# MPLS Deployment – Best practices

Yogesh Jiandani, Consulting Engineer, India and SAARC

Vagish Dwivedi, Consulting Systems Engineer

# Classroom basics



- **This session is for people who already have an understanding of how MPLS, MPLS L3 VPNs and MPLS TE work**
- **Cellphones off or in vibrate mode. If important, calls to be taken strictly outside the classroom**
- **No Internet/Email/Chat etc access while in the class. In short, please shut off your laptops.**
- **PLEASE ASK QUESTIONS. THERE IS NO SUCH THING AS A STUPID OR DUMB QUESTION.**

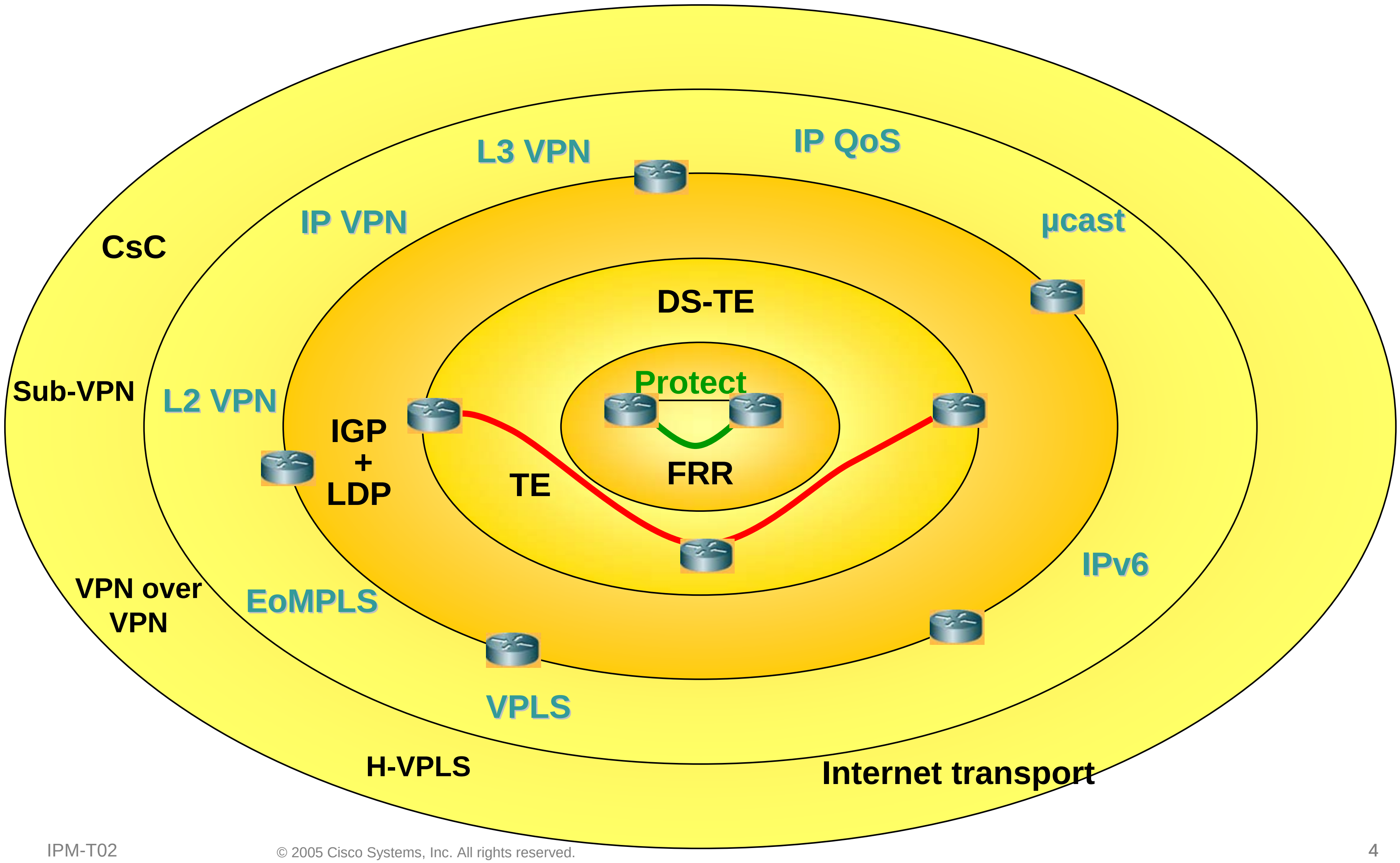
# Session objectives



- Main new developments
- Set-up services
- Best practices & designs

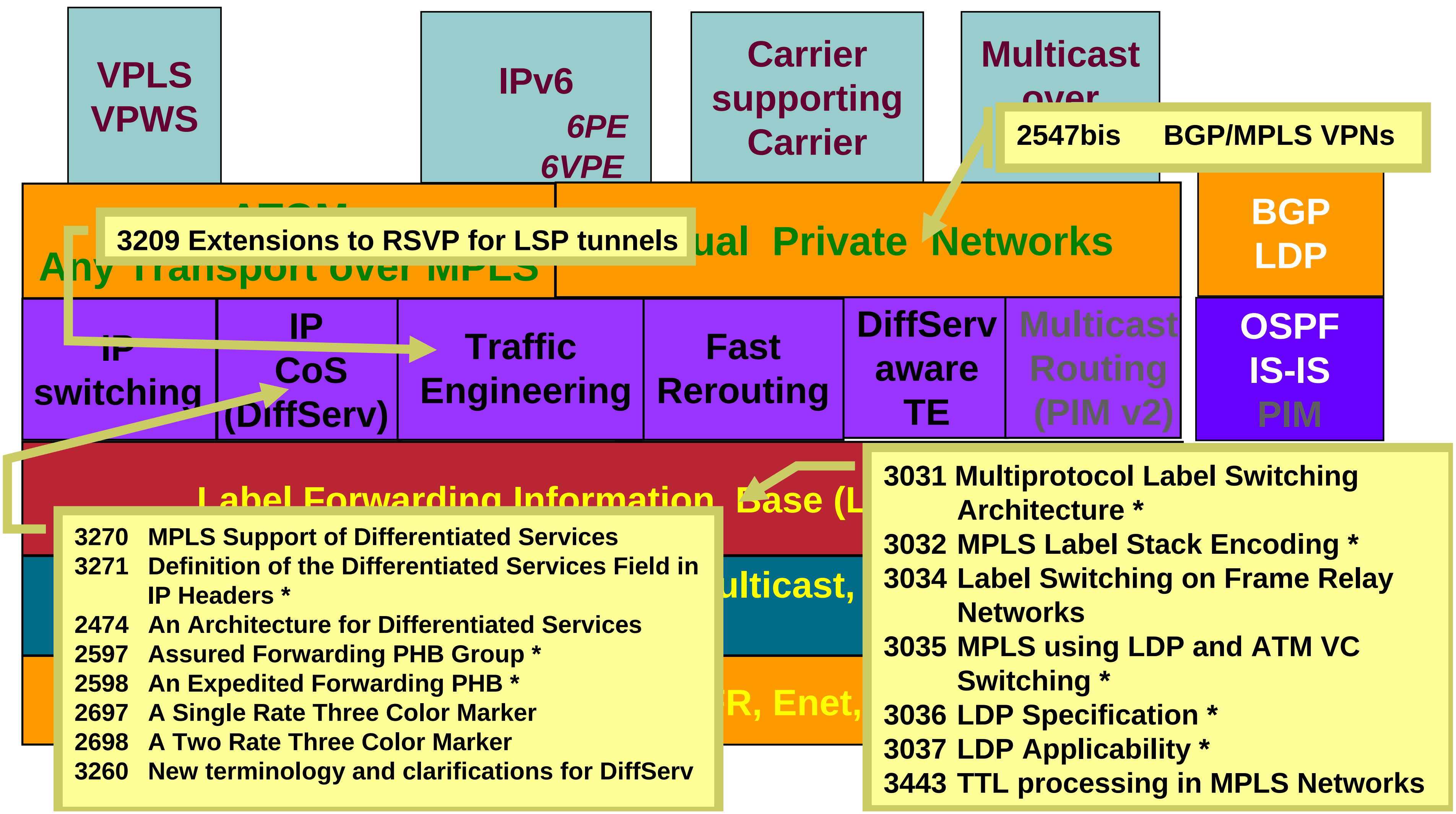
# MPLS - The Big Picture





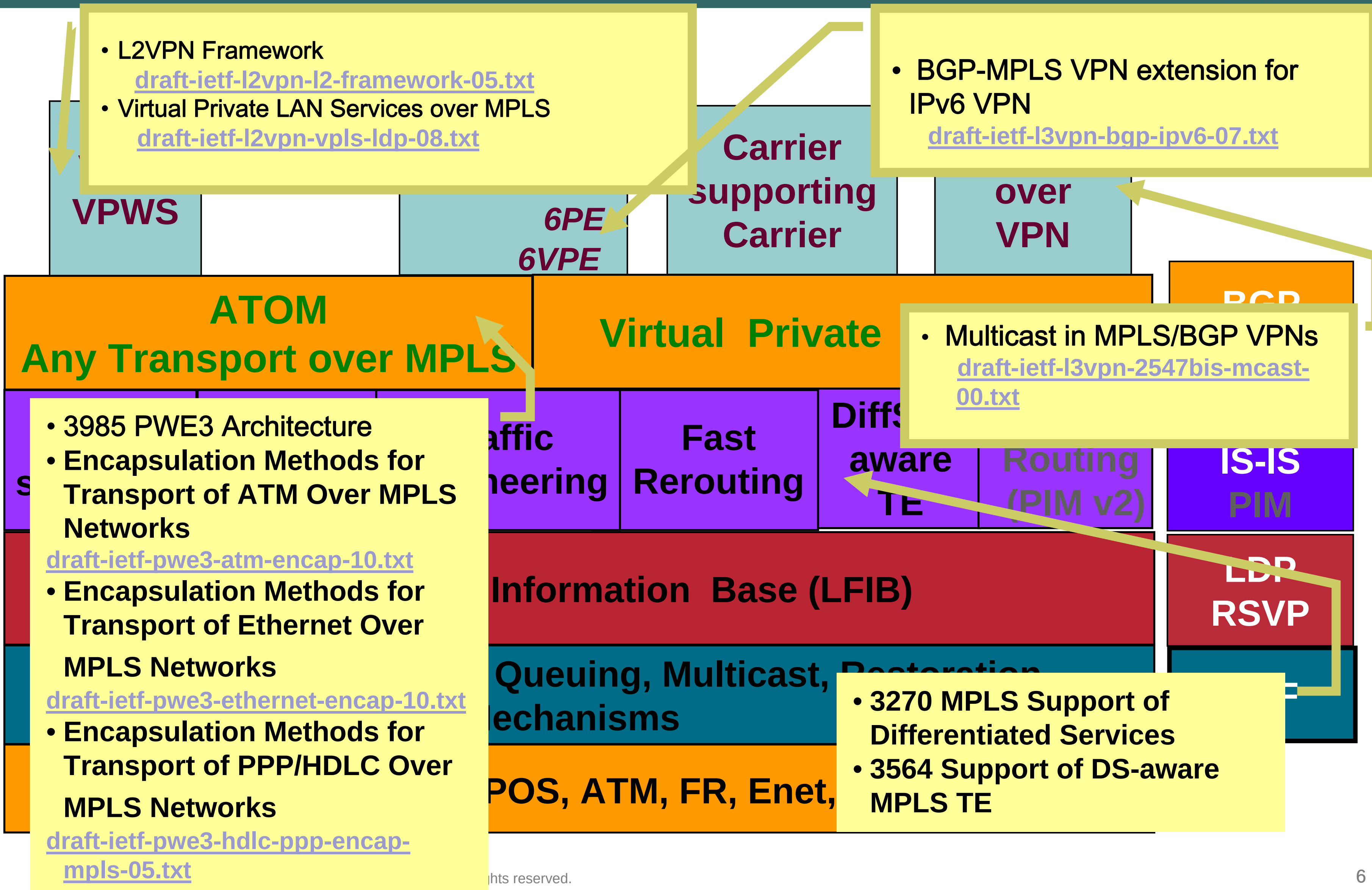
# MPLS
# Innovation & Standards

Cisco.com



# MPLS
# Innovation-in-Progress





# IETF at work
# Working Groups related to MPLS

Cisco.com

| | | |
|---|---|---|
| MPLS | = Multi-Protocol Label Switching | 39 RFCs / 19 Drafts |
| L3vpn | = Layer 3 Virtual Private Networks | 6 RFCs / 19 Drafts |
| Pwe3 | = Pseudo Wire Emulation Edge to Edge | 3 RFCs / 24 Drafts |
| L2vpn | = Layer 2 Virtual Private Networks | 11 Drafts |
| L1vpn | = Layer 1 Virtual Private Networks | (VPN over GMPLS) |
| Ccamp | = Common Control and Measurement Plane (GMPLS) | |
| Isis | = IS-IS for IP Internets | |
| Ospf | = Open Shortest Path First IGP | |
| Pim | = Protocol Independent Multicast (as consultant for MPLS) | |
| BFD | = Bi-directional Forwarding Detection | |

**+ Personal contributions** (draft & Informational RFC)

**TE Internet Traffic Engineering** (Concluded, transferred to MPLS WG)

# Agenda



- **L3 VPNs**
  - Create VPN
  - Internet access
  - Security considerations
  - BGP advanced features
- **Manage L3-VPN services**
  - Troubleshooting / Diagnostics
  - Configuration
- **MPLS L2 Transport**
  - Virtual Private Wired Services
  - Virtual Private LAN Services
- **MPLS OAM and Traffic Management**
  - OAM
  - Fast-convergence
  - Traffic Engineering

# Agenda

Cisco.com

- **Create L3-VPN service**

  **Create VPN**

  **Internet access**

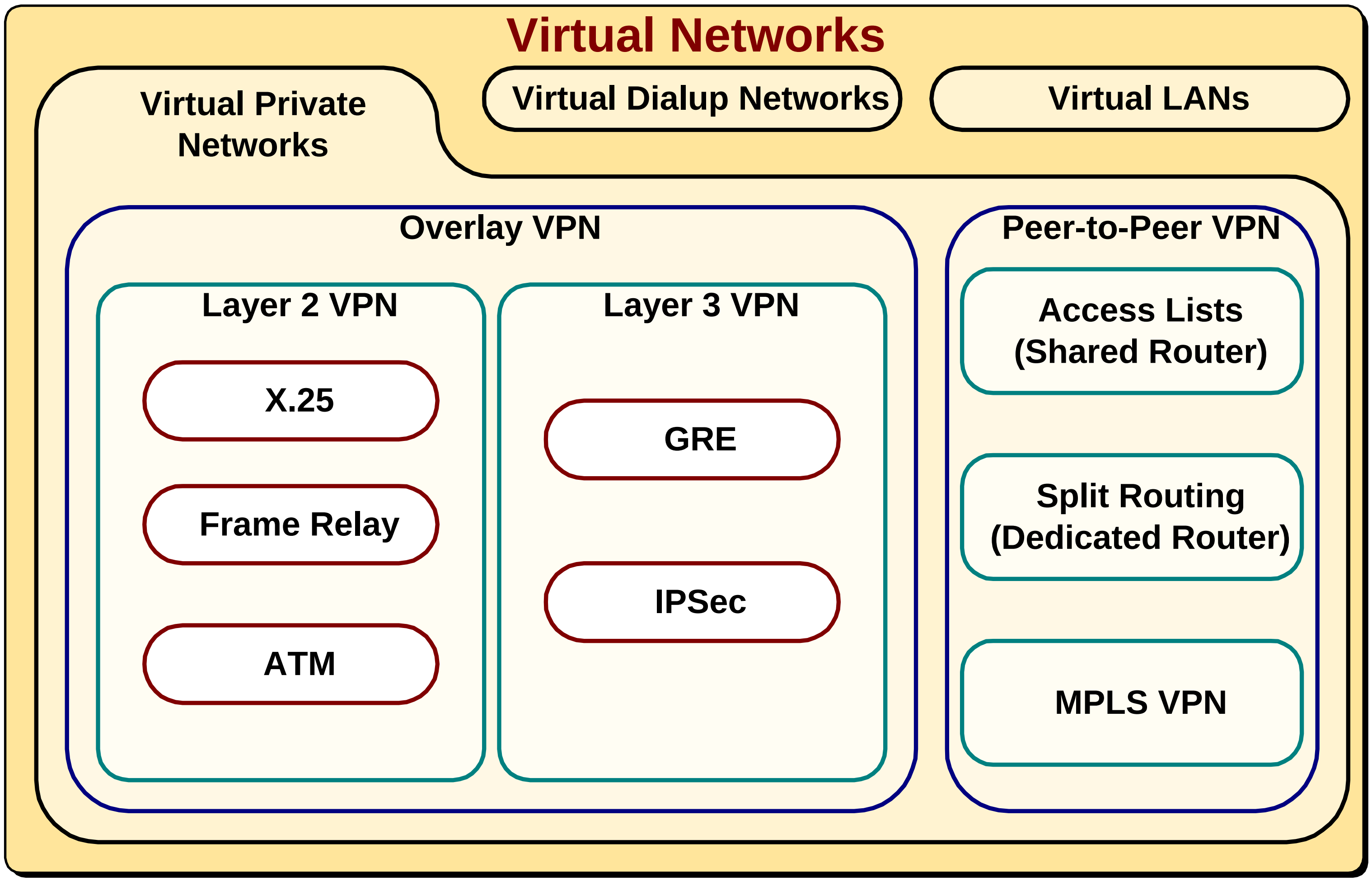
  **Security considerations**

  **BGP advanced features**

***Vagish Dwivedi***

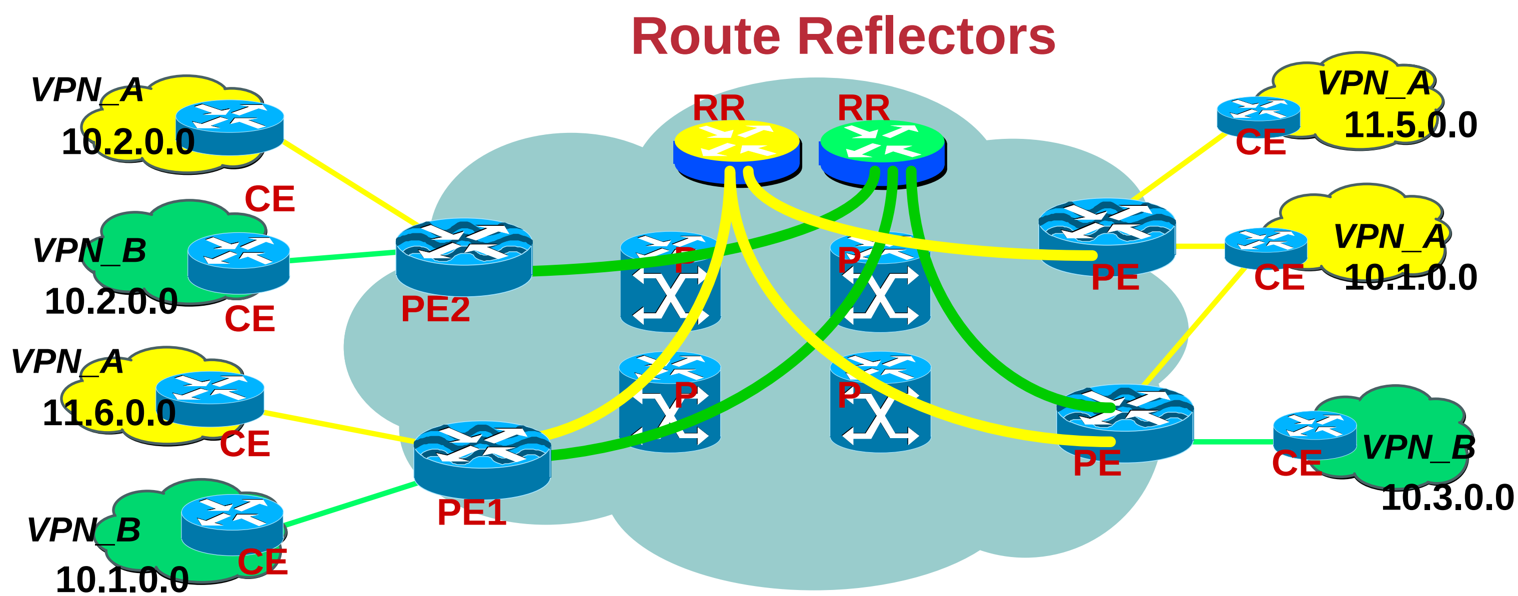
# VPN Taxonomy





# MPLS-VPN
# BGP peering design

Cisco.com




**Do not build a fully meshed iBGP network:**

- **Use Route Reflector (RR)**
- **Easy add of new site / Central point for routing security**
- **Stability**

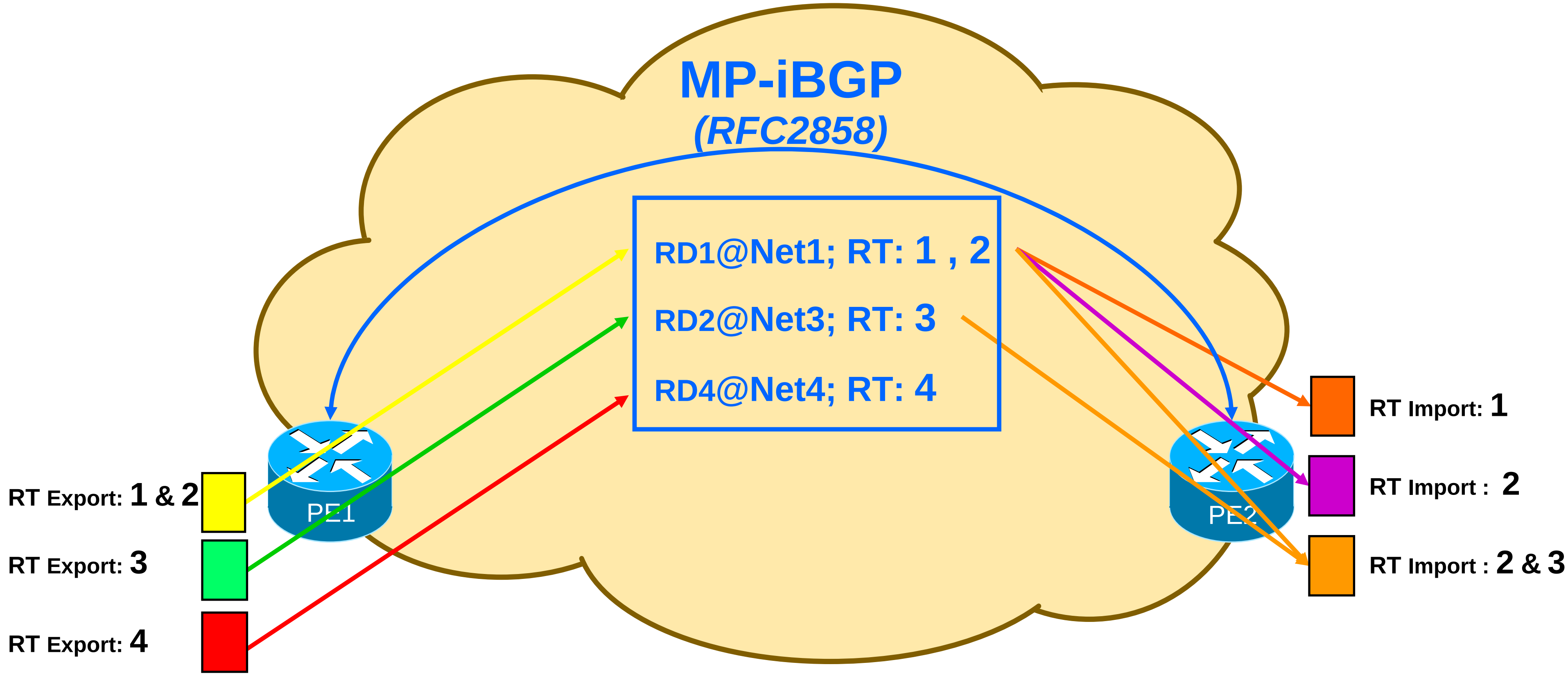
**Two RR (or three) are enough in general in Enterprise**

- **Each RR can store easily up-to 200.000 Routes**
- **Larger design is easily feasible**

# VRF Route Distribution control

## Interconnect Private Virtual Routers across the network





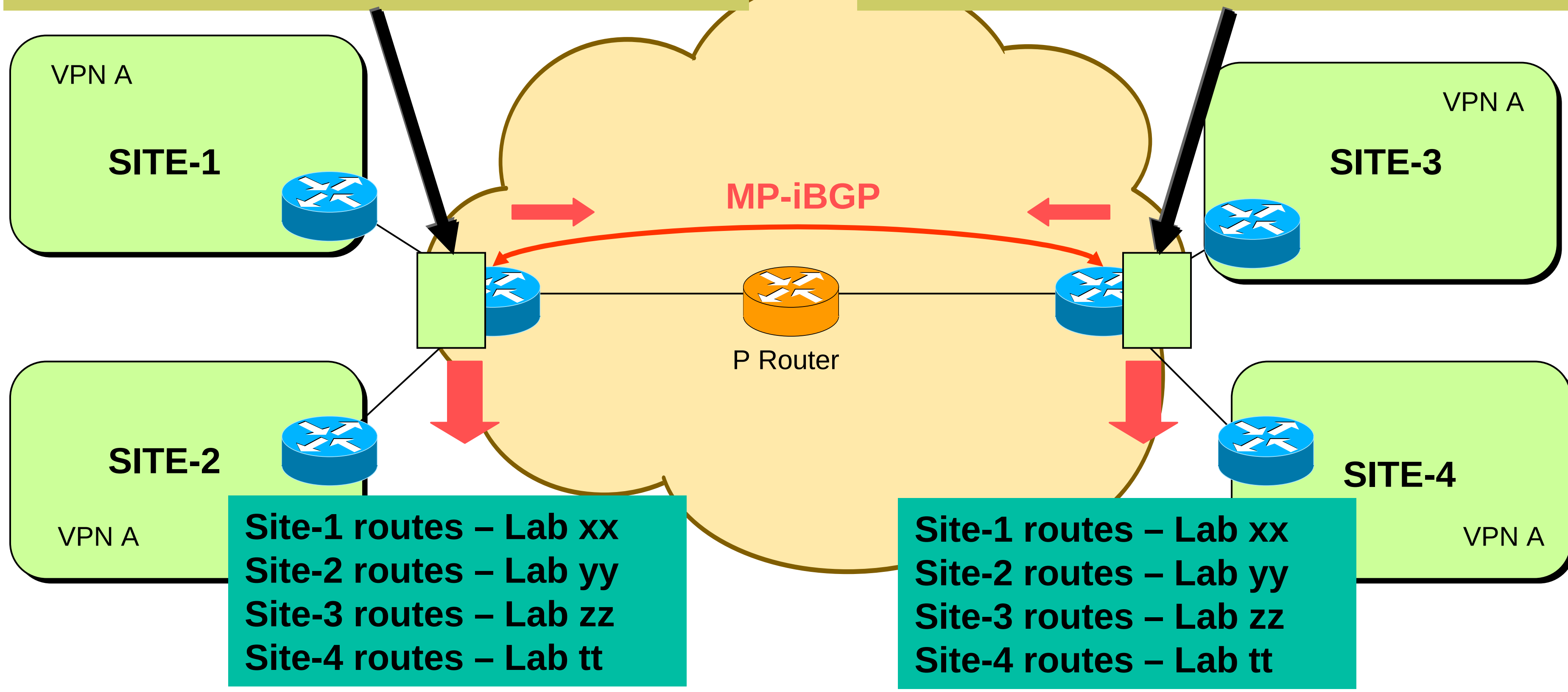

- The RD is prepended to IP address to make it globally unique.
- The RD serves as a VPN identifier for simple VPN topologies but may not be true for complex topologies.
- Route-Target (RT) are acting as Import/Export filters
- *No limitation on number of RT per VRF ➔ a VRF may belong to multiple VPN*

# Intranet Model - Simple

Cisco.com

```
ip vrf VPNA
 rd 500:18
 route-target export 500:18
 route-target import 500:18
```

```
ip vrf VPNA
 rd 500:18
 route-target export 500:18
 route-target import 500:18
```

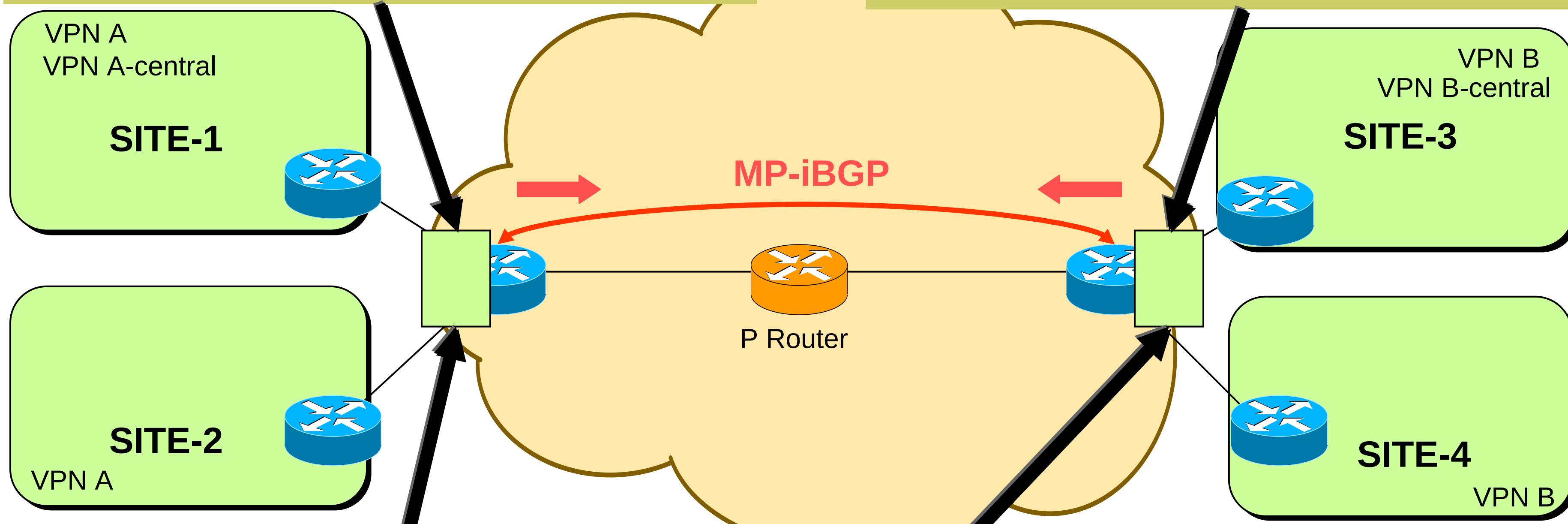


# Intranet Model - Complex



```
ip vrf VPNA-central
 rd 500:28
 route-target export 500:28
 route-target import 500:28
 route-target import 500:18
```

```
ip vrf VPNB-central
 rd 500:28
 route-target export 500:28
 route-target import 500:28
 route-target import 500:18
```

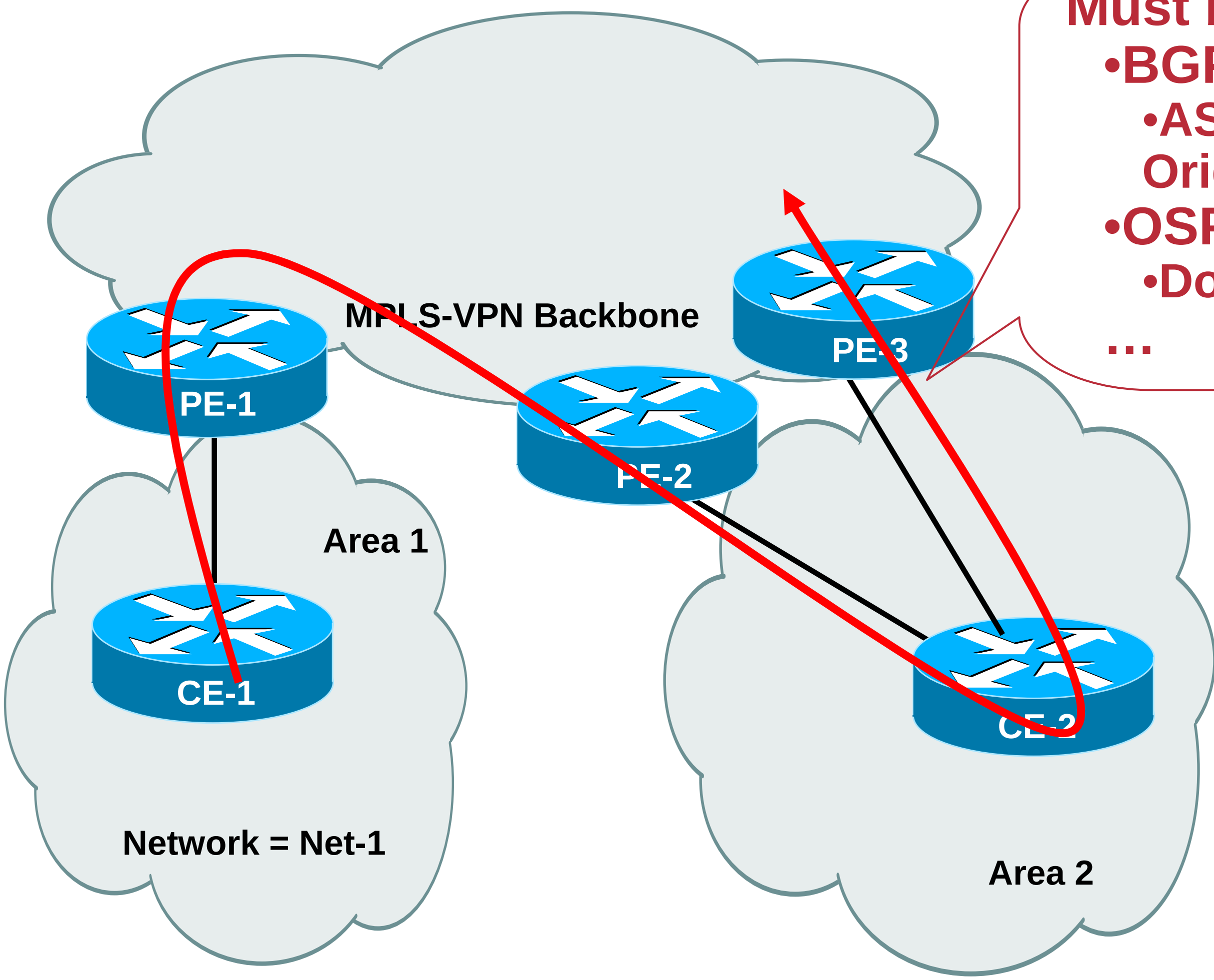



```
ip vrf VPNA
 rd 500:18
 route-target export 500:18
 route-target import 500:18
```

```
ip vrf VPNB
 rd 500:18
 route-target export 500:18
 route-target import 500:18
```

