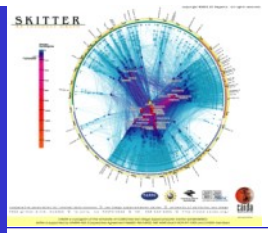
ISP and NSP Security Workshop

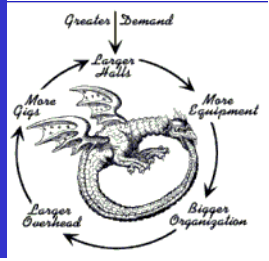
APRICOT 2013

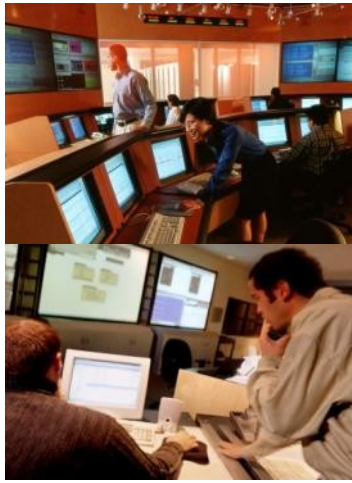
Day 4

MPLS/L3VPN Security

MPLS / L3VPN Security







Before we start…

- Mainly of interest to providers/ISP/Carriers
  - Large enterprises, verticals also using MPLS internally
- To support MPLS in your network you **MUST** have:
  - Fully working IP network. If it's broken MPLS won't fix it
  - Hardware and Software support. Depends on vendors
    - Alcatel-Lucent
      - Service Routers (7705, 7210, 7450, 7750, 7950)
    - Juniper
      - M-series, T-series, J-series, E-series, MX-series
    - Cisco
      - Most platforms with CEF support (1800/2800/3600/6500/7200/7600/9000 etc)

Things I want you to know

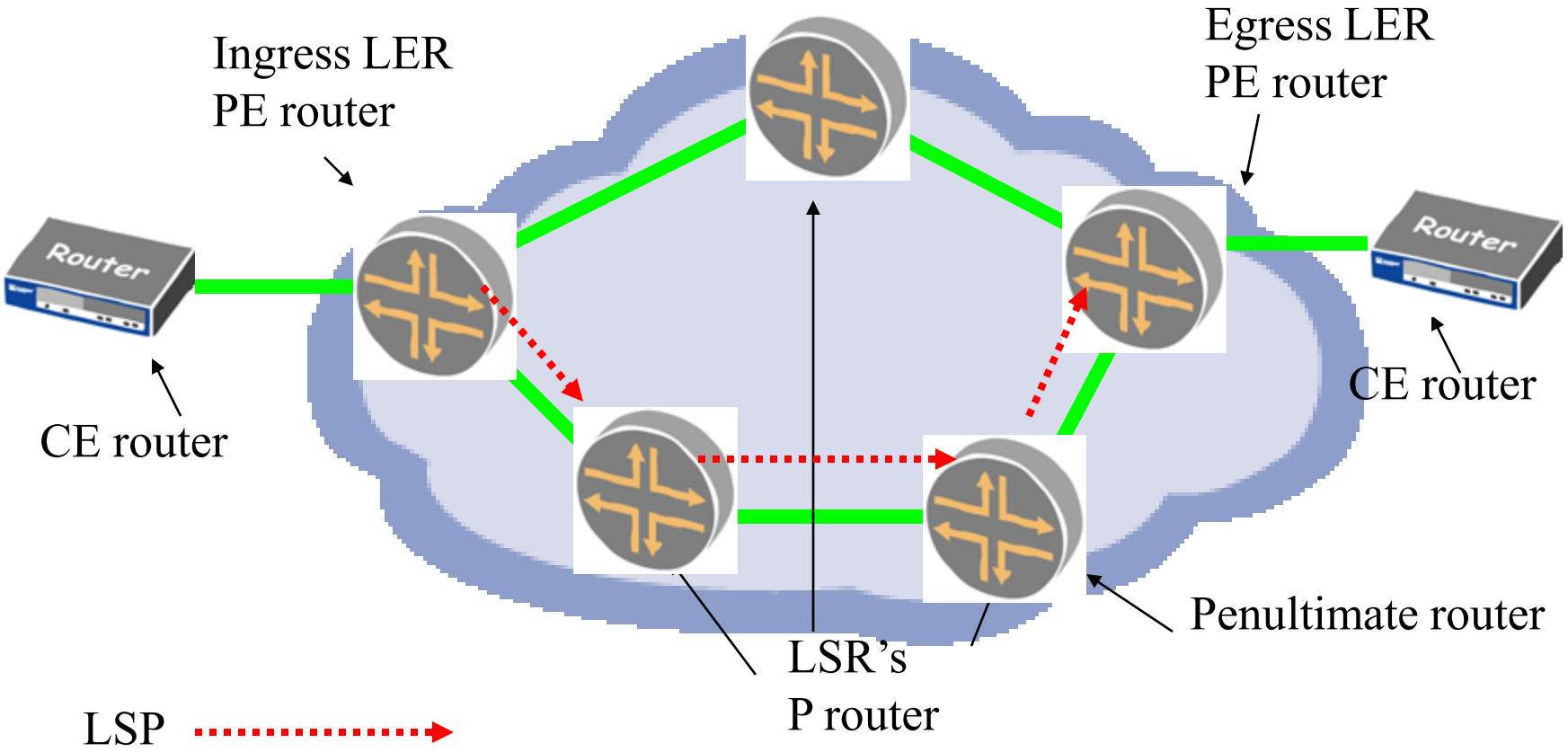
- MPLS is a tool to solve problems
  - Not everyone has the same problems or pain
- In other words reason to deploy (choose 1+)
  - Traffic Engineering
  - Traffic Protection
  - Provider provisioned VPN's
    - Layer 3 and/or Layer 2
- Or in other words
  - Save money
  - Make money

What is MPLS?

- Multiprotocol Label Switching
- Connection Orientated Virtual Circuits over IP imple with label switching
- Grew out of
  - Cisco's Tag switching
  - Ipsilon (Nokia) IP switching
  - IBM ARIS
  - 3Com's FAST IP
- Expanding areas of application
  - Cost savings
  - New services
- Promise of Multiprotocol Unification (Core NOT edge)
- Defined by RFC 3031, RFC 3032

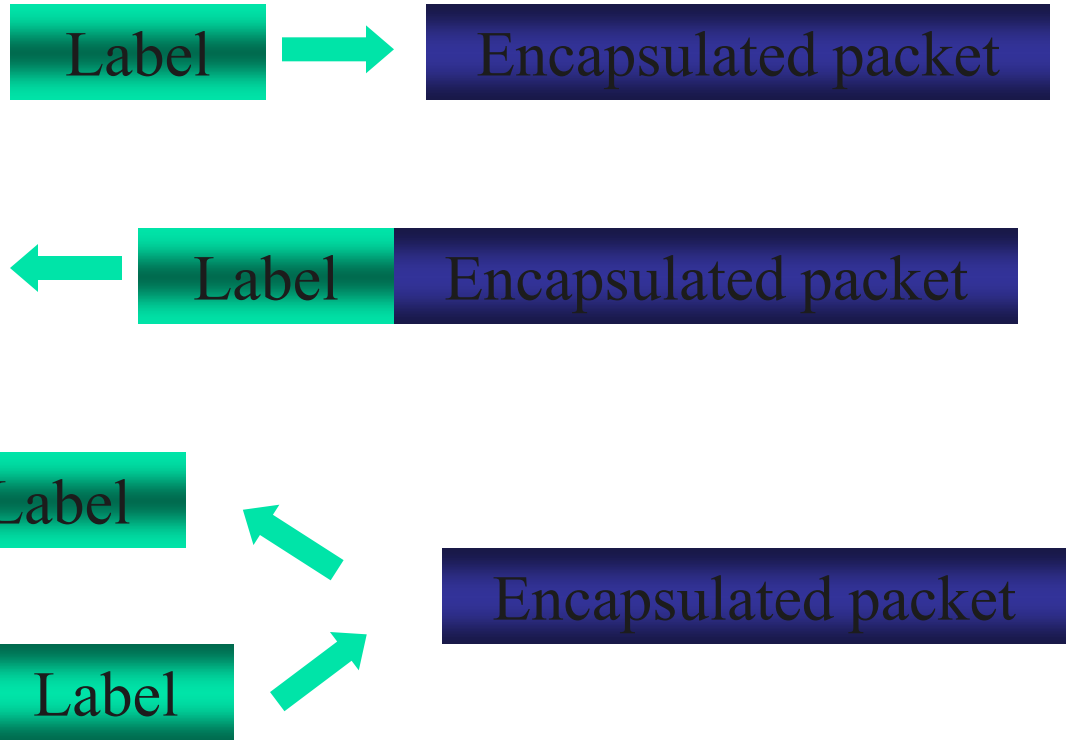
MPLS Terminology

- An LSP is a unidirectional flow of traffic

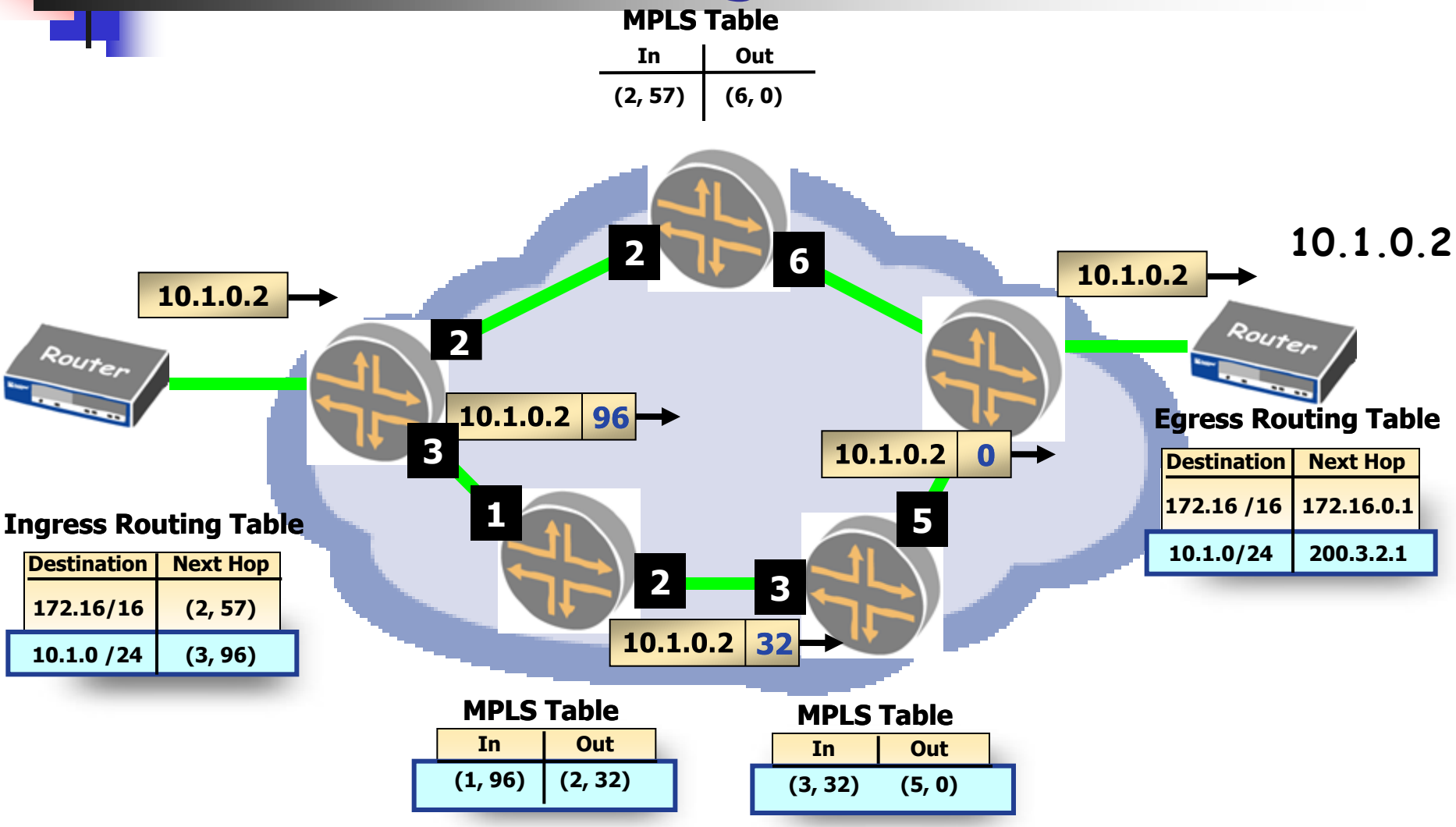


Push, Pop, Swap

- Push
- Pop
- Swap

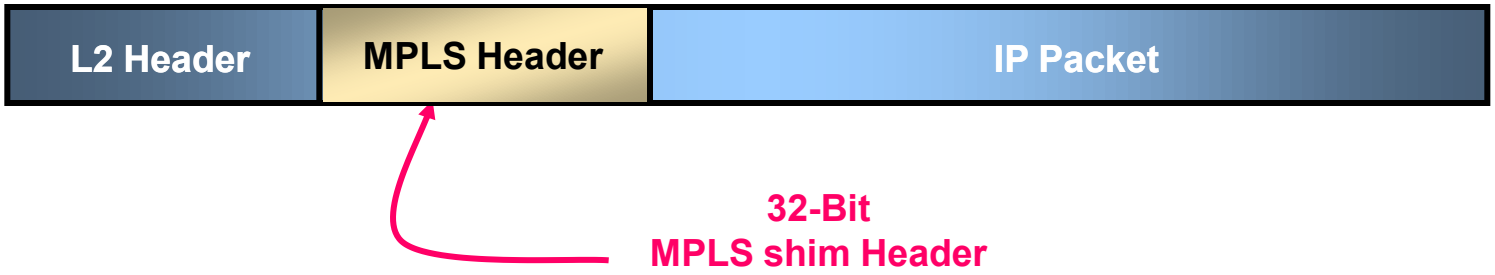


MPLS Forwarding Plane



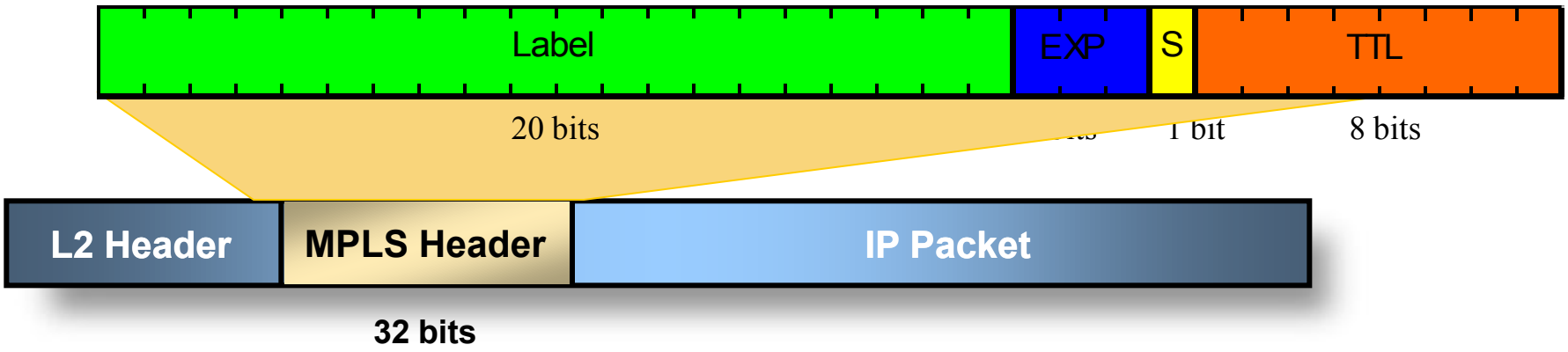
Labeled Packets

- MPLS header is prepended to packet with a *push* operation at ingress node
  - Label is added immediately after Layer 2 encapsulation header



- Packet is restored at the end of the LSP with a *pop* operation
  - Normally the label stack is popped at penultimate node

The Label



- Label
  - Used to identify virtual circuit
- EXP
  - Experimental. Currently this is used to identify class of servce (CoS)
- S (Stack Bit)
  - Used to indicate if there is another label inside this packet or is it the original encapsulated data
- TTL
  - Time to live, functionally equivalent to IP TTL.

