- Helps support *back door* links



VRF Example: OSPF with Domain ID

```
test@HK-pe# show routing-instances
vpna {
    instance-type vrf;
    interface fe-0/0/0.0;
    route-distinguisher 192.168.16.1:1;
    vrf-import vpna-import;
    vrf-export vpna-export;
    routing-options {
        router-id 192.168.16.1;
    }
    protocols {
        ospf {
            domain-id 1.1.1.1;
            export bgp;
            area 0.0.0.0 {
                interface all;
            }
        }
    }
}
```

OSPF Domain ID Policy Example

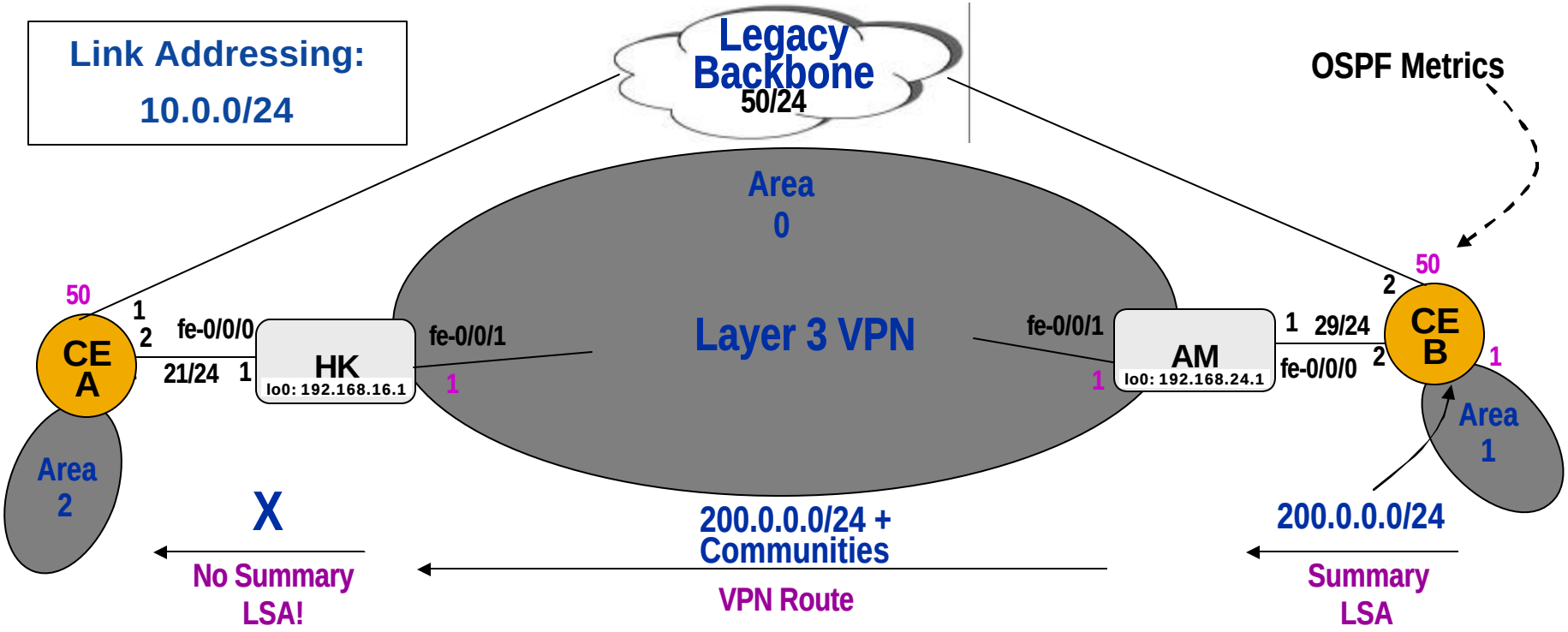
```
lab@Amsterdam-pe# show policy-options

. . .
policy-statement vpna-export {
    term 1 {
        from protocol ospf;
        then {
            community add vpna;
            community add domain;
            accept;
        }
    }
    term 2 {
        then reject;
    }
}
community domain members domain:1.1.1.1:0;
community vpna members target:65412:100;
```

Mismatched OSPF Domain IDs

- All remote routes are now presented as external LSAs
  - Makes back-door links problematic
  - Externals may be desired for extranet support

```
lab@ce-a> show ospf database

    OSPF link state database, area 0.0.0.0
 Type       ID               Adv Rtr           Seq      Age  Opt  Cksum  Len
Router   10.0.21.1        10.0.21.1       0x80000012     9  0x2  0xf2ca  36
Router  *192.168.20.1     192.168.20.1    0x80000028     8  0x2  0xa9b2  48
Network *10.0.21.2       192.168.20.1    0x80000010     8  0x2  0x1eee  32
    OSPF external link state database
 Type       ID               Adv Rtr           Seq      Age  Opt  Cksum  Len
Extern  *10.0.21.0       192.168.20.1    0x80000018     8  0x2  0x9987  36
Extern   10.0.29.0       10.0.21.1       0x80000007     9  0x2  0x9b97  36
Extern  *172.20.0.0      192.168.20.1    0x80000015     8  0x2  0x6619  36
Extern   172.20.4.0      10.0.21.1       0x80000007     9  0x2  0x8e04  36
Extern   192.168.28.0    10.0.21.1       0x80000004     9  0x2  0x8f45  36
Extern   192.168.28.1    10.0.21.1       0x80000002     9  0x2  0x9341  36
Extern   200.0.0.0       10.0.21.1       0x80000002     9  0x2  0x5c35  36
```