# Avoiding Routing Loop

Cisco.com



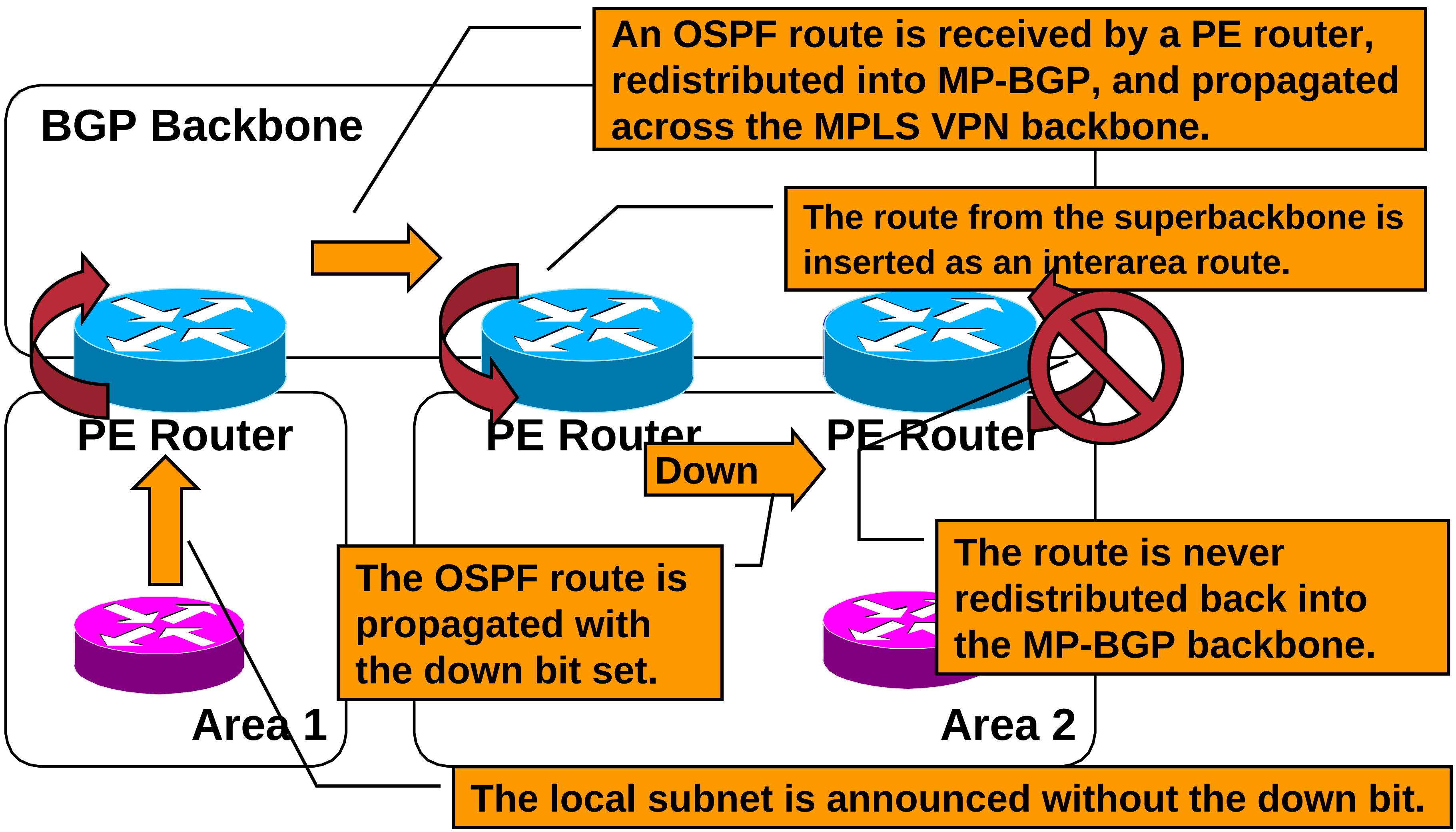

Must block loops:

- BGP:
  - AS number or Site of Origin route-map
- OSPF:
  - Down-bit or External Tags

…

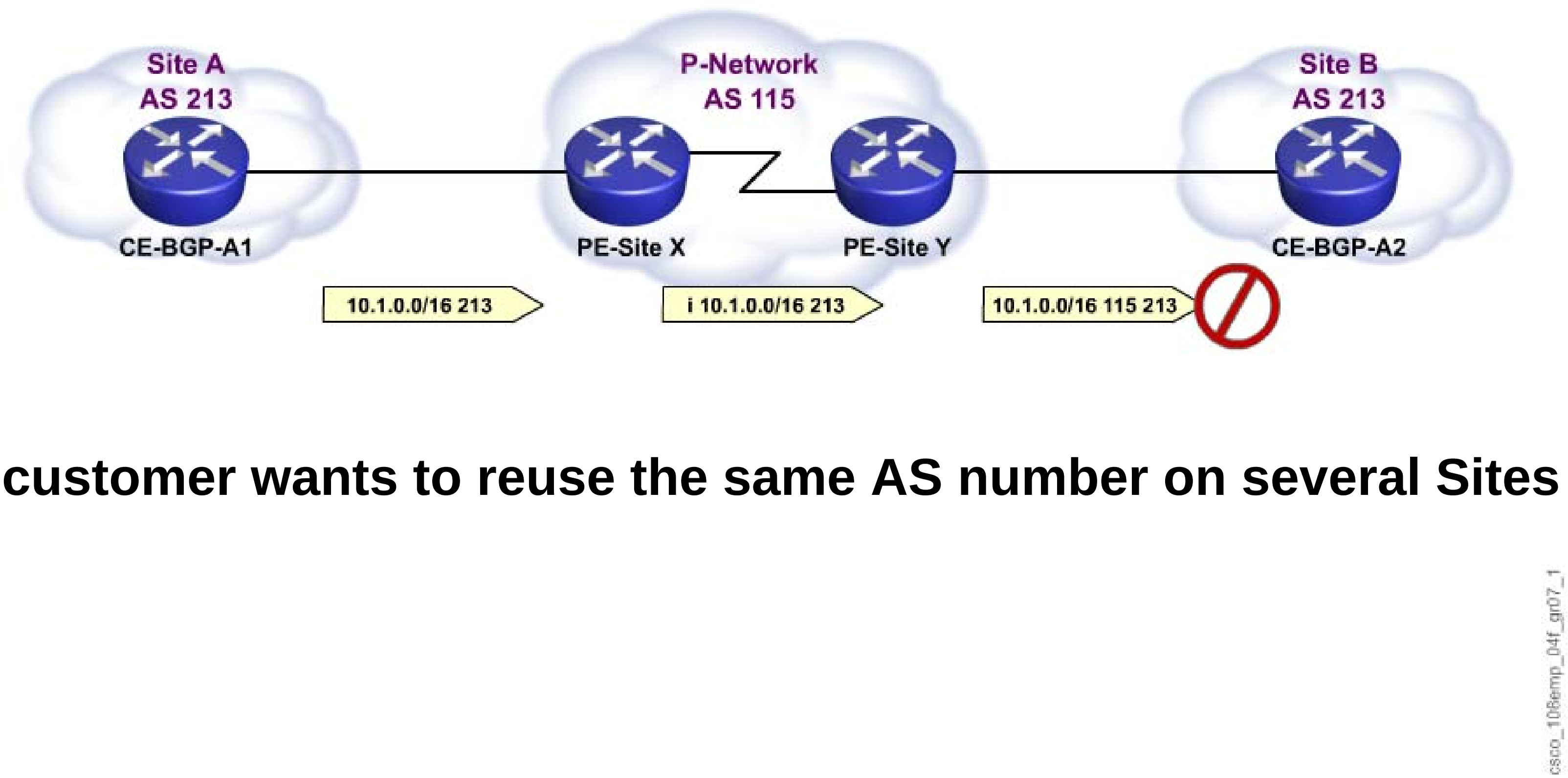
# Loop Avoidance

Cisco.com



# AS Override






**The customer wants to reuse the same AS number on several Sites**

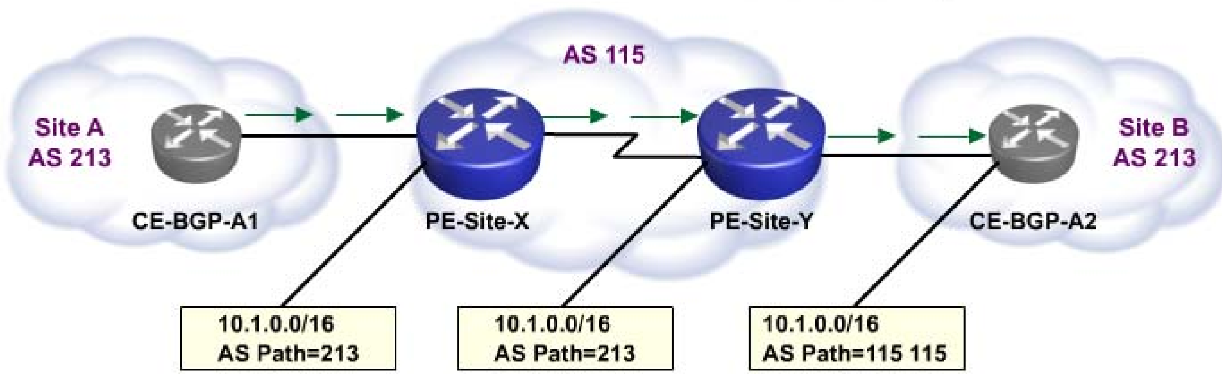
# AS Override



Console

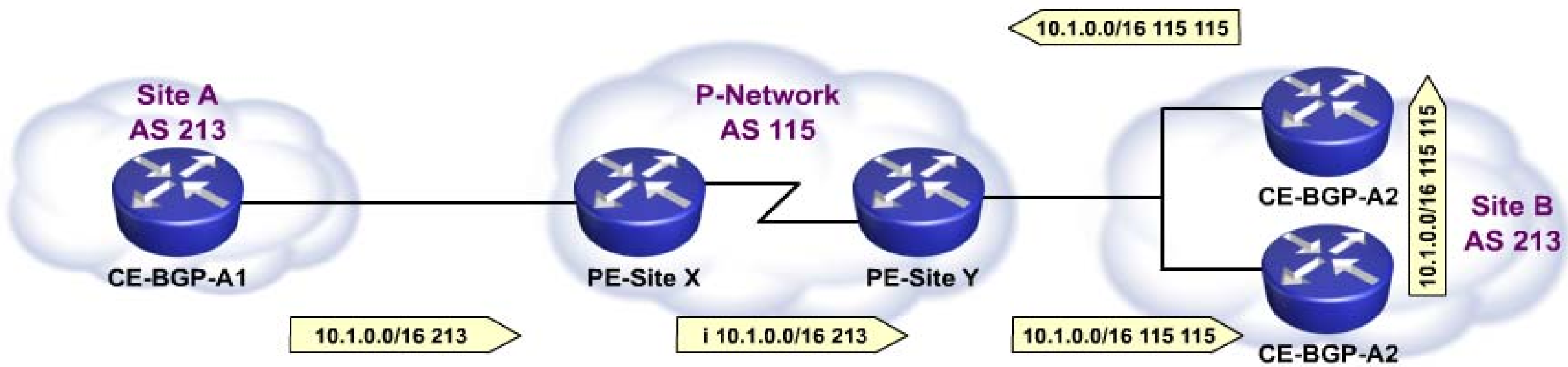
File Edit View Call Transfer Help

```
router bgp 115
address family ipv4 Customer_A
neighbor 10.200.2.1 remote-as 213
neighbor 10.200.2.1 activate
neighbor 10.200.2.1.as-override
```



# Site of Origin






- **AS path-based BGP loop prevention is bypassed with AS-override.**

**Implementing SOO with route-map helps in Loop Prevention**

# Insertion of MPLS VPN into an existing network



- As MPLS VPN relies on BGP

    A smooth insertion into an existing IGP network is not by default

- Two approaches:

    External Autonomous System insertion ( from Edge )

        Insert MPLS VPN at boundary points- requires GRE

        Splitting of existing network – Core does not run MPLS

        Core Still carries normal traffic

    Transparent insertion ( from Core )

        Core runs MPLS and exchanges labels –

        no configuration change of existing CE router

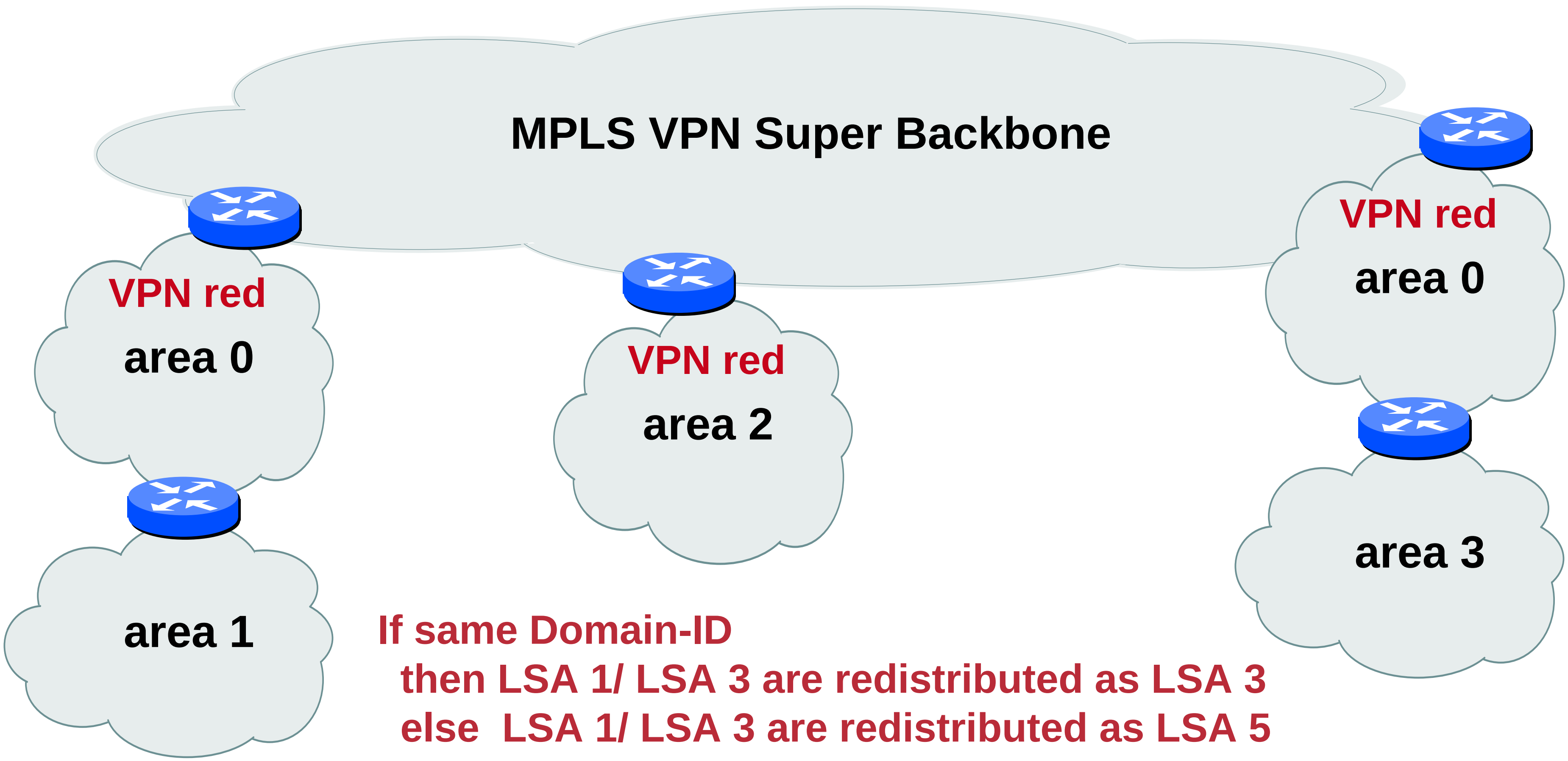
        Smooth integration into area or AS

        More complex definition

# MPLS VPN insertion in an OSPF network: PE acts as ABR or ASBR



## Network splitting

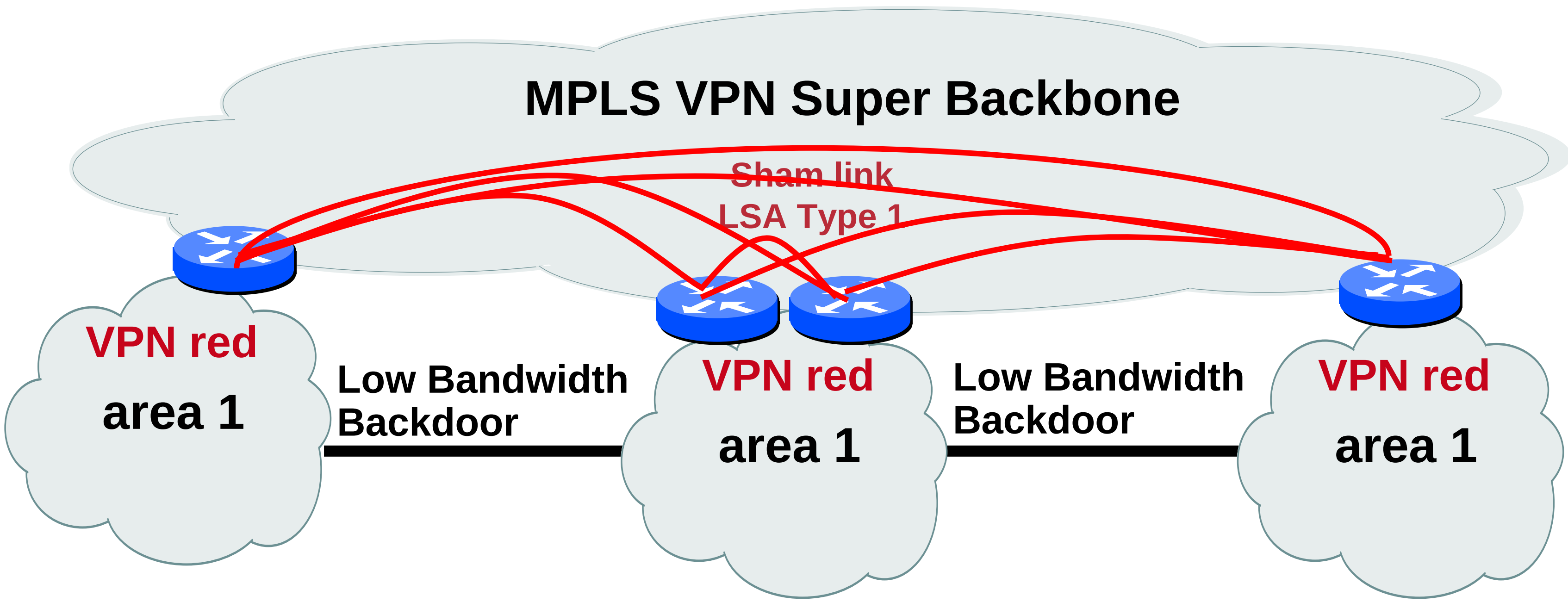



If same Domain-ID
then LSA 1/ LSA 3 are redistributed as LSA 3
else LSA 1/ LSA 3 are redistributed as LSA 5
(By default domain-id is equal to OSPF process-id)

# MPLS VPN inserted into any Area
# PE is intra-area router

Cisco.com

## *Use Sham-link to keep LSA type 1*




**More transparency, easy migration, but add more complexity**

***OSPF link in parallel of the MPLS network are supported***

***The PE acts like an intra-area router, and the MPLS network is seen as an intra-area link***

# OSPF Sham Link
## PE config. - Sham-link Connection



```
ip vrf OSPFCustomer
 rd 10:1
 route-target both 10:1
interface Serial0/1
ip vrf forwarding OSPFCustomer
 ip address 100.10.146.4 255.255.255.0

interface Loopback44
 desc Sham-link interface
 ip vrf forwarding OSPFCustomer
 ip address 100.10.44.4 255.255.255.255

router ospf 12 vrf OSPFCustomer
 area 1 sham-link 100.10.44.4 100.10.55.5 cost 2
 redistribute bgp 1 subnets
 network 100.10.146.0 0.0.0.255 area 1
!
router bgp 1
 address-family ipv4 vrf OSPFCustomer
 redistribute ospf 12
```

**Warning:**
**Sham-link loopback must be learned thru Core MP-BGP**

# Sub-VPN on a Site



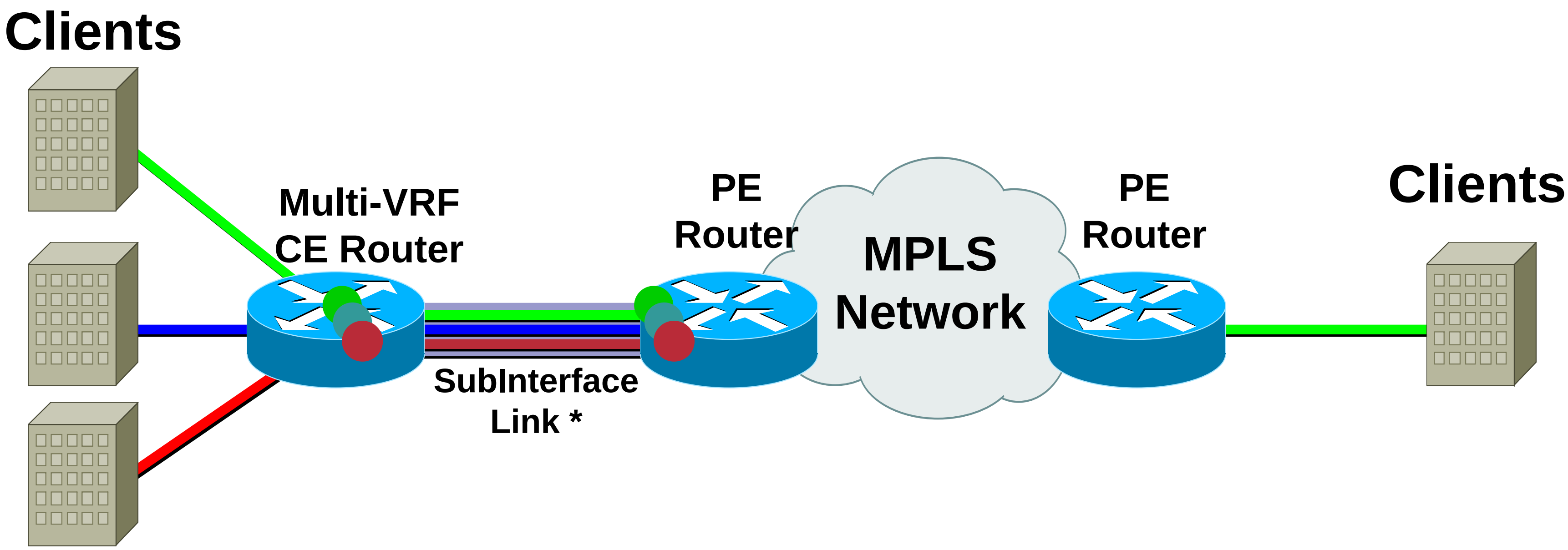
- Separate Intranet / Internet
  - Need one VRF and Global access
- Enterprise is composed of independent groups
  - à la Holding, add & remove easily entity
  - Need a few VRFs, and Centralized Services VRF
  - Common QoS
- Zoning for Security management
  - Raising need: Some VRF, but may increase in future
  - VRF are used to group users with same rights
  - Firewalling is acting between VPN (users & Servers)
  - Common QoS
- Enterprise acts like an Internal-SP
  - Requires security and per VPN QoS
- Customer is an SP

# Multi-VRF CE - Extending MPLS-VPN
# Ability to create VRF without MPLS switching



*SubInterface Link*

*Any Interface type that supports Sub Interfaces, FE-Vlan, Frame Relay, ATM VC's, GRE*



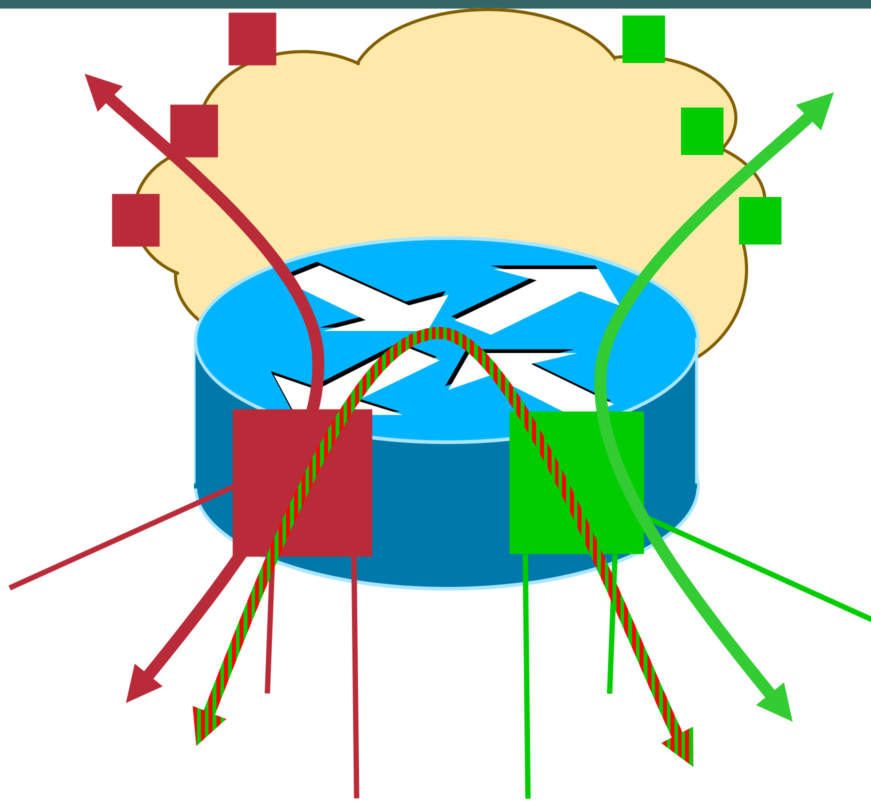

**Allows to push 'PE-like' function to CE**

- **Independance of core versus edge (no peering between CE & all PEs)**
- **Using simple CE : no MP-BGP / no LDP**

**Most of CE functions are supported into a VRF**

# Local Inter Multi-VRF routing





Route-target export 10:12
Route-target import 10:21

Route-target export 10:21
Route-target import 10:12

**Local Red is routed with local Green, but VPN are still separated**
**Requires MP-BGP to be enabled on CE**
**(you can control export/import using Route-map)**

# Agenda



- **Create L3-VPN service**
  - **Create VPN**
  - **Internet access**
  - **Security considerations**
  - **BGP advanced features**

***Vagish Dwivedi***

# Internet Access



- **Facts:**
  - **Basic MPLS switching allows not to distribute Internet routes into the core**
    - **No label is given to external BGP routes**
    - **One label is given to Next-Hop**
  - **Some customer requires optimum access to Internet @**
  - **The Internet table is too big to be populated in multiple VRF**
    - **Ex:  100 VRF * 130,000routes = 13,000,000 !!!**
    - **And even 130,000 VPNv4 @ is consuming …**

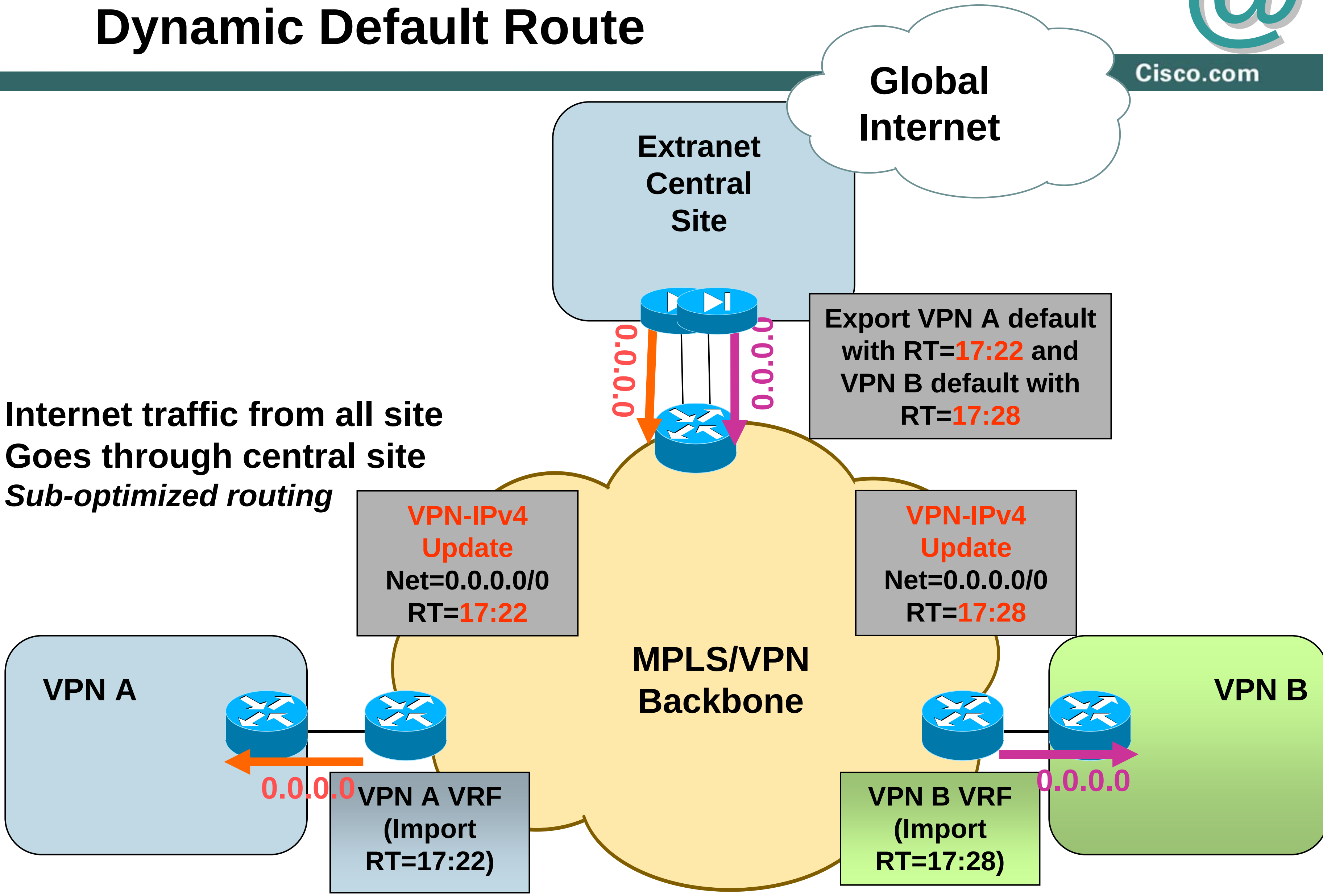
# VPN access to Internet



1. **Default route into a VRF**
   - Most common Internet access, only single point need to be secured
   - All traffic from all site goes across the central site.
2. **Full routing to a dedicated Internet-PE**
   - Insecurity is physically restrained, additional security needed on CE
   - Shared Internet/VPN core (no need of full-routing into core)
3. **Full routing into the PE global routing table**
   - Shared Internet/VPN core (no need of full-routing into core)
4. **Full routing into dedicated Internet-VRF**
   - Shared Internet/VPN core (no need of full-routing into core)
   - Access thru VRF-lite concept
5. **Full routing only into the CE**
   - MPLS to the edge (CsC)