Example - Ethernet

0 0 1 0 1 1 1 1 0 1 0 0 0 1 0 1 1

My Web Page

TCP | port = 80 (www)

IP Header | Protocol = TCP

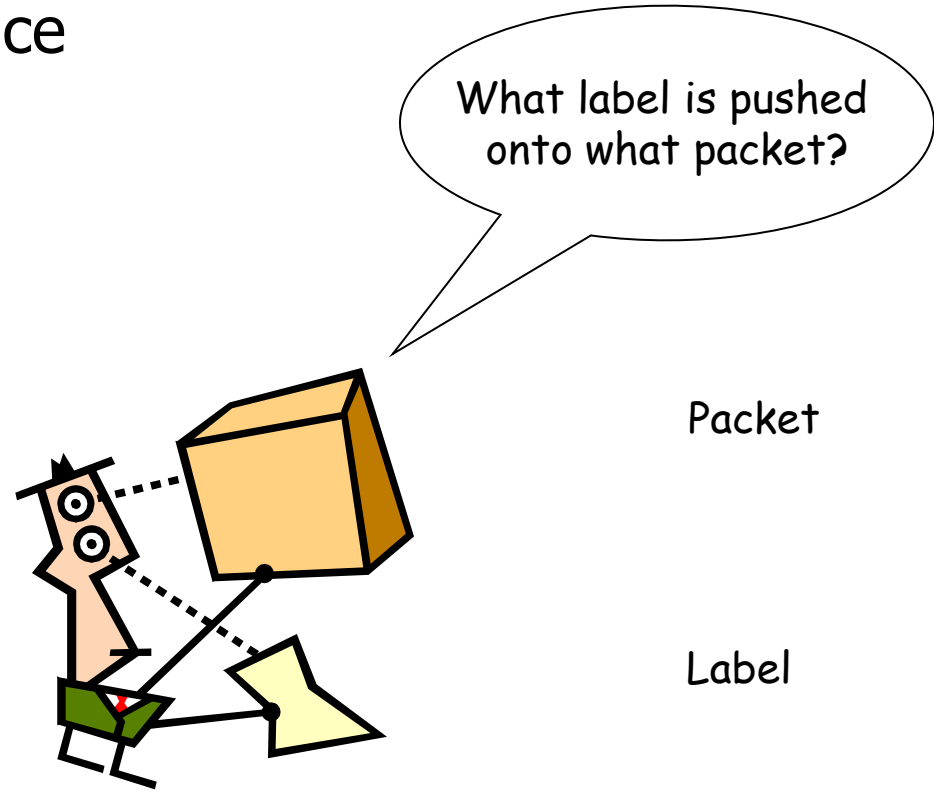
Label = 23 | EXP = BE | S = 0 | TTL = 254

Label = 47 | EXP = BE | S = 1 | TTL = 240

Dest. MAC Src. MAC Type = 8347

FEC – Forwarding Equivalency class

- All traffic with the same FEC will follow the same path and experience same level of service
- E.g. of FEC
  - Destination IP address
  - BGP next hop
  - VPN membership
  - Source address
  - Any combination of above



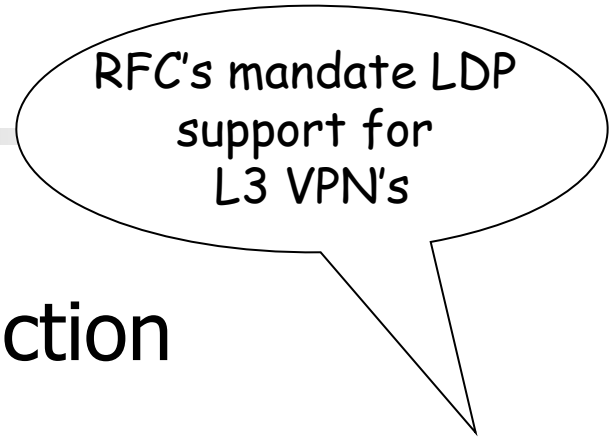
Signaling

- Protocols that are used to setup maintain and tear down LSP's.
- Can behave differently depending on function
- Let's describe a language / concepts to understand these differences in operati

Signalling Protocols

- LDP
  - Label Distribution Protocol
- RSVP-TE
  - Resource Reservation Protocol with Traffic Engineering Extensions
- MBGP
  - Multi-protocol BGP

Which to choose…

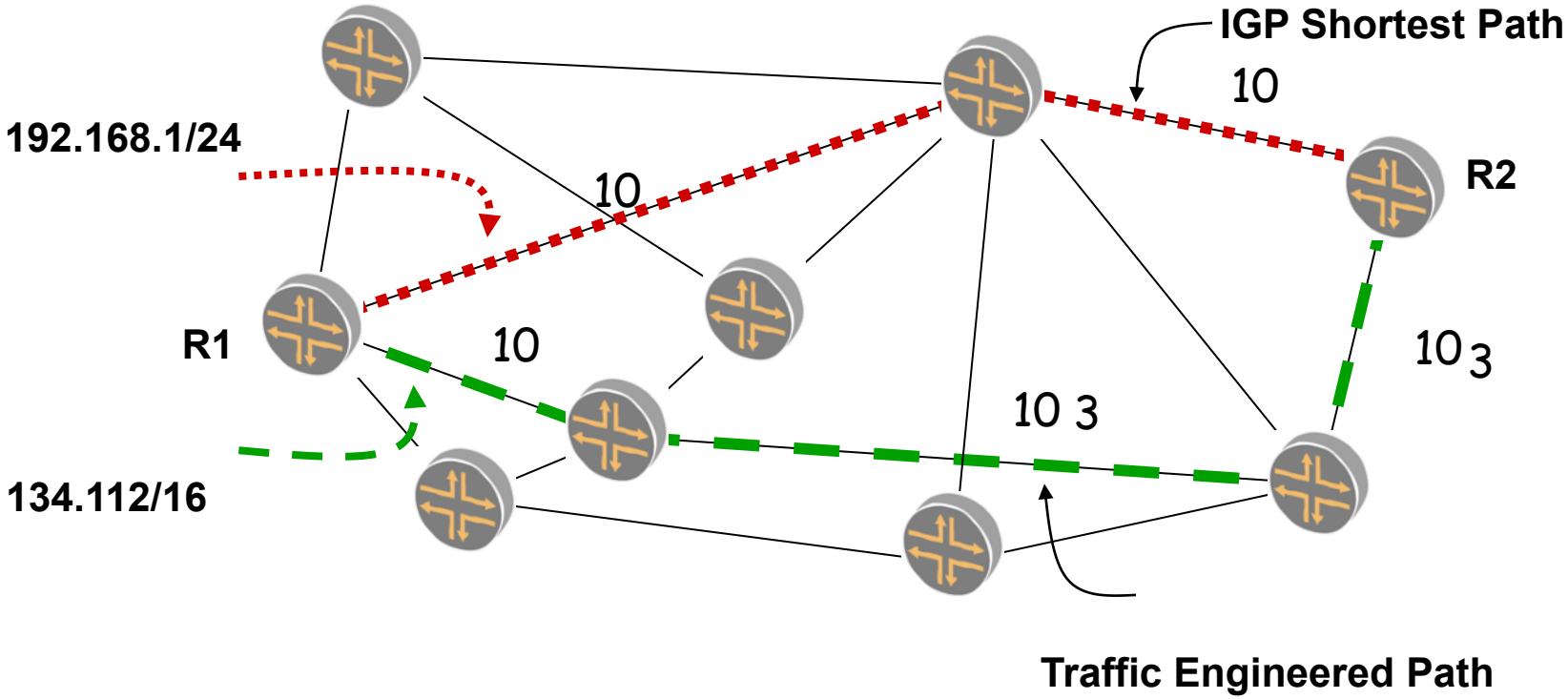


- Traffic Engineering, Traffic Protection
  - RSVP
  - Link State protocol
- VPN's
  - LDP or RSVP (all LSR's)
  - MBGP (PE's only)
- Why use LDP at all?
  - Configuration scaling
  - LDP configuration is "per box"
  - RSVP configuration is "per LSP"

Traffic Engineering Defined

- Sub Optimal routing
- Network Engineering is putting bandwidth where the traffic is. Traffic Engineering is putting the traffic where the bandwidth is!
- To meet one of two requirements
  - To better utilize network capacity and resources.
  - To put traffic on a path that can support it's requirements
- Incorporate Traffic Protection to achieve SONET like failure recovery.

MPLS-Based Traffic Engineering



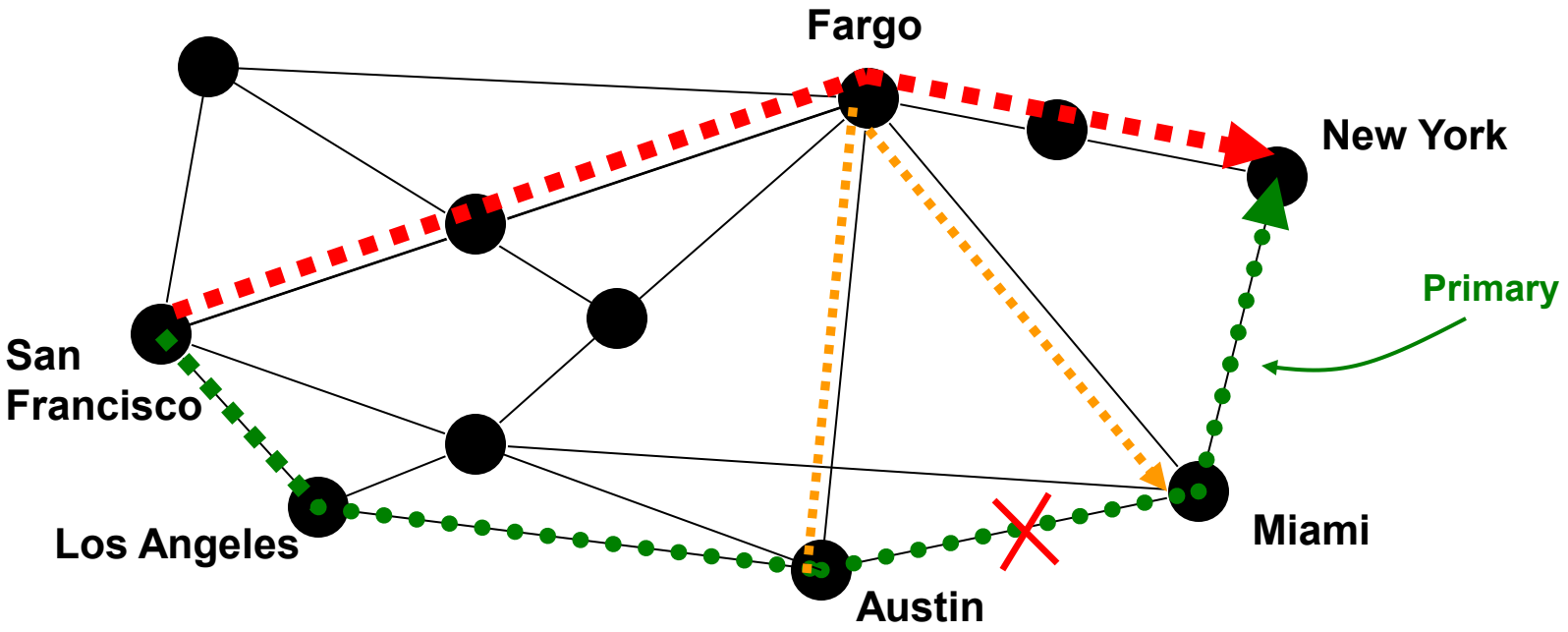
Traffic Engineering Options

- Can we do this another way
  - IGP metrics ☹
  - Flow = all traffic with same destination
- MPLS because
  - Granularity of flows
  - Flow = all traffic with same FEC
  - One network for all services
  - Less expensive

Traffic Protection

- Working definition
  - Reduce time of disruption
  - Reduce Packet Loss
  - "SONET like" sub millisecond recovery under failure conditions
- Can we do this another way
  - SONET/SDH
  - Lower IGP timers
- MPLS because
  - No extra capital – config change only
  - Pick which traffic needs it
  - One network for all services
  - Less expensive

