Introduction to MPLS

Based on MPLS tutorial from
Tim Griffin
and
MPLS/VPN tutorial from
Chris Chase

1

What's all this talk about MPLS?

- "MPLS is going to solve all of our problems"
- "MPLS is a solution in search of a problem"
- "MPLS is all about traffic engineering"
- "MPLS is what I wish on all of my competitors"
- "MPLS is all about virtual private networks"
- "MPLS solves network operations problems"
- "MPLS creates network operations problems"
- "MPLS is all about lowering operational costs"
- "MPLS is going to cost more than its worth"
- "MPLS is the natural next step in Internet evolution"
- "MPLS is too complicated to survive in the Internet"

But what is MPLS anyway?

2

Goals of this Tutorial

- To understand MPLS from a purely technical point of view
  - avoid the hype
  - avoid the cynicism
- To understand the broad technical issues without getting lost in the vast number of details
  - the gains
  - the costs
  - the tradeoffs

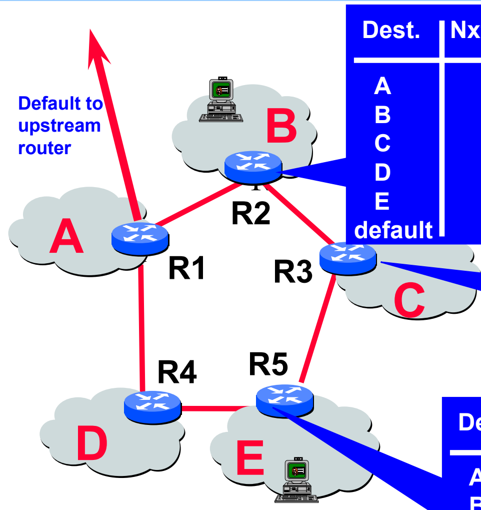
3

Outline

- Why MPLS?
  - Problems with current IP routing and forwarding
  - Complexity of overlay model
- What is MPLS?
  - Label swapping
  - Label distribution
  - Constraint based routing
- What applications could exploit MPLS?
  - Traffic Engineering
  - Virtual Private Networks
    - Both Layer 2 and Layer 3 VPNs

4

IP forwarding paths are implemented with destination-based next hop tables



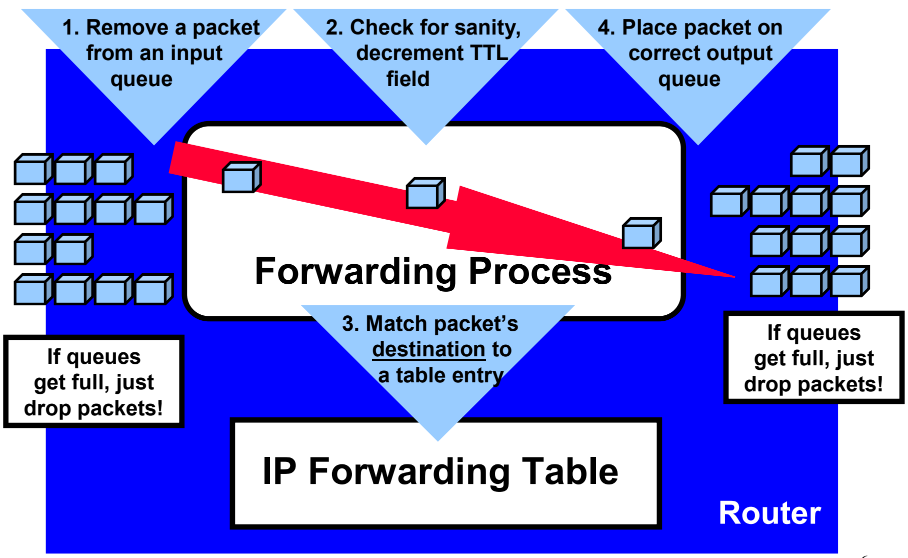
Dest.	Nxt Hop
A	R1
B	Direct
C	R3
D	R1
E	R3
default	R1