# MPLS/VPN Internet Connectivity Dynamic Default Route

Internet traffic from all site
Goes through central site
*Sub-optimized routing*

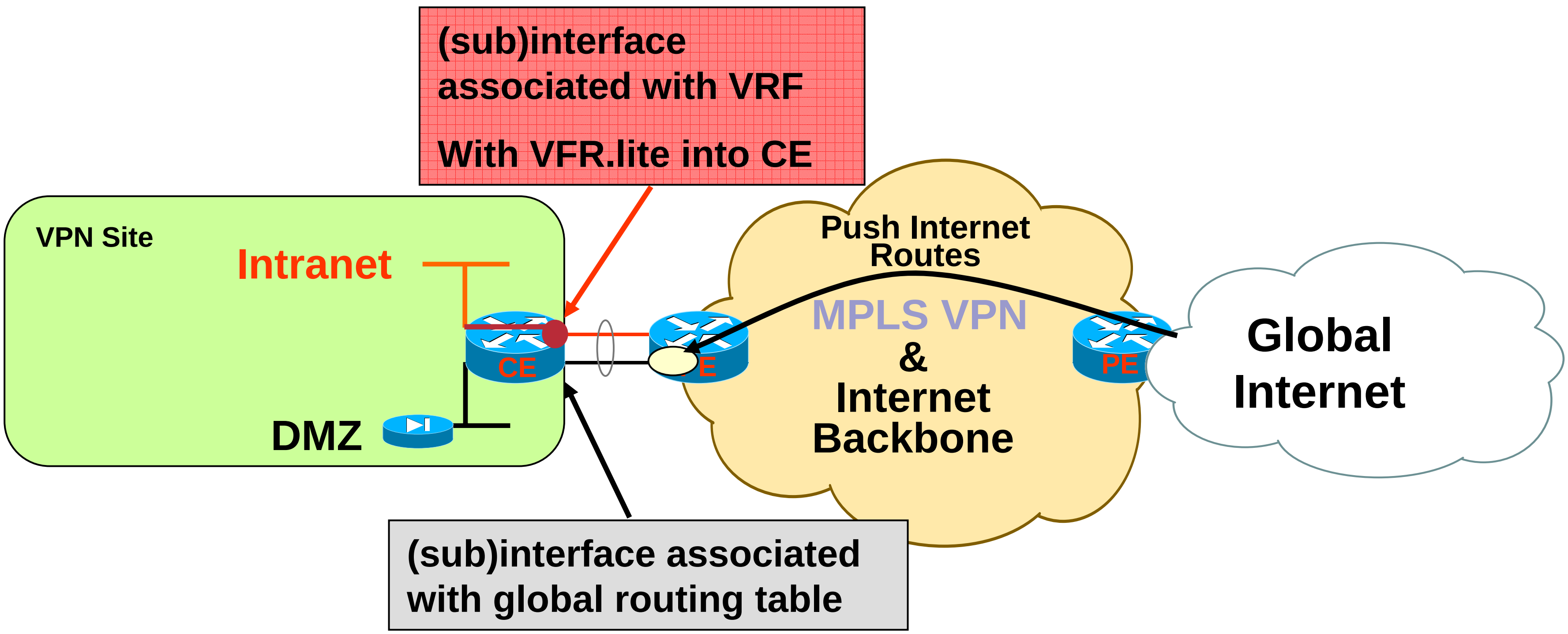


# MPLS/VPN Internet Connectivity
# Dual parallel access using VRF.lite

Cisco.com

**Achieved by using a second interface to the client site**

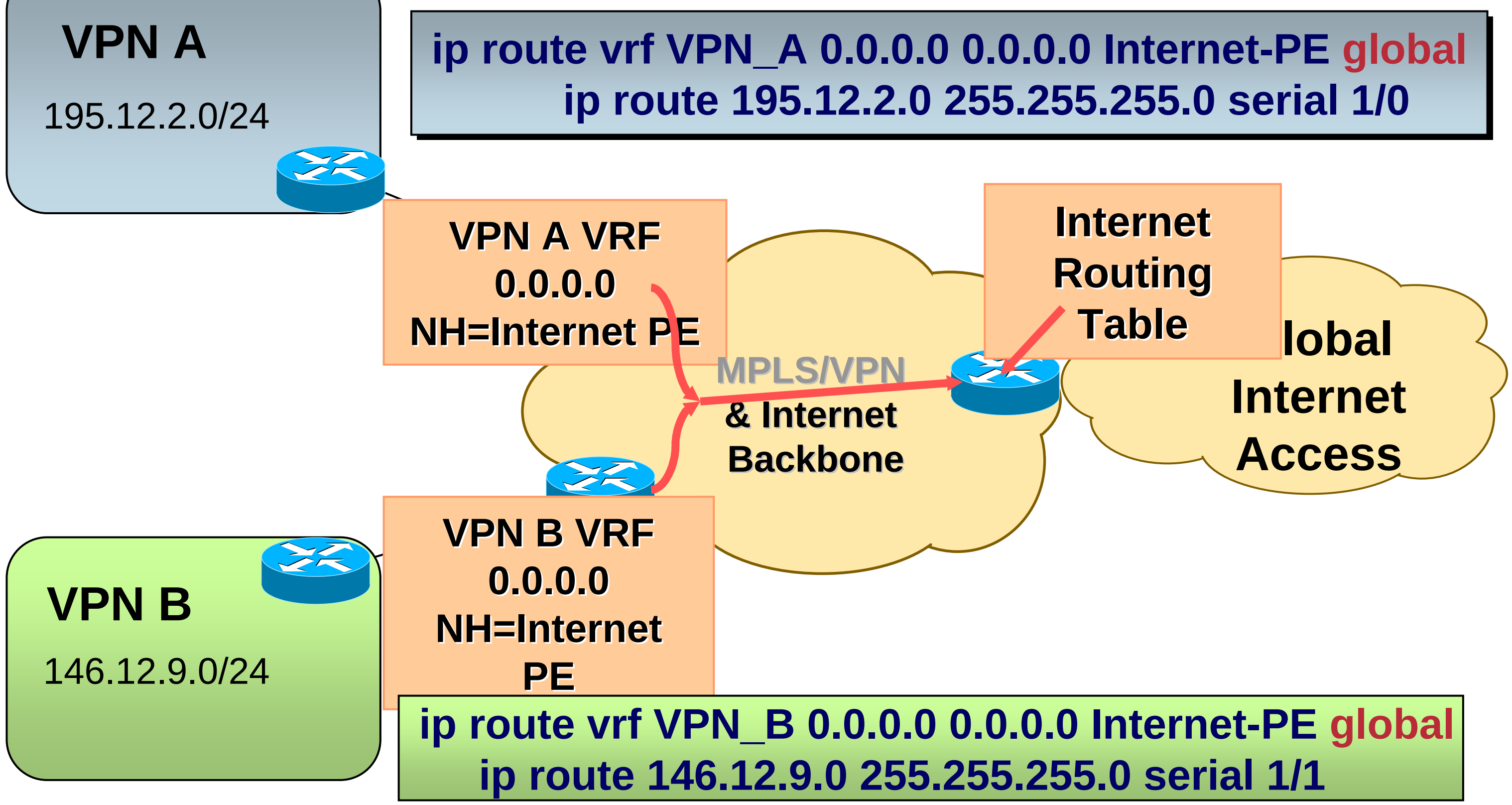
*either physical, or logical such as sub-interface or tunnel*



# MPLS/VPN Internet Connectivity (Packet Leaking) Static Default Route to Global Internet gateway

Cisco.com

## The PE router is also the Internet access router




**Drawbacks:** Internet and VPN packets are mixed on the same link; security issues arise. Packets toward temporarily unreachable VPN destinations might leak into the Internet.

**Benefits:** A PE does not need Internet routes, only an IGP route toward the Internet gateway.

# Import from global table to VRF



Feature allowing for dynamic import from global table to vrf

Use uRPF check, Internet routes leaking

Should be used with max-prefix per vrf

CLI:

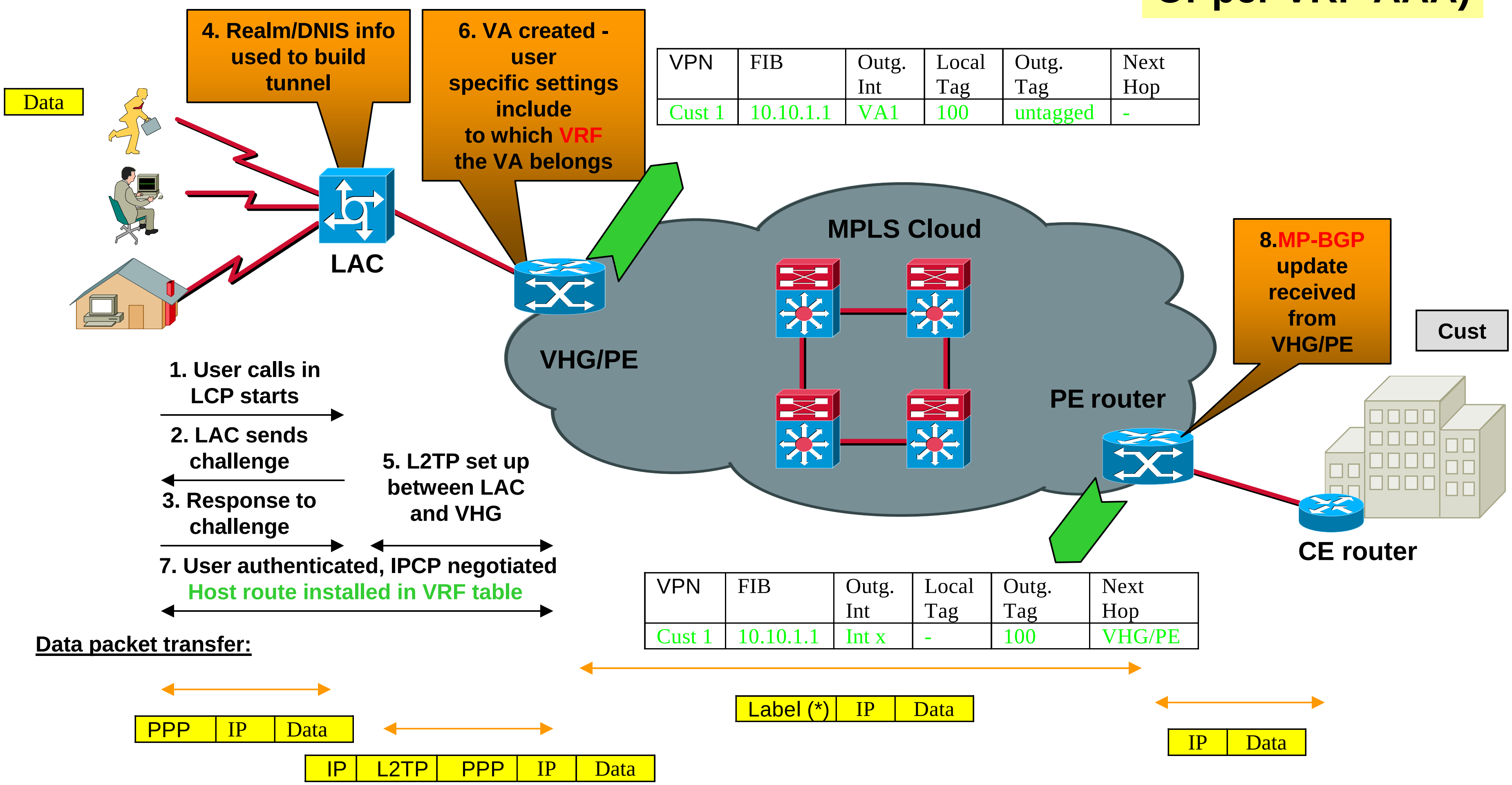
```
import [ ipv4 unicast|multicast [<prefix limit>] ] map <name>
```

e.g. `Router(config-vrf)# import ipv4 unicast 1000 map UNICAST`

- Creates an import map to import IPv4 prefixes from the global routing table to a VRF table.
- imports up to 1000 unicast prefixes that pass through the route map named UNICAST.

# Remote-access
# L2TP to MPLS VPN

**Global AAA
Or per VRF AAA)**



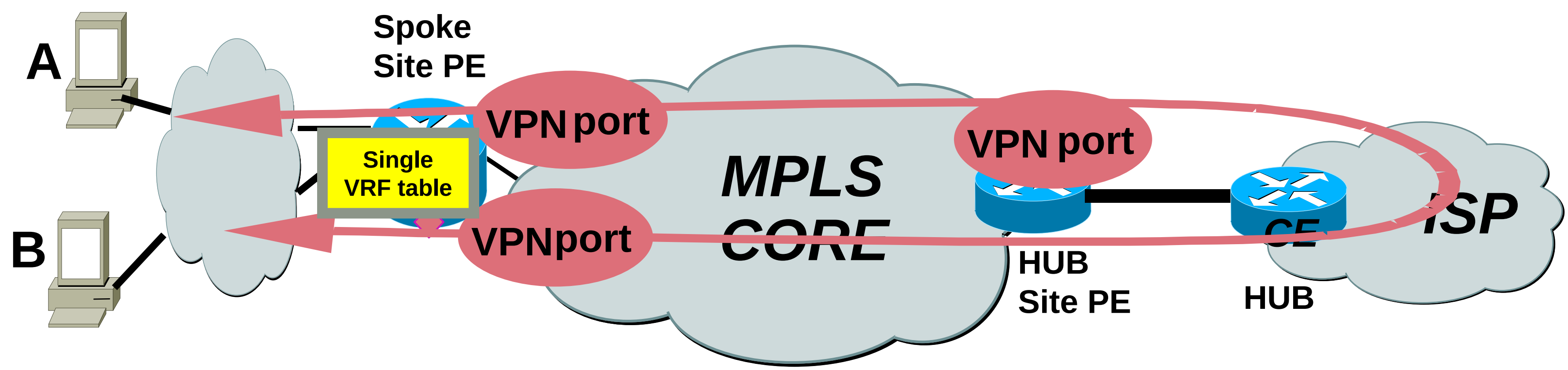

| VPN | FIB | Outg. Int | Local Tag | Outg. Tag | Next Hop |
|---|---|---|---|---|---|
| Cust 1 | 10.10.1.1 | VA1 | 100 | untagged | - |

| VPN | FIB | Outg. Int | Local Tag | Outg. Tag | Next Hop |
|---|---|---|---|---|---|
| Cust 1 | 10.10.1.1 | Int x | - | 100 | VHG/PE |

**(*): at least 2 labels inside (1 label where we do pen-ultimate hop popping)**

# Hub & Spoke Connectivity With Half-Duplex-VRF

Cisco.com




**If two subscribers of the same service terminate on the same PE-router, traffic between them must not be switched locally !**

**Upstream VRF only requires a route-target import statement**

**Imports the default route from the hub PE router (@WholeSale Provider)**

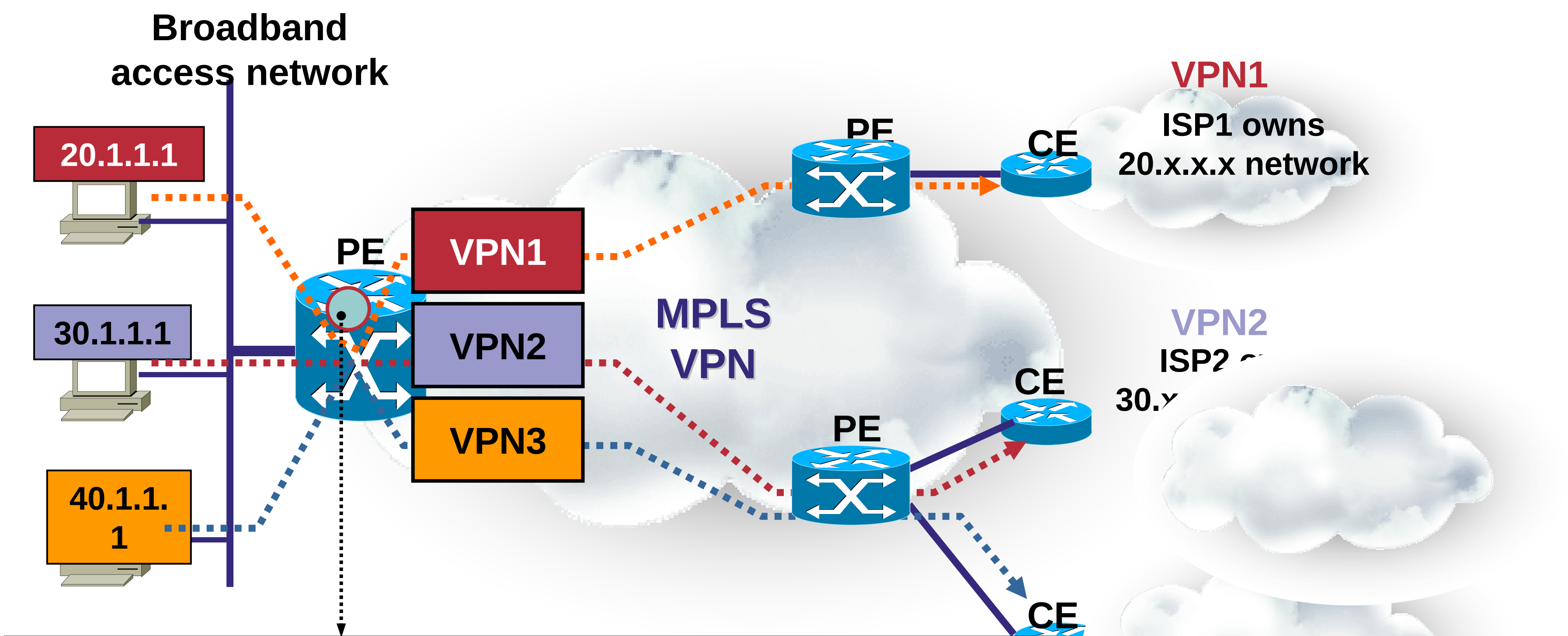
**Downstream VRF only requires a route-target export command**

**Used to export all of the /32 (virtual-access ints) addresses toward the hub PE-router**

- This feature prevents situations where the PE router locally switches the spokes without passing the traffic through the upstream ISP, which causes the wholesale service provider to lose revenue.

# Remote-access
# Source based - VPN Selection



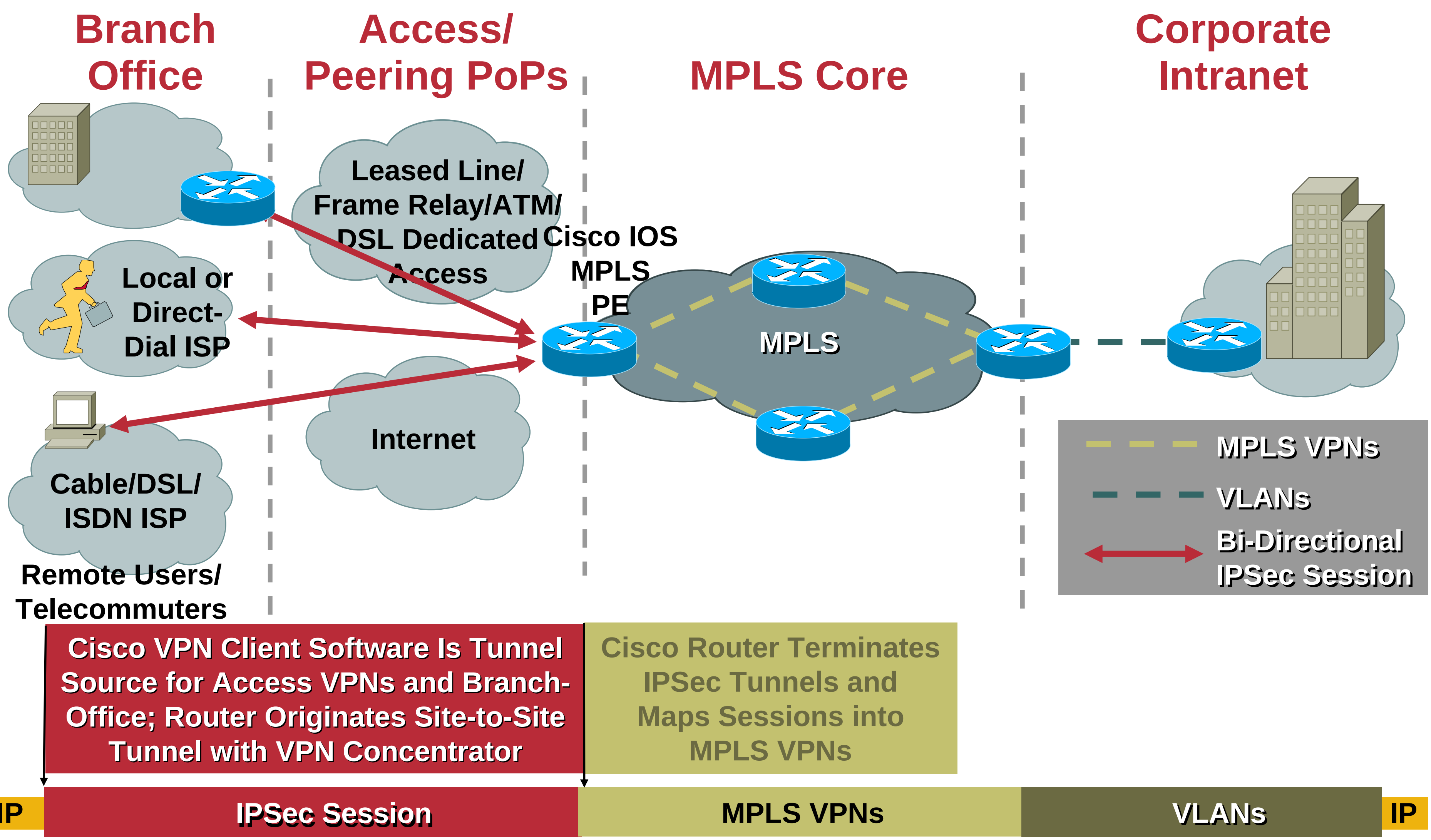



- Customers with high speed (broadband) links can connected to any ISP providing VPN capabilities
- New market: Service Providers can offer corporate VPN connectivity to broadband users

Traditionally physical interface was associated with one VRF table, in situations where multiple customers are connected over single link this provide
vrf selection source 196.7.25.0 255.255.255.128 vrf Trading
vrf selection source 196.7.25.128 255.255.255.128 vrf Retail
!

# Remote-access IPSec + MPLS PE

Cisco.com

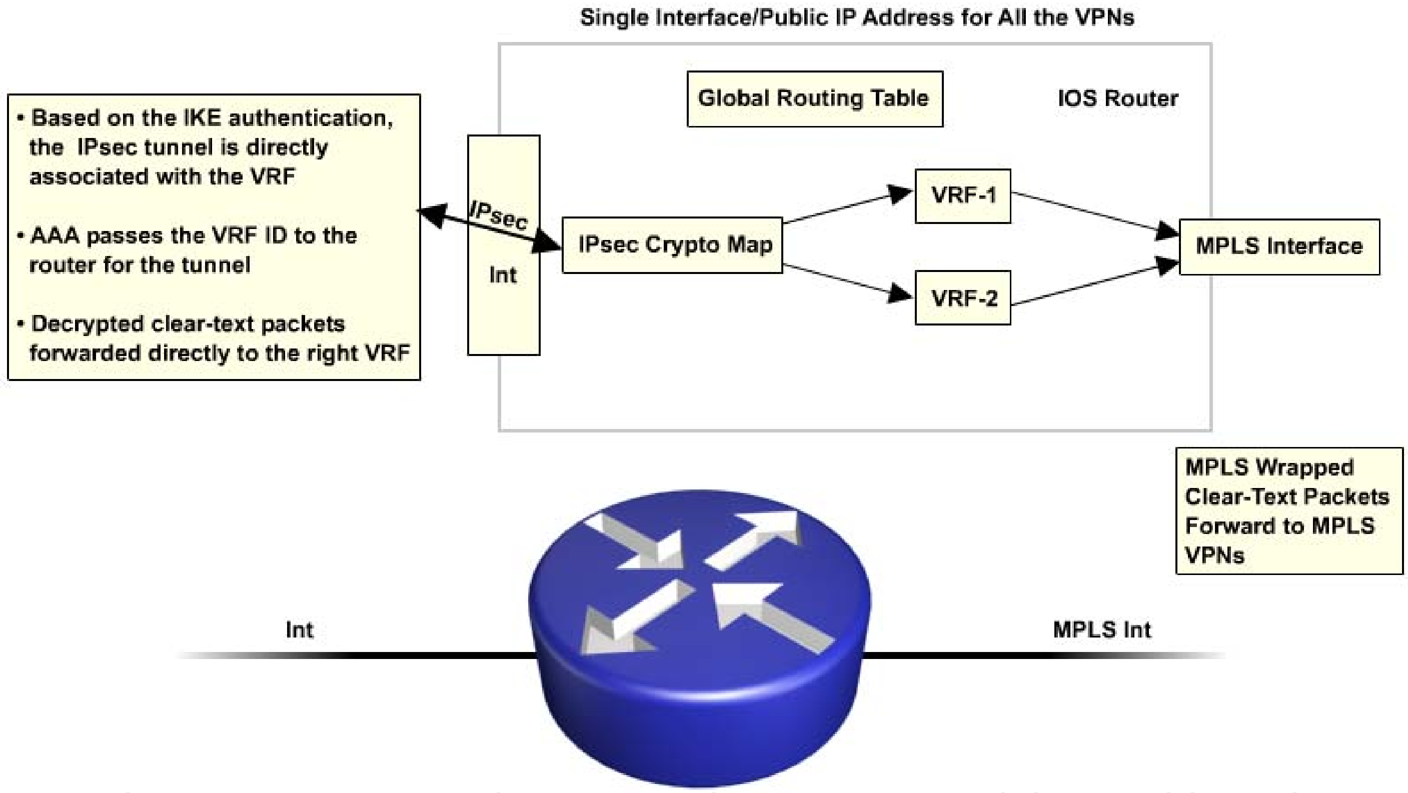


# VRF-Aware IPsec



## Packet Flow

**No Limitations! Works for Both Site-to-Site and Client-to-Concentrator Type of IPsec Tunnels.**




By implementing VRF-aware IKE/IPsec solution only one public IP address is needed to terminate IPsec tunnels from different VPN customers without penalty of additional encapsulation overhead. Based on the IKE authentication, the IPsec tunnel is directly associated with the VRF. AAA passes the VRF ID to the router for the tunnel.

# Agenda



- **Create L3-VPN service**
  - **Create VPN**
  - **Internet access**
  - **Security considerations**
  - **BGP advanced features**

***Yogesh Jiandani***

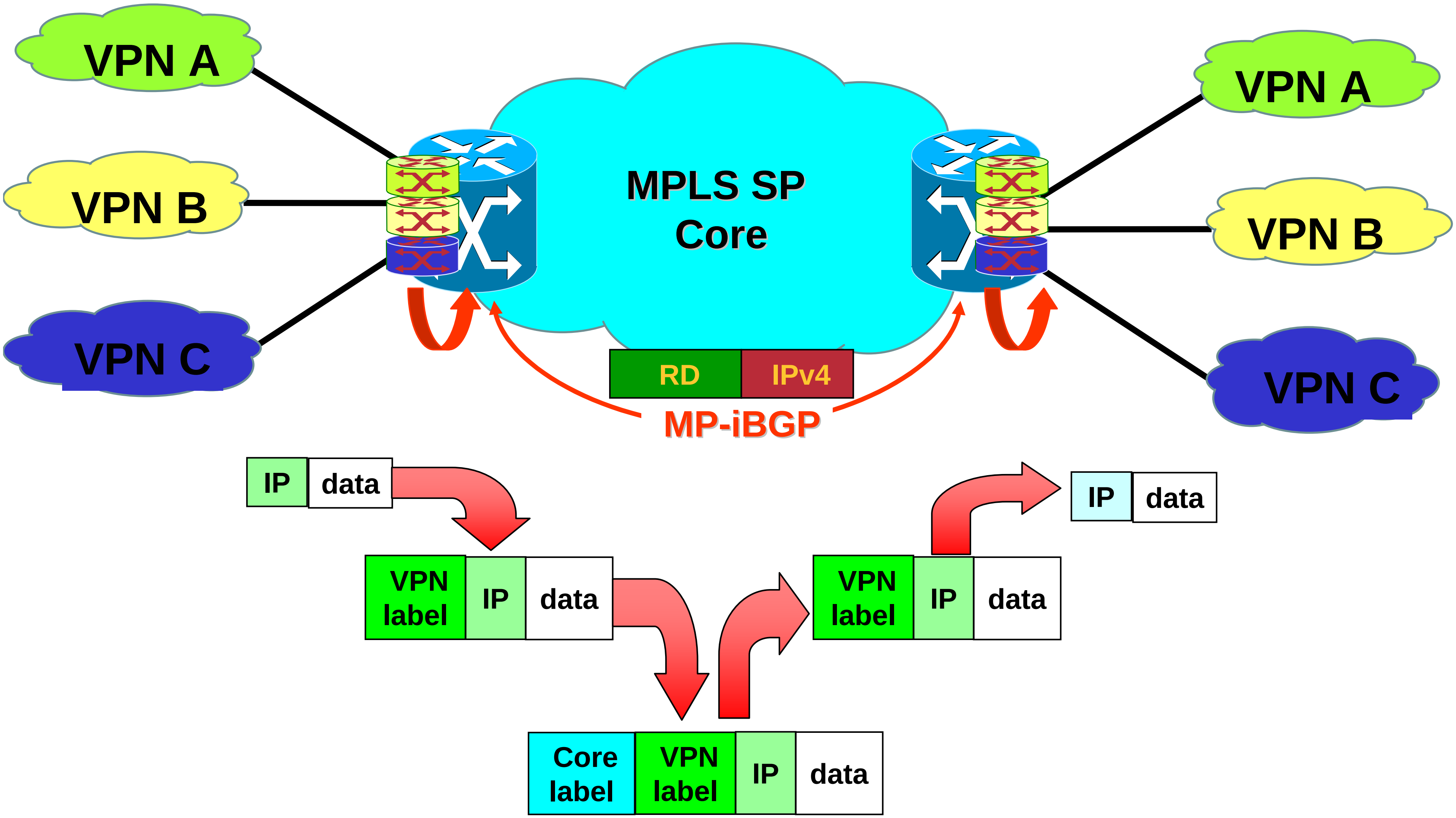
# General VPN Security Requirements

Cisco.com

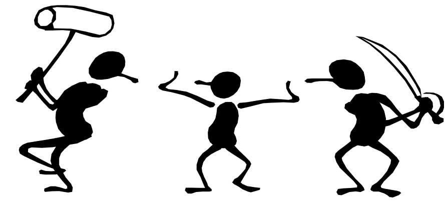
- Address Space and Routing Separation
- Hiding of the MPLS Core Structure
- Resistance to Attacks
- Impossibility of VPN Spoofing

# A hierarchy of Labels

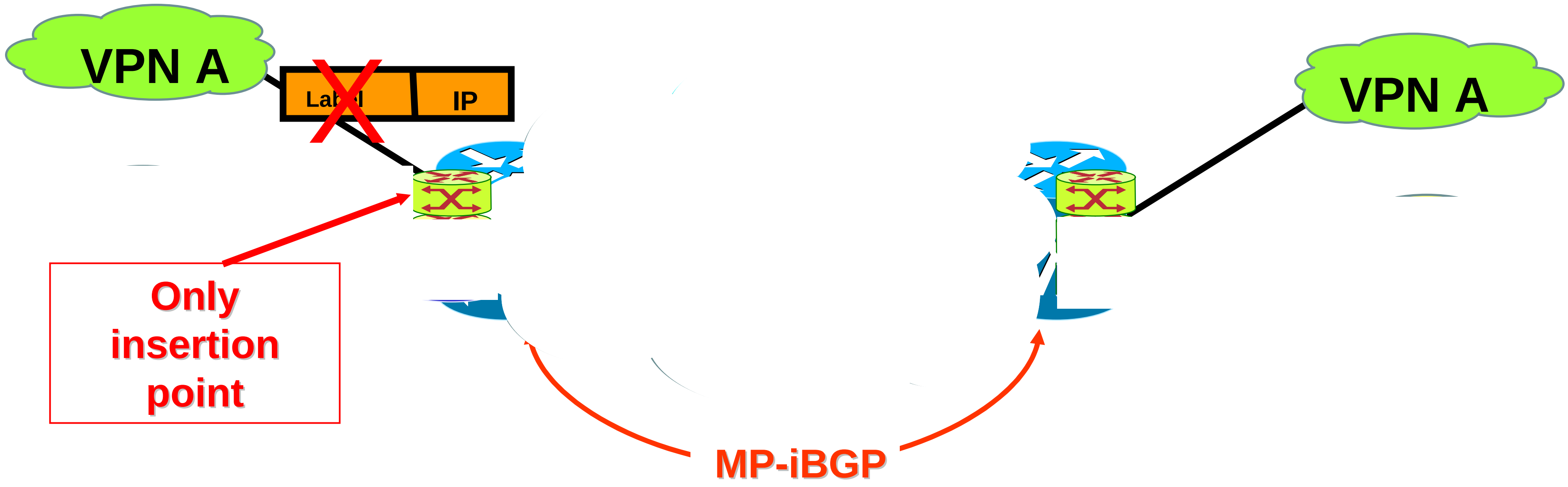




# Attack MPLS VPN








- **Where can you attack?**
  **Address and Routing Separation, thus:**
  ***Only Attack point: peering PE***
- **How?**
  **- Intrusions**
  **(telnet, SNMP, …, routing protocol)**
  **- DoS**

# Security Recommendations Conformity to “draft-behringer-mpls-security-10”



- Protect routing protocols from CE to PE:
  - Use static when possible
  - Using ACL (source is only CE)
  - MD5 authentication
  - BGP [RFC2385], OSPF [RFC2154], RIP2 [RFC2082], EIGRP
  - BGP dampening, filtering, maximum-prefix
- Protect PE ressources
  - VRF number of routes limitation
  - CAR (Commited Access Rate) to control traffic (specially UDP)
- Validate CE-CE exchanges thru PEs
  - CE may write/check BGP-Community with customer identification

***MPLS is as secured as Frame/Relay or ATM (Miercom / Gartner / …)***

# Limiting the Number of Prefixes Received from a BGP Neighbor



router(config-router-af)#

**neighbor *ip-address* maximum-prefix *maximum* [*threshold*] [warning-only]**

- **Controls how many prefixes can be received from a neighbor**
- **Optional threshold parameter specifies the percentage where a warning message is logged (default is 75%)**
- **Optional warning-only keyword specifies the action on exceeding the maximum number (default is to drop neighborship)**

# Configuring VRF Route Limit



router(config-vrf)#

**maximum route *number* { *warning-percent* | warn-only}**

- This command configures the maximum number of routes accepted into a VRF:
  - *Number* is the route limit for the VRF.
  - *Warning-percent* is the percentage value over which a warning message is sent to syslog.
  - With *warn-only* the PE continues accepting routes after the configured limit.
- Syslog messages generated by this command are rate-limited.