Traffic Protection – example



Traffic Protection Variations

- Fast reroute
- Link Protection
- Link-Node Protection

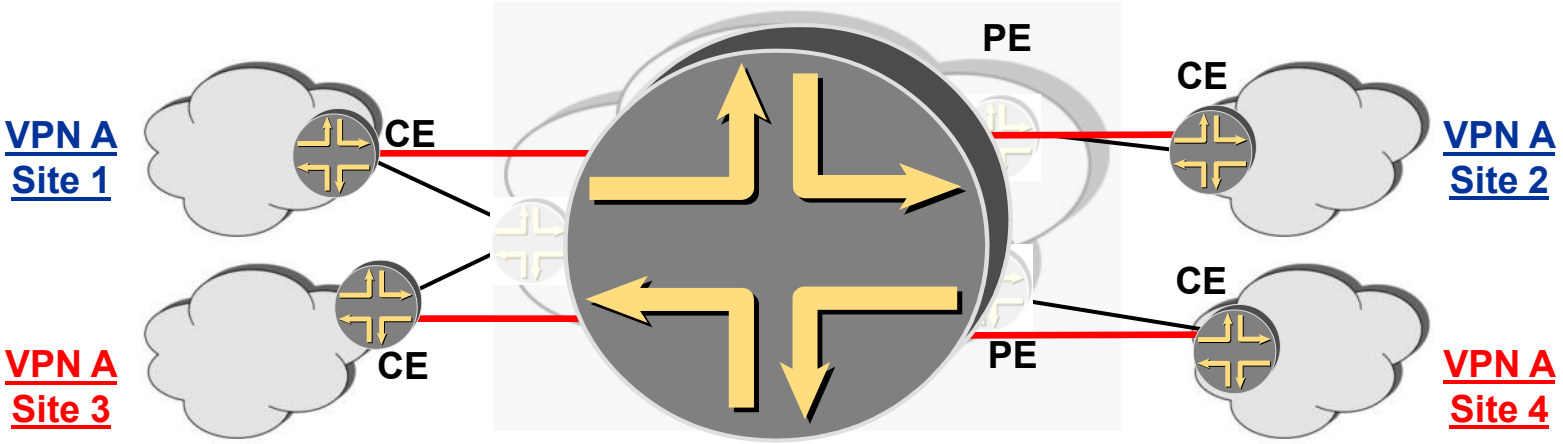
Layer 3 VPN (4364 BGP/MPLS VPN)

Provider provisioned VPN

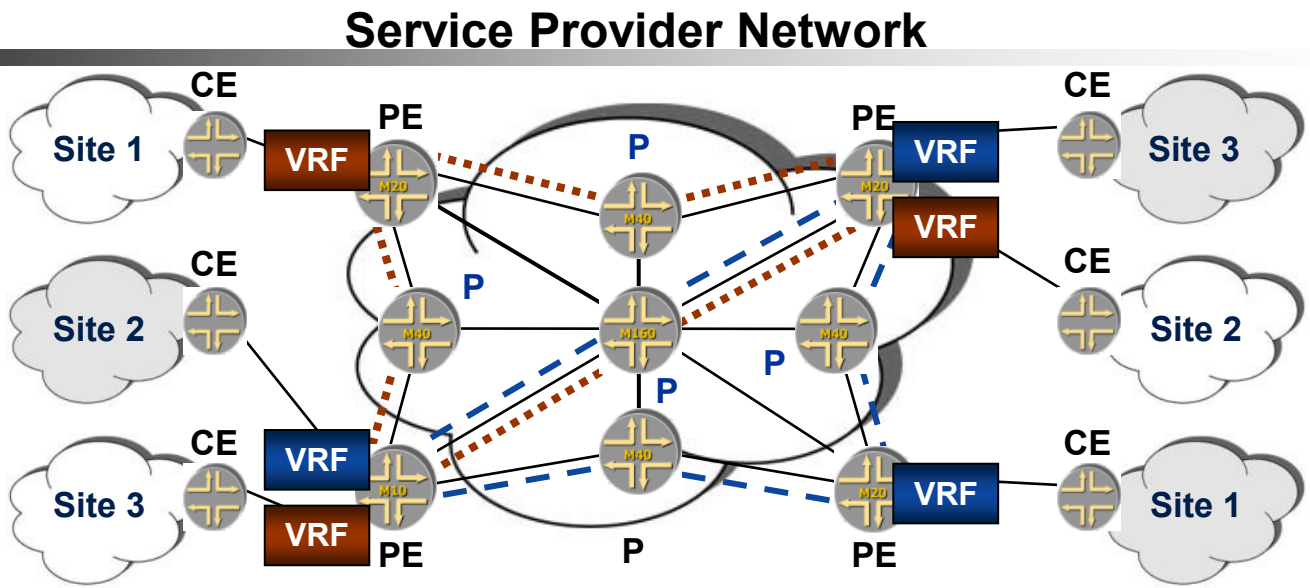
- ISP runs backbone for customer
  - Customer can be another ISP!
- Attractive to
  - Customer who do not want to run their own backbone
- Not attractive to
  - Customer who doesn't trust carrier
  - Customers who's jobs are threatened

Customer View of L3VPN

- Make the cloud look like a router
- Single site provisioning



Layer 3 PP-VPNs: RFC 4364 (1 of 2)



- Application: Outsource VPN
  - PE router maintains VPN-specific forwarding tables for each of its directly connected VPNs
  - Conventional IP routing between CE and PE routers
  - VPN routes distributed using MP-BGP
    - Uses extended communities
  - VPN traffic forwarded across provider backbone using MPLS

Layer 3 PP-VPNs: RFC 4364 (2 of 2)

- LDP or RSVP is used to set up PE-to-PE LSPs
- MP-BGP is used to distribute information about the VPN
  - Routing and reachability for the VPN
  - Labels for customer sites (tunneled in PE-PE LSP)
- Constrain connectivity by route filtering
  - Flexible, policy-based control mechanism

L3 VPN Options

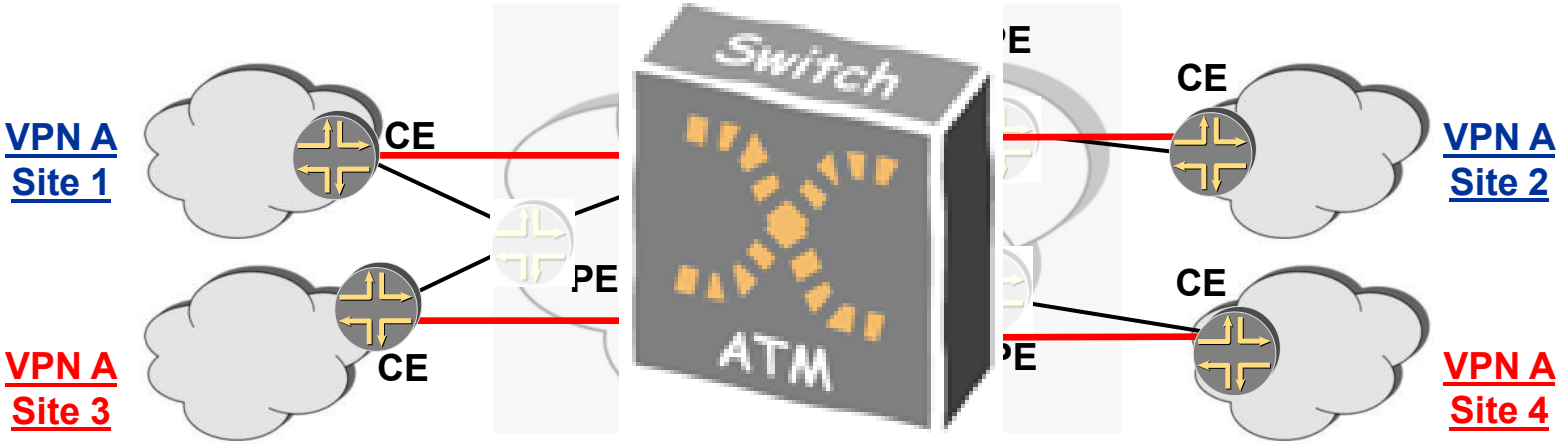
- Can we do it another way
  - Separate Physical routers
  - Separate Logical Routers
- MPLS because
  - Scaling
  - Single site provisioning
  - Less expensive

Layer 2 VPN's

- Provider provisioned VPN
  - ISP runs backbone for customer
    - Customer can be another ISP!
- Attractive to
  - Customers who want to preserve current CE technology
  - Customers who don't trust provider with L3
  - Carriers who want to offer another service
- Not Attractive to
  - Customers who do not want to run their own backbone

Customer View of L2VPN

- Make the cloud look like a ATM/FR network



L2 VPN Options

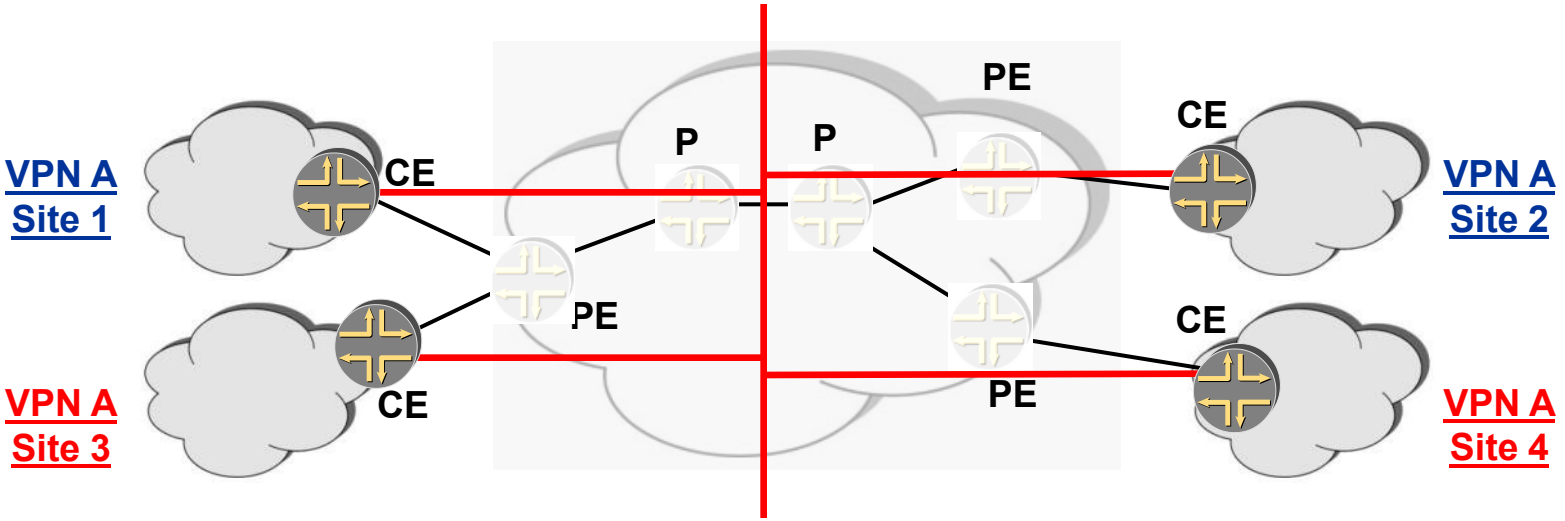
- Can we do it another way?
  - Traditional ATM/FR/leased line infrastructure
- MPLS because
  - One network for all services
  - Less expensive
  - Scaling
  - Single site provisioning *

VPLS

- Virtual Private LAN Service
- Attractive to
  - Customers who like ethernet as CE
  - Lots of locations close together with 'high' WAN bandwidth requirements (kiosks)
  - No routing required
- Not attractive to
  - Customers who like control and visibility of core. "what can I ping to identify fault-domain?"
  - Controlling broadcasts

VPLS

- Make the cloud look like an ethernet switch

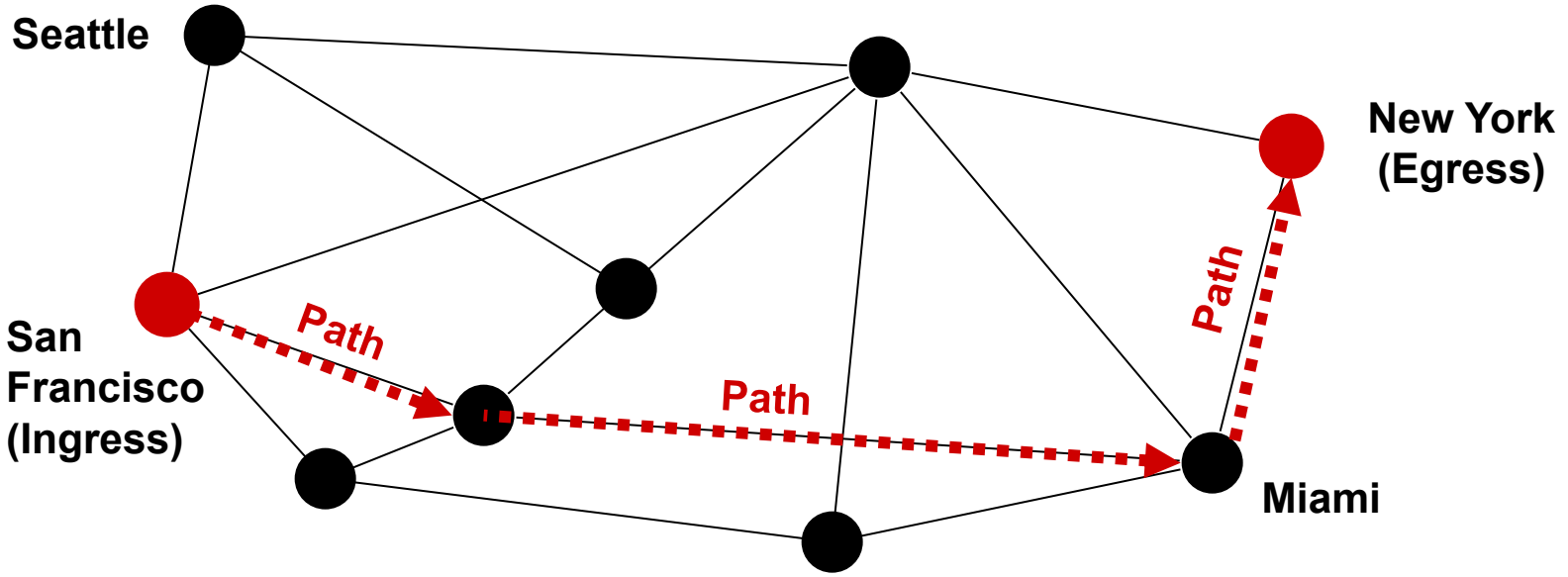


VPLS Options

- Can we do it another way?
  - Separate physical switches tying all customer sites
  - VLAN's over layer 2 backbone
- MPLS because
  - Scaling
  - One network for all services
  - Less expensive