Dest.	Nxt Hop
A	R2
B	R2
C	Direct
D	R5
E	R5
default	R2

Dest.	Nxt Hop
A	R4
B	R3
C	R3
D	R4
E	Direct
default	R4

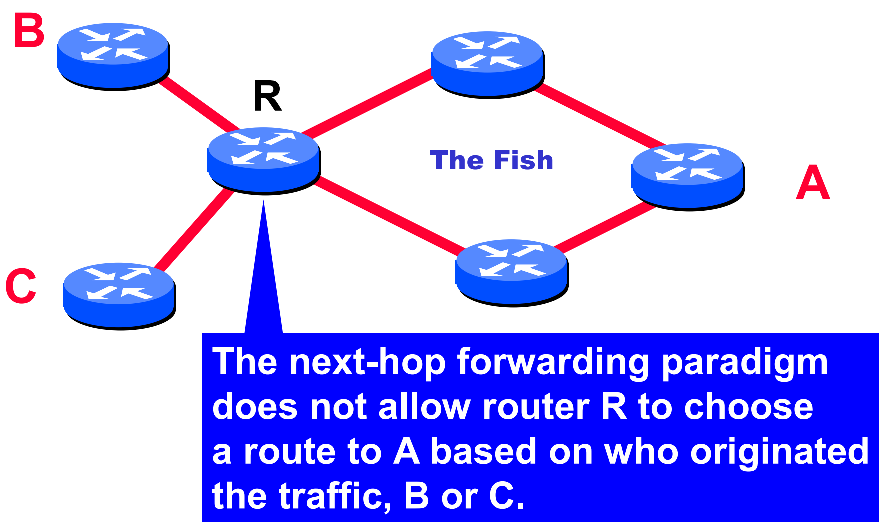
5

IP Forwarding Process



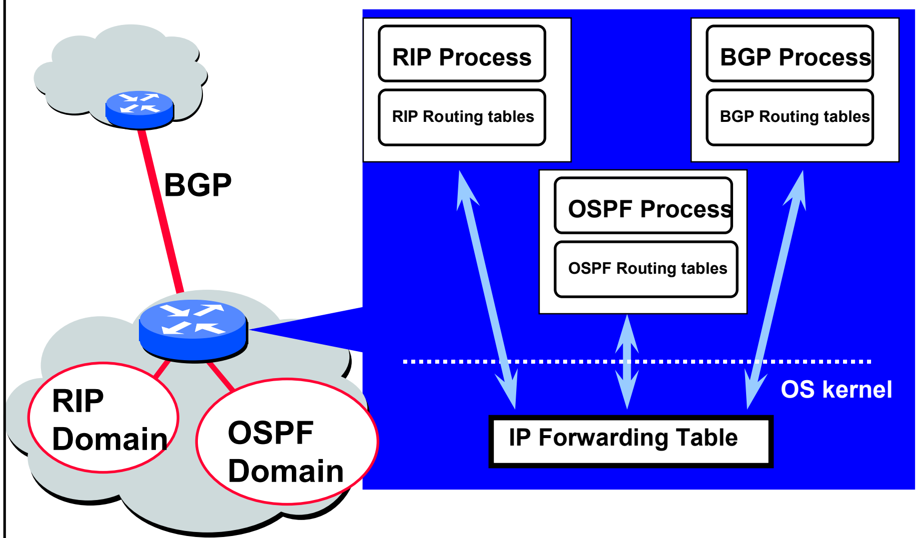
6

IP routing protocols assume all forwarding is destination-based



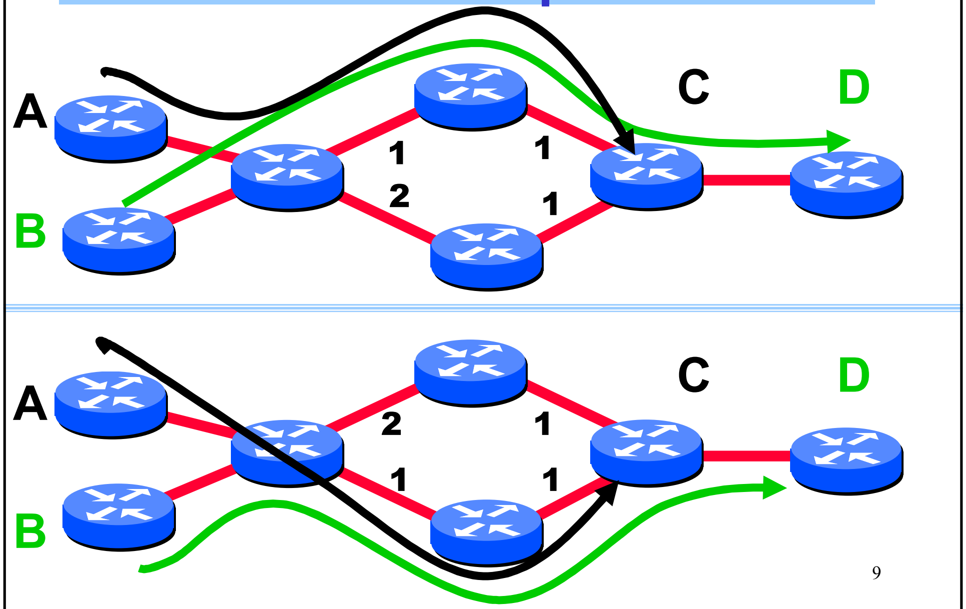
7

IP forwarding tables are maintained by dynamic routing protocols



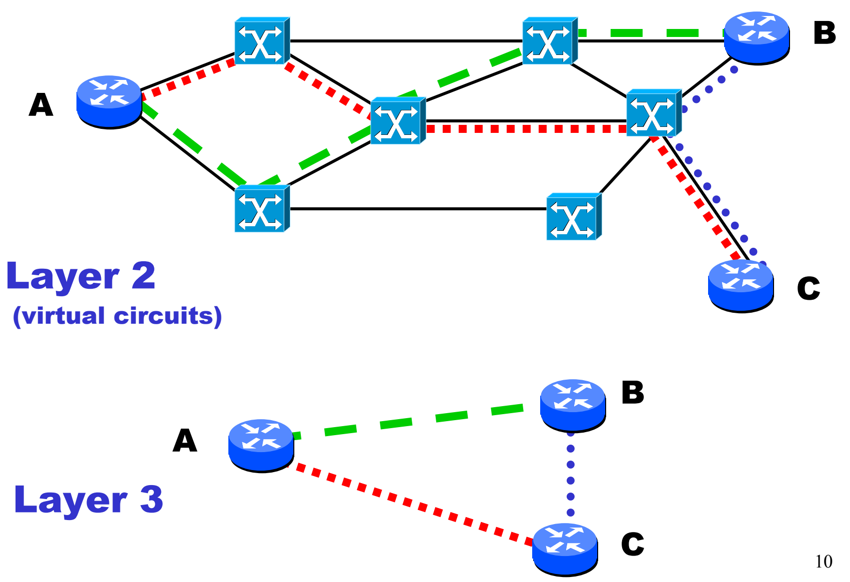
8

Shortest Path Routing: Link weights tend to attract or repel all traffic



9

Overlay Networks



10

Advantages of Overlay Networks

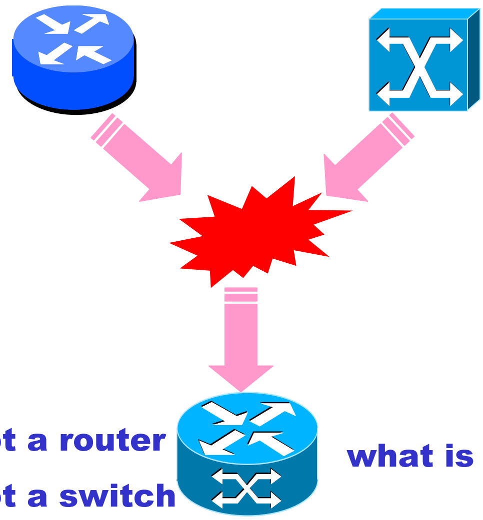
- ATM and Frame Relay switches offer high reliability and low cost
- Virtual circuits can be reengineered without changing the layer 3 network
- Large degree of control over traffic
- Detailed per-circuit statistics
- Isolates layer 2 network management from the details of higher layer services

Problems with Overlay Networks

- Often use proprietary protocols and management tools
- Often requires "full meshing" of statically provisioned virtual circuits
- ATM cell tax ---- about 20% of bandwidth
- If layer 3 is all IP, then the overlay model seems overly complicated and costly
- Advances in optical networking cast some doubt on the entire approach