# MD5 Authentication – OSPF and BGP

Cisco.com

- **MD5 authentication between two peers**

  **password must be known to both peers**

- **OSPF authentication**

  `area <`*`area-id`*`> authentication message-digest`
  **(whole area)**

  `ip ospf message-digest-key 1 md5 <`*`key`*`>`

- **BGP neighbor authentication**

  `neighbor 169.222.10.1 password v61ne0qkel33&`

# BGP Dampening

Cisco.com

- **Fixed damping**

```
router bgp 100

bgp dampening [<half-life> <reuse-value> <suppress-
penalty> <maximum suppress time>]
```

- **Selective and variable damping**

```
bgp dampening [route-map <name>]

route-map <name> permit 10

 match ip address prefix-list FLAP-LIST

 set dampening [<half-life> <reuse-value> <suppress-
penalty> <maximum suppress time>]

ip prefix-list FLAP-LIST permit 192.0.2.0/24 le 32
```

# IP TTL Propagation – Stop discovery of the network



router(config)#

```
no mpls ip propagate-ttl
```

- **By default, IP TTL is copied into label header at label imposition and label TTL is copied into IP TTL at label removal.**
- **This command disables IP TTL and label TTL propagation.**
  - **TTL value of 255 is inserted in the label header.**
- **The TTL propagation has to be disabled on ingress and egress edge LSR.**

# IP TTL Propagation—Extended Options



router(config)#

```
no mpls ip propagate-ttl [forwarded | local]
```

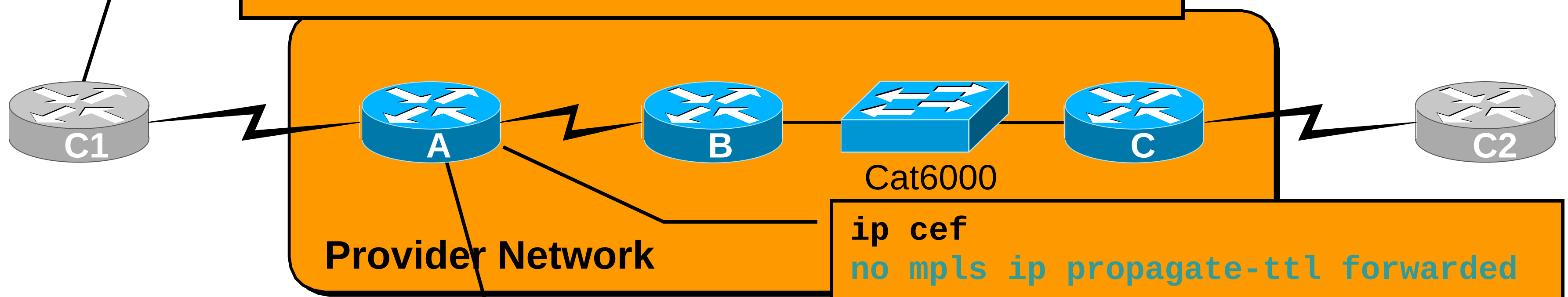
**Selectively disables IP TTL propagation for:**

- **Forwarded traffic (traceroute does not work for transit traffic labeled by this router)**
- **Local traffic (traceroute does not work from the router but works for transit traffic labeled by this router)**

# Disabling IP TTL Propagation for Customer Traffic



```
C1#trace C2.cust.com
Tracing the route to C2.cust.com
  1 A.provider.net   44 msec 36 msec 32 msec
  2 C2.cust.com     180 msec *   181 msec
```



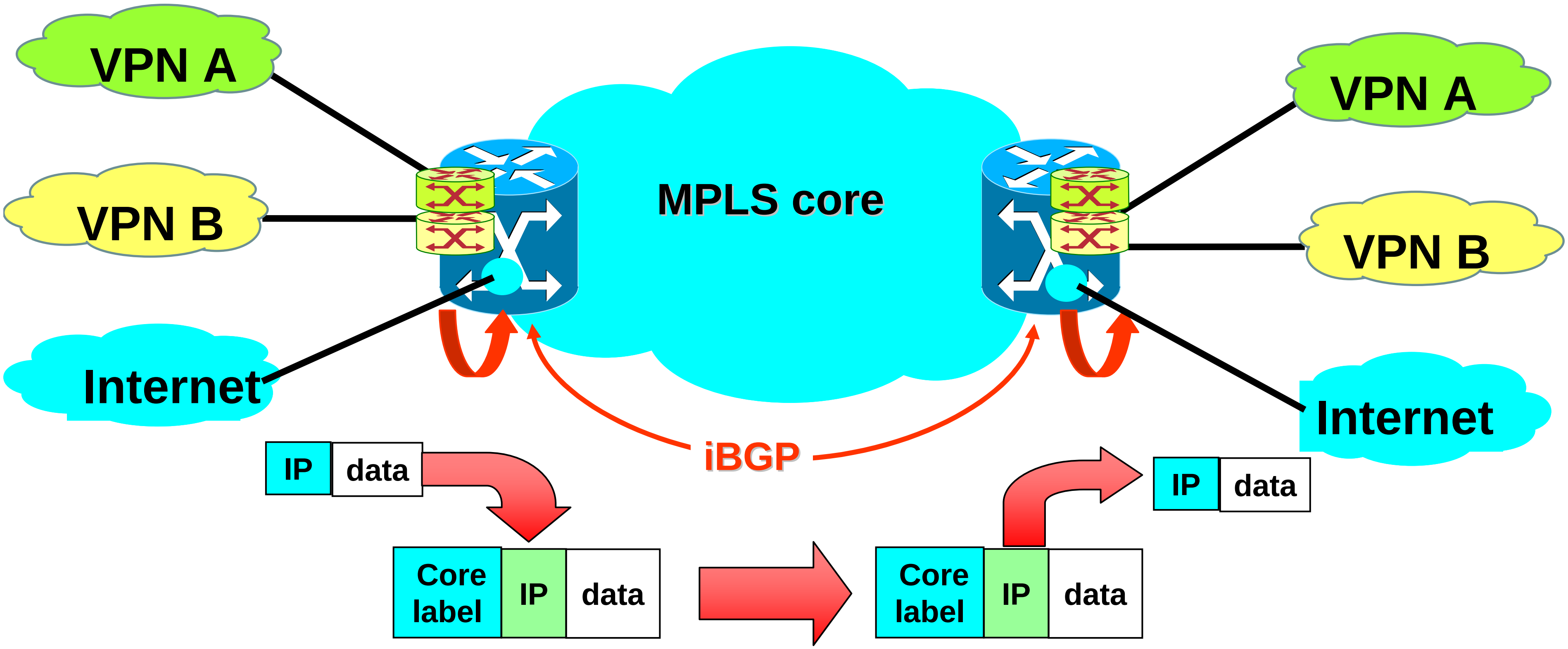

```
ip cef
no mpls ip propagate-ttl forwarded

interface hssi 1/0
 mpls ip
```

```
A#trace C2.cust.com
Tracing the route to C2.cust.com
 1 B.provider.net 164 msec 132 msec 128 msec
 2 C.provider.net 148 msec 156 msec 152 msec
 3 C2.cust.com    180 msec *   181 msec
```

**Selective IP TTL propagation hides the provider network from the customer but still allows troubleshooting.**

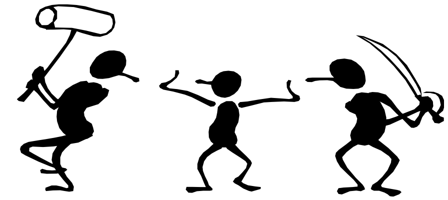
# Internet core transport isolation
# A switching of Labels



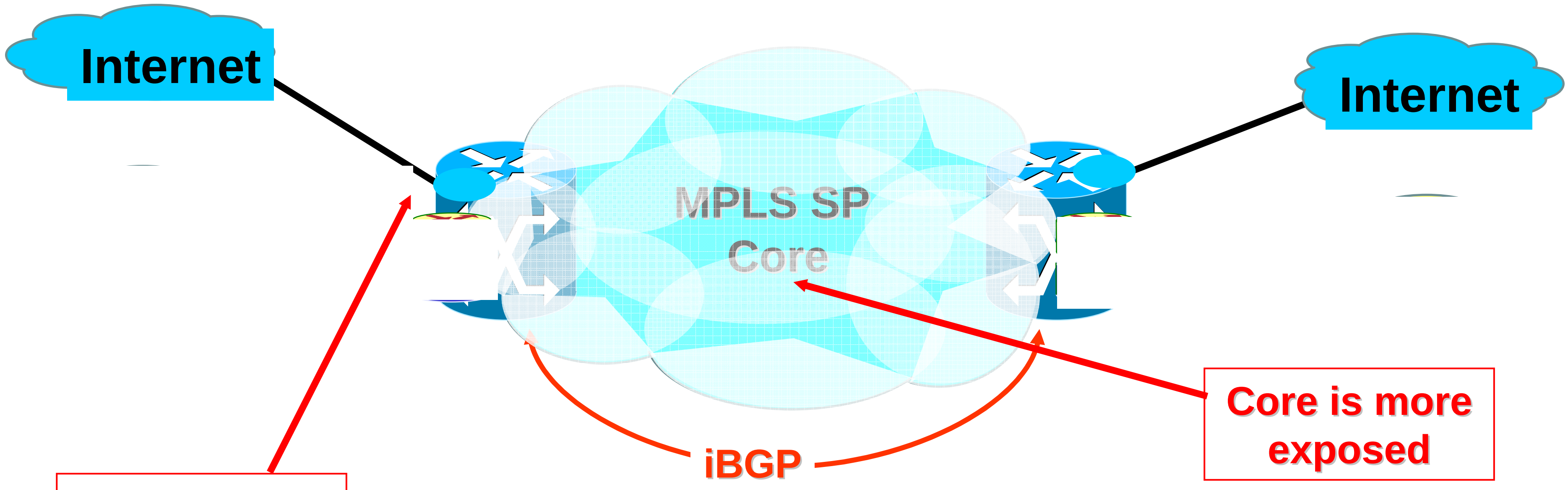



**P routers do not participate in IP routing**

# Attack MPLS VPN from Internet



Cisco.com




- **Where can you attack?**

  **Address and Routing Separation, thus:**

  ***Main Attack point: peering PE***

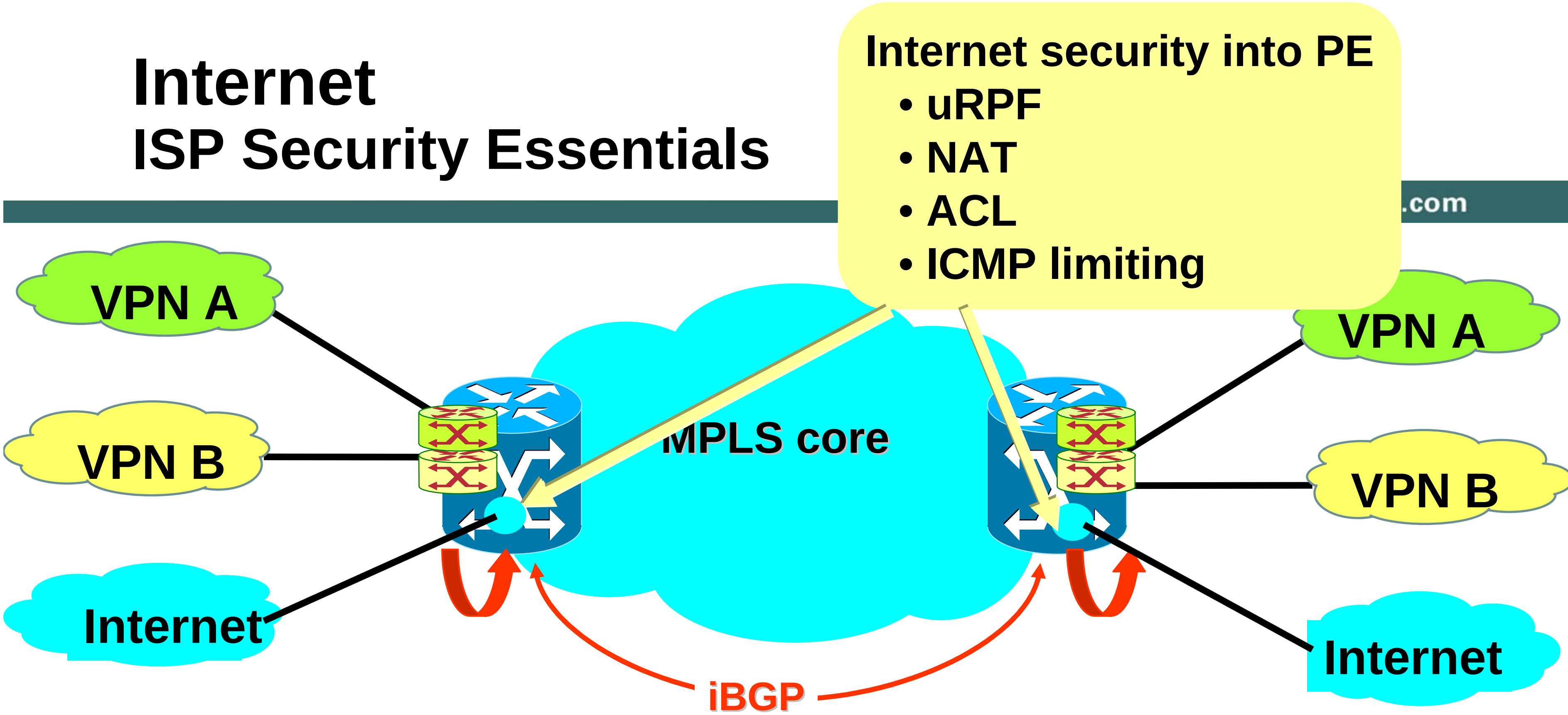
  ***P should be safe if not into iBGP full-routing***

- **How?**

  **- Intrusions**
  **(telnet, SNMP, …, routing protocol)**

  **- DoS**

# Internet
## ISP Security Essentials




- The “bible” for Core Security
- Available as book, and on FTP:

  ftp://ftp-eng.cisco.com/cons/isp/security

- How to secure the core

  Security for devices, routing, traffic, management, …

# Summary
# MPLS as secure as any L2



- **Meircom performed testing that proved that MPLS-VPNs have met or exceeded all of the security characteristics of a comparable layer two based VPN such as Frame-Relay or ATM**
- **Address space and routing separation**
  - **Unique addressing utilizing VPN-IPv4 addresses**
  - **Routing separation by the use of VRFs**
- **Service Providers core structure is not revealing**
  - **Only information shared is already part of the VRF**
- **The network is resistant to attacks**
  - **Mechanisms in place to limit the impact of DoS attacks**

# If more security than L2 is required
# Use IPsec if you don't trust enough the core

Cisco.com

- Address space separation
- Traffic separation
- Routing separation
- Authentication
- Confidentiality
- Data integrity

"VPN"

"Secure VPN"

# Agenda

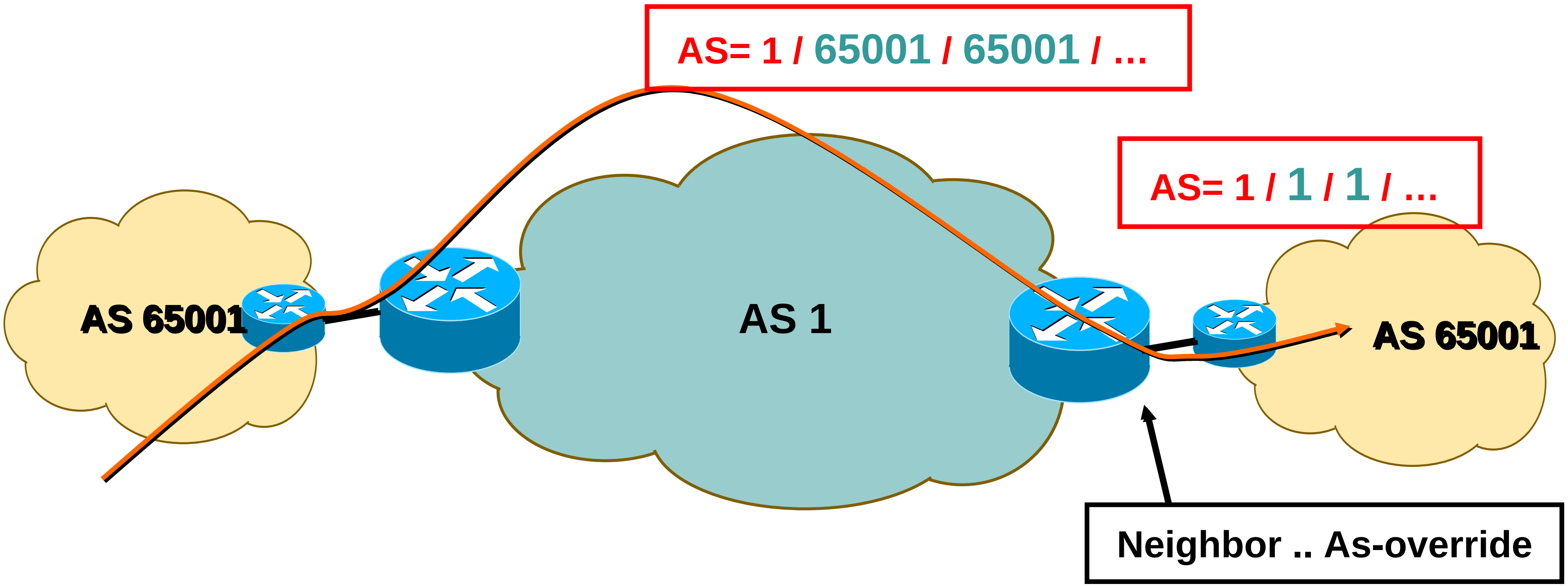
Cisco.com

- **Create L3-VPN service**
  - **Create VPN**
  - **Internet access**
  - **Security considerations**
  - **BGP advanced features**

***Vagish Dwivedi***

# AS-Override

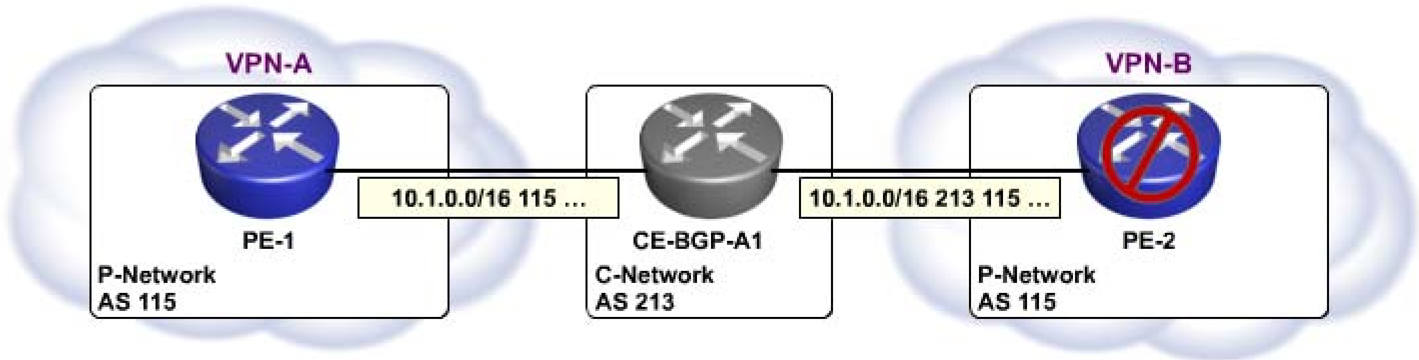





**Allows all sites of the same customer to be into the same splitted AS**

# Allowas-in

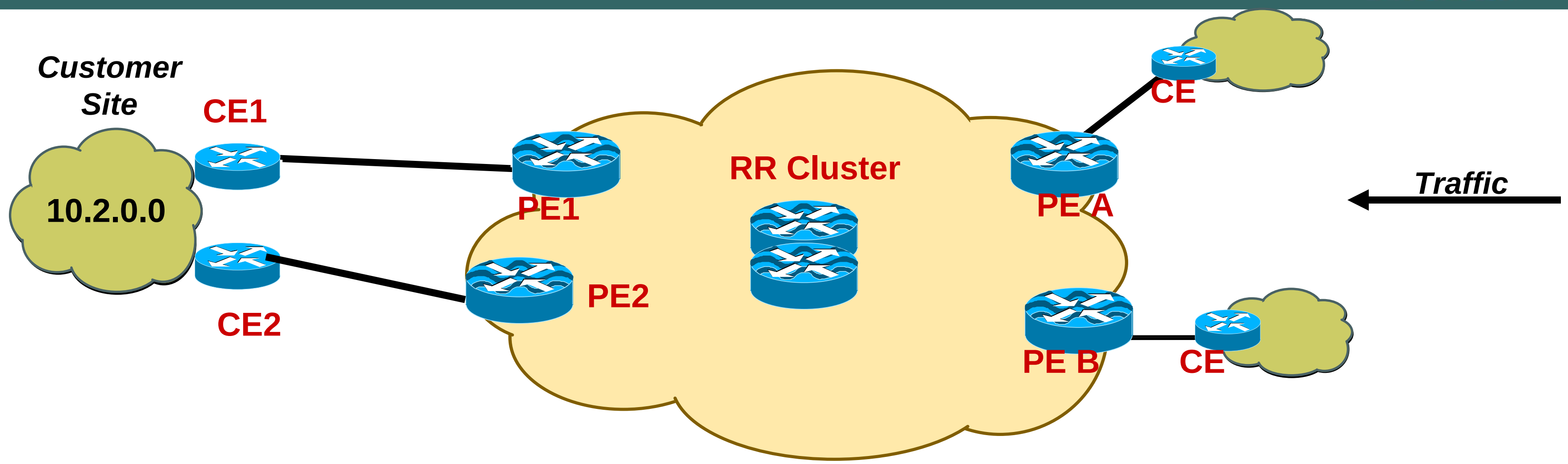





- VPN A  connected to VPN-B via CE-BGP-A1
- CE router has two connections to AS 115
- PE-1 announces network 10.1.0.0/16 to CE-BGP-A1
- PE-2 drops the update because it's AS number is already in the AS-Path
- AS-Override is needed on CE-BGP-A1
- This modifies BGP loop prevention mechanism

# Load Balancing – iBGP Multipath for MPLS/VPN

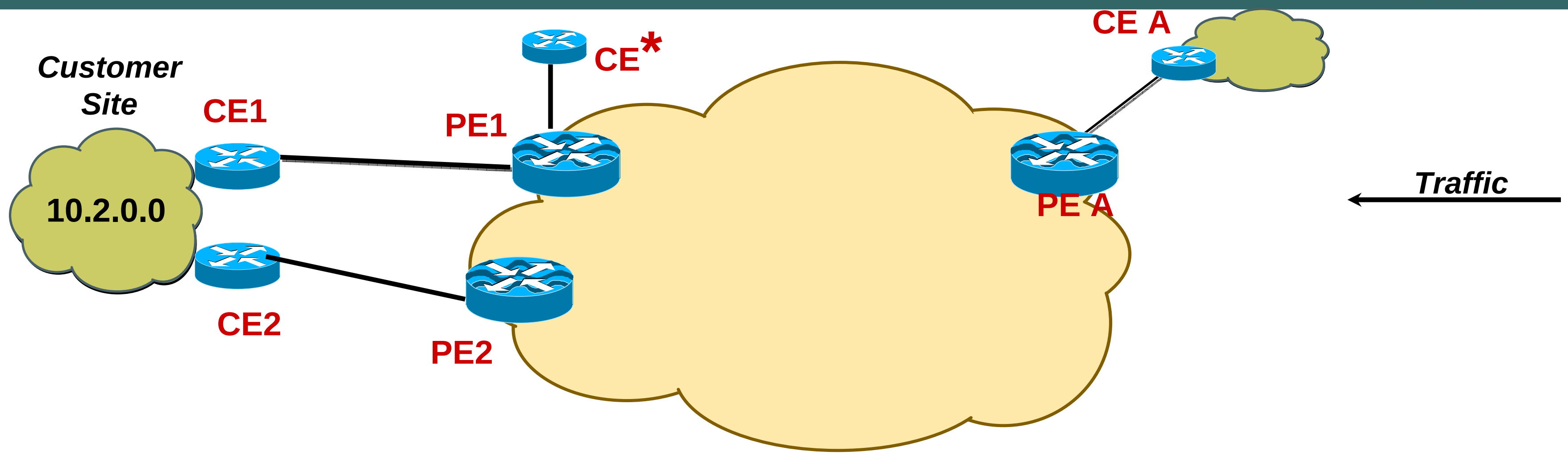





- iBGP Multipath
  - PE A (and PE B) see two next-hop PEs towards 10.2/16
  - Installs all two next-hops in the BGP table
  - Must use different RDs on PE1, PE2, as RR will select vpnv4 best path
  - CEF resolves next-hops and load balances over them (via src/dest hash)
    - Limit of 6 paths for BGP multipath !!!
- Note – IGP cost should be equal by default.
  - Can use MPLS TE forwarding adjacency to impose equal IGP cost
  - Can do unequal IGP load balancing (max-paths ibgp unequal-cost)

# PE to CE Load Balancing
# eiBGP Multipath for MPLS/VPN

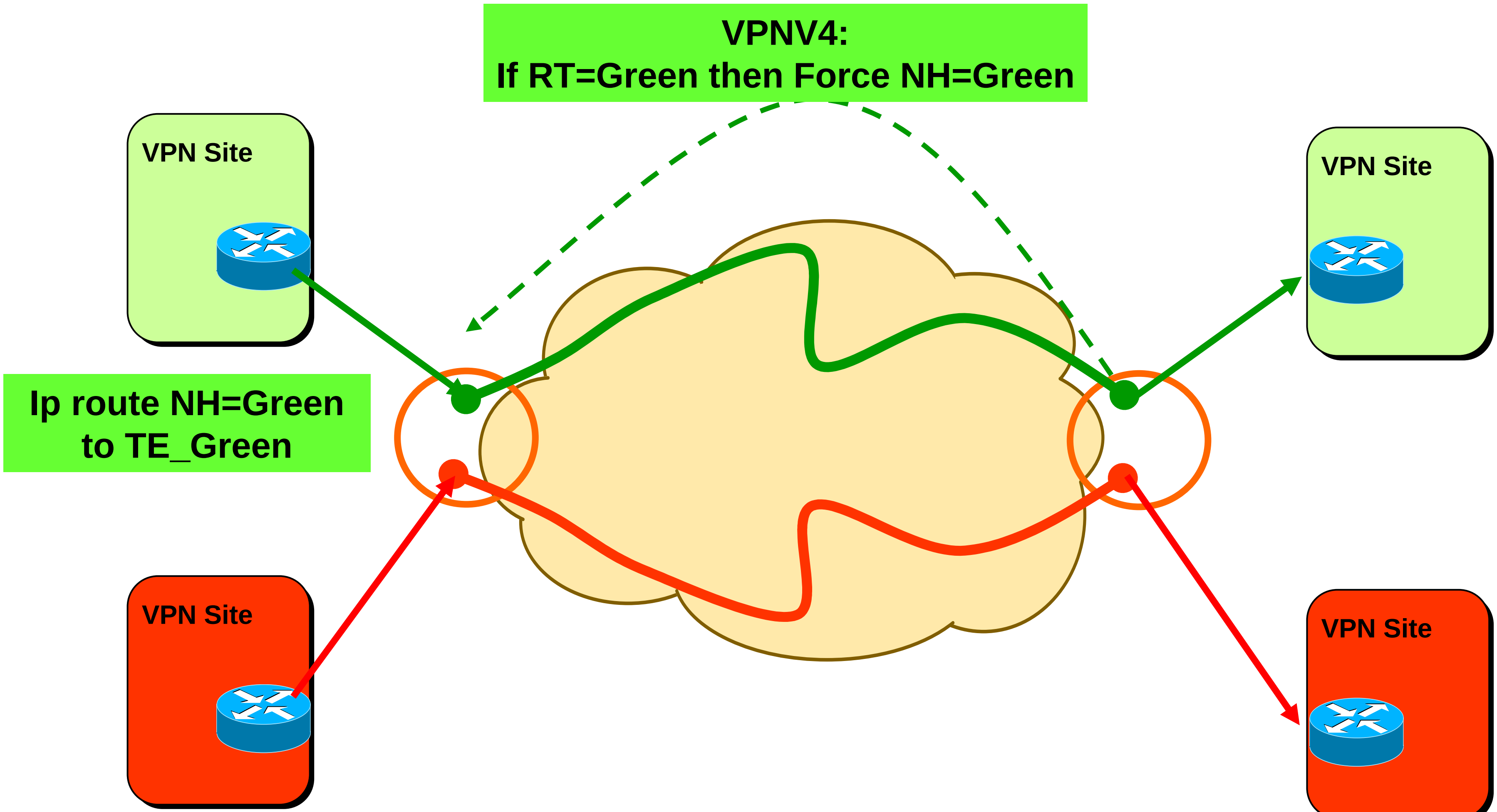





- Consider implications of load balancing traffic from CE* towards 10.2/16
- PE1 will have learnt 10.2/16 from CE1 and from PE2
  - CE1 via eBGP; PE2 via MP-iBGP
- eBGP path will always be preferred (lower admin distance)
- eiBGP multipath (on PE1) allows load balancing thru PE1 and PE2 toward 10.2/16

# Per VPN TE

Cisco.com



# Per VPN TE
## *Or even inside VPN TE*



**One one side:**

```
address-family vpnv4
  neighbor 1.1.12.1 activate
  neighbor 1.1.12.1 send-community extended
  neighbor 1.1.12.1 route-map set-pref-nh out

ip extcommunity-list 70 permit rt 10:70
route-map set-pref-nh permit 10
 match extcommunity 70
 set ip next-hop 10.52.52.52
```

**And in addition:**

```
ip vrf green
 rd 10:2
 export map Set_RT70
 route-target both 10:2
!
access-list 1 permit 100.10.2.12
!
route-map Set_RT70 permit 10
 match ip address 1
 set extcommunity rt  10:70 additive
```

« Or even per Subnet into the VRF »

**On the other side:**

```
ip route 10.52.52.52 255.255.255.255 Tunnel70
```

**And why not to use Source-@ VPN selection !**

# Agenda



- **Manage L3-VPN service a.k.a L3 OAM**

***Yogesh Jiandani***

# NMS Best Practices for MPLS VPN Operational Efficiency and Effectiveness



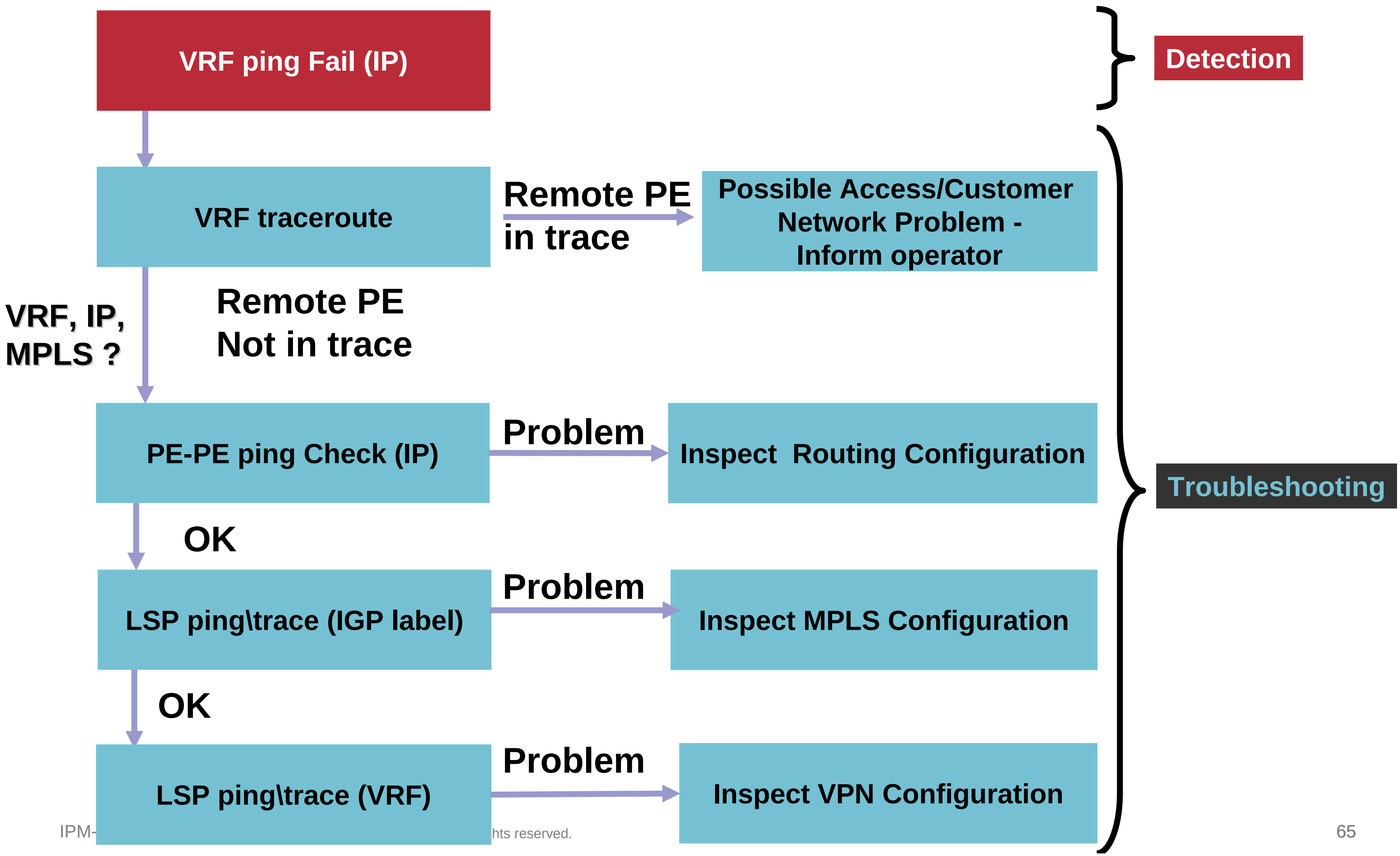
- **Automate, Automate, Automate … however……**
- **The MPLS VPN Assurance/Troubleshooting Process**

  **The Problems**

  **Troubleshooting Workflows**

# Troubleshooting Workflow - VRF data plane

Cisco.com



# show ip vrf



```
Router#show ip vrf
 Name                      Default RD           Interfaces
 SiteA2                    103:30               Serial1/0.20
 SiteB                     103:11               Serial1/0.100
 SiteX                     103:20               Ethernet0/0
Router#
```

# show ip vrf detail

Cisco.com

```
Router#show ip vrf detail
VRF SiteA2; default RD 103:30
  Interfaces:
    Serial1/0.20
  Connected addresses are not in global routing table
  No Export VPN route-target communities
  Import VPN route-target communities
    RT:103:10
  No import route-map
  Export route-map: A2
VRF SiteB; default RD 103:11
  Interfaces:
    Serial1/0.100
  Connected addresses are not in global routing table
  Export VPN route-target communities
    RT:103:11
  Import VPN route-target communities
    RT:103:11                RT:103:20
  No import route-map
  No export route-map
```

# show ip vrf interfaces



```
Router#show ip vrf interfaces
Interface         IP-Address       VRF          Protocol
Serial1/0.20      150.1.31.37      SiteA2       up
Serial1/0.100     150.1.32.33      SiteB        up
Ethernet0/0       192.168.22.3     SiteX        up
```