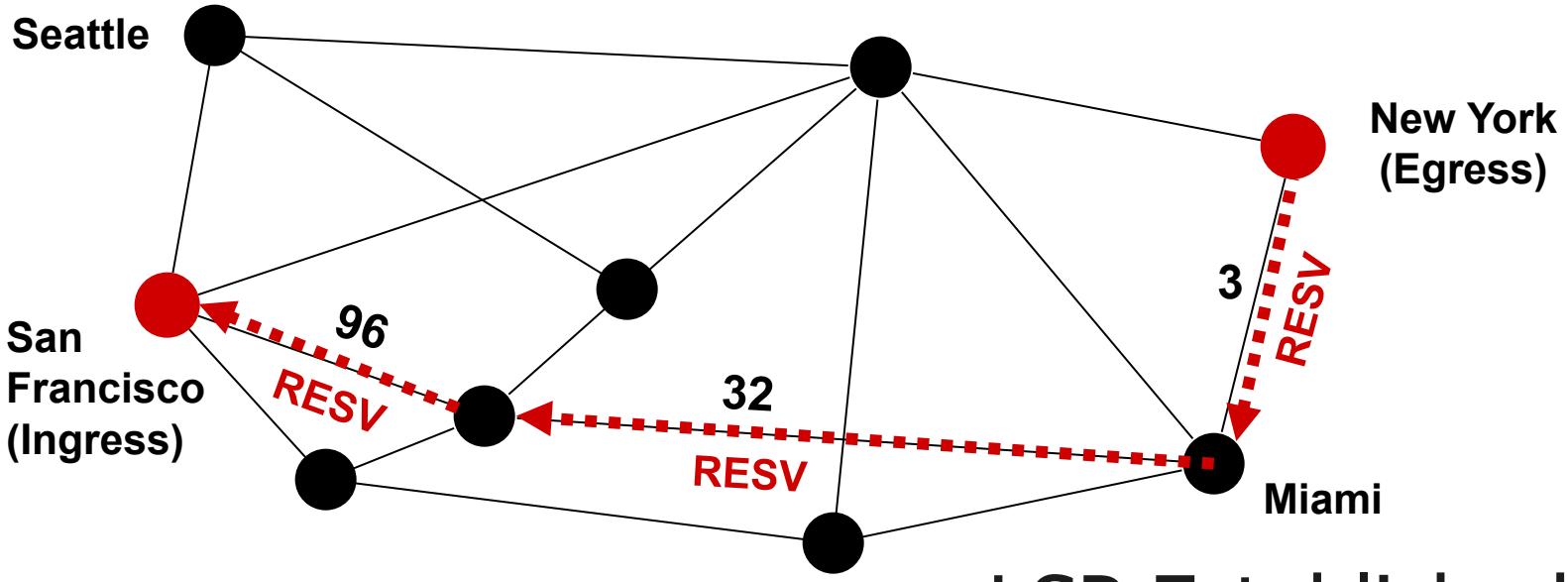
RSVP Signaling Example: Path

RSVP sets up path from San Francisco to New York



RSVP Signaling Example: Reservation

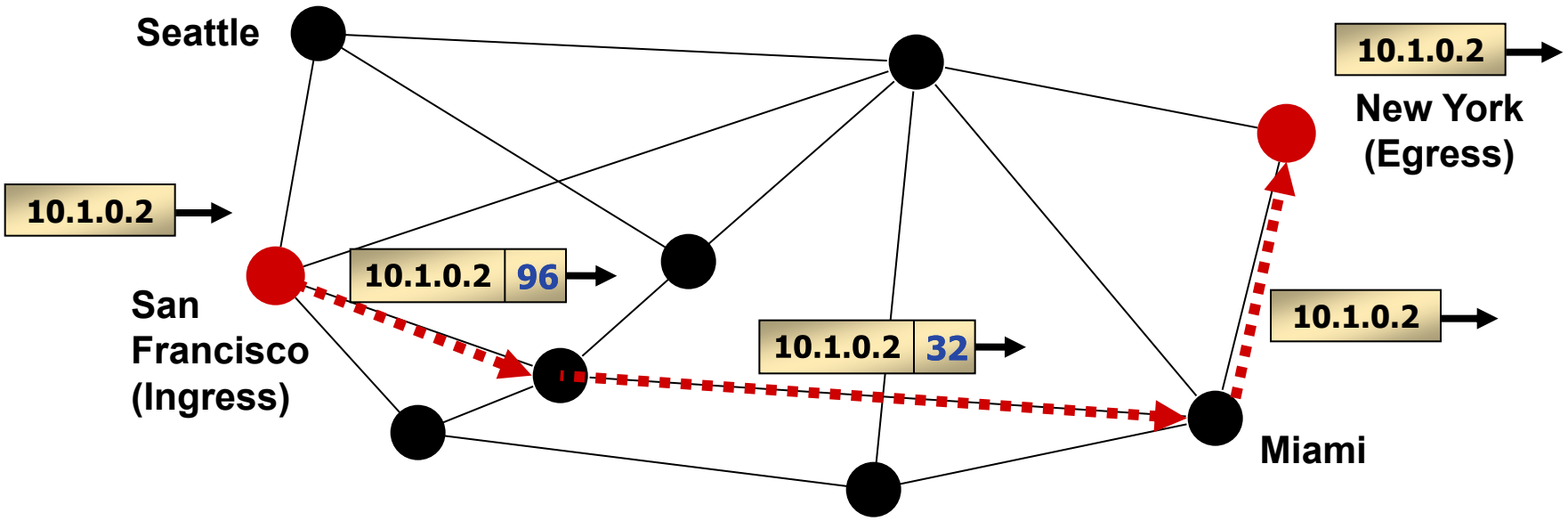
- The resv message visits each router on the path in reverse order
  - Labels assigned hop to hop in the upstream direction



LSP Established!

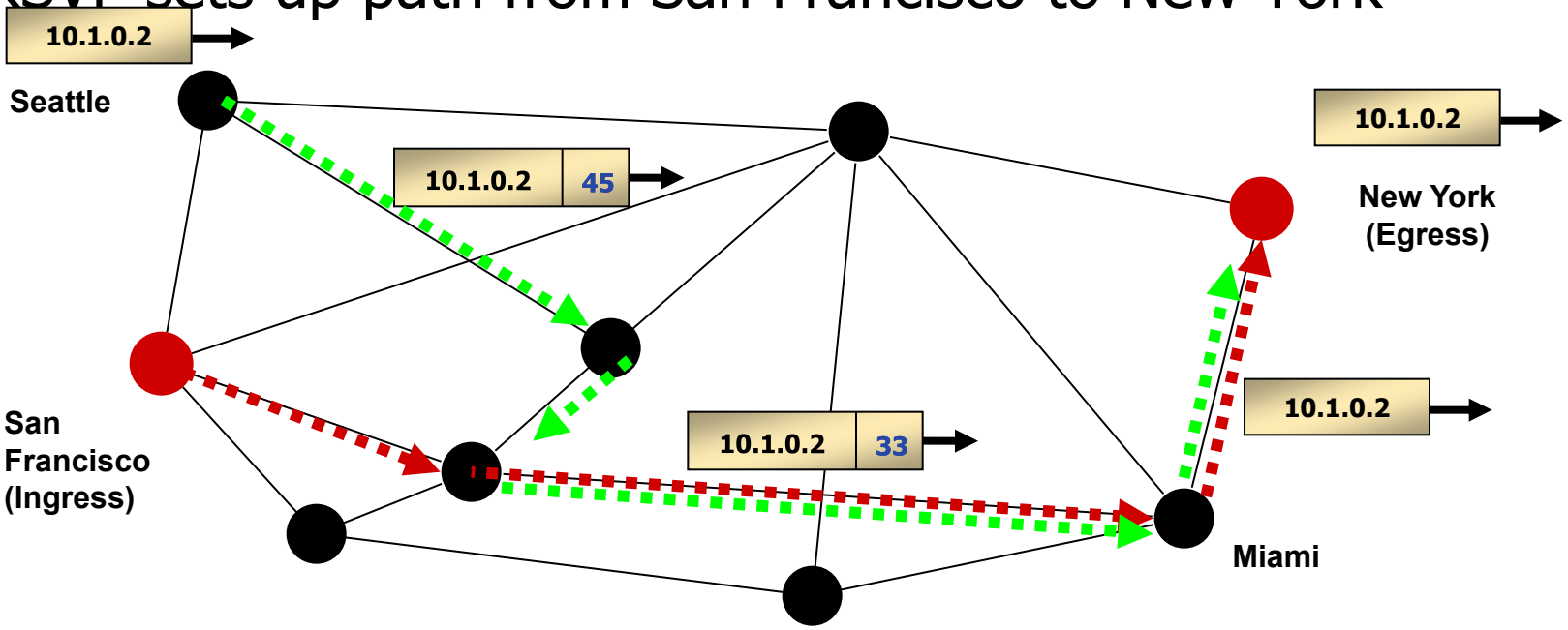
RSVP Signaling Example: Forwarding

RSVP sets up path from San Francisco to New York



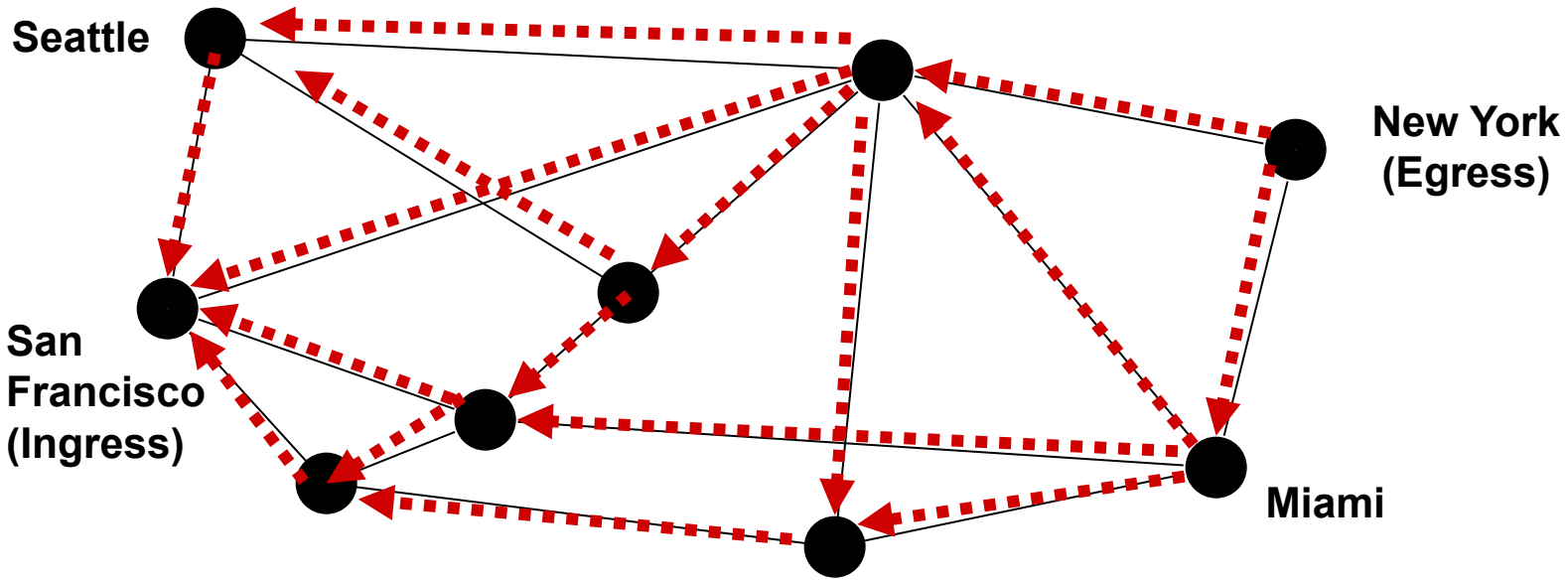
RSVP Signaling Example: Forwarding 2

RSVP sets up path from San Francisco to New York



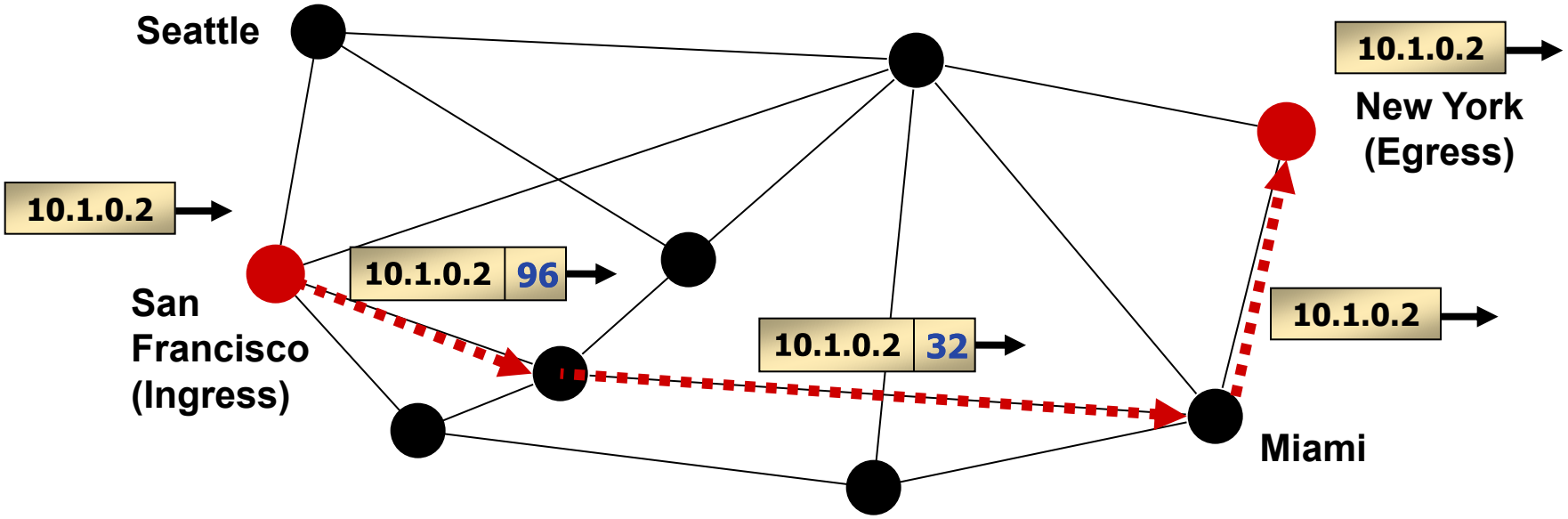
LDP Signaling Example: Label Binding

- Label Mappings are made for entries in the routing table
  - Labels assigned hop to hop in the upstream direction