OSPF Back Door Links: A Case Study



- CE A forwards to 200.0.0.0/24 over the legacy backbone with a metric of 51
  - Downing the legacy backbone causes CE A to use the Layer 3 backbone, now with a metric of 3
- HK does not generate a summary LSA for 200.0.0/24 when the legacy backbone is operational

A Vital Clue

```
test@HK-pe> show route 200.0.0.0

vpna.inet.0: 14 destinations, 14 routes (14 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

200.0.0.0/24        *[OSPF/10] 00:01:39, metric 52
                     > to 10.0.21.2 via fe-0/0/0.0
                     [BGP/170] 00:01:40, MED 2, localpref 100, from 192.168.24.1
                       AS path: I
                     > to 10.0.16.2 via fe-0/0/1.0, label-switched-path AM
```

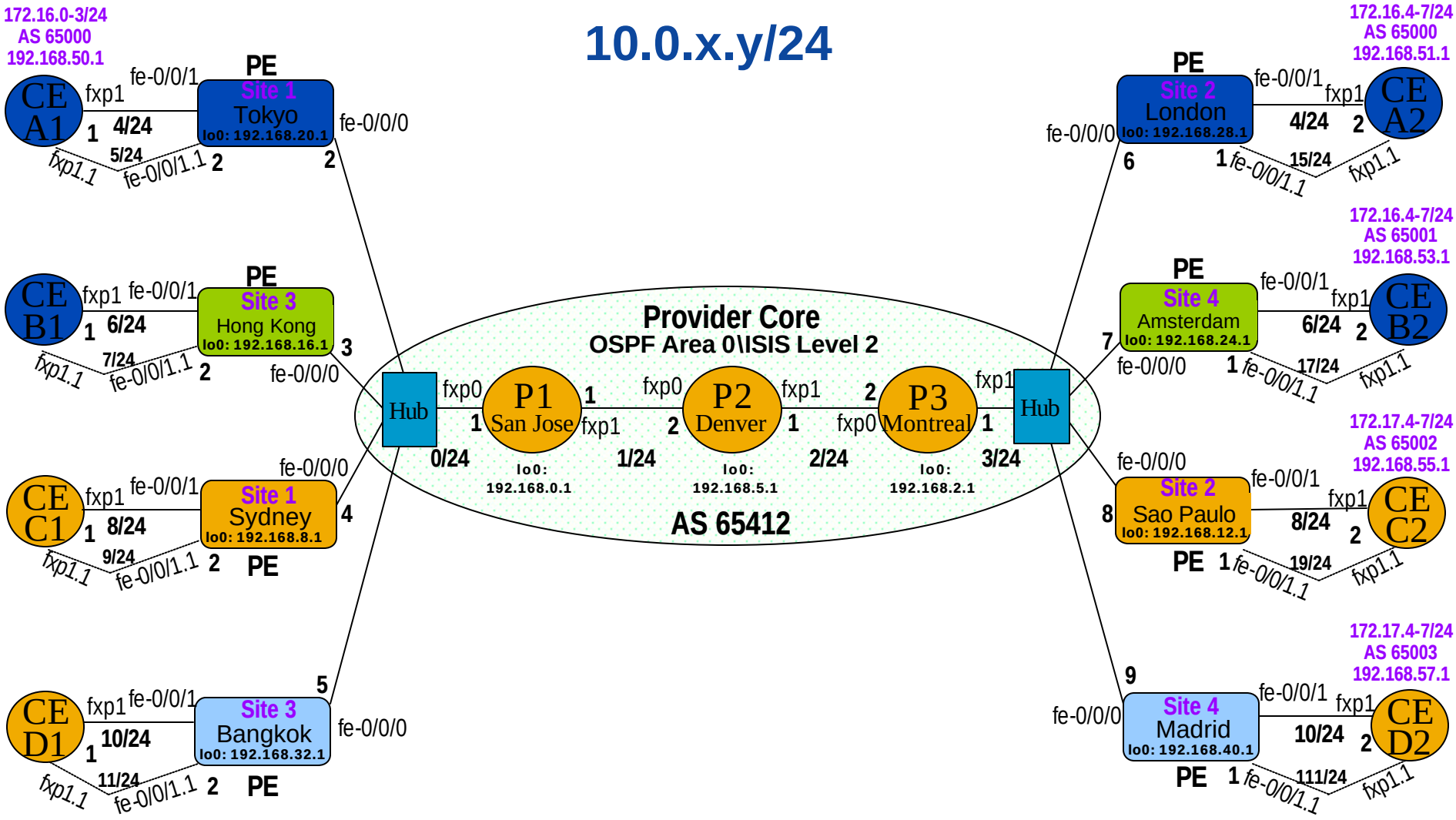
- JUNOS software policy only affects *active* routes
  - Default route preference causes the PE to choose the OSPF route received, learned from CE-A
  - The route learned from BGP cannot be sent until it becomes active

A Solution

```
[edit routing-instances vpna]
test@HK-pe# set protocols ospf preference 180

[test@HK-pe# commit and-quit

test@HK-pe> show route 200.0.0.0

vpna.inet.0: 14 destinations, 14 routes (14 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

200.0.0.0/24        *[BGP/170] 00:00:21, MED 2, localpref 100, from 192.168.24.1
                      AS path: I
                     > to 10.0.16.2 via fe-0/0/1.0, label-switched-path AM
                     [OSPF/180] 00:00:20, metric 52
                     > to 10.0.21.2 via fe-0/0/0.0
```

- Change the preferences associated with this routing instance
  - Allows the BGP route to become active, even when receiving the OSPF route from CE-A

Lab 2:
Point-to-Point VPN with RSVP Signaling



Module Review

- Can you now:
  - Create VRFs?
  - Write and apply VRF policy?
  - Configure BGP extended communities?
  - Configure a point-to-point Layer 3 VPN topology using RSVP?