# show ip route vrf

Cisco.com

```
Router#show ip route vrf SiteA2
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2,
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route

Gateway of last resort is not set

O    203.1.20.0/24 [110/782] via 150.1.31.38, 02:52:13, Serial1/0.20
     203.1.2.0/32 is subnetted, 1 subnets
O       203.1.2.1 [110/782] via 150.1.31.38, 02:52:13, Serial1/0.20
     203.1.1.0/32 is subnetted, 1 subnets
B       203.1.1.1 [200/1] via 192.168.3.103, 01:14:32
B    203.1.135.0/24 [200/782] via 192.168.3.101, 02:05:38
B    203.1.134.0/24 [200/1] via 192.168.3.101, 02:05:38
B    203.1.10.0/24 [200/1] via 192.168.3.103, 01:14:32

… rest deleted …
```

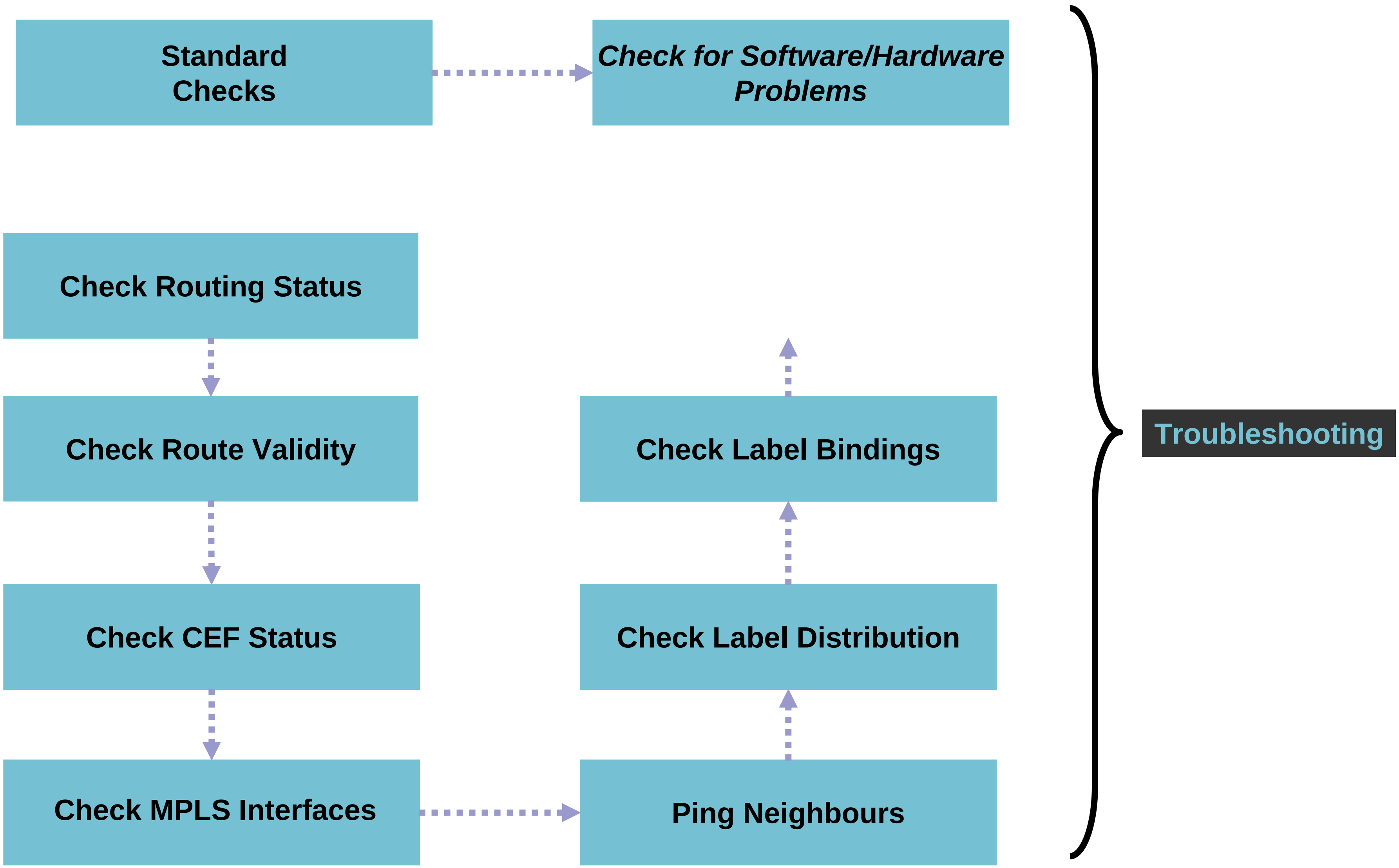
# show ip bgp vpnv4 vrf neighbor



```
Router#show ip bgp vpnv4 vrf SiteB neighbors
BGP neighbor is 150.1.32.34,  vrf SiteB,  remote AS 65032, external link
  BGP version 4, remote router ID 203.2.10.1
  BGP state = Established, up for 02:01:41
  Last read 00:00:56, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 549 messages, 0 notifications, 0 in queue
  Sent 646 messages, 0 notifications, 0 in queue
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 30 seconds

 For address family: VPNv4 Unicast
  Translates address family IPv4 Unicast for VRF SiteB
  BGP table version 416, neighbor version 416
  Index 4, Offset 0, Mask 0x10
  Community attribute sent to this neighbor
  2 accepted prefixes consume 120 bytes
  Prefix advertised 107, suppressed 0, withdrawn 63

… rest deleted …
```

# Example Workflow
# Troubleshooting the MPLS core





# show mpls interface

Cisco.com

```
Router#show mpls interfaces [interface] [detail]
```

| Interface | IP | Tunnel | Operational |
|---|---|---|---|
| Ethernet1/1/1 | Yes (tdp) | No | No |
| Ethernet1/1/2 | Yes (tdp) | Yes | No |
| Ethernet1/1/3 | Yes (tdp) | Yes | Yes |
| POS2/0/0 | Yes (tdp) | No | No |
| ATM0/0.1 | Yes (tdp) | No | No (ATM labels) |
| ATM3/0.1 | Yes (ldp) | No | Yes (ATM labels) |
| ATM0/0.2 | Yes (tdp) | No | Yes |

# show cef interface



```
Router#show cef interface serial 1/0.20
Serial1/0.20 is up (if_number 18)
  Internet address is 150.1.31.37/30
  ICMP redirects are always sent
  Per packet loadbalancing is disabled
  IP unicast RPF check is disabled
  Inbound access list is not set
  Outbound access list is not set
  IP policy routing is disabled
  Interface is marked as point to point interface
  Hardware idb is Serial1/0
  Fast switching type 5, interface type 64
  IP CEF switching enabled
  IP CEF VPN Fast switching turbo vector
  VPN Forwarding table "SiteA2"
  Input fast flags 0x1000, Output fast flags 0x0
  ifindex 3(3)
  Slot 1 Slot unit 0 VC -1
  Transmit limit accumulator 0x0 (0x0)
  IP MTU 1500
```

# show mpls forwarding-table

Cisco.com

```
Router#show mpls forwarding-table detail
Local  Outgoing    Prefix            Bytes label  Outgoing   Next Hop
label    label or VC   or Tunnel Id       switched    interface
26     Unlabeled    192.168.3.3/32    0           Se1/0.3    point2point
     MAC/Encaps=0/0, MTU=1504, label Stack{}
27     Pop label     192.168.3.4/32    0           Se0/0.4    point2point
     MAC/Encaps=4/4, MTU=1504, label Stack{}
     20618847
28     29          192.168.3.4/32    0           Se1/0.3    point2point
     MAC/Encaps=4/8, MTU=1500, label Stack{29}
     18718847 0001D000
```

# show mpls ldp bindings

Cisco.com

```
Router#show mpls ldp bindings
  lib entry: 192.168.3.1/32, rev 9
     local binding:  label: 28
     remote binding: lsr: 19.16.3.3:0, label: 28
  lib entry: 192.168.3.2/32, rev 8
     local binding:  label: 27
     remote binding: lsr: 19.16.3.3:0, label: 27
  lib entry: 192.168.3.3/32, rev 7
     local binding:  label: 26
     remote binding: lsr: 19.16.3.3:0, label: imp-
null(1)
  lib entry: 192.168.3.10/32, rev 6
     local binding:  label: imp-null(1)
     remote binding: lsr: 19.16.3.3:0, label: 26
```

# Agenda

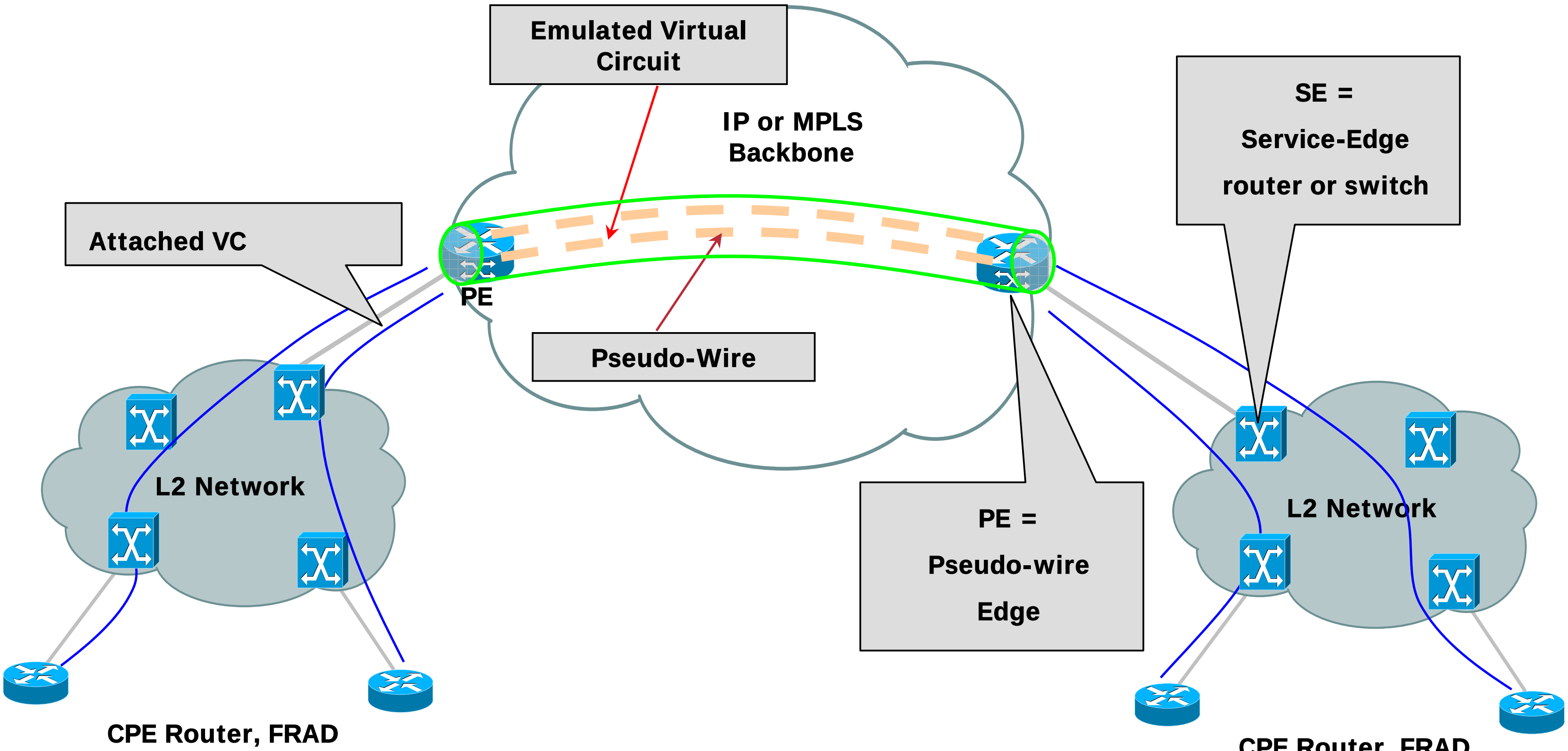


- **L2 transport over MPLS**
  - **AToM**
  - **EoMPLS**
  - **VPLS**
  - **H-VPLS**

***Yogesh Jiandani***

# IETF : L2transport (Pseudo-Wire Emulation Edge to Edge)






**L2transport over IP = L2TPext or L2TPv3**

**L2transport over MPLS = pwe3-mpls-cp**

# PWE3 MPLS based drafts



draft-ietf-pwe3-control-protocol-xx.txt

draft-ietf-pwe3-atm-encap-xx.txt
draft-ietf-pwe3-frame-relay-xx.txt
draft-ietf-pwe3-ethernet-encap-xx.txt
draft-ietf-pwe3-hdlc-ppp-xx.txt

draft-ietf-pwe3-cesopsn-xx.txt
draft-ietf-pwe3-satop-xx.txt
draft-ietf-pwe3-sonet-xx.txt

draft-ietf-pwe3-vccv-xx.txt
draft-ietf-pwe3-mib-xx.txt

**Control plane**

**Data plane (L2 emulation)**

- ATM AAL5 PDU
- ATM cells (non AAL5 mode)
- FR PDU
- Ethernet
- 802.1Q (Ethernet VLAN)
- Cisco-HDLC (LAPD)
- PPP

**Circuit Emulation (L1 emulation)**

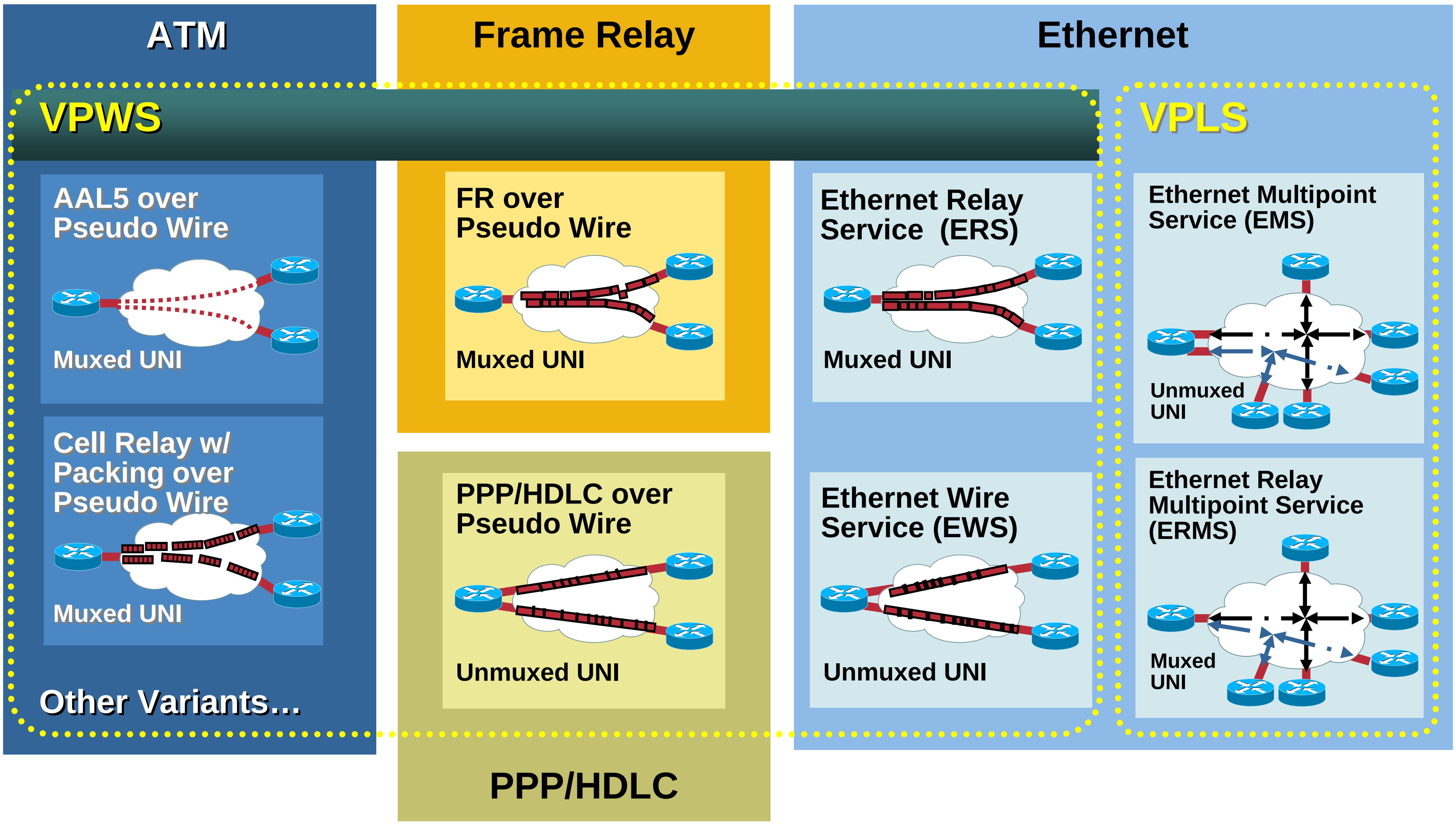
- Structured E1/T1
- Unstructured E1/T1/E3/DS3
- Unstructured STMx/OCx

**Management Plane**

- OAM's
- MIB's

# Layer 2 VPN Taxonomy

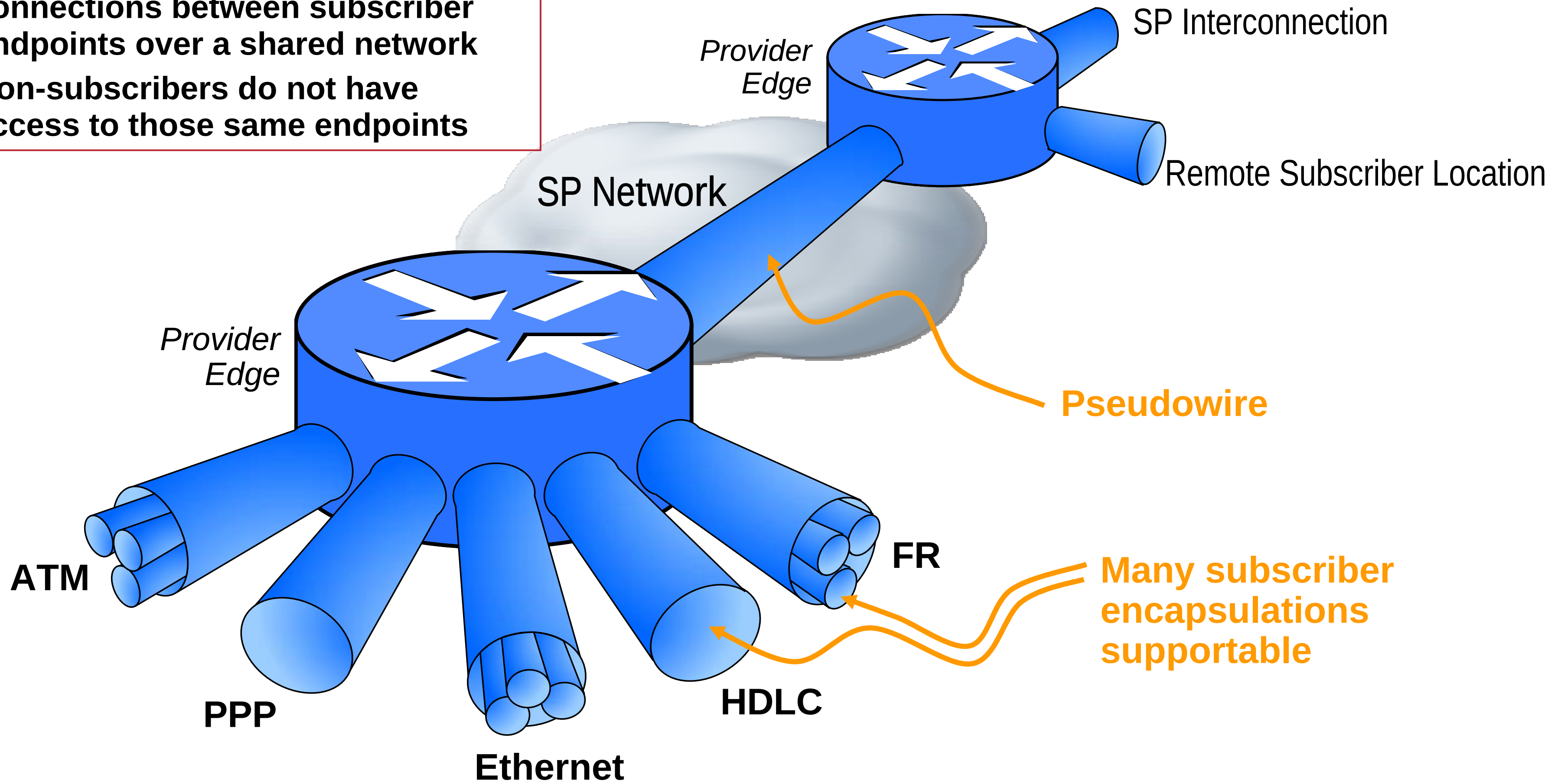




# What is an L2VPN?
# IETF's L2VPN Logical Context



- An L2VPN is comprised of switched connections between subscriber endpoints over a shared network
- Non-subscribers do not have access to those same endpoints




**Some Layer 1 frame encapsulations are transportable under the framework of L2VPN. This is acceptable because (unlike native L1) Frames can be dropped due to congestion.**

# Pseudo Wire – IETF Working Groups



| Internet Area | | Transport Area |
|---|---|---|
| **L2TPEXT** | **L2VPN** | **PWE3** |

**L2TP(v2 & v3)**

- Extensions to RFC2661
- Control Plane Operation
- AVPs
- Updated data plane
- Relevant MIBs

**VPLS, VPWS, IPLS**

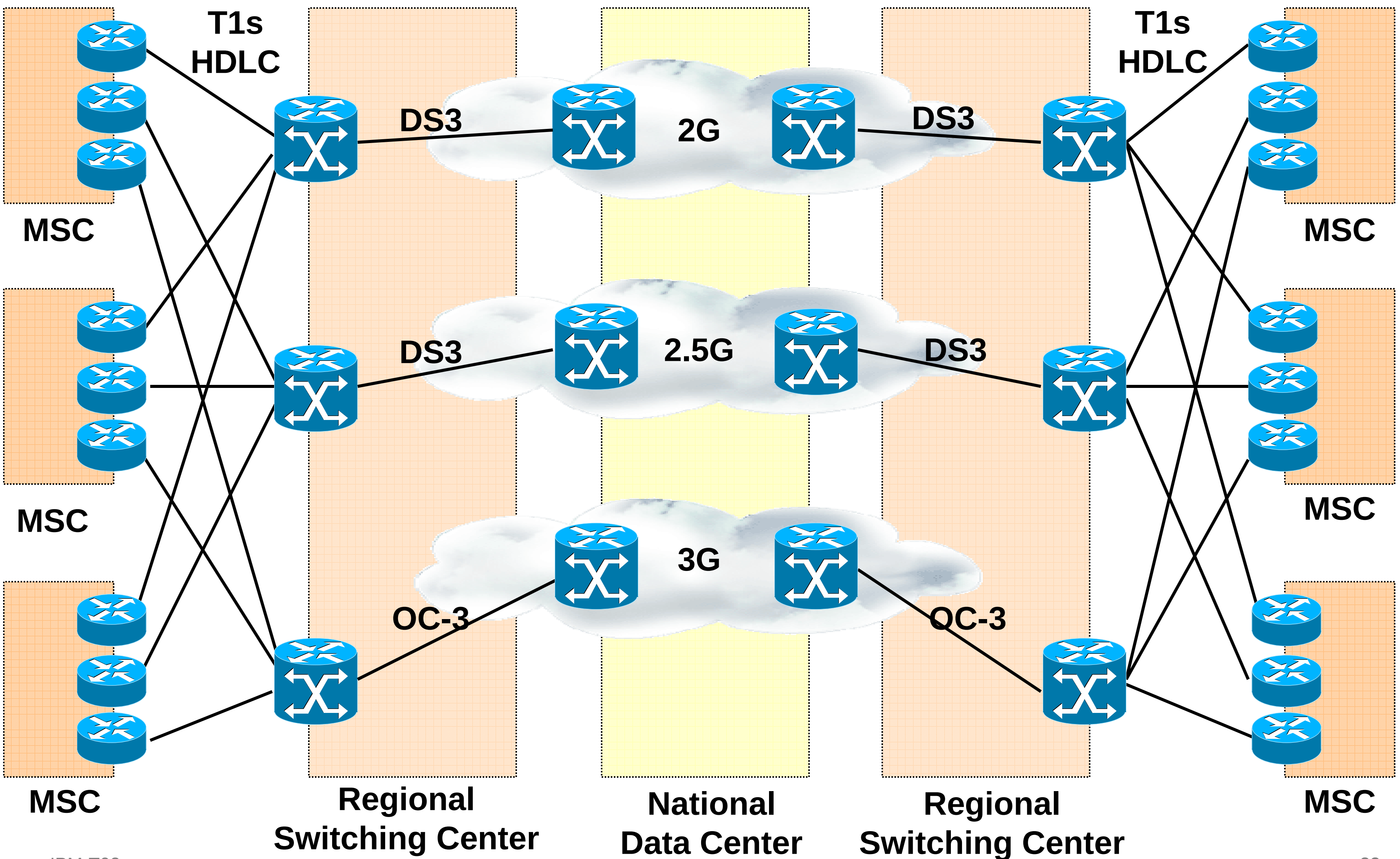
- Solution Architectures
- PE Discovery
- Signaling (with PWE3)
- L2VPN OAM extensions
- Relevant MIBs

**AToM**

- PWE3 Architecture
- PWE3 Requirements
- LDP Control Channel
- L2 Service Encap Specifics
- TDM, CES, etc.
- Relevant MIBs

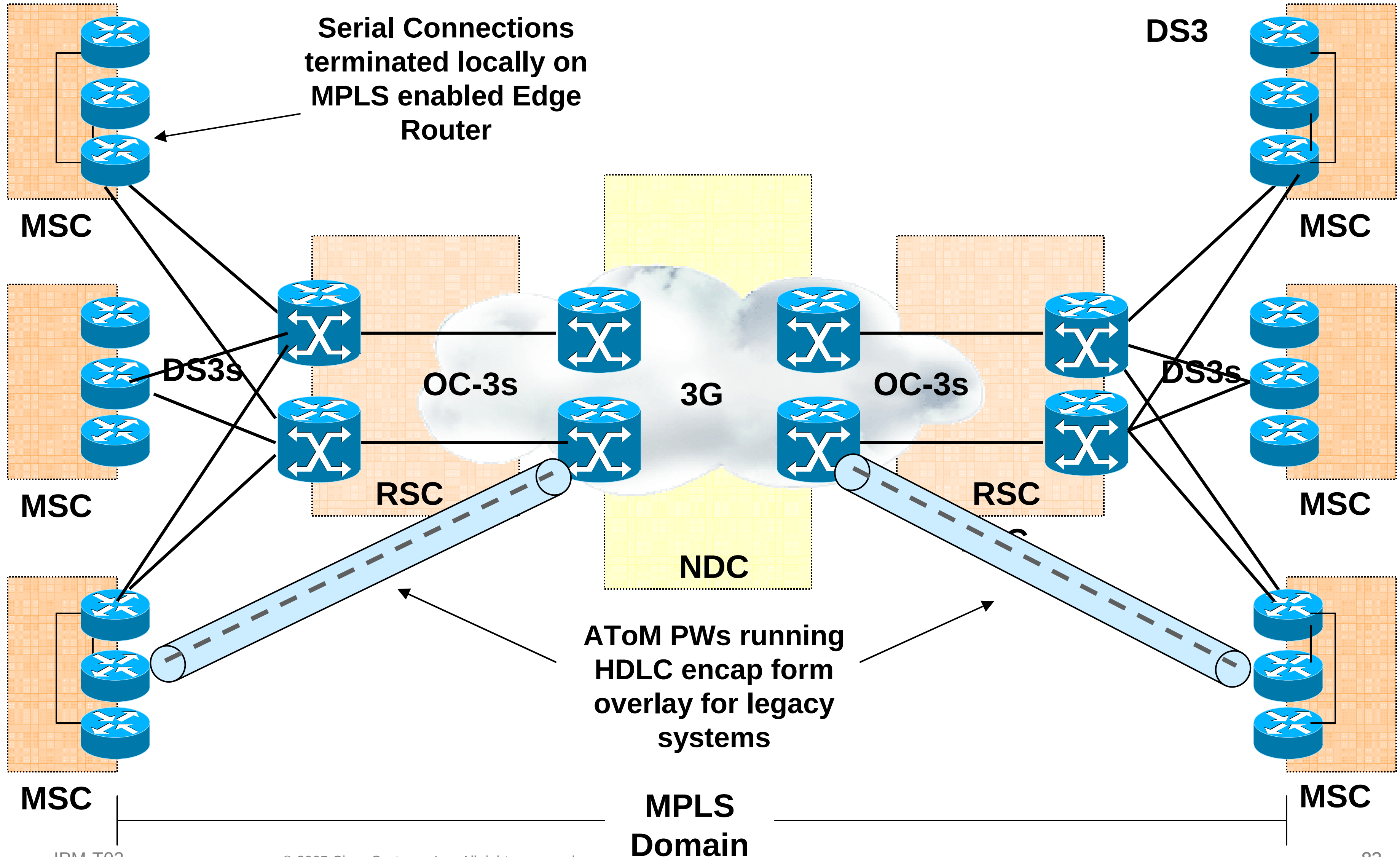
# L2VPNs: Pre-Network Consolidation





# L2VPNs: Post-Network Consolidation





# Agenda

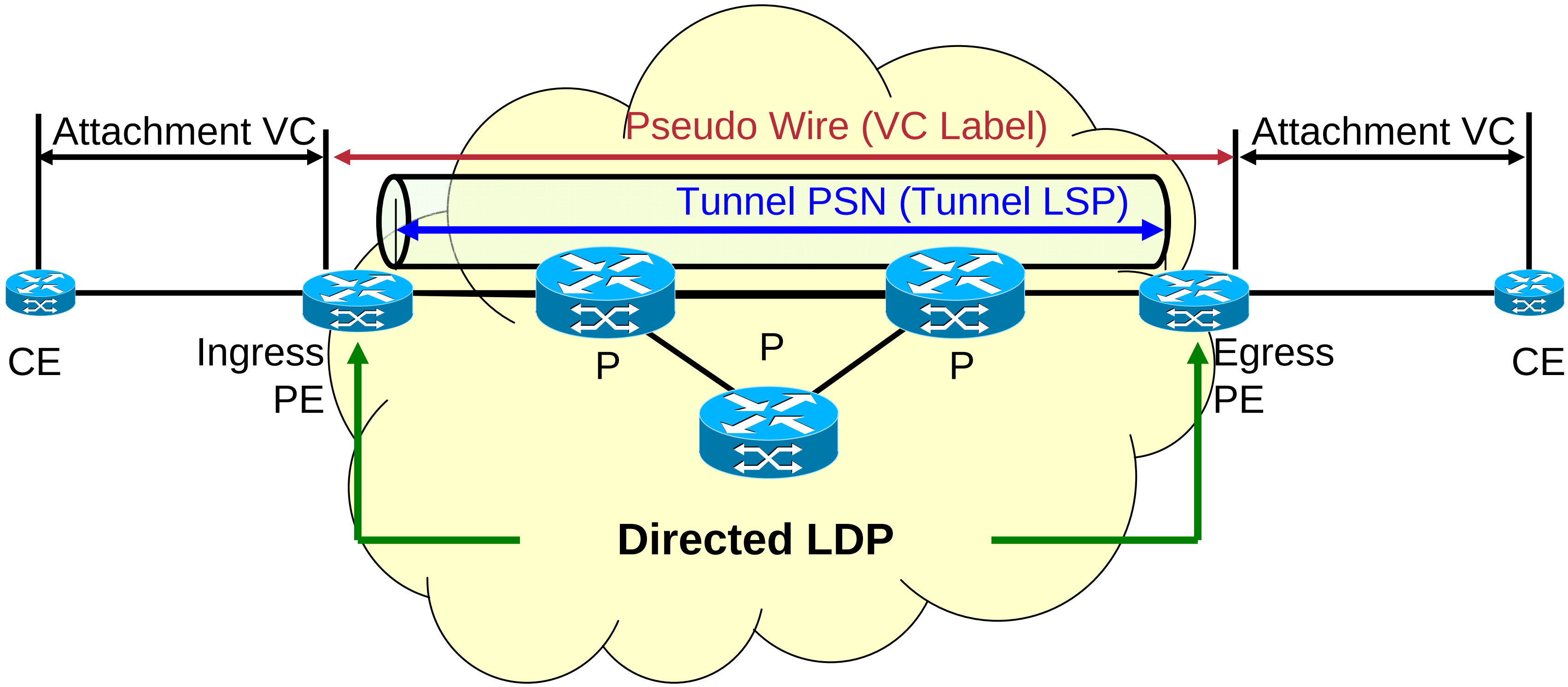
Cisco.com

- **L2 transport over MPLS**
  - **AToM**
  - **EoMPLS**
  - **VPLS**
  - **H-VPLS**

***Vagish Dwivedi***

# AToM semantic






- Ingress and egress interfaces (attachment VCs) are non-MPLS interfaces
- Ingress PE encapsulates into MPLS, egress PE de-encapsulates
- Label stack of two labels is used
  - Top-most label (“tunnel-label”) used for LSP PE to PE.
  - Second label (“VC-label”) identifies outgoing interface in egress PE
- LDP has been extended to carry VC-FEC
- A directed LDP session is used from PE to PE to exchange VC labels

# AToM Idea



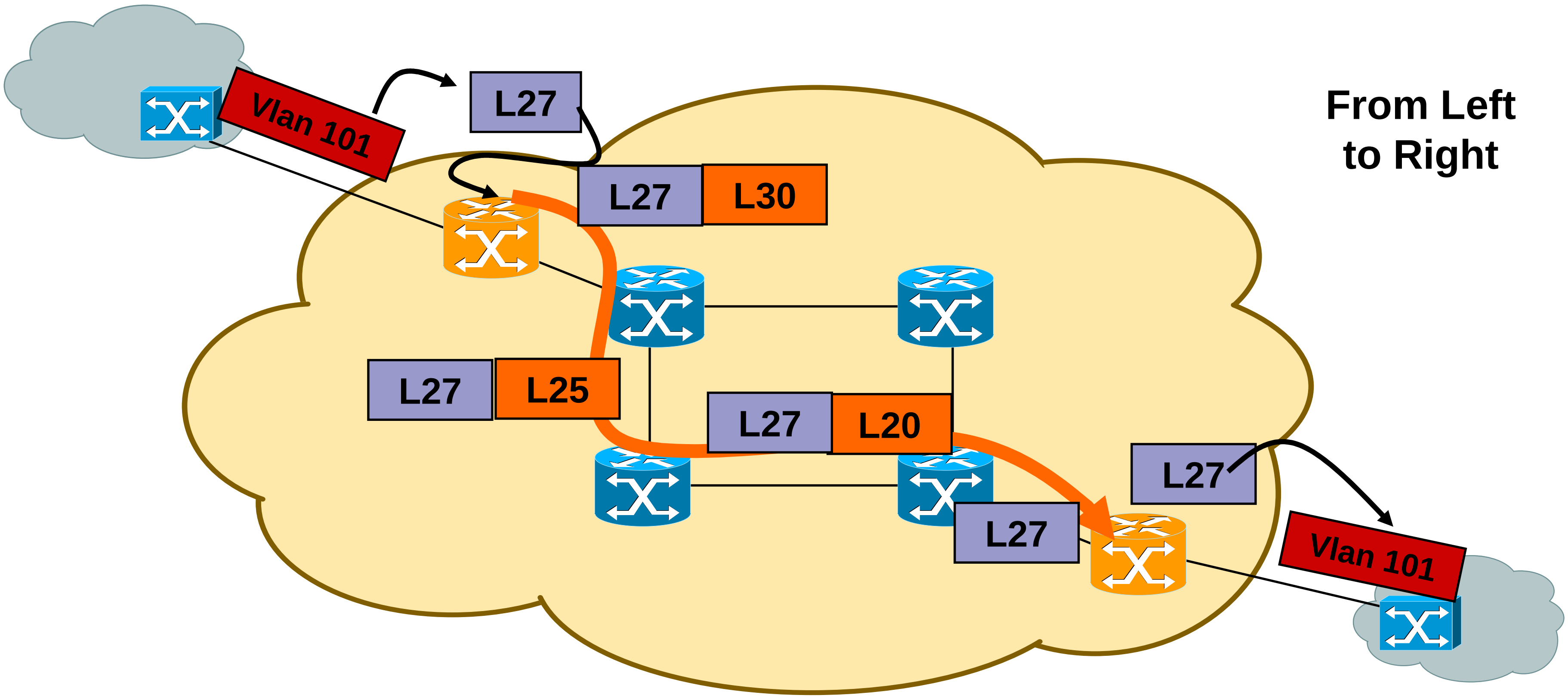
- The layer 2 transport service over MPLS is implemented through the use of two level label switching between the edge routers.

    very similar to RFC2547 (MPLS-VPN)

- The label used to route the packet over the MPLS backbone to the destination PE is called the "tunnel label".
- The label used to determine the egress interface is referred to as the VC label.
- The egress PE creates a VC label and binds the Layer 2 egress interface to this VC, then sends this label to the ingress PE using the directed LDP session.