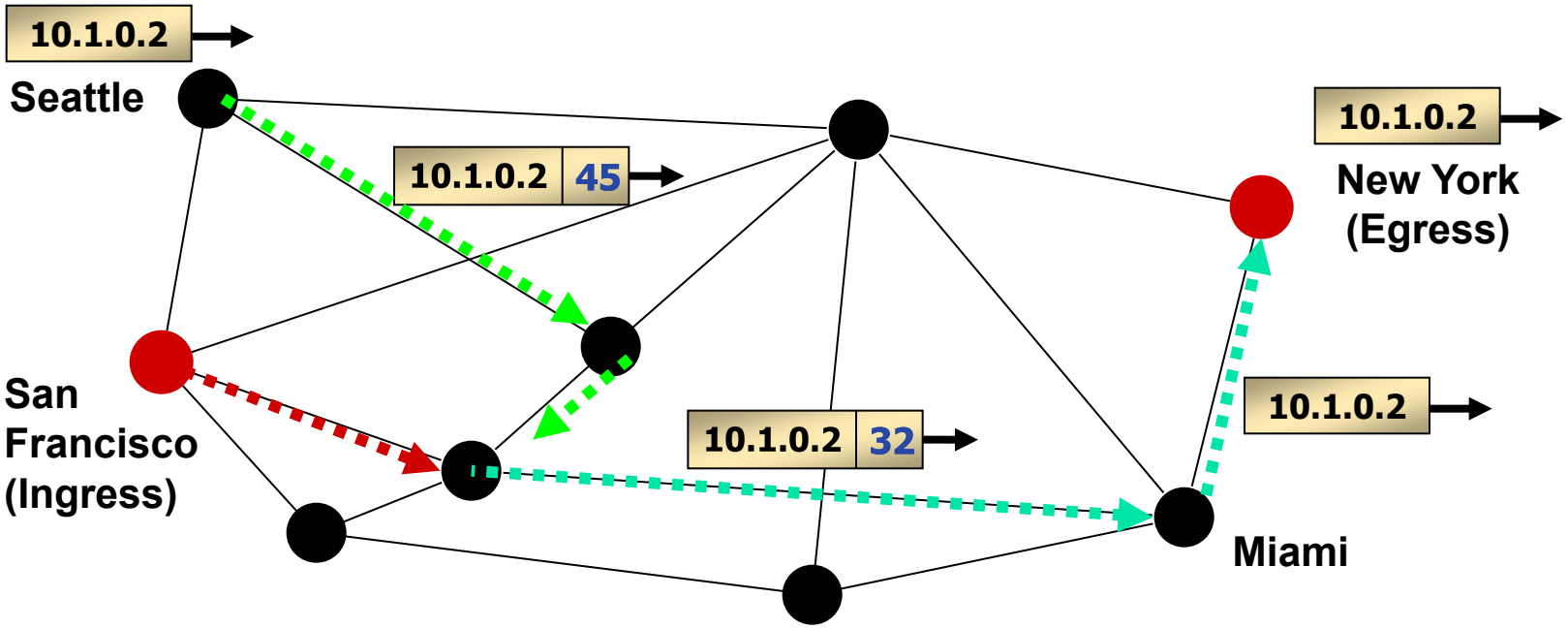
LDP Signaling Example: Forwarding

LDP path available to egress



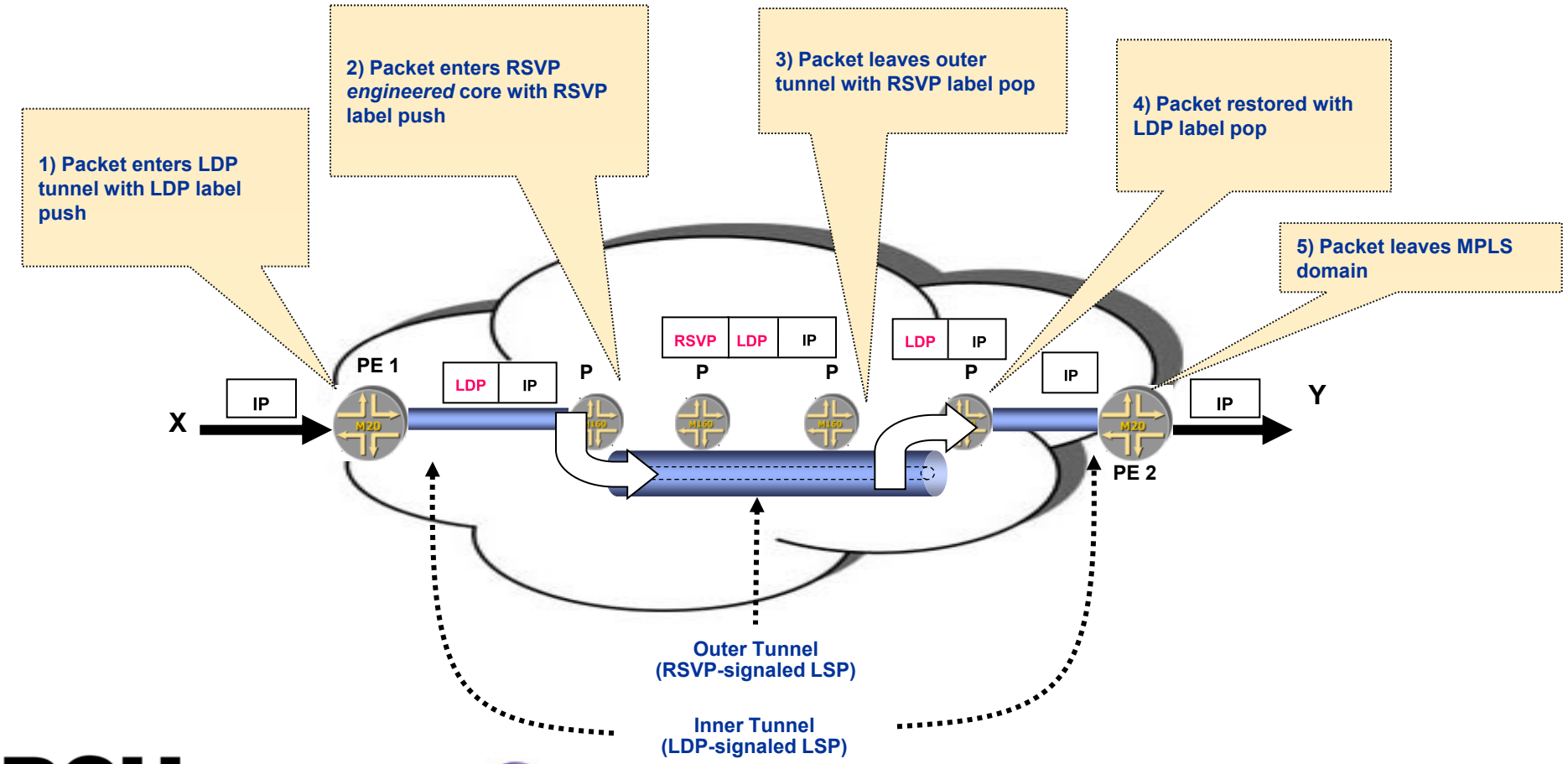
LDP Signaling Example:Forwarding 2

LSP Merging occurs



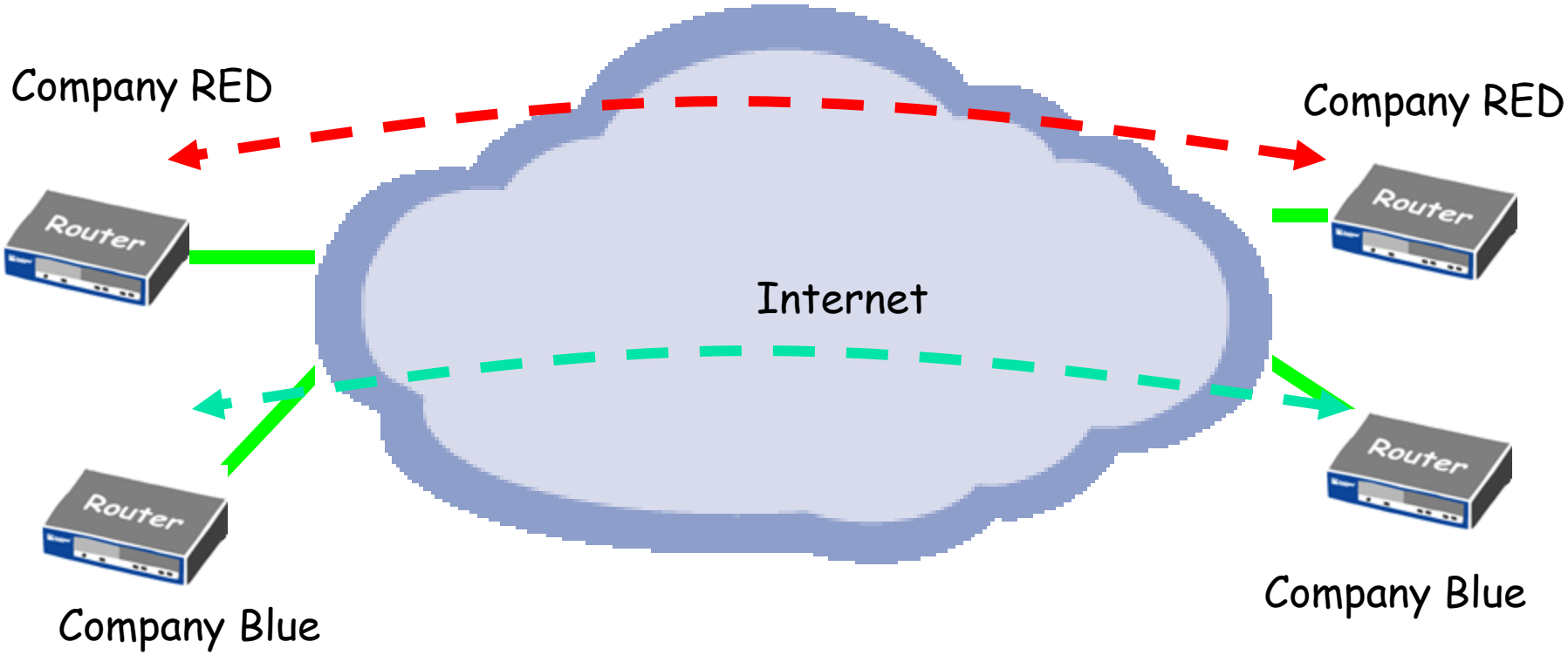
Label Stacking

- Label stacking improves scalability
  - Similar to ATM's VP and VC hierarchy



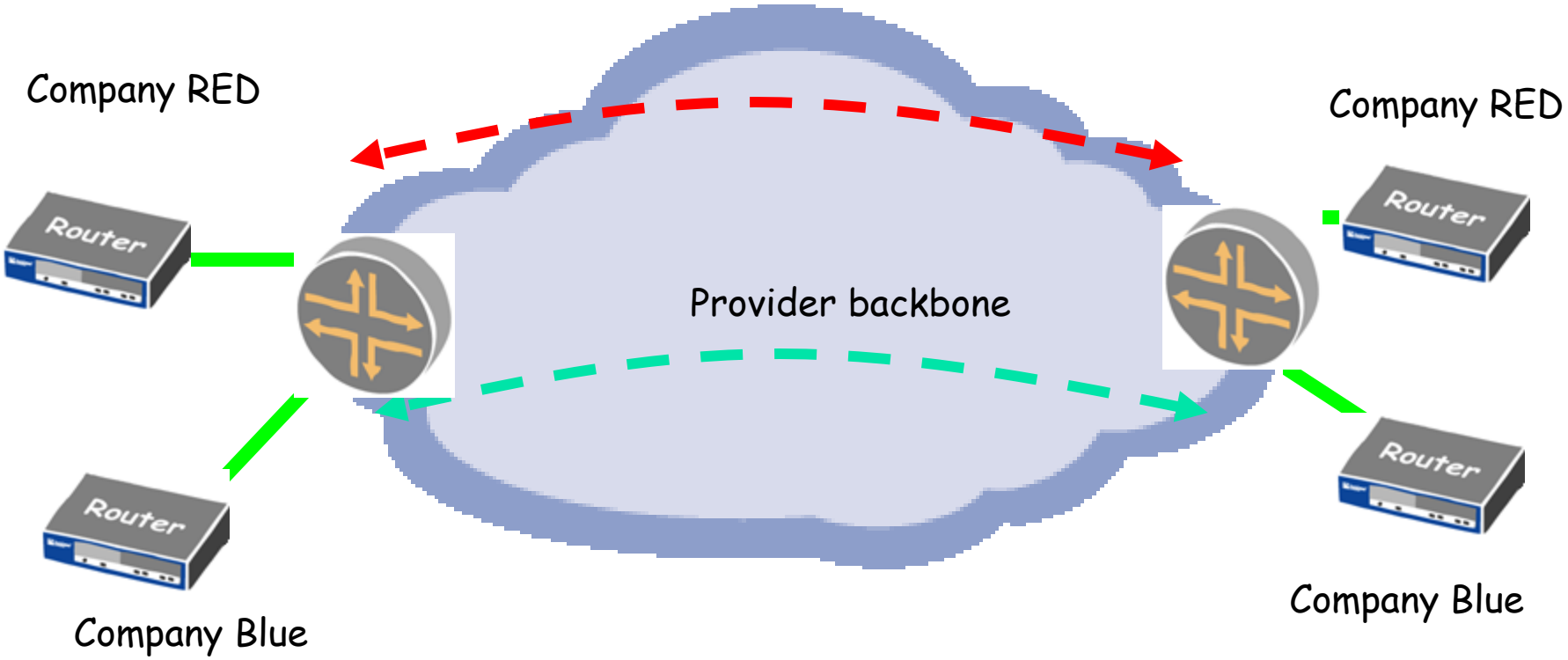
Traditional VPN's

- CPE based
- Customer controlled
- No value add for provider



Provider provisioned VPN's - PPVPN

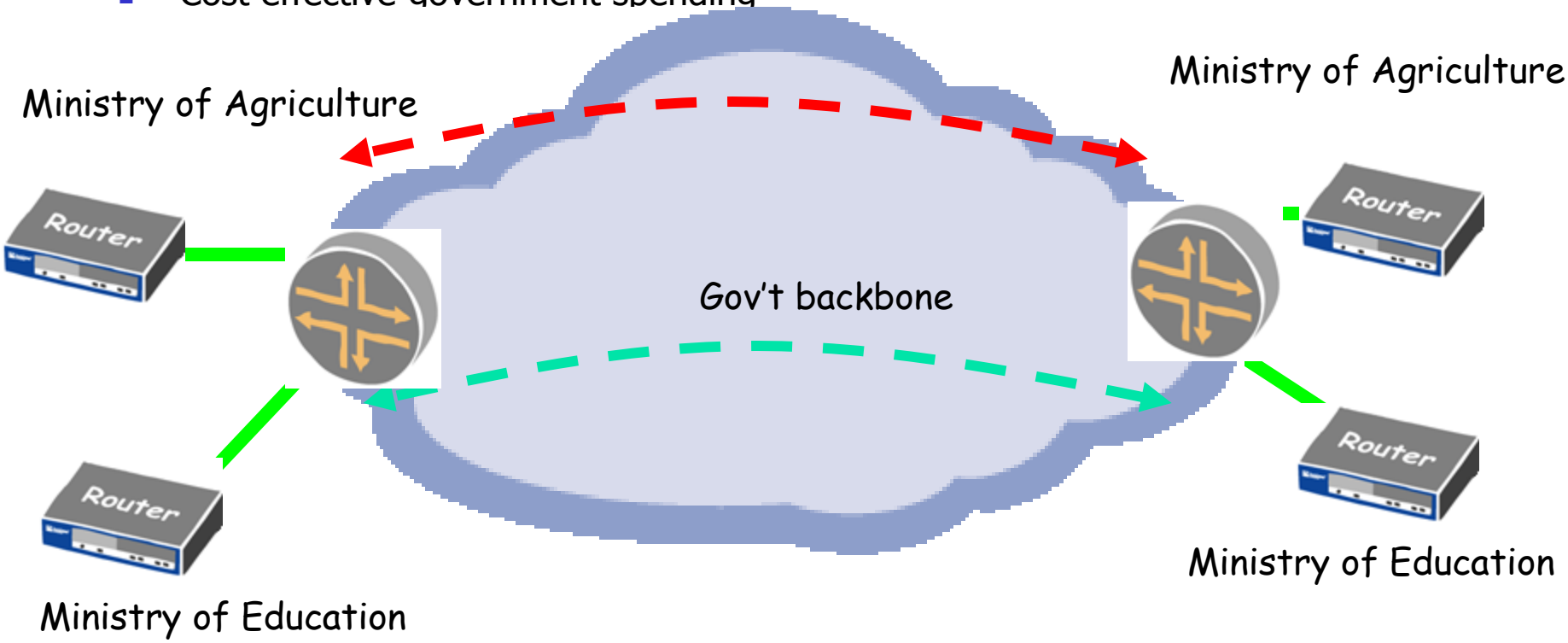
- PE based
- Customer outsource backbone
- Value add for provider
- Single Site Provisioning (BGP, + Route refresh + Route Target Filtering)