Overlay model is just fine when layer 2 network provides diverse non IP services (e.g., IPv6, AppleTalk, IPX, …)

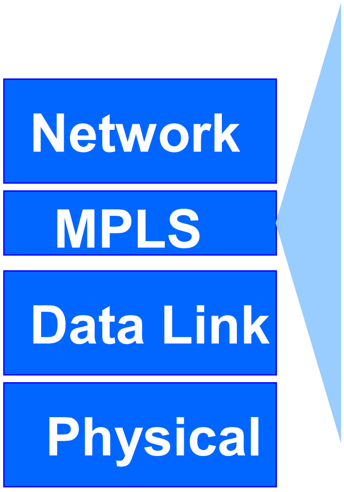
Blur Layer 2 and 3?



Sanity Check?

- **The problems with IP forwarding and routing do not *require* technologies like MPLS**
  - Many can be addressed with simple solutions. Like the design of simple networks!
  - The problems are not "show stoppers"
  - The MPLS cure will have side effects
  - For many applications, TCP/IP handles congestion very well
- **Technologies like MPLS may be very valuable if they can enable new services and generate new revenue**

MPLS = MultiProtocol Label Switching



- A "Layer 2.5" tunneling protocol
- Based on ATM-like notion of "label swapping"
- A simple way of labeling each network layer packet
- Independent of Link Layer
- Independent of Network Layer
- Used to set up "Label-switched paths" (LSP), similar to ATM PVCs

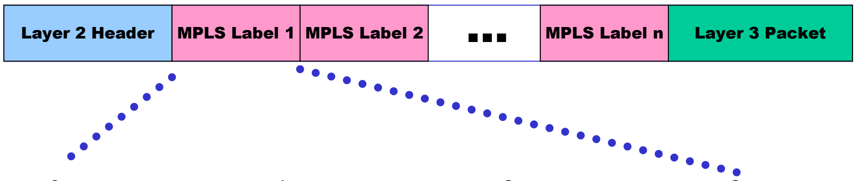
RFC 3031 : Multiprotocol Label Switching Architecture