# AToM : Label forwarding





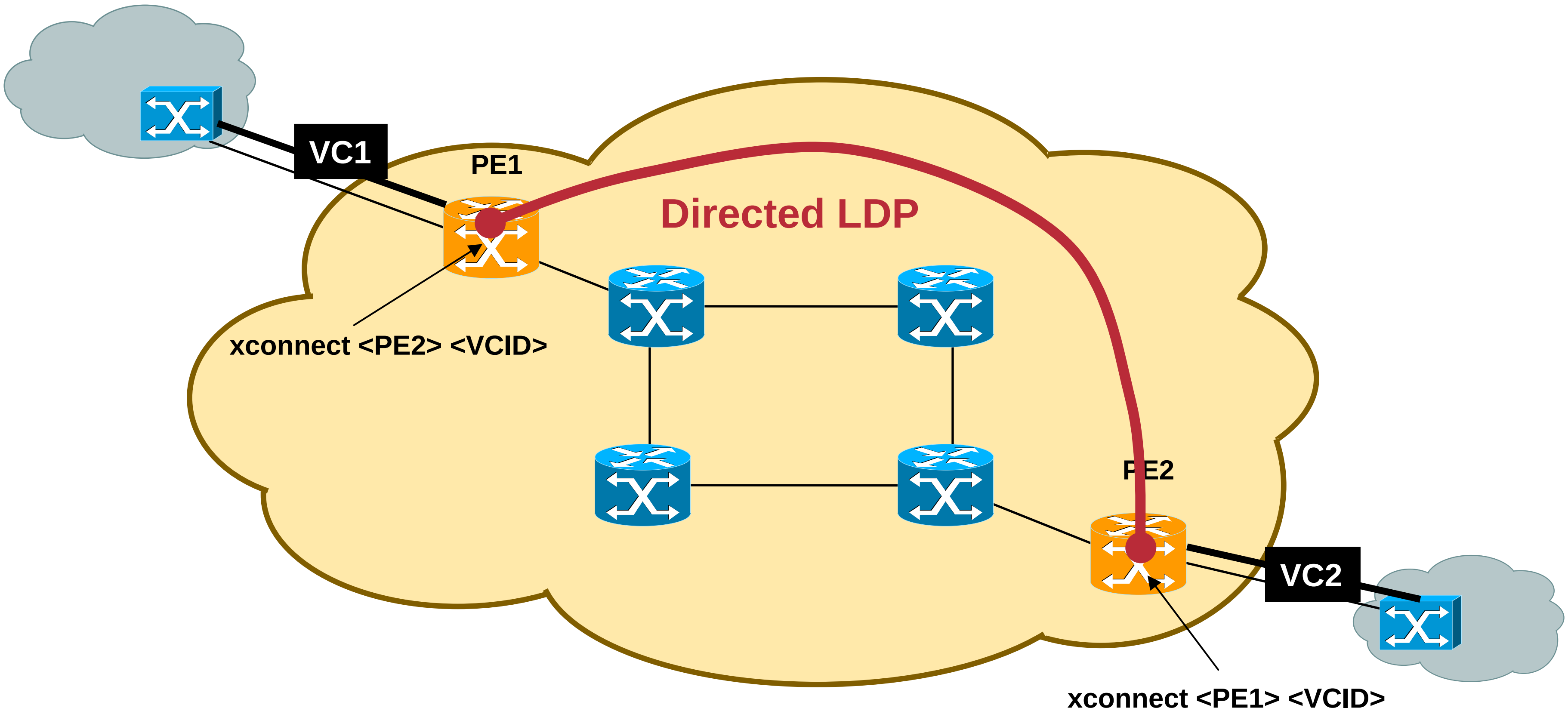
# Method to distribute VC Labels



- Static assigned label
- LDP with PWid FEC TLV
- LDP with Generalized FEC TLV

# AToM: PWid FEC signaling



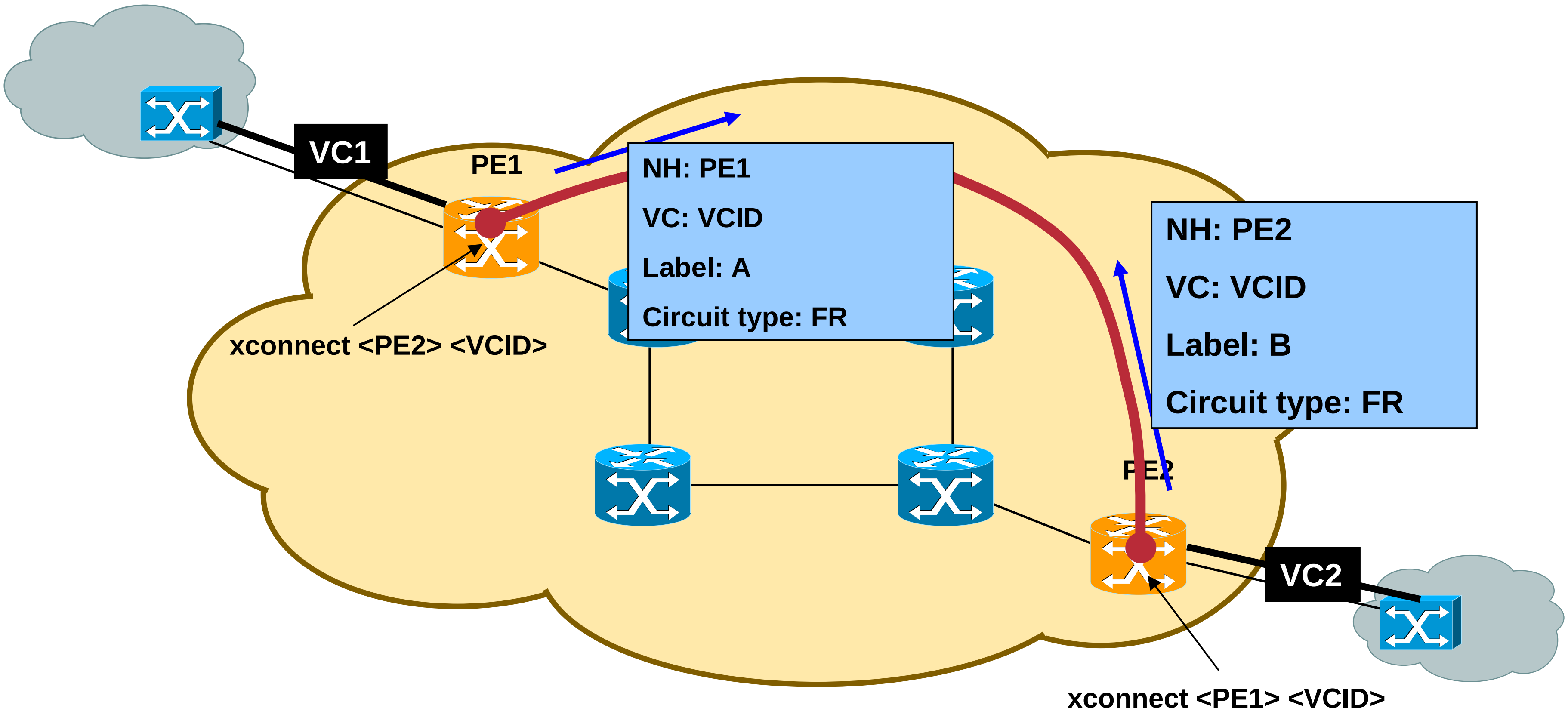



**Based on xconnect command, both PE's will create directed LDP session if doesn't exist already**

- **LSP can be LDP only or TE**

# AToM: VC Label distributed through directed LDP session






# PWid FEC TLV

# LDP: PWid FEC TLV



<table>
<tr><td>VC TLV</td><td>C</td><td>VC Type</td><td>VC info length</td></tr>
<tr><td colspan="4">Group ID</td></tr>
<tr><td colspan="4">VC ID</td></tr>
<tr><td colspan="4">Interface Parameter</td></tr>
</table>

**VC TLV = 128 or 0x80**

**VC Type: FR, ATM, E802.1Q, Eth…**

**C: 1 control word present**

**Group ID: If for a group of VC, useful to withdraws many labels at once**

**VC ID : ID for the transported L2 vc**

**Int. Param: MTU…**

# AToM : Control word



| | | | |
|---|---|---|---|
| Label (LSP) | EXP | 0 | TTL |
| Label (VC) | EXP | 1 | TTL |

| | | | |
|---|---|---|---|
| 0000 | Flags | Length | sequence number |
| L2 PDU | | | |

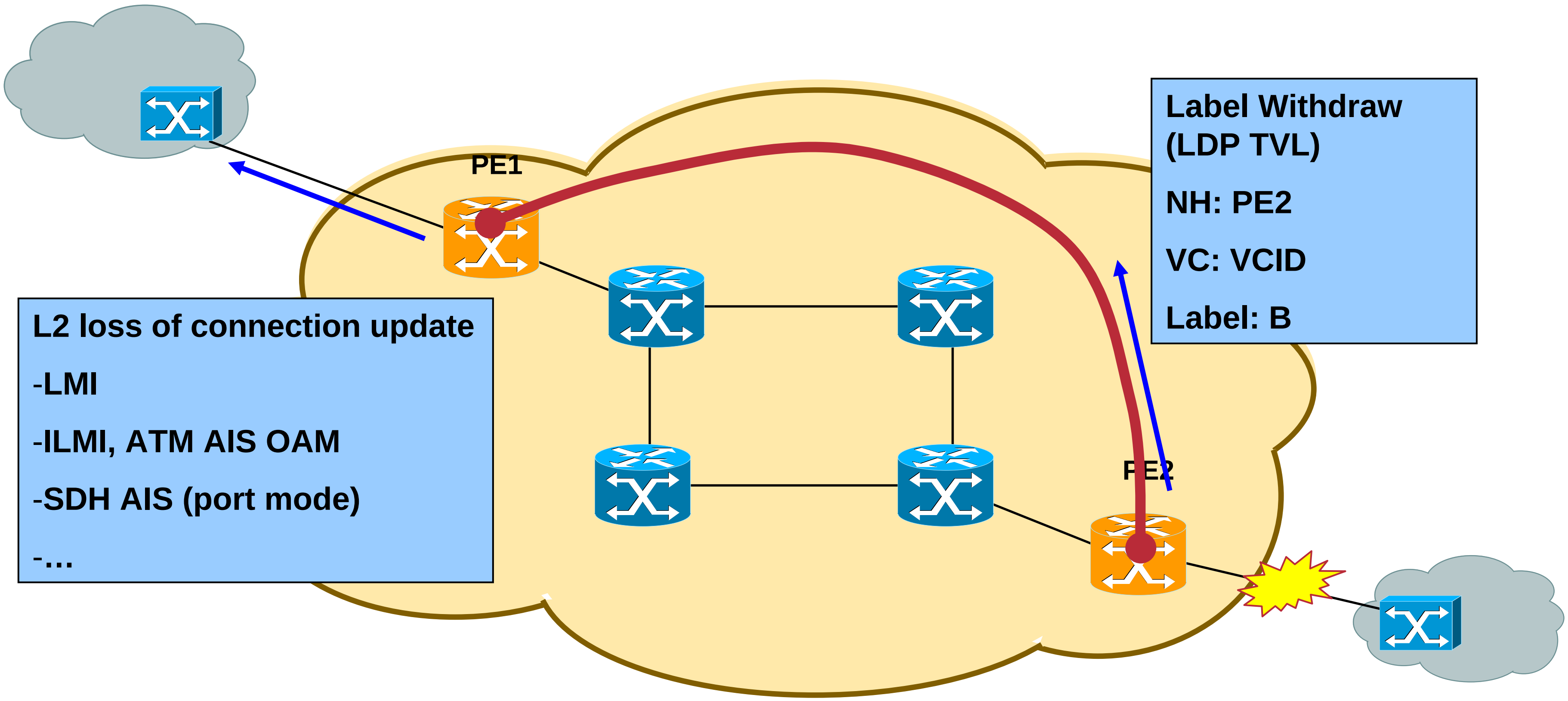
| | | |
|---|---|---|
| ATM | TELC | Transport type, EFCI, CLP, C/RC/R |
| FR | BFDC | BECN, FECN, DE, C/R |

**When transporting L2 protocols over an MPLS backbone:**

- The sequence of the packets should be preserved;
- sequence number 0 indicates that no sequencing is done.
- Small packets may need padding if MTU of medium is larger than packet size
- Control bits carried in header of Layer-2 frame may need to be transported in Flag fields:
  - F/R: FECN, BECN, DE, C/R
  - ATM: AAL5 or cell, EFCI, CLP, C/R

# AToM: Lost of connectivity and Label Withdraw





# Agenda



- **L2 transport over MPLS**
  - **AToM**
  - **EoMPLS**
  - **VPLS**
  - **H-VPLS**

***Yogesh Jiandani***

# Transport of Ethernet over MPLS



**2 main requirements for transport of Ethernet frames**

1. 802.1q VLAN to 802.1q VLAN transport;
2. Ethernet port to port transport

# EoMPLS Transport Formats



## Ethernet II Encapsulation

| Preamble | SFD | DA | SA | TPID | TCI | Ethertype | Data | FCS |
|---|---|---|---|---|---|---|---|---|
| <7 octets> | <1 octet> | <6 octets> | <6 octets> | <2 octets> | <2 octets> | <2 octets> | <46-1500> | <4 octets> |

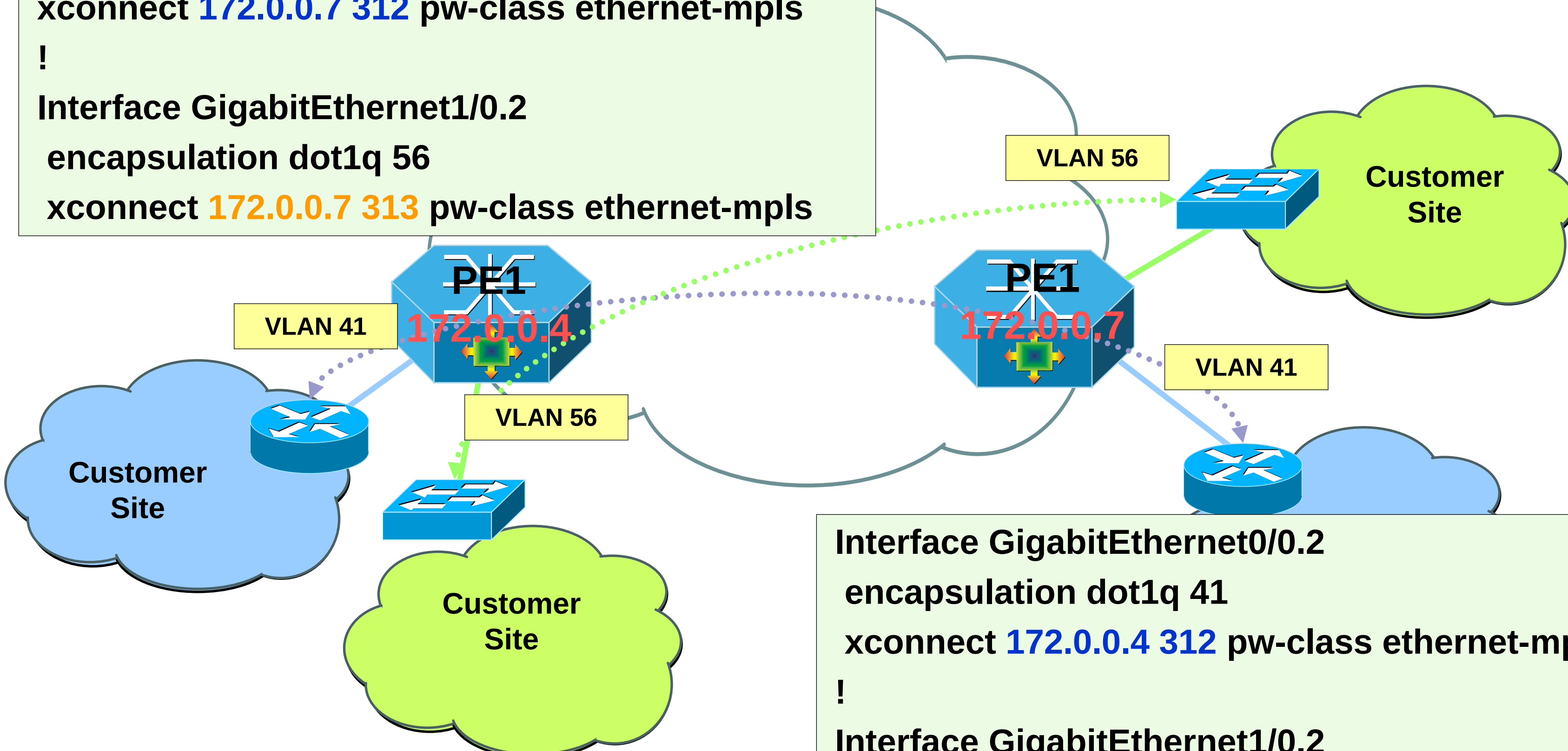
**Transported using AToM**

| Preamble | SFD | DA | SA | TPID | TCI | Length | AA-AA-03 | OUI 0x00-00-00 | Ethertype | Data | FCS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| <7 octets> | <1 octet> | <6 octets> | <6 octets> | <2 octets> | <2 octets> | <2 octets> | <3 octets> | <3 octets> | <2 octets> | <46-1492> | <4 octets> |

## 802.3/802.2/SNAP Encapsulation

# Ethernet 802.1q VLAN Transport

```
Interface GigabitEthernet0/0.2
encapsulation dot1q 41
xconnect 172.0.0.7 312 pw-class ethernet-mpls
!
Interface GigabitEthernet1/0.2
 encapsulation dot1q 56
 xconnect 172.0.0.7 313 pw-class ethernet-mpls
```




```
Interface GigabitEthernet0/0.2
 encapsulation dot1q 41
 xconnect 172.0.0.4 312 pw-class ethernet-mpls
!
Interface GigabitEthernet1/0.2
 encapsulation dot1q 56
 xconnect 172.0.0.4 313 pw-class ethernet-mpls
```

# Agenda

Cisco.com

- **L2 transport over MPLS**
  - **AToM**
  - **EoMPLS**
  - **VPLS**

***Vagish Dwivedi***

# Spanning Tree Protocol - Reminder Primer and general issue.(1)



**Standard Spanning Tree Protocol (IEEE802.1d)**

- **Sent BPDU (info about root of the tree) every 2s to peer bridges onto native VLAN.**
- **Wait 15s to change from (Blocking, Listening, Learning, Forwarding states).**
- **On missing hello's, wait 20s before considering peer bridge down.**

**Standard STP (IEEE802.1d) with default timers allow to have NO more than 7 switches/bridges from the root.**

**In Some implementation those timer a tunable.**

# Spanning Tree Protocol - Reminder Primer and general issue.(2)



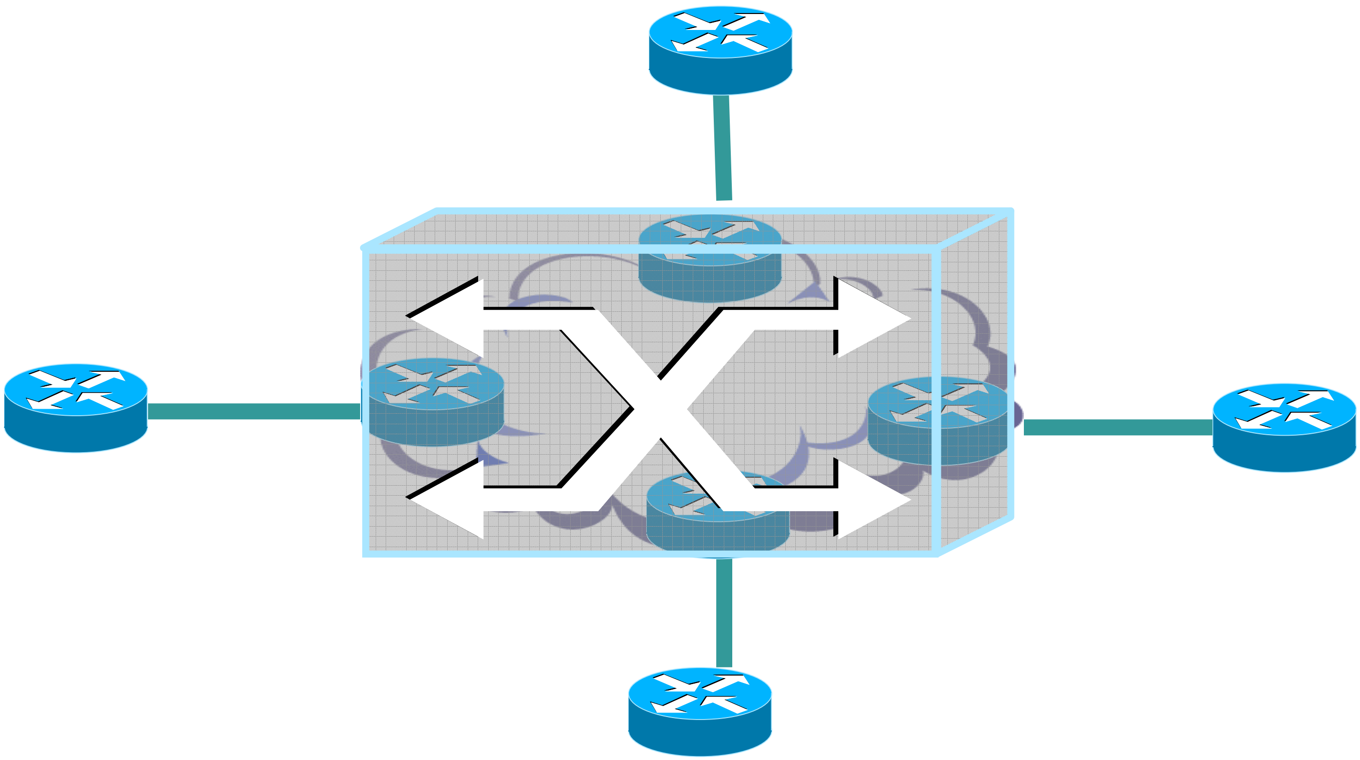
**Rapid STP (IEEE802.1w) sent BPDU every 2s or triggered when receiving new BPDU from peer switches.**

- **This allow to increase convergence and number of switches allowed from root.**
- **Still need to be in 15s time-frame end-end to prevent loops.**

**Recommendation to keep STP domain within a MAN area and not to create inter MAN domain.**

# VPLS (Transparent LAN Services)

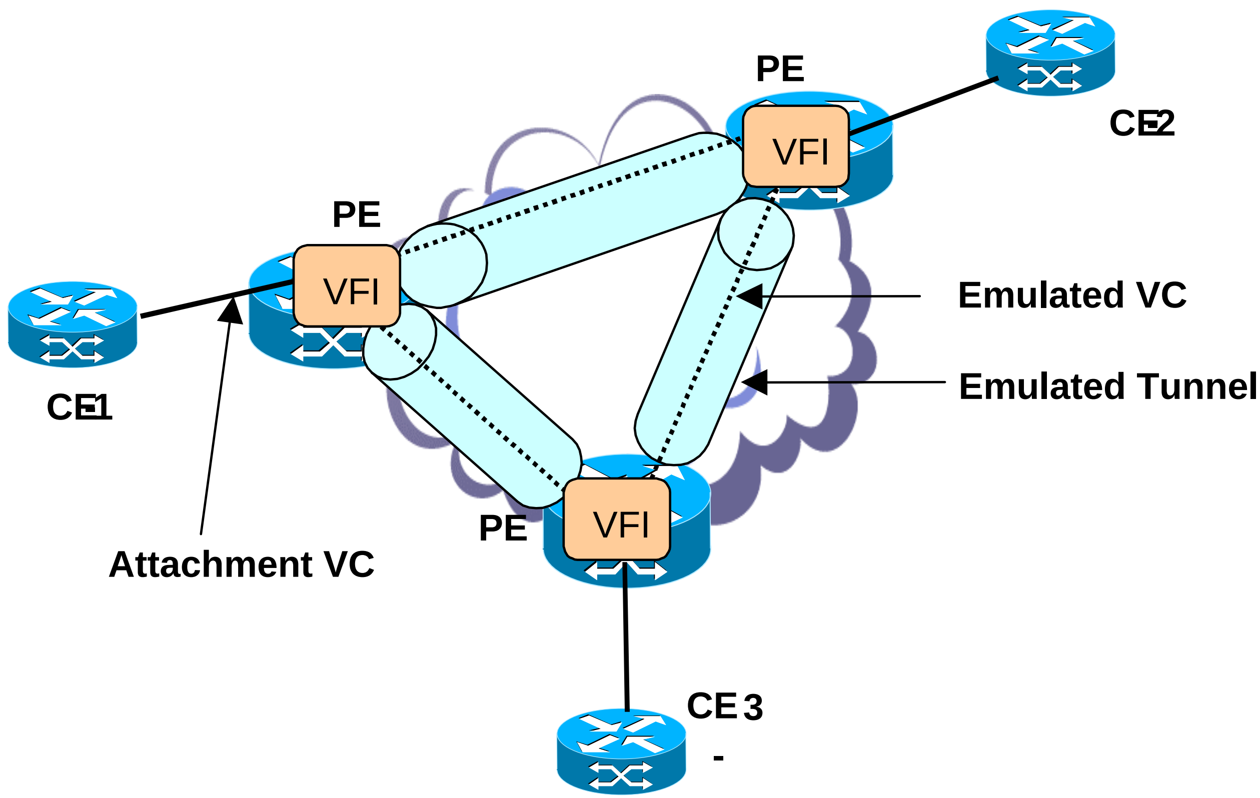




- The network will simulate a L2 switch

# VPLS – An Example

Cisco.com




**In VPLS - transport is provided by pseudo-wires (emulated VCs).**
**In ATOM - transport is provided over Label Switched Paths (LSPs).**
**An MPLS-enabled core, Attachment VCs (native Ethernet or 802.1Q VLAN), and full mesh directed LDP sessions between PE routers are prerequisite for VPLS services**

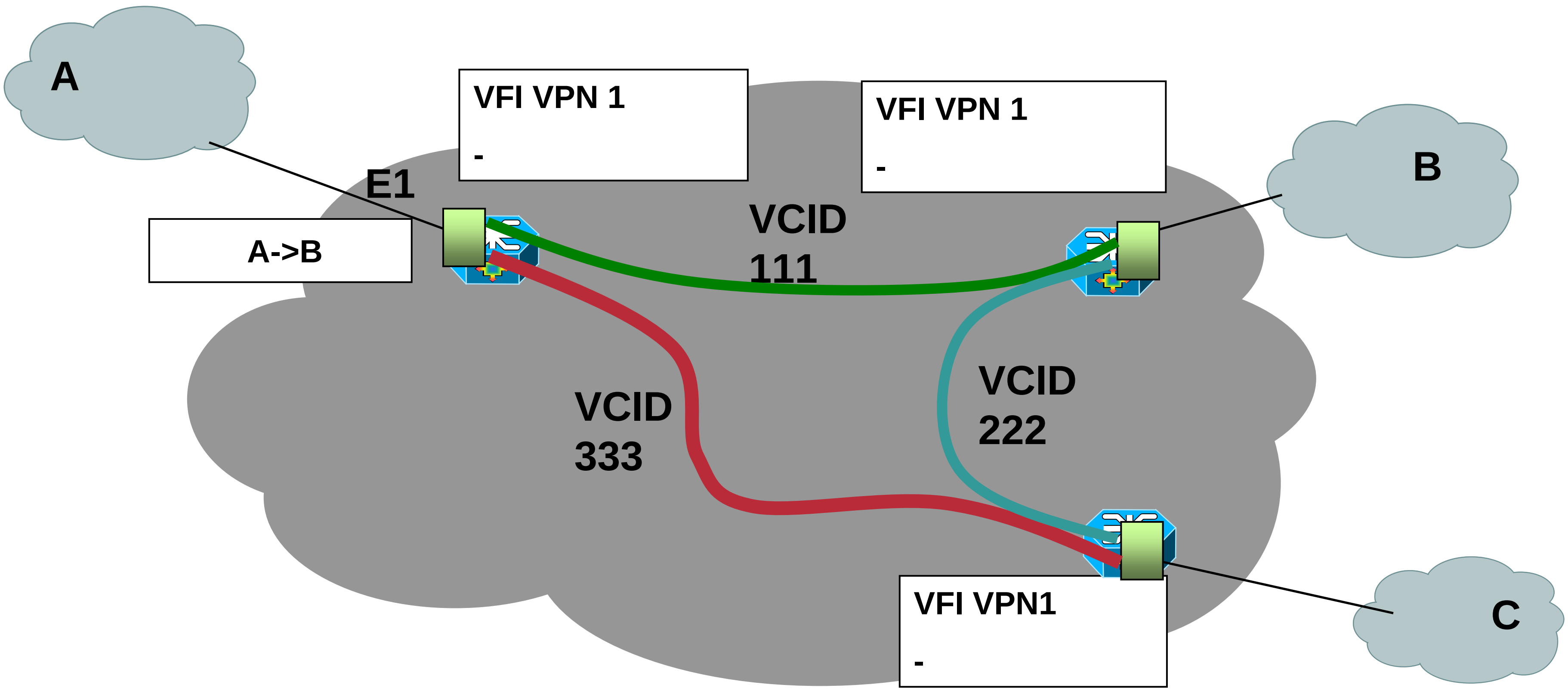
# VPLS: VPN L2forwarding instance



- **Requirement for this solution**
  - **MAC table instances per customer and per Customer VLAN (L2-VRF idea) for each PE.**
    - **Called Virtual Forwarding Instance (VFI)**
    - **VFI will participate to learning, forwarding process.**
  - **Create full-mesh (or partial mesh) of emulated VCs per VPLS.**
  - **Usage of “network split-horizon” to prevent loops in VPLS domain.**
  - **Use of VLAN and Ethernet VC-Type defined PWE3-WG**
  - **(Optional) New MAC TLV (LDP) to accelerate MAC withdraw equivalent function to IEEE Rapid Spanning Tree (IEEE 802.1w)**