Sharing Network backbones

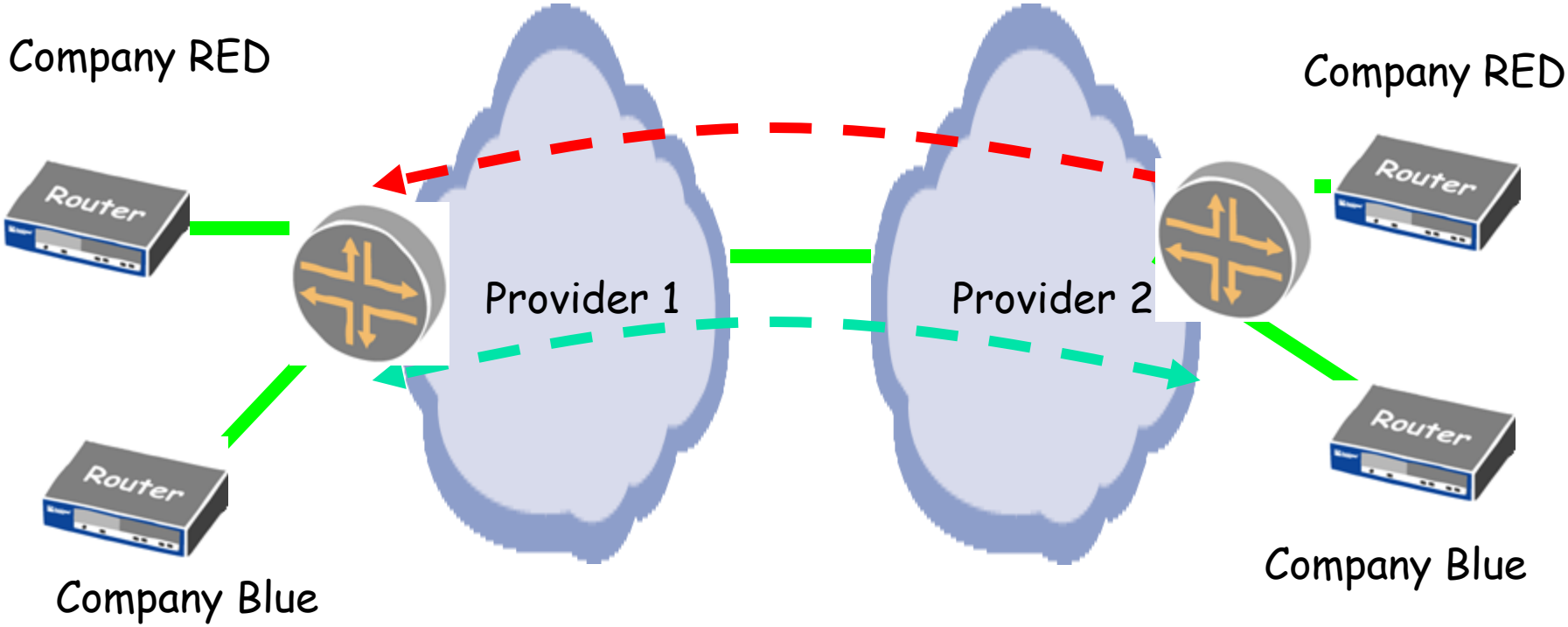
- Infrastructure built by one department
- Shared by other departments
- Cost effective government spending

- Examples
  - Gov't backbones
  - Industry Aligned



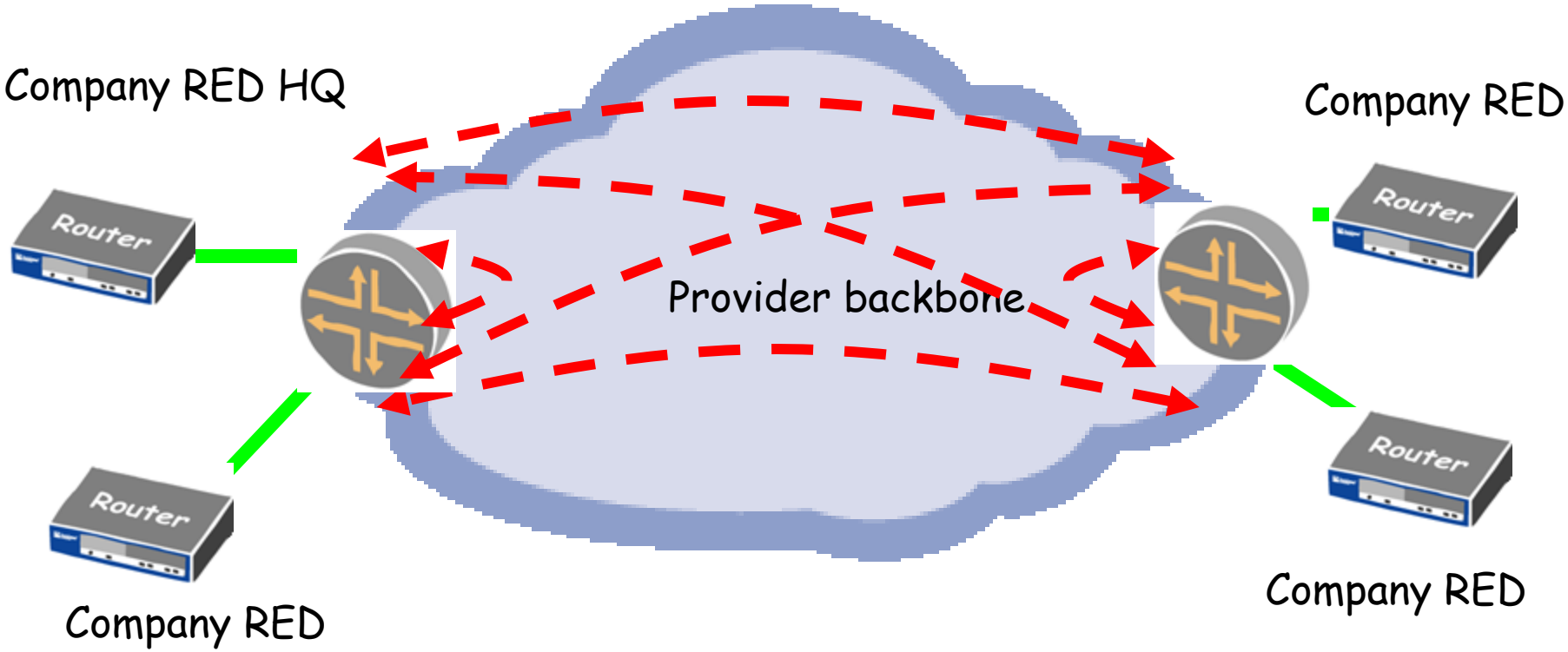
InterAS VPN's

- Requires Co-operation
- Opportunity for global coverage



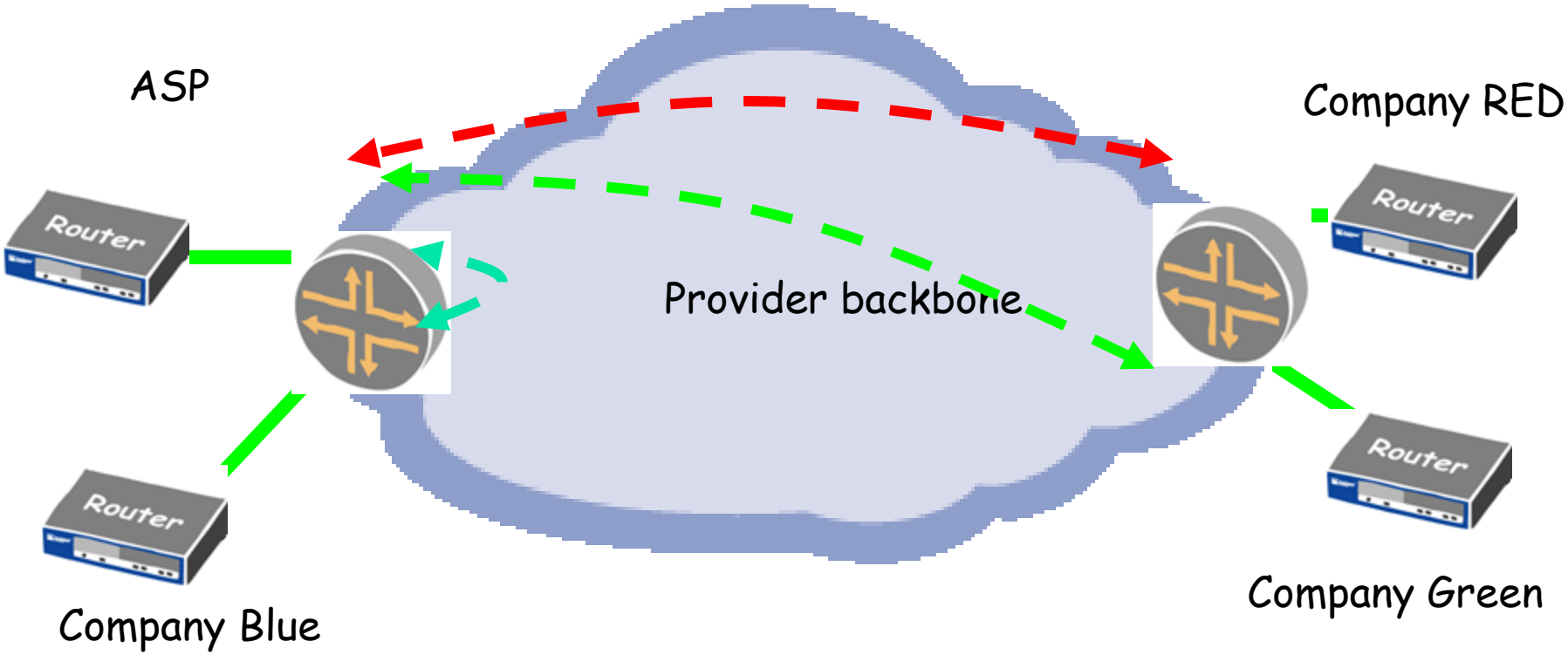
Site Connectivity

- Partial or Full Mesh is supported
- Full Mesh is more cost effective and competitive with traditional solutions