15

MPLS Data Plane

16

Generic MPLS Encapsulation




```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Label                  | Exp |S|       TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

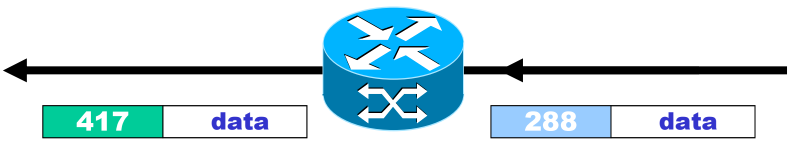
Often called a "shim" (or "sham") header

RFC 3032. MPLS Label Stack Encoding

- **Label: Label Value, 20 bits**
- **Exp: Experimental, 3 bits**
- **S: Bottom of Stack, 1 bit**
- **TTL: Time to Live, 8 bits**

17

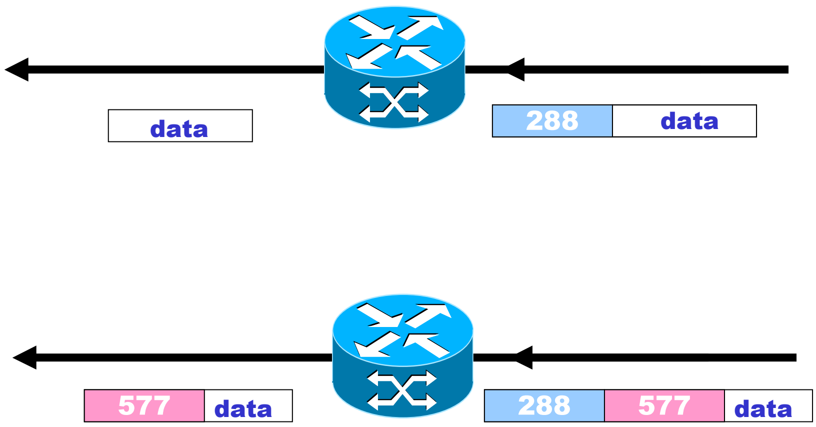
Forwarding via Label Swapping



Labels are short, fixed-length values.

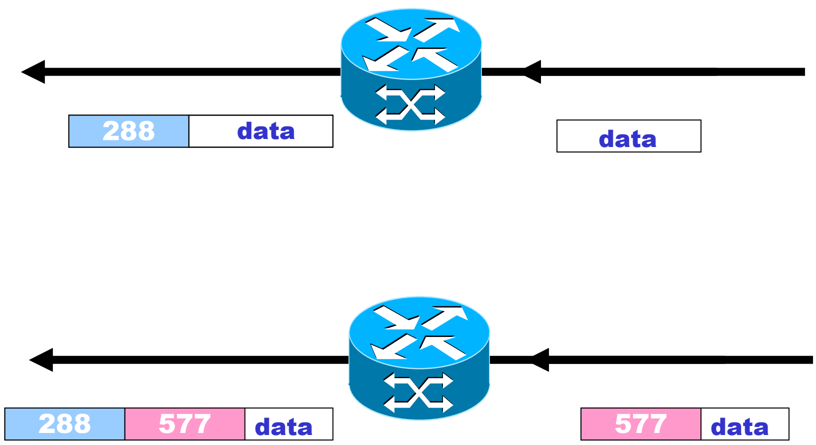
18

Popping Labels



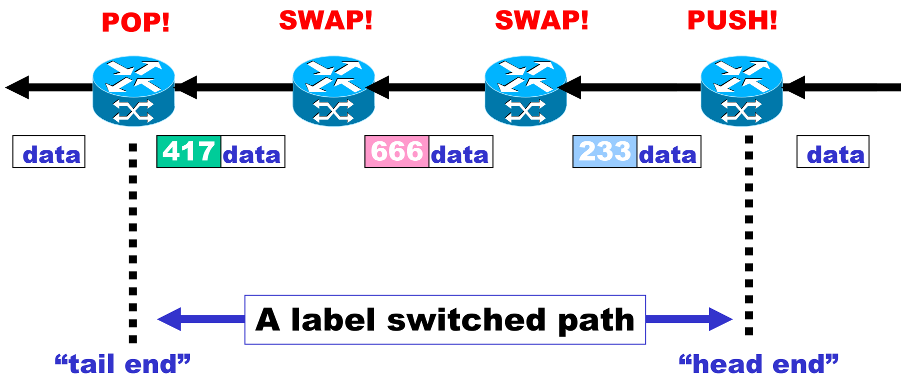
19

Pushing Labels



20