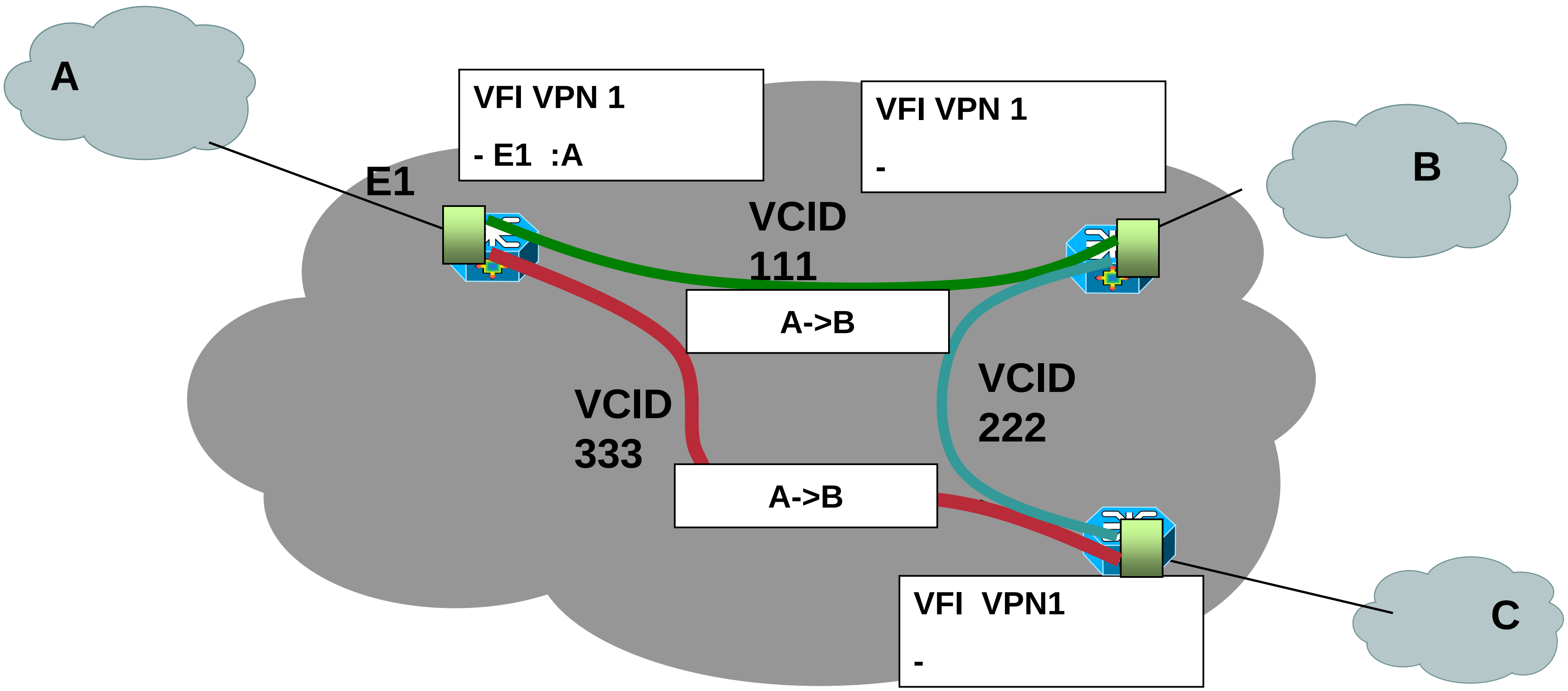
# VPLS L2signalling and forwarding





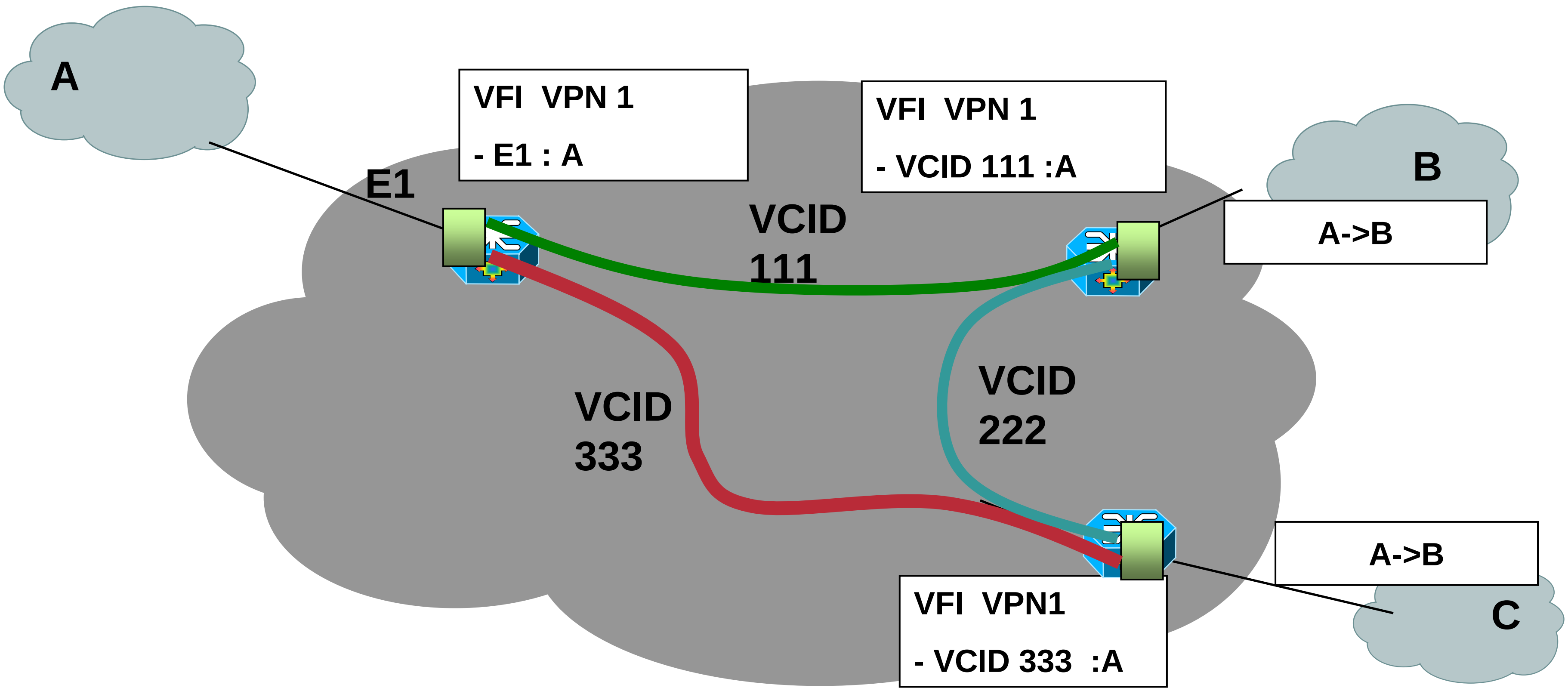
# VPLS L2signalling and forwarding

Cisco.com



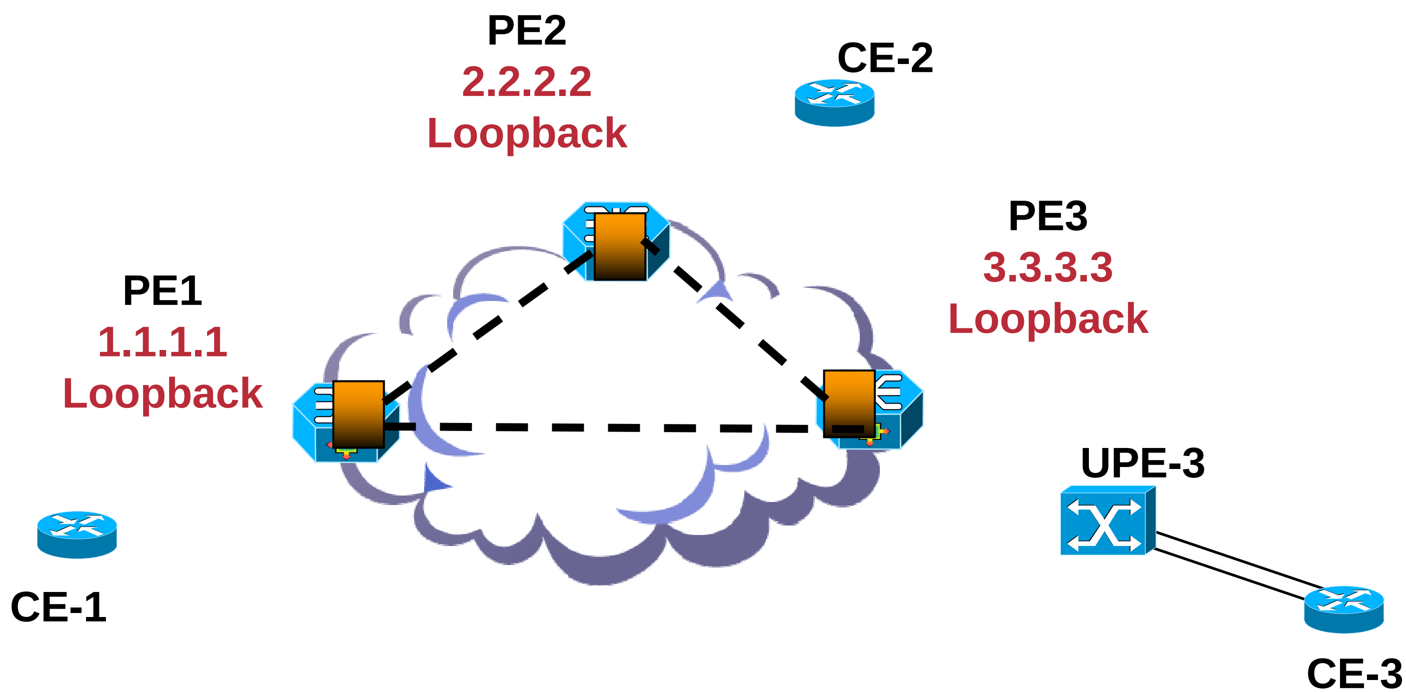
# VPLS L2signalling and forwarding





# Sample of CLI…






**VFI: green**

**VPN-ID: 100:1**

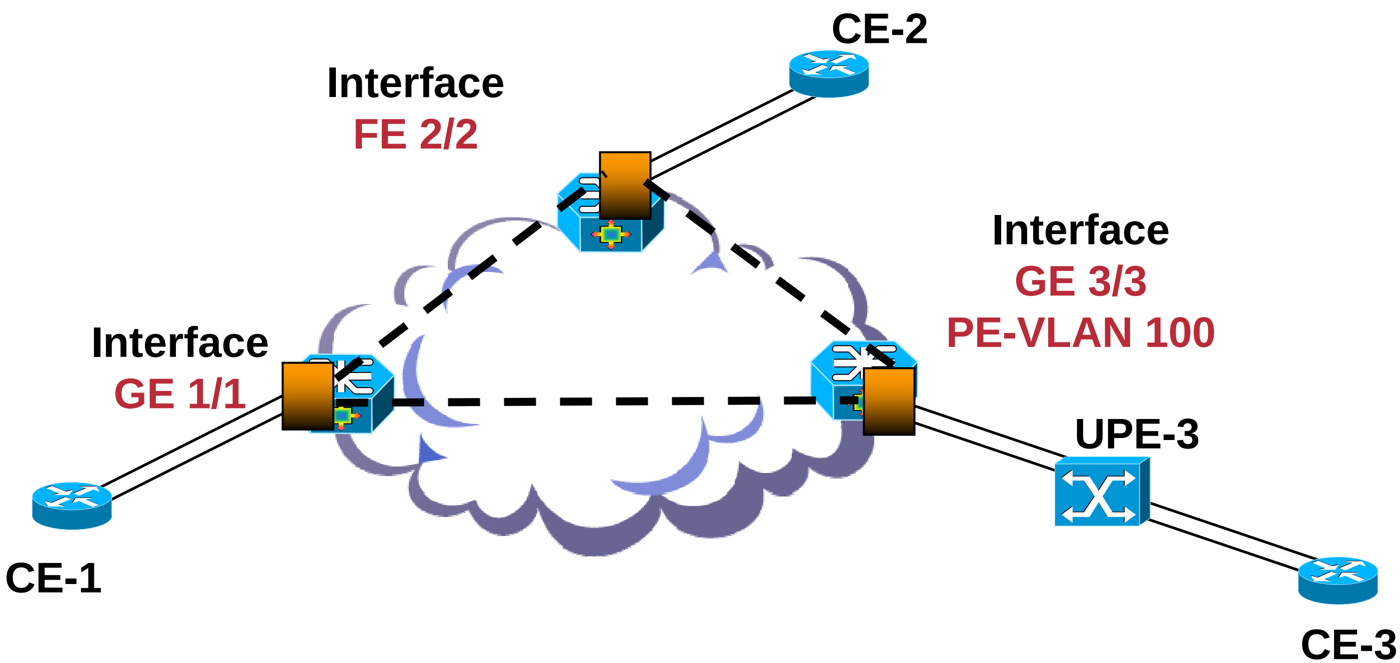
# 1. Create a VFI and attach neighbour VFI

Cisco.com

```
C7600-1(config)# l2 vfi green-vfi manual
C7600-1(config-vfi)# vpn id 100:1
C7600-1(config-vfi)# neighbor 2.2.2.2 encapsulation mpls
C7600-1(config-vfi)# neighbor 3.3.3.3 encapsulation mpls

C7600-2(config)# l2 vfi green-vfi manual
C7600-2(config-vfi)# vpn id 100:1
C7600-2(config-vfi)# neighbor 1.1.1.1 encapsulation mpls
C7600-2(config-vfi)# neighbor 3.3.3.3 encapsulation mpls

C7600-3(config)# l2 vfi green-vfi manual
C7600-3(config-vfi)# vpn id 100:1
C7600-3(config-vfi)# neighbor 2.2.2.2 encapsulation mpls
C7600-3(config-vfi)# neighbor 1.1.1.1 encapsulation mpls
```

# 2. Configure direct attached CPE and UPE





# 2. Configure direct attached CPE and UPE

Cisco.com

```
C7600-3(config)# interface GigEthernet3/3
C7600-3(config-inf)# switchport
C7600-3(config-inf)# switchport mode trunk
C7600-3(config-inf)# switchport trunk encap dot1q
C7600-3(config-inf)# switchport trunk allow vlan 100,105,1002-1005

C7600-3(config)# interface vlan 100
C7600-3(config-inf)# xconnect vfi green-vfi

C7600-3(config)# interface vlan 105
C7600-3(config-inf)# xconnect vfi red-vfi
```

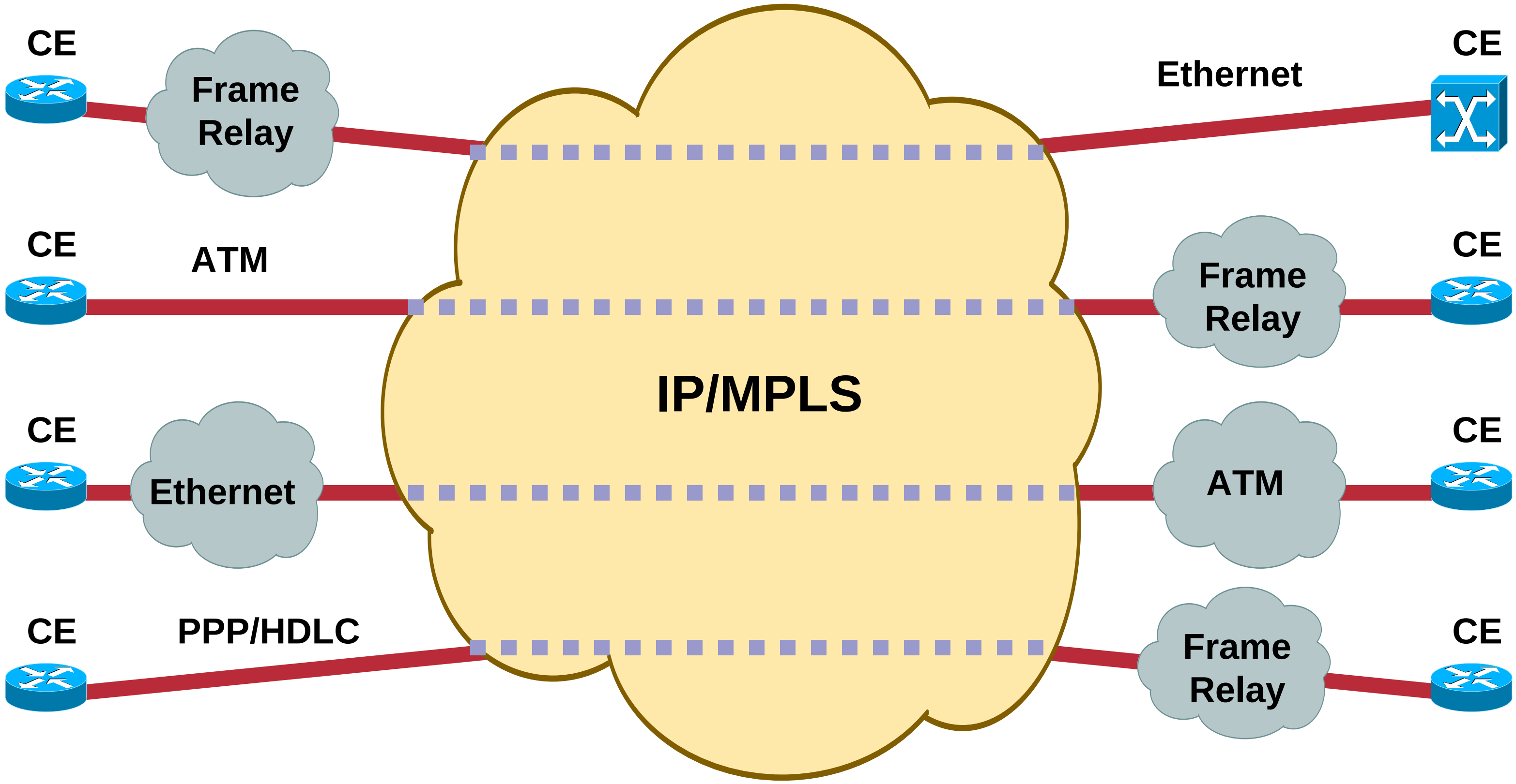
# MPLS L2-VPNs
# Today's market acceptance



- **Is widely deployed**
  - **Ethernet, Frame Relay**
- **Is fairly deployed**
  - **ATM (Cell and AAL5)**
  - **VPLS**
- **Is sparsely deployed**
  - **PPP**
  - **Interworking**

# Layer 2 Service Interworking

Cisco.com




**How to connect different encapsulations and retain a Layer 2 service…**

# Agenda

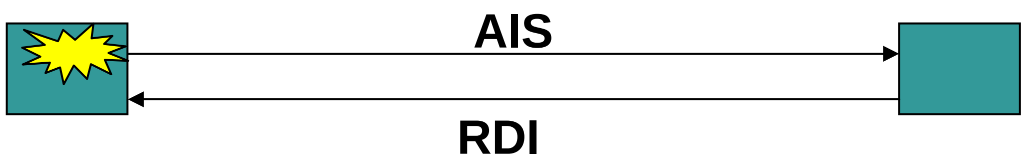
Cisco.com

- **Build an MPLS core**
    - **OAM**
    - **Fast-convergence**
    - **Traffic Engineering**

***Yogesh Jiandani***

# AIS/RDI
## Proactive error signalling

Cisco.com

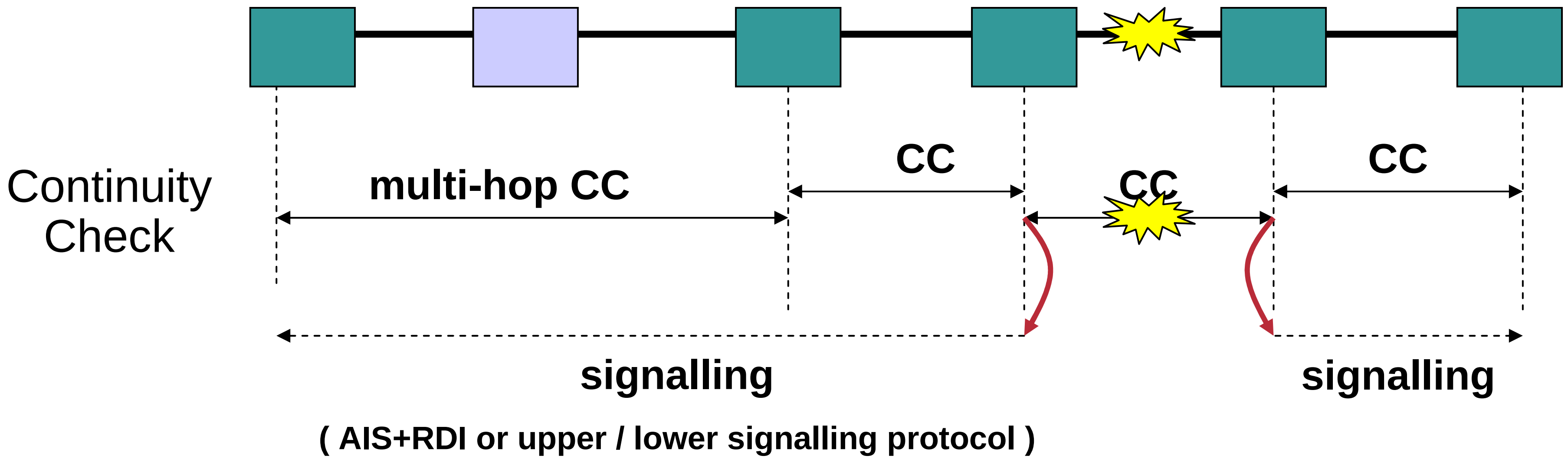



- **AIS: Alarm Indication Signal**
- **RDI: Remote Defect Indicator**

**Media: SDH, SONET, ATM,…**

# Continuity Check (Data Plane, Control Plane)
## Connectivity Aliveness

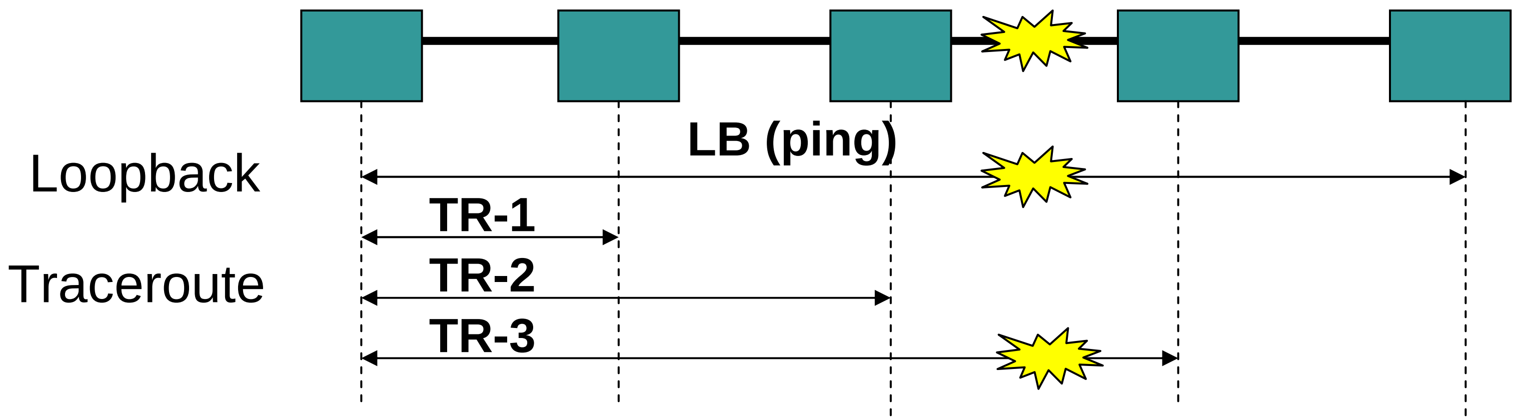





**Media: FR, ATM, IP IGP/EGP, RSVP,…**

# Loopback & Traceroute
## Path Troubleshooting

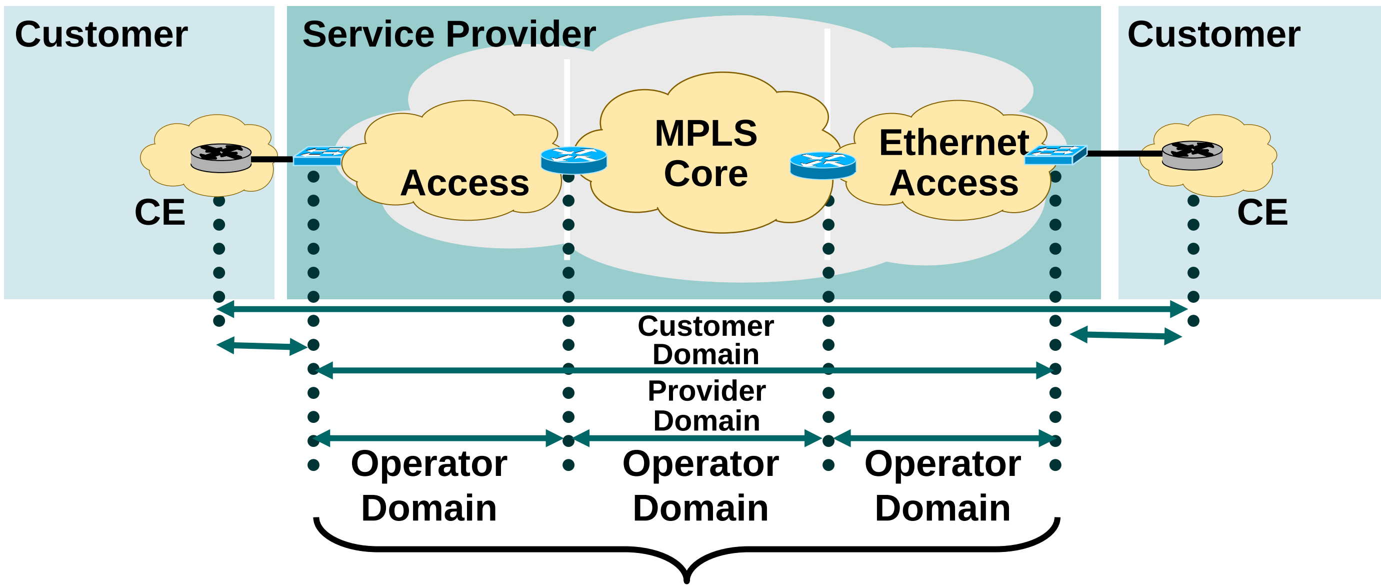





**Media: ATM, MPLS, IP ICMP, …**

# The OAM Landscape

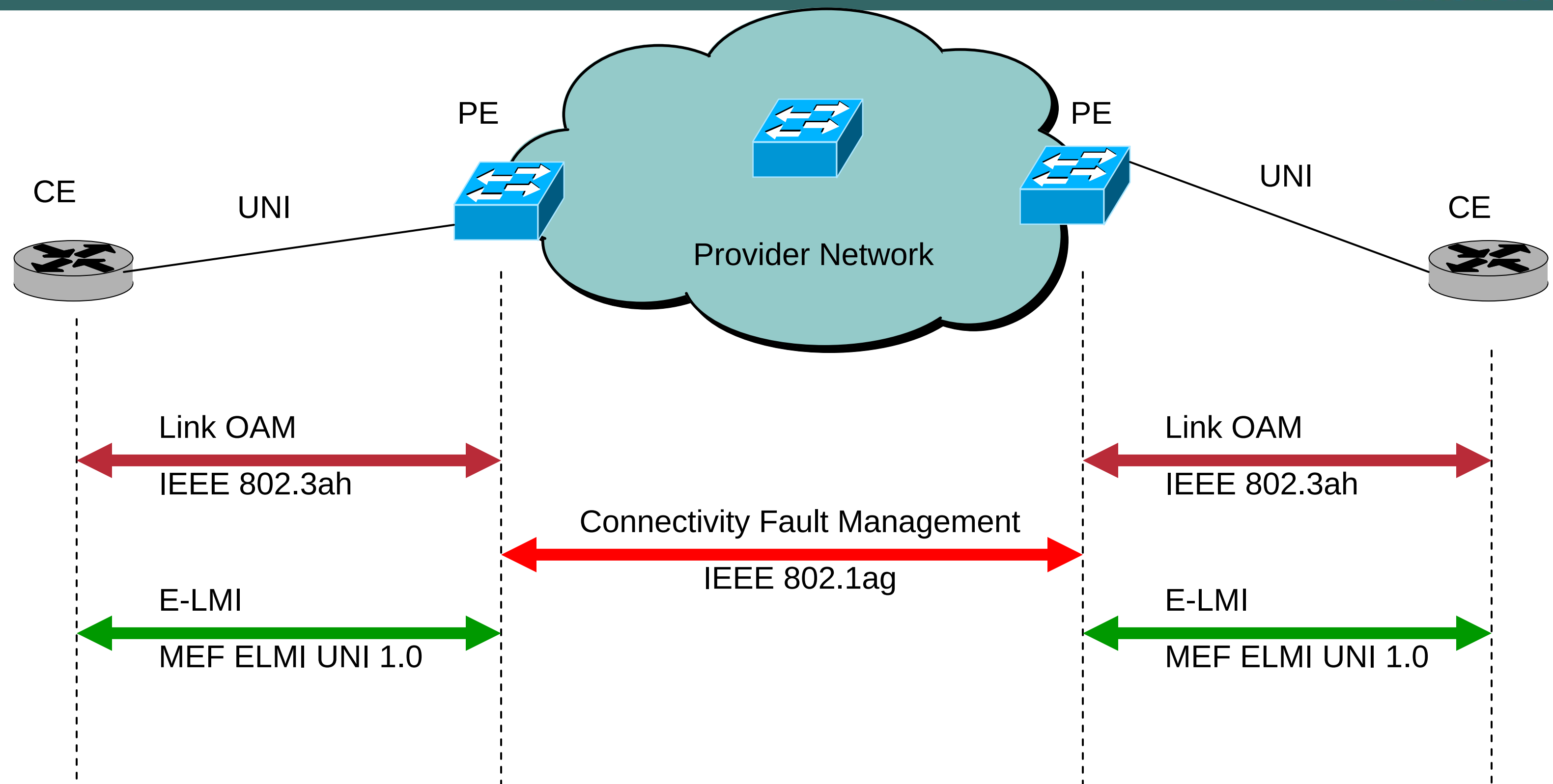





- Customer contracts with Provider for end-to-end service.
- Provider contracts with Operator(s) to provide equipment and networks.
- Provider and Operator(s) may or may not be the same company or same division.

# Ethernet OAM's
## Example: Metro Ethernet Architecture






- **IEEE 802.3ah — Ethernet in the First Mile Specification (Port Ethernet Keepalive, AIS/RDI)**
- **E-LMI — Draft Ethernet Local Management Interface Specification**
- **IEEE 802.1ag — Connectivity Fault Management Specification (MAC Ping, TR, CC)**

# IP OAM's

Cisco.com

- **Data Plane layer**
  - IP Bidirectional Forwarding Detection (BFD)
  - IP Ping
  - IP Traceroute
  - IP SLA (RTR, SAA)
  - …
- **Control Plane layer**
  - IGP Hello's : OSPF Hello's, ISIS Hello's, RIP Hello's
  - EGP Hello's : BGP TCP keepalive,…

# IP VPNv4 OAM's



- **Data Plane layer**
  - **VRF aware IP Ping**
  - **VRF aware IP Traceroute**
  - **VRF aware IP SLA (RTR, SAA)**
  - **…**
- **Control Plane layer**
  - **EGP Hello's : MP-BGP TCP keepalive,…**

# Bi-directional Forwarding Detection (BFD)



## Goal

A sub hundred milliseconds forwarding plane failure detection using single, common, standardized mechanism, which is independent of media and routing protocols.