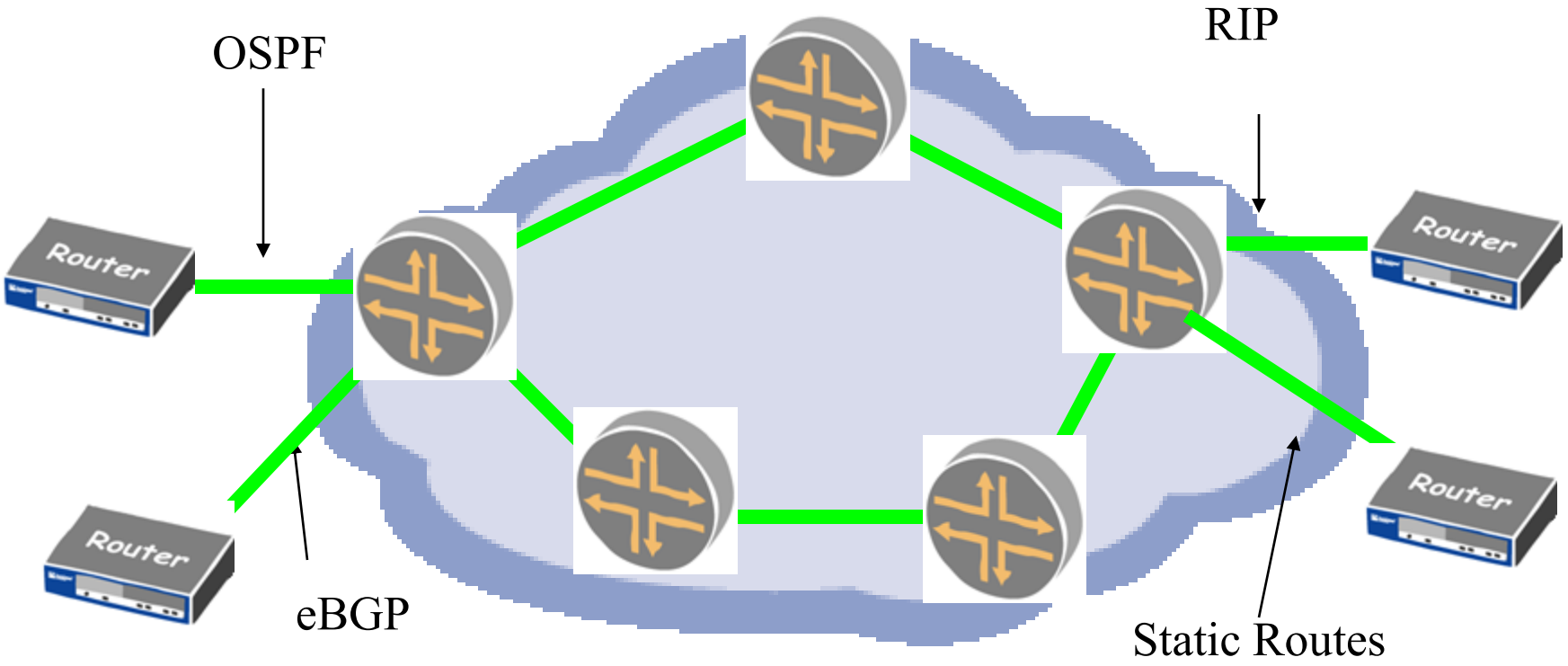
Overlapping VPN's

- Suites application / service providers



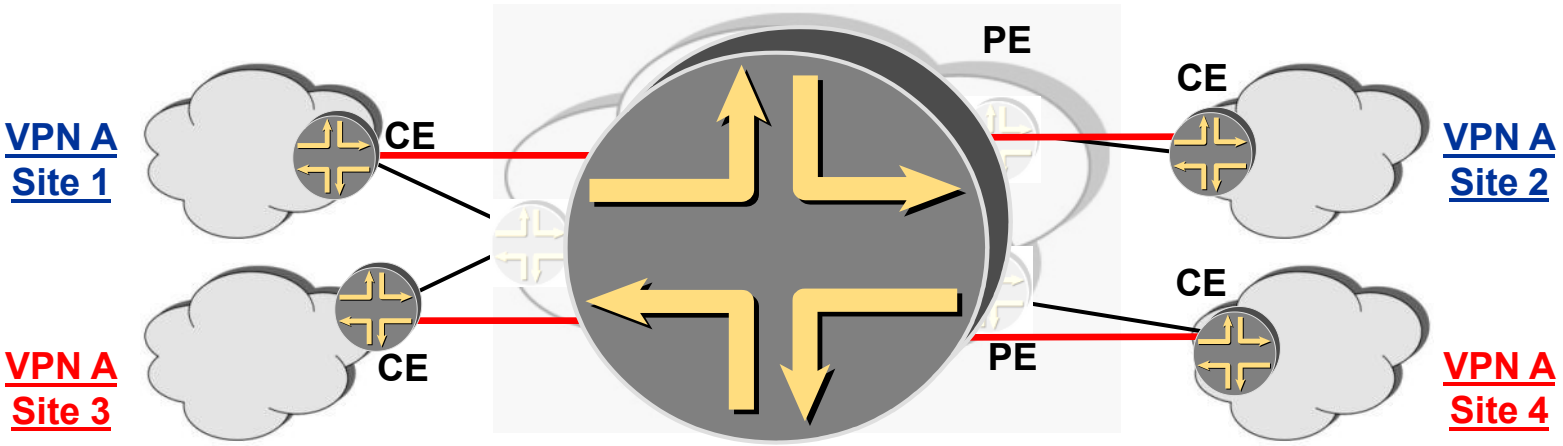
CE-PE interaction

- Any L2 connection, Any routing protocol
- CE peers at layer 3 with PE



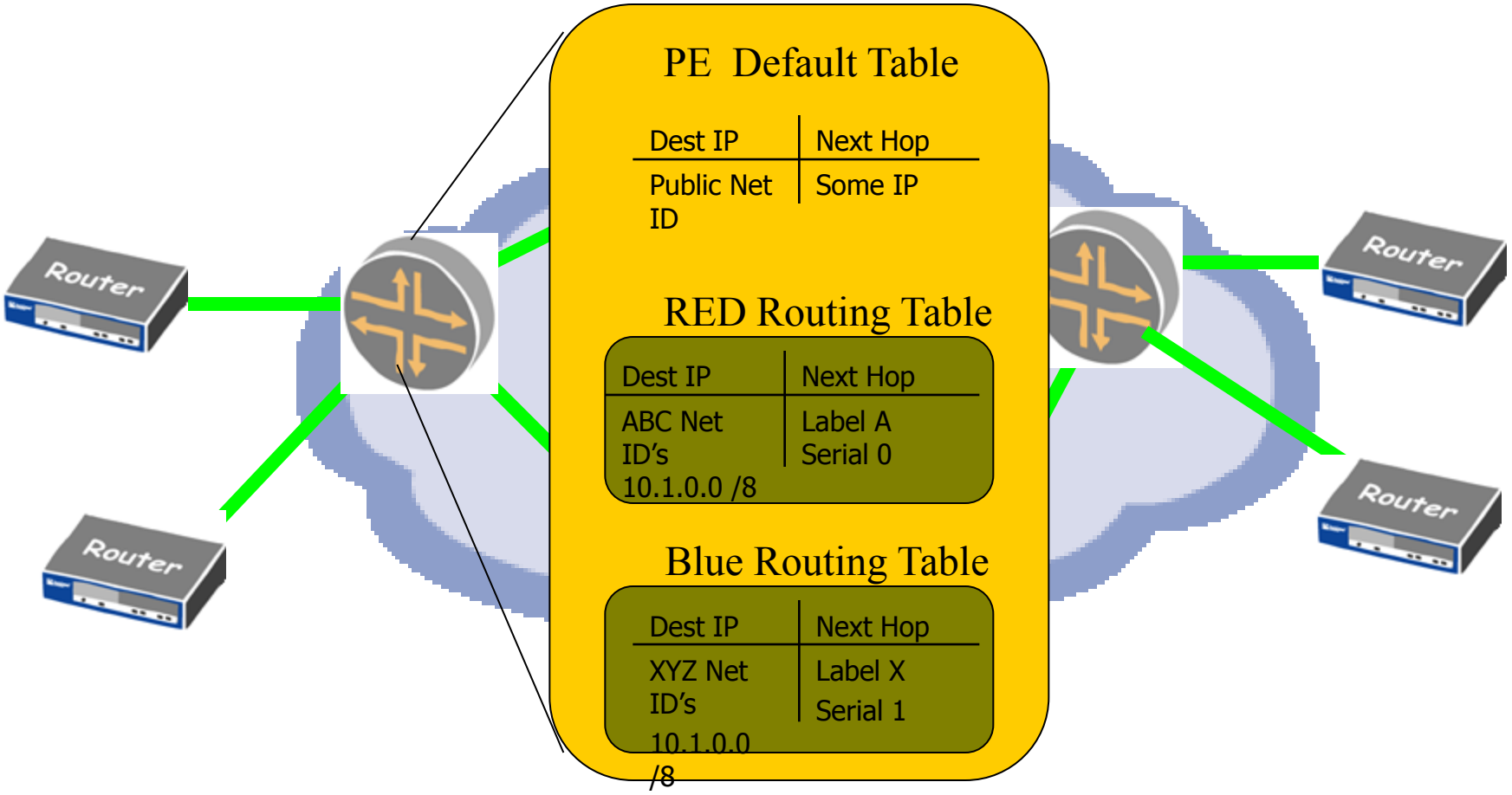
Customer View of L3VPN

- Make the cloud look like a router
- Single site provisioning



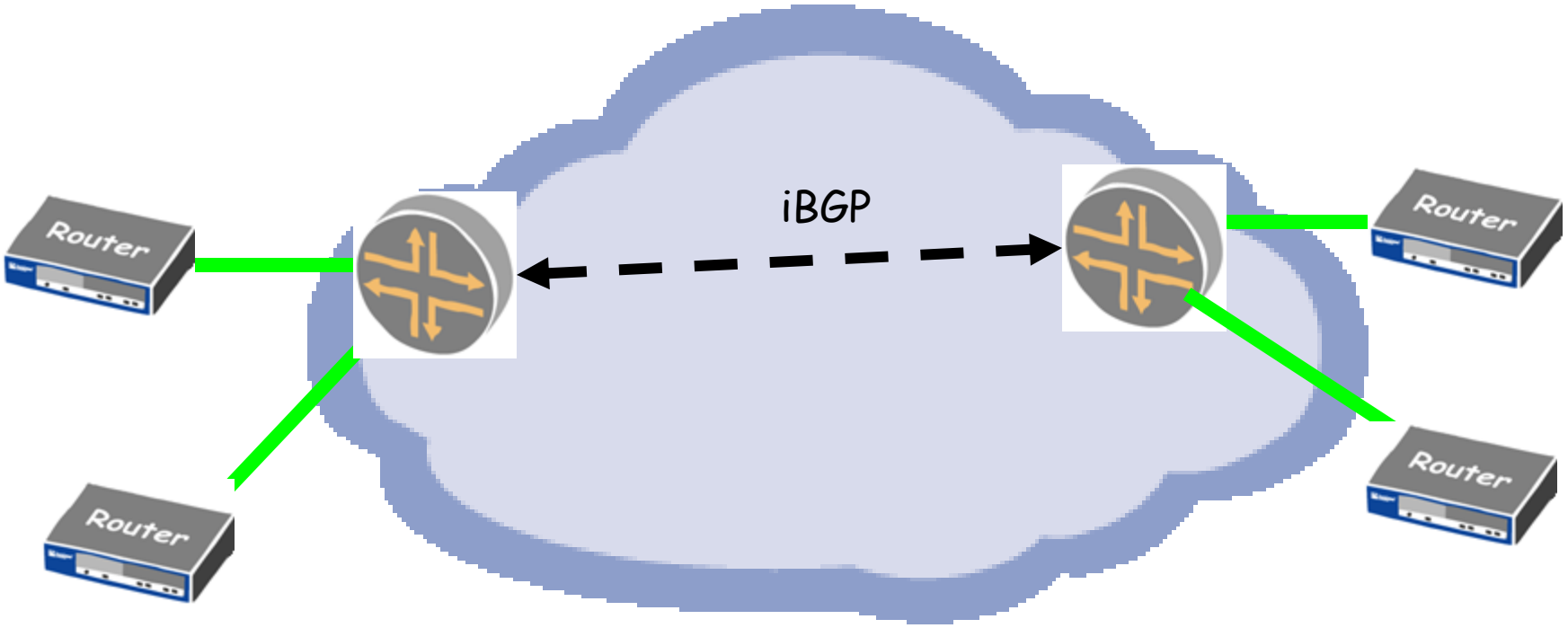
VRF – Virtual Routing and Forwarding instance

- VRF per VPN on PE
- Logical Interface packet arrives on defines the VRF used



PE-PE interaction

- iBGP between PE's carries routing information
- Assigns label per VPN

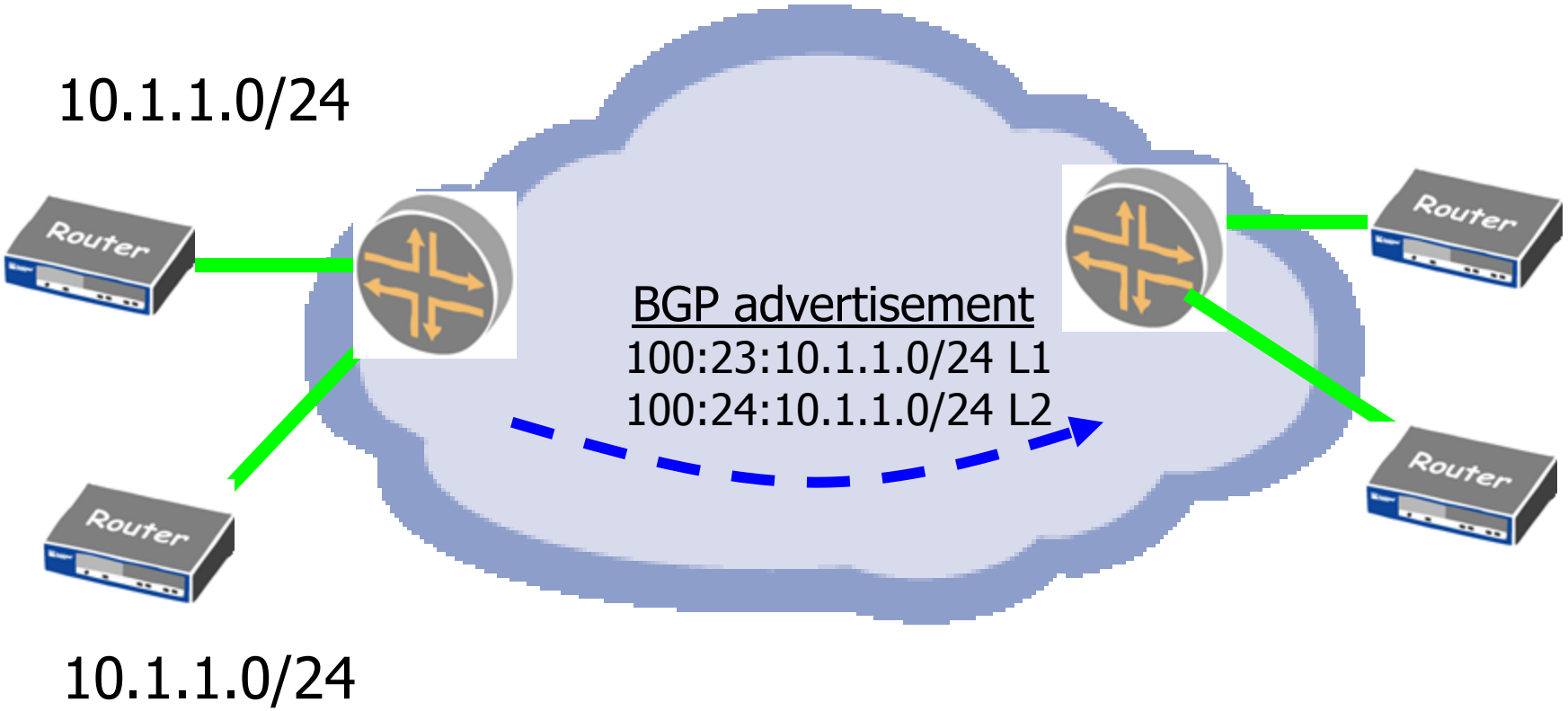


Route Distinguishers

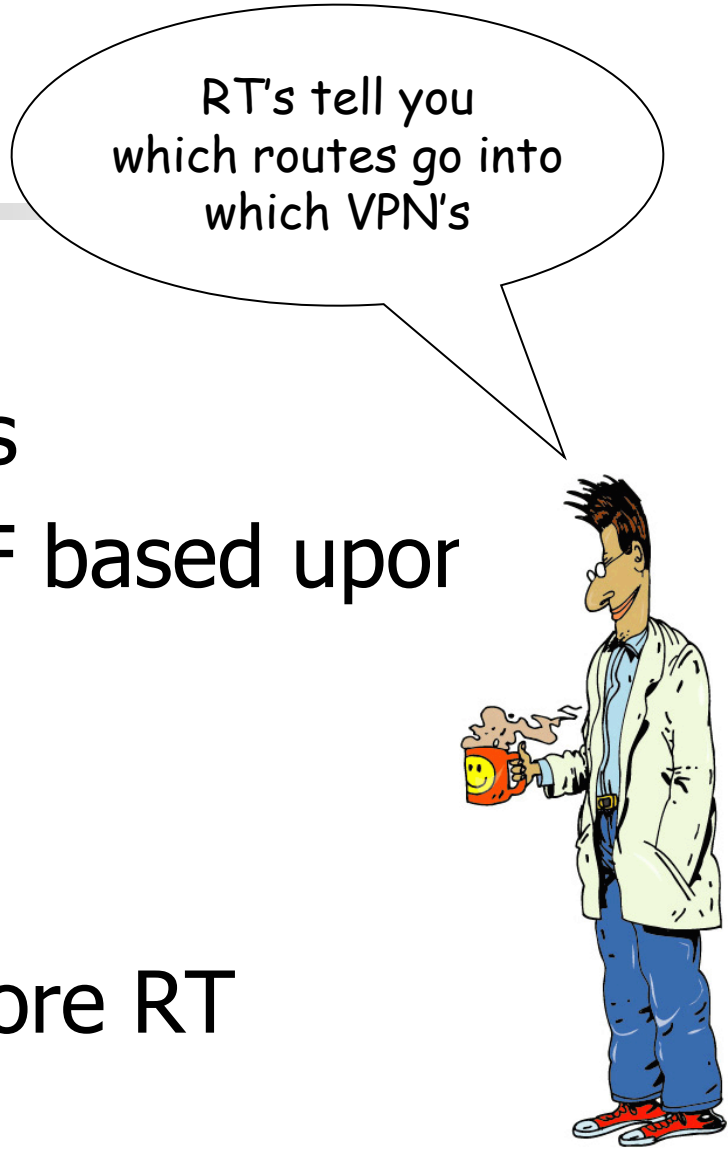
- Used to disambiguate possibly duplicate routes from VRF's
  - i.e. guarantee unique addressing space
  - AS:nn e.g. 100:23
  - IPv4:nn e.g. 192.168.1.1:23
- Creates a guaranteed unique address th BGP can advertise in a single database
- VPNIPv4 addresses