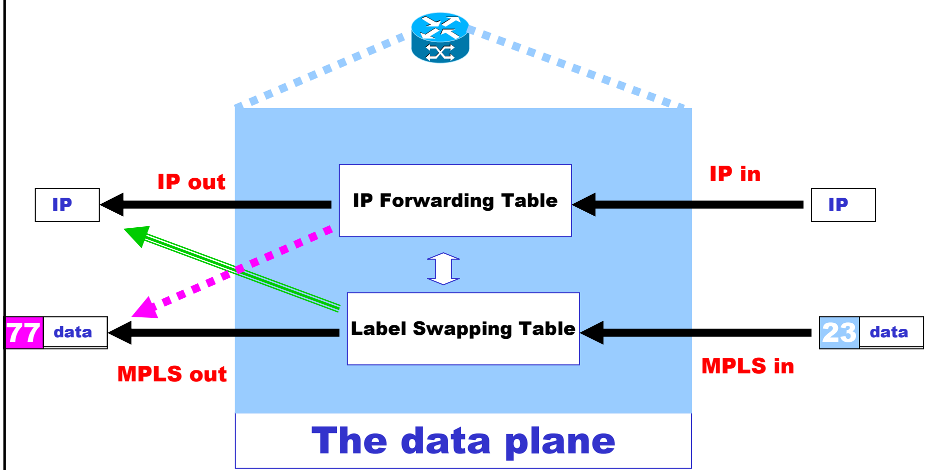
A Label Switched Path (LSP)



Often called an MPLS tunnel: payload headers are not Inspected inside of an LSP. Payload could be MPLS ...

21

Label Switched Routers

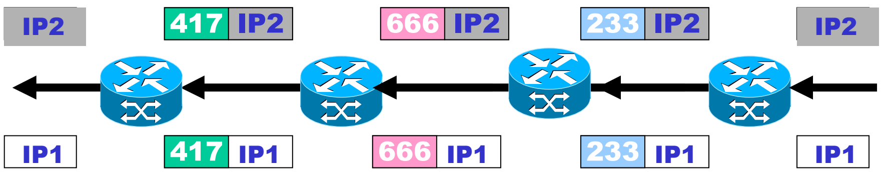


represents IP Lookup + label push

represents label pop + IP lookup

22

Forwarding Equivalence Class (FEC)

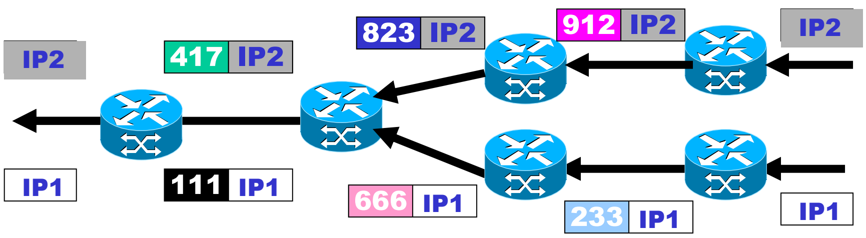


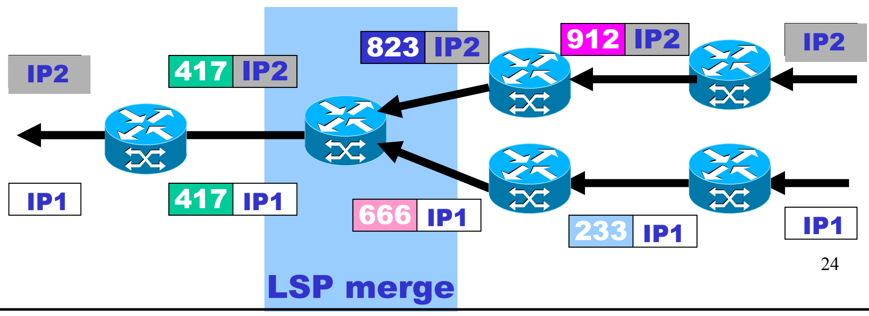
Packets IP1 and IP2 are forwarded in the same way --- they are in the same FEC.

Network layer headers are not inspected inside an MPLS LSP. This means that inside of the tunnel the LSRs do not need full IP forwarding table.

23

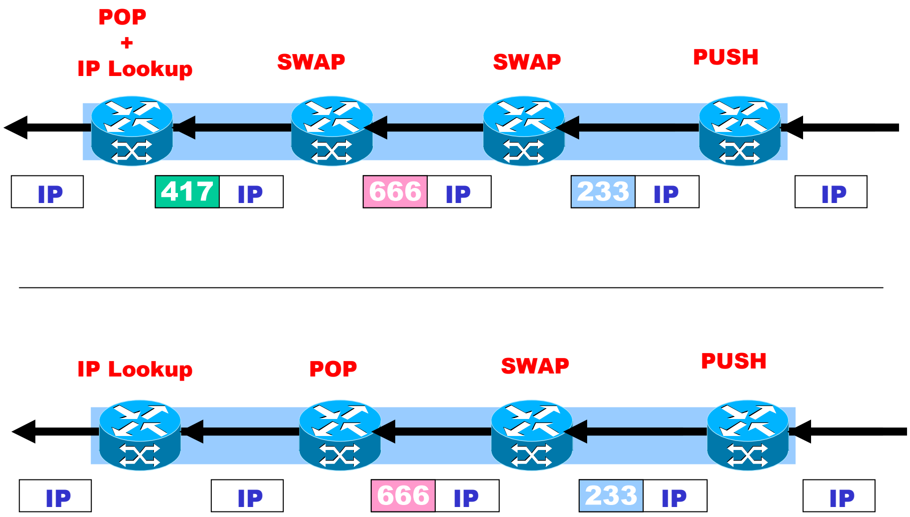
LSP Merge