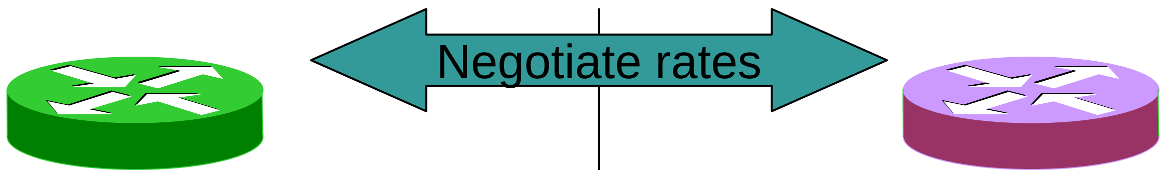
# BFD Protocol Overview



- **BFD control packets will be encapsulated in UDP datagram.**
- **Destination port 3784 and source port between 49252 to 65535.**
- **Because of the fast nature of the protocol, all output features are bypassed for locally generated BFD control packets.**
- **Active & Passive modes.**

# Timer negotiation

- **Neighbors continuously negotiate their desired transmit and receive rates in terms of microseconds.**
- **The system reporting the slower rate determines the transmission rate.**




Desired Receive rate = 50 ms
Desired Transmit rate = 100 ms

Desired Receive rate = 60 ms
Desired Transmit rate = 40 ms

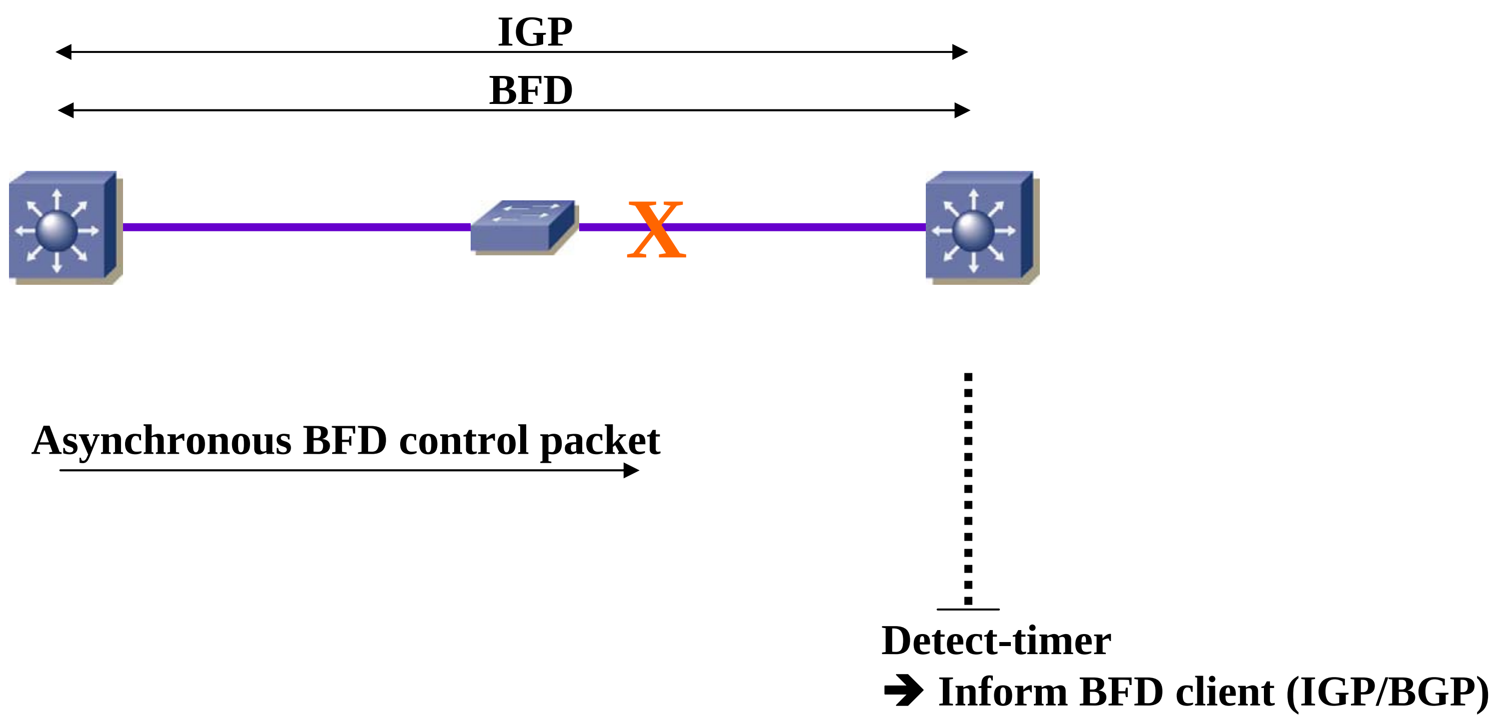
Green Transmits at 100ms

Purple transmits at 50ms

# BFD – Asynchronous Bi-directional mode



BFD will detect in a few hundred of milliseconds Layer 3 neighboring failure, faster than any IGP hellos



# BFD Application



- Forwarding plane liveliness
- Tunnel liveliness detection
- IP/MPLS FRR
- BFD over Ethernet
- MPLS LSP data plane failure

# MPLS OAM's



- **Data Plane layer**
  - **MPLS Bidirectional Forwarding Detection (BFD)**
  - **MPLS Ping (IPv4 + LDP / Traffic Engineering Tunnel)**
  - **MPLS Traceroute (IPv4 +LDP / Traffic Engineering Tunnel)**
  - **Virtual Circuit Connection Verification (AToM Ping)**
- **Control Plane layer**
  - **LDP Hello's : TCP keepalive**
  - **BGP Hello's : TCP keepalive,…**
  - **RSVP Hello's & fast RSVP Hello's**

# Pseudowires a.k.a L2 OAM's
## VC Connection Verification (VCCV)



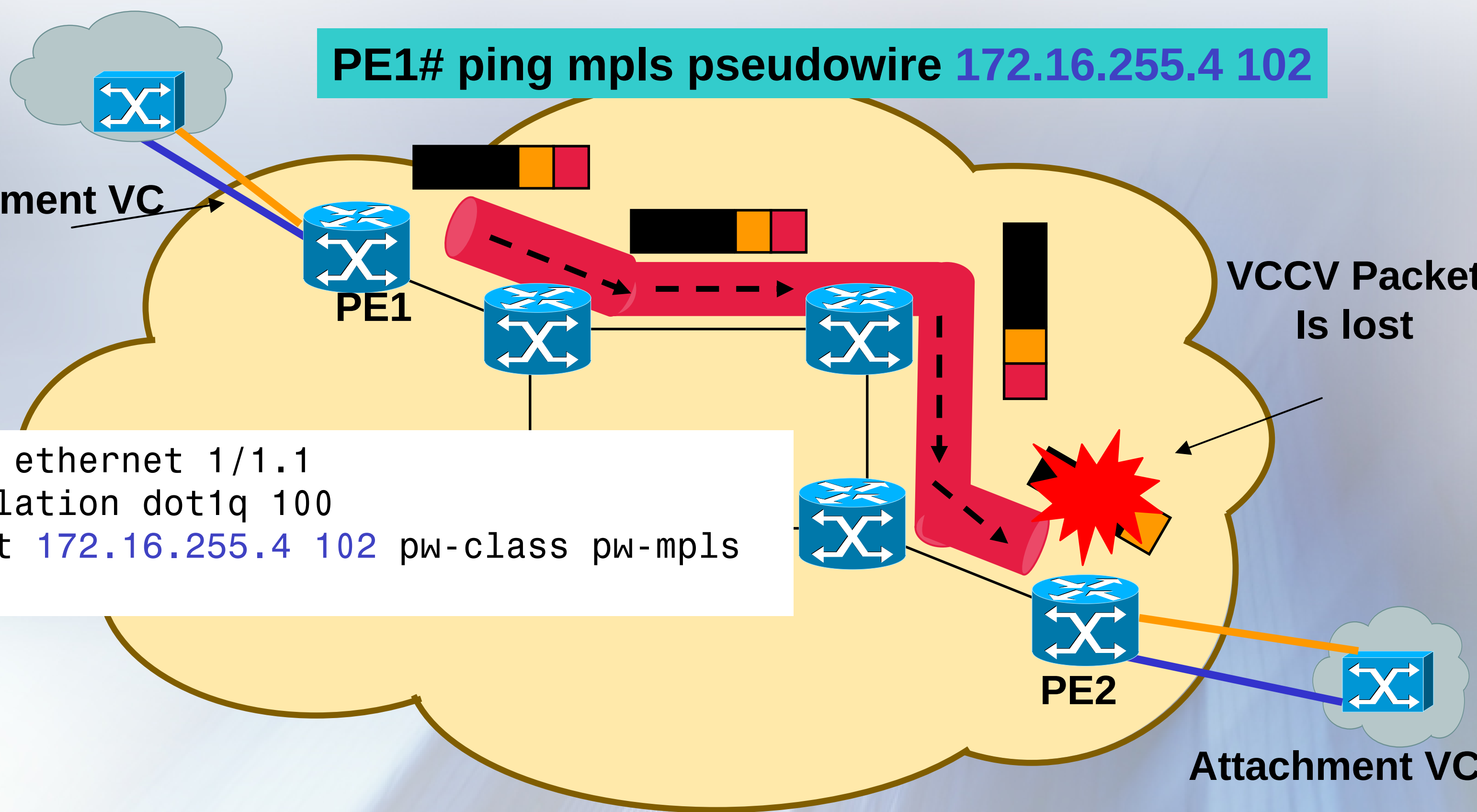
- **VCCV goal is to verify aliveness, integrity of defined pseudowire**
- **VCCV capability is negotiated when the AToM tunnel is brought up**
- **A new pseudowire interface parameter is defined**
- **2 data plane methods defined**

  **Inband : One bit from pseudowire Control-Word is defined VCCV bit, egress PE are going to intercept all packets with VCCV bit set 1**

  **outband : An additional VCCV label is defined, egress PE are going to intercept all packets with this label.**

# Connectivity Trace Using VCCV

Cisco.com

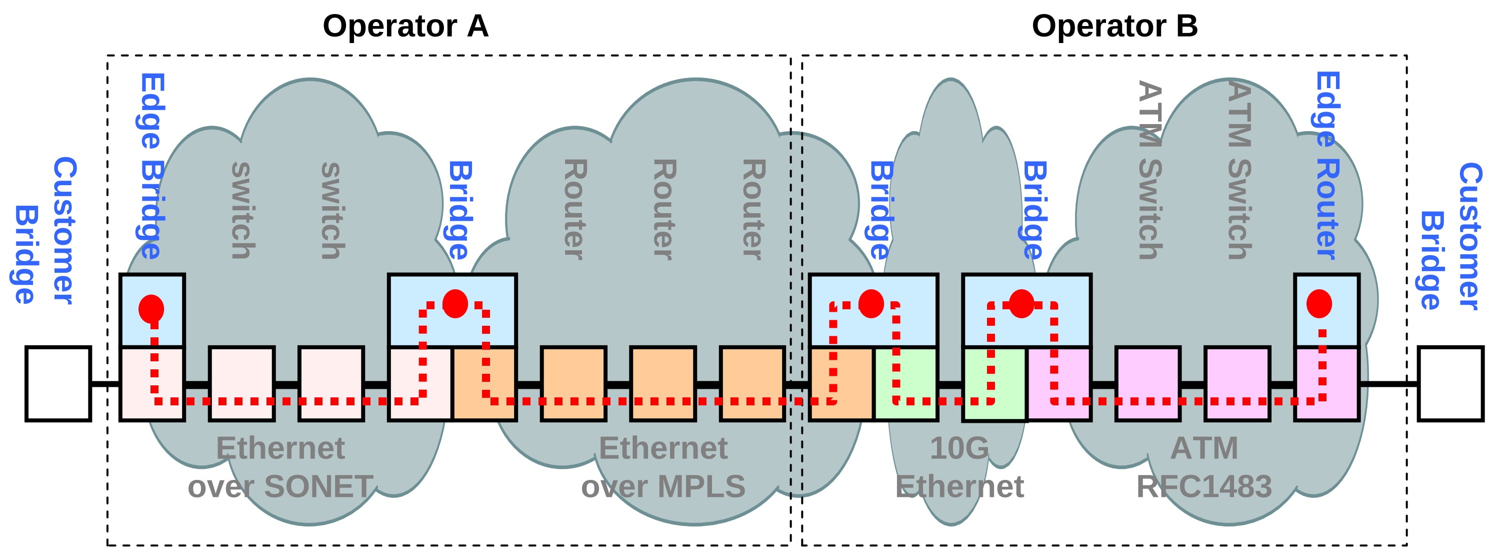



```
interface ethernet 1/1.1
  encapsulation dot1q 100
  xconnect 172.16.255.4 102 pw-class pw-mpls
```

# What is OAM Inter-working?



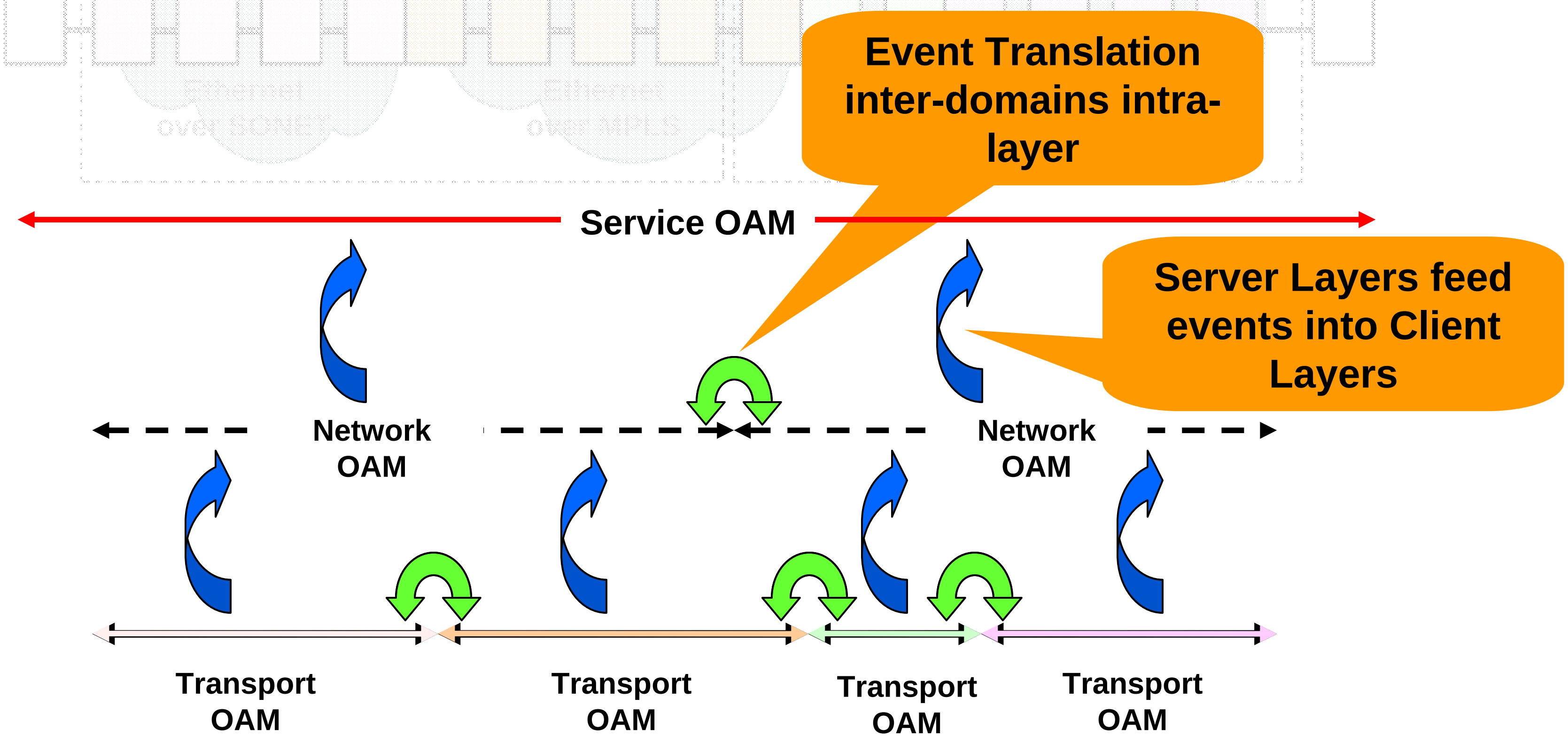
## This is NOT OAM inter-working!




- Transport OAM message with embedded MAC address carried from bridge to bridge, visible to ETH layer (when present), and translated to new transport's OAM format when crossing physical media boundaries.
- Creates dependency on Physical layer and inter-operability issues.

# What is OAM Inter-working?

- Strict OAM layering should be honored: messages should not cross layers
- OAM Messages should not leak outside domain boundaries within a layer
- Inter-working is event translations & not necessarily 1:1 message mapping
- Inter-working may be inter-layer and intra-layer



# Summary on OAM’s

| Media type | Signalling | Aliveness | | Troubleshooting | | |
|---|---|---|---|---|---|---|
| | AIS/RDI | CC CP | CC DP | Loopback | Performance | Traceroute |
| ATM VP | F4 | ILMI | F4 (VC-3) | F4 (VC-4) | | |
| ATM VC | F5 | | F5 (PT 100) | F5 (PT 101) | | |
| FR | | LMI | Keepalive | | | |
| Ethernet last mile | | **E-LMI** | **IEEE 802.1ag** | | | |
| Ethernet provider bridge | | **IEEE 802.1ag** | | **IEEE 802.1ag** | | **IEEE 802.1ag** |
| MPLS LDP | LDP | LDP Hello | **MPLS BFD** | LSP Ping | | LSP TR |
| MPLS TE | RSVP | RSVP Hello | | | | |
| MPLS PW | LDP | LDP Hello | **VCCV BFD** | VCCV Ping | | |
| IPv4 | | IGP/BGP Hello | BFD | IP Ping | IP SLA | IP TR |
| IPv4 VPN | | IGP/BGP Hello | BFD | IP Ping (VRF) | IP SLA (VRF) | IP TR (VRF) |

# Agenda

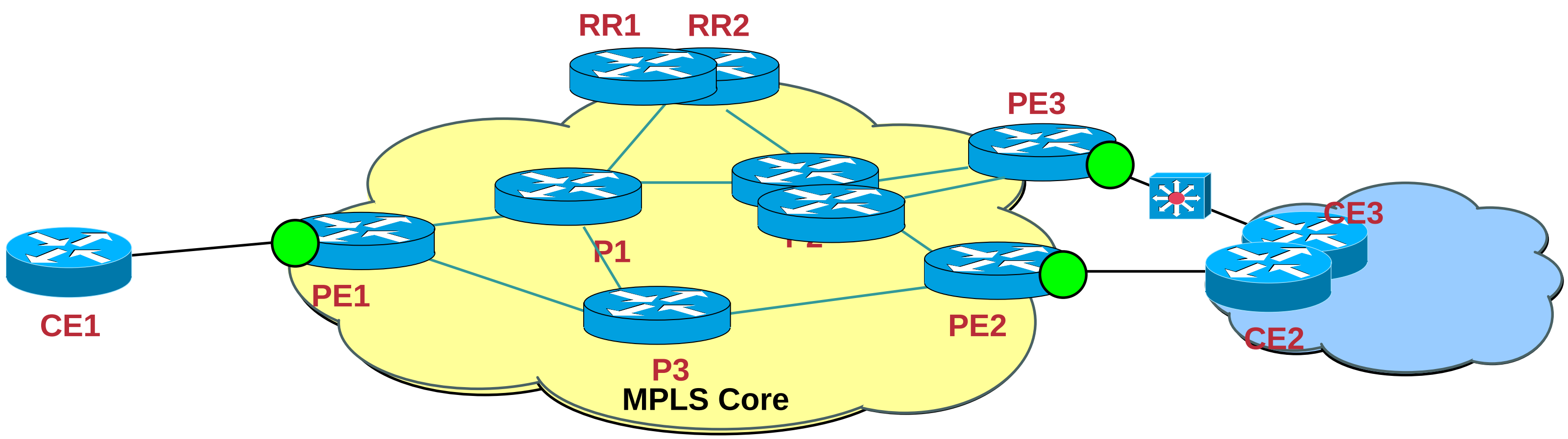


- **Build an MPLS core**
  - **OAM**
  - **Fast-convergence**
  - **Engineering of traffic**

***Vagish Dwivedi***

# MPLS Core convergence






**Technologies to consider for convergence**

- **Core options:**
  - P Links/Node protection → FRR
  - PE/P IGP restoration → Fast IGP + LDP
- **Edge:**
  - PE-PE iMP-BGP → Fast BGP via RR
  - CE-PE edge routing → IGP or BGP / OAM

# Core Fast convergence

Cisco.com

1. **Detection of Link / Node failure**
   - Link down (ms detection)
   - Neighboring failure
     - IGP Hello detection = 1 second detection
     - BFD = Sub-second detection (or below)

2. **Alternate path computation**
   - IGP flooding / SPF
     - (sub-second or even 200ms computation)

   - Alternate path pre-computation:
     - MPLS Fast-ReRoute
       - (sub-100ms or even sub-50ms convergence)

# Fast Detection of Link / Node failure



```
interface FastEthernet1/1
 ip address …
  carrier-delay msec 0
 ip router isis

 isis network point-to-point
 dampening
```

**Link down**

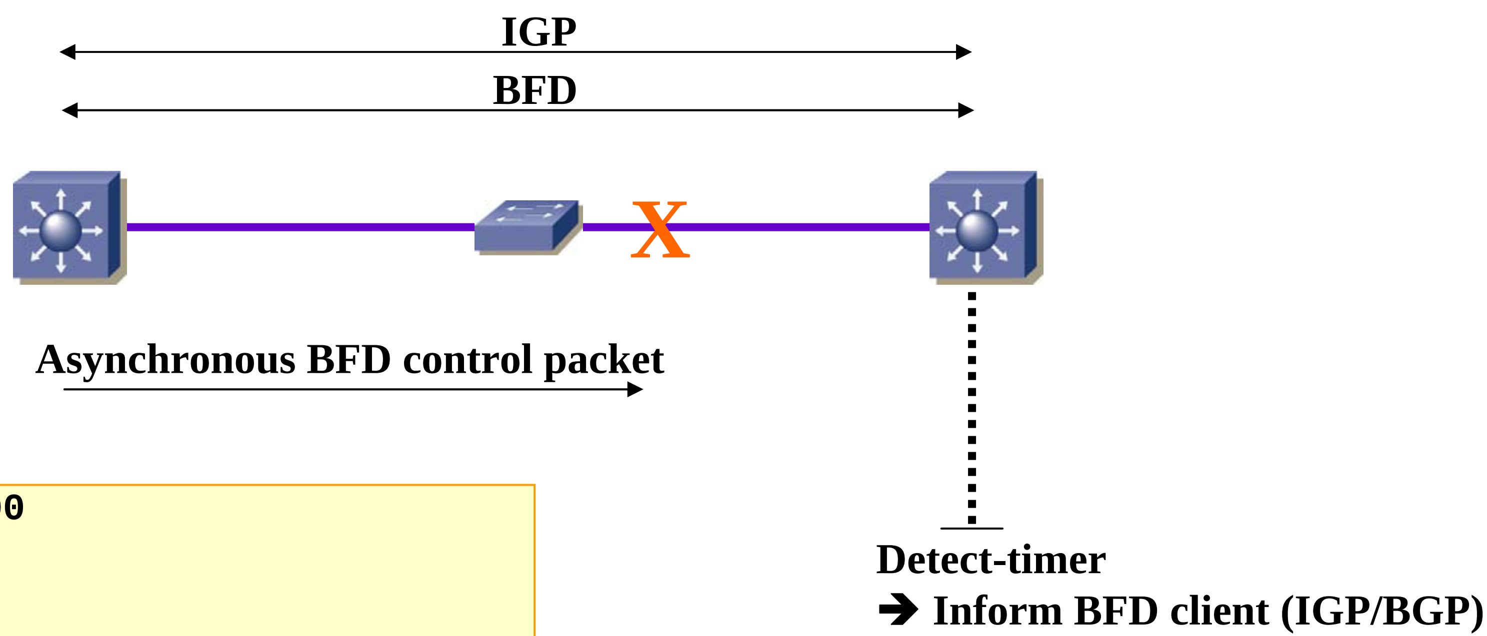
```
interface FastEthernet1/1
 ip address …
 ip router isis
 …
 isis circuit-type level-1
 isis hello-multiplier 10 level-1
 isis hello-interval minimal level-1
```

**Node down**
**IGP detection**

# Bi-directional Forwarding Detection



BFD will detect in a few hundred of milliseconds Layer 3 neighboring failure, faster than any IGP hellos




```
interface Vlan600
 ip address …
 ip router isis
 bfd interval 10 min_rx 10 multiplier 3
 bfd neighbor 10.10.0.18
 dampening
```

# IGP alternate path re-computation thru SPF



```
router isis
 net 49.0001.0000.6500.5555.00
 is-type level-1
 metric-style wide

 spf-interval 20 100 20
 prc-interval 20 100 20
 lsg-gen-interval 1 1 20

 fast-flood 15
```

Allows Traffic-Engineering attribute propagation

Fast reaction, but backoff protection:

- SPF computation
- IP addresses changes
- LSP advertisement

Fast flooding for first LSP

# Agenda

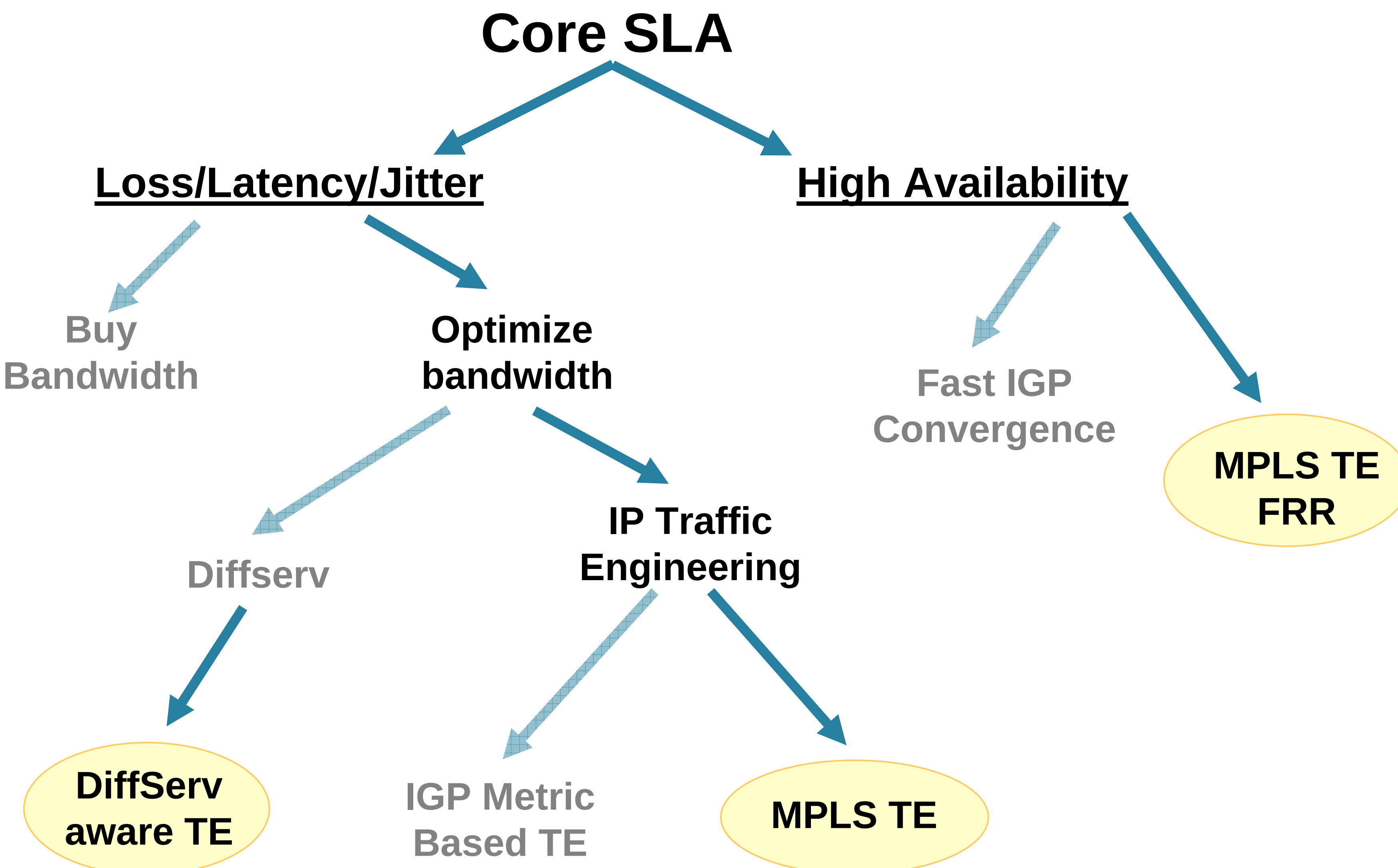
Cisco.com

- **Build an MPLS core**
  - **OAM**
  - **Fast-convergence**
  - **Traffic Engineering**

***Yogesh Jiandani***

# Which problems are we trying to solve ?

Cisco.com



# IP Traffic Engineering: The Benefit for traffic SLA



- **The more effective use of backbone bandwidth potentially allows:**

  **Either …**

    **higher SLA targets (higher availability, lower loss, lower delay) to be offered with the existing backbone bandwidth**

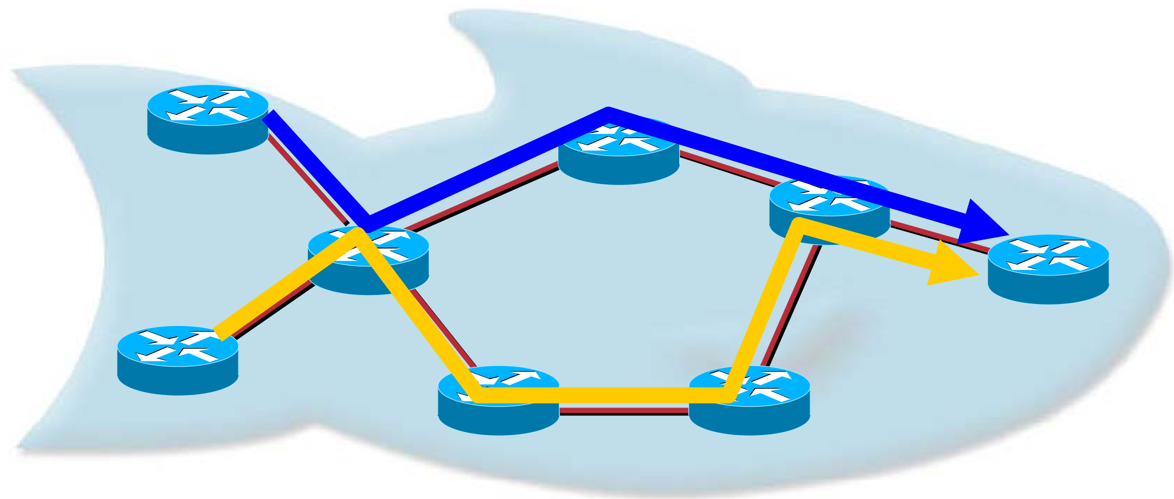
    **Meet very tight SLA requirement (refer to MPLS QoS part)**

  **Or …**

    **the existing SLA targets to be achieved with less backbone bandwidth or with delayed time to bandwidth upgrades**

# MPLS Traffic Engineering

Cisco.com



- MPLS Traffic Engineering gives us an “explicit” routing capability (a.k.a. “source routing”) at Layer 3
- Lets you use paths other than IGP shortest path
- Allows unequal-cost load sharing
- MPLS TE label switched paths (termed “traffic engineering tunnels”) are used to steer traffic through the network

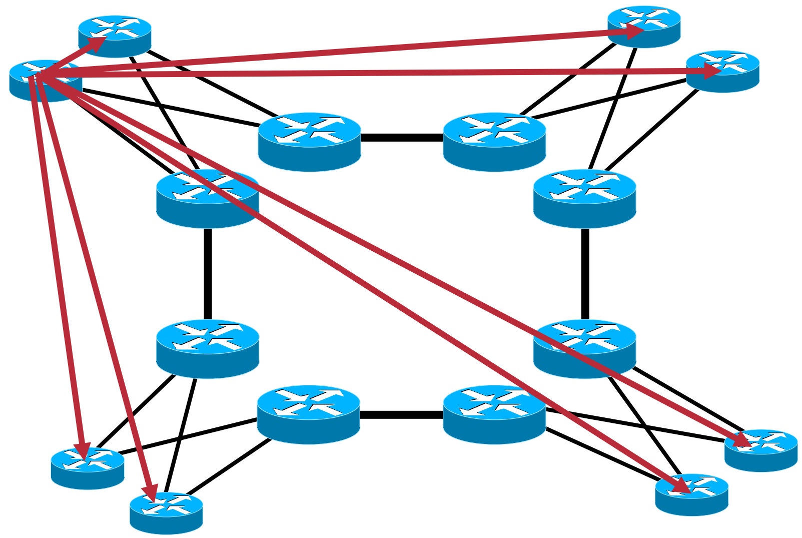
# MPLS TE Components – Refresher



1. Resource / policy information distribution
2. Constraint based path computation
3. RSVP for tunnel signaling
4. Link admission control
5. LSP establishment
6. TE tunnel control and maintenance
7. Assign traffic to tunnels

# Strategic Deployment: Full Mesh





- **Requires n * (n-1) tunnels, where n = # of head-ends**
- **Reality check: largest TE network today has ~100 head-ends**

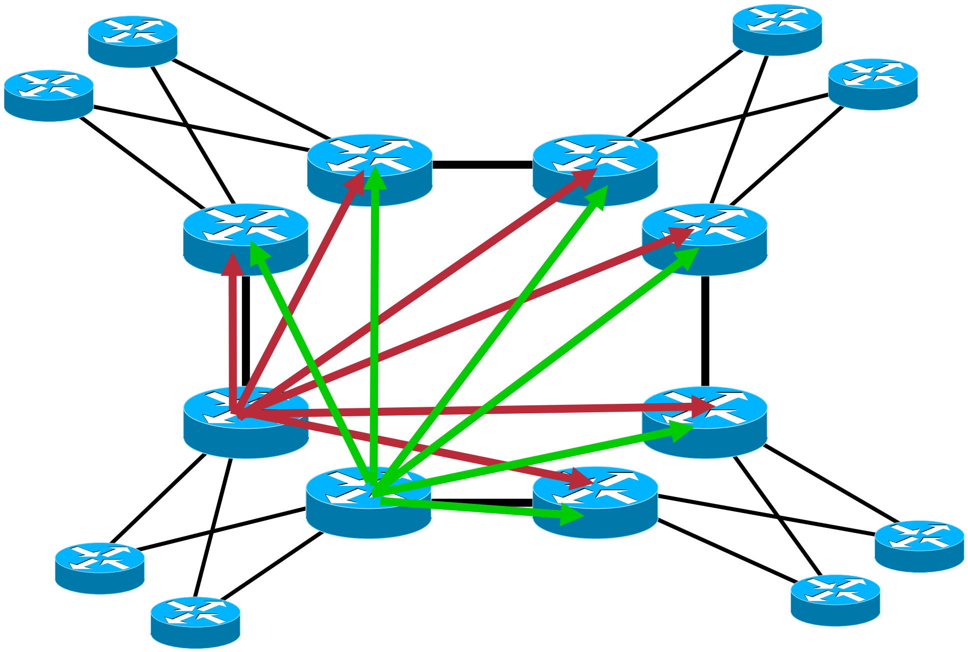
  **~9,900 tunnels in total**

  **max 99 tunnels per head-end (may go up to 600)**

  **max ~1,500 tunnels per link (may go up to 5000)**
- **Provisioning burden may be eased with AutoTunnel Mesh**

# Strategic Deployment: Core Mesh





- **Reduces number of tunnels required**
- **Can be susceptible to “traffic-sloshing”**

# MPLS TE Deployment Considerations



## Statically (explicit) or dynamically established tunnels

- **Dynamic**
  - **must specify bandwidths for tunnels**
    - **Otherwise defaults to IGP shortest path**
  - **Dynamic tunnels introduce indeterminism**
    - **Can be addressed with explicit tunnels or prioritisation scheme – higher priority for larger tunnels**
- **Static (explicit)**
  - **More deterministic**
  - **If strategic approach then computer-aided tools can ease the task**

# Tunnel Sizing



- Determine core traffic demand matrix
  - Full mesh of TE tunnels and Interface MIB
  - NetFlow or BGP Policy Accounting
- Also key is the relationship of tunnel bandwidth to QoS
- Offline sizing
  - Statically set reservation to percentile (e.g. P95) of expected max load
  - Periodically readjust – not in real time, e.g. weekly, monthly
- Dynamic Sizing: autobandwidth
  - Router automatically adjusts reservation (up or down) potentially in near real time based on traffic observed in previous time slot:
  - Tunnel bandwidth is not persistent (lost on reload)

# Summary

- **MPLS Technology has evolved to enable provisioning of many services**
- **MPLS accommodates both connection-oriented and connectionless environments**
- **MPLS provides many technological advances to enable a reliable, controlled, stable infrastructure**
- **All of these advances have been leveraged with standards driven through appropriate bodies – Cisco a major contributor in these efforts**

# MPLS, The Foundation for the NGN
# A quick recap of the Benefits

Cisco.com

- MPLS is a Services Creation & Convergence platform
  - Layer3 MPLS based IP-VPN Services (RFC 2547bis)
  - Layer2 VPN Services (VPWS & VPLS)
    - Legacy Frame-relay and ATM Services
    - New Ethernet based Wire and LAN Services
  - Wide range of Value Added Services (Voice, Video…)
- Quality of Service and Traffic Engineering
- Network Reliability via Link and Node Protection and Restoration
- IP & ATM Integration (Routers and Switches)
- IP & Optical Integration (G-MPLS)
- Large End-user acceptance as enabler of Business IP Services

# Q and A



© 2005, Cisco Systems, Inc. All rights reserved.

Cisco Systems