RD's in action

- Per VPN via BGP label assignment
- PE – PE set up via LDP or RSVP (saves state)



Route Targets

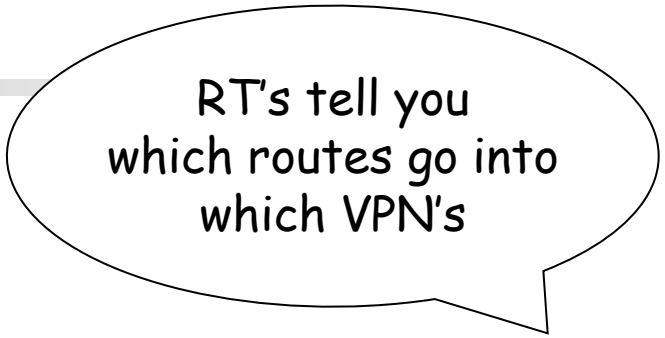


- PE receives VPN IPv4 NLRI's
- Routes then placed into VRF based upor
  - Extended BGP community,
  - AS:nn 100:45
  - IPv4:nn e.g. 192.168.1.1:45
- A route may have one or more RT

Route Targets in action

- When routes are advertised, they are exported with one or more RT's
- A VRF can import routes with matching RT's
- Security of this architecture depends on YOUR provisioning integrity

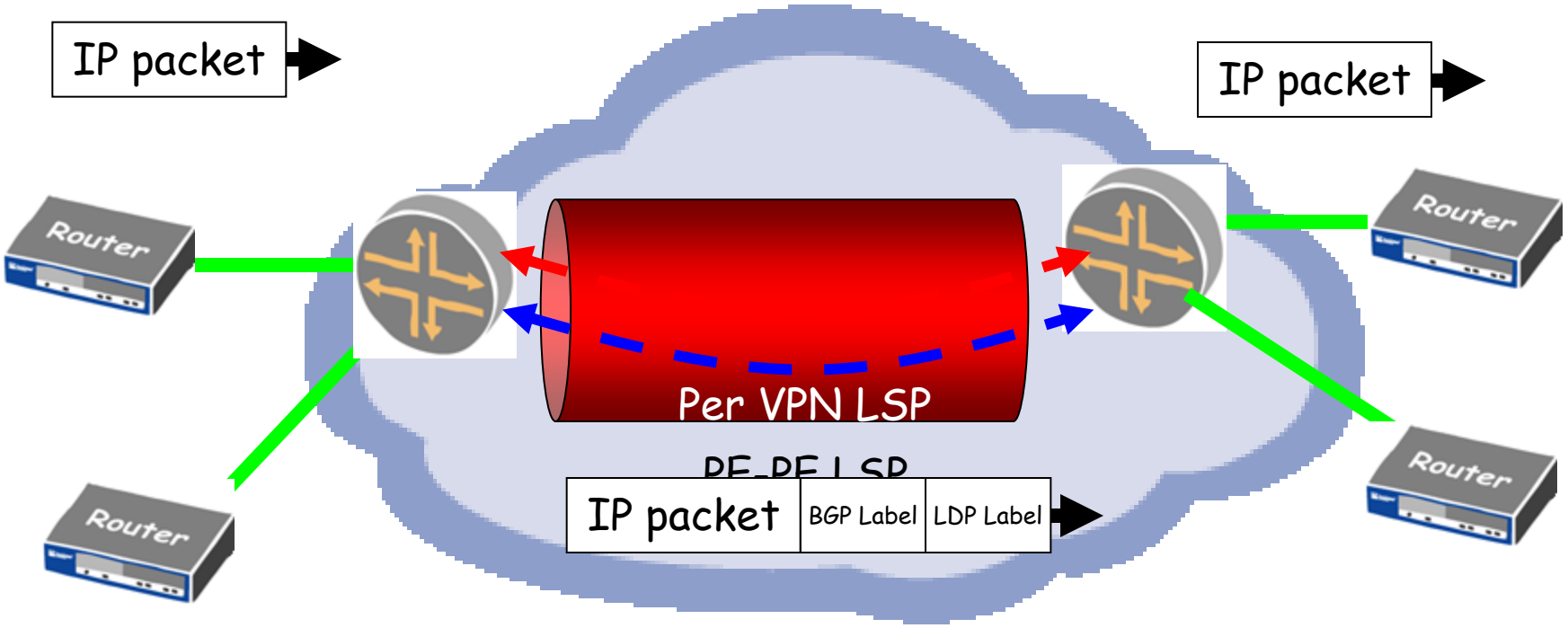
Why RD's and RT's?



- Overhead is better when
  - Advertisements get bigger, as opposed to
  - More advertisements
- Allows for overlapping VPN's
- Can be the same
  - But don't lock yourself in

LSP establishment

- Per VPN via BGP label assignment
- PE – PE set up via LDP or RSVP (saves state)

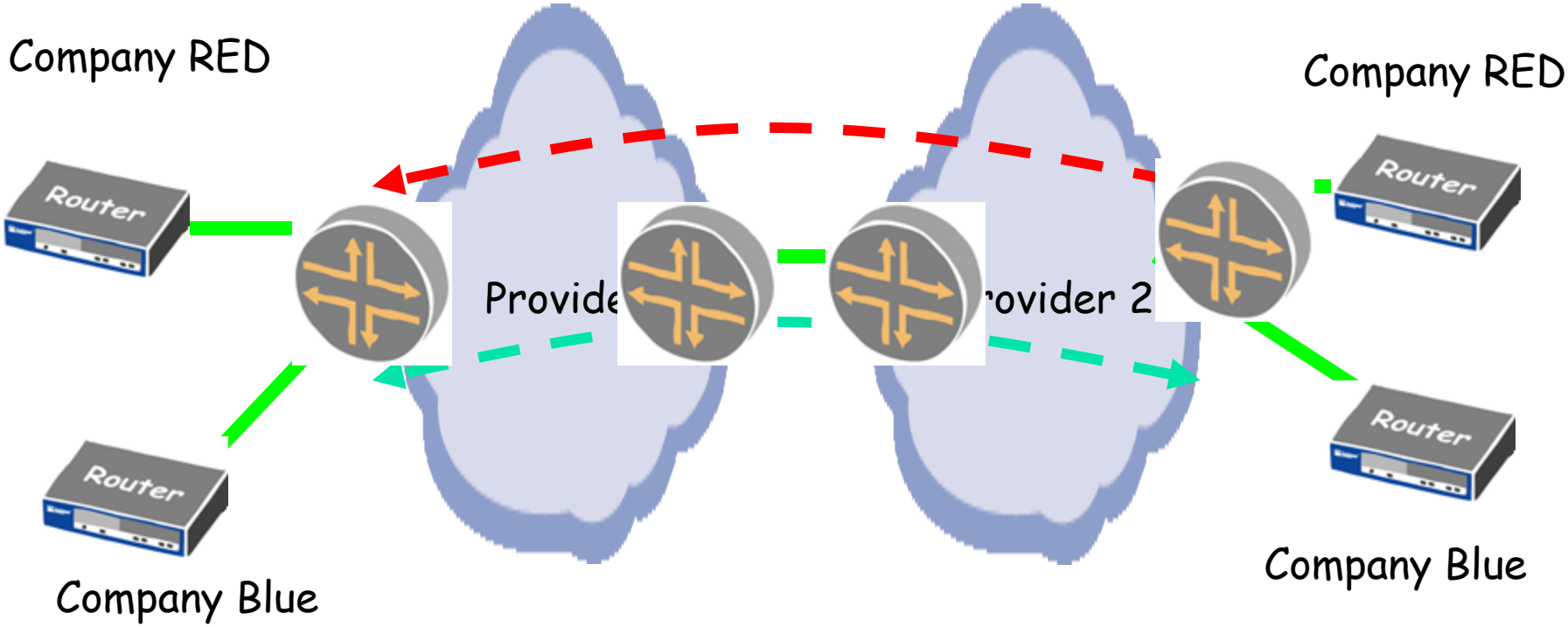


Connectivity

- Hub and spoke
  - Outsourcing internet access and Applications
- Full Mesh

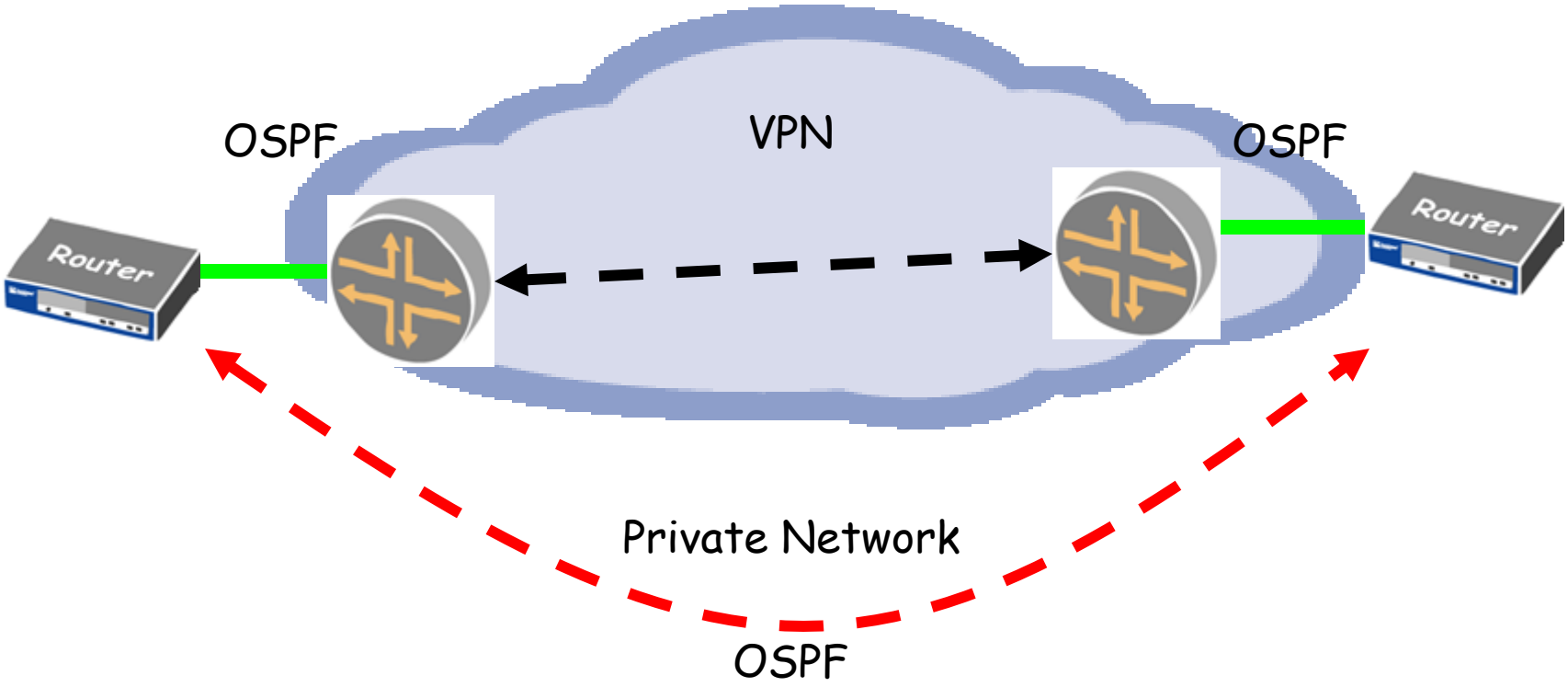
InterAS VPN's

- VRF-to-VRF
- MBGP between ASBR (not OSPF)
- MBGP between PE's



VPN as backup

- Do you want PE to appear as
  - Intra Area Router (Sham Links)
  - ABR
  - ASBR



Issues

- BGP scaling
  - RR, often separate from IP RR
- Inter-AS scaling
  - MBGP between PE's is desirable
- Management
  - Usual MPLS, OAM, root cause automation.
  - Overlap NOC with VPN? Addressing?
- QoS
  - Carriers mapping 4+ queues

Security

- Routing protocol security is just as critical in an MPLS VPN environment as it is in the Internet environments
- PE-CE security is important
  - RIP
  - OSPF
  - BGP
- When using BGP, we need to ensure customers are not injecting malicious communities to hijack VPN traffic
- Labels
  - Label injection exploits are possible
  - Malicious PE injects traffic with a label for a different service
  - Misconfigured PE may accept labeled traffic from a CE

Configuring L3VPN's

Enable MPLS and LDP

```
ip cef
mpls ip
mpls label protocol ldp
!
interface fast 0/1
mpls ip
mpls label protocol ldp
!
```

PE-PE MP-IBGP Peering

- PE-to-PE MP-IBGP sessions require VPN-IPv4 NLRI

```
!
router bgp 150
neighbor 192.168.16.1 activate
!
address-family vpnv4
neighbor 192.168.16.1 activate
neighbor 192.168.16.1 send-community extended
!
```

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 vpn-a.inet.0 Bit: 40000
    Send state: in sync
    Active prefixes: 7
    Received prefixes: 8
```

Assigning the Route Distinguisher

- Manually assign the RD per VRF table

```
IOS
---
ip vrf ODD_Customer
rd 150:101
…
```

- Automatic RD assignment is possible on some platforms

A Sample VRF Table Configuration

Create a VRF table called *vpn-a* with BGP running between the PE and CE routers using the `vrf-target` statement:

```
ip vrf vpn-a
rd 3:101

interface fastethernet 0/0
ip vrf forwarding vpn-a
ip address 200.1.9.1 255.255.255.0


ip vrf vpn-a
route-target export 3:111
route-target import 3:111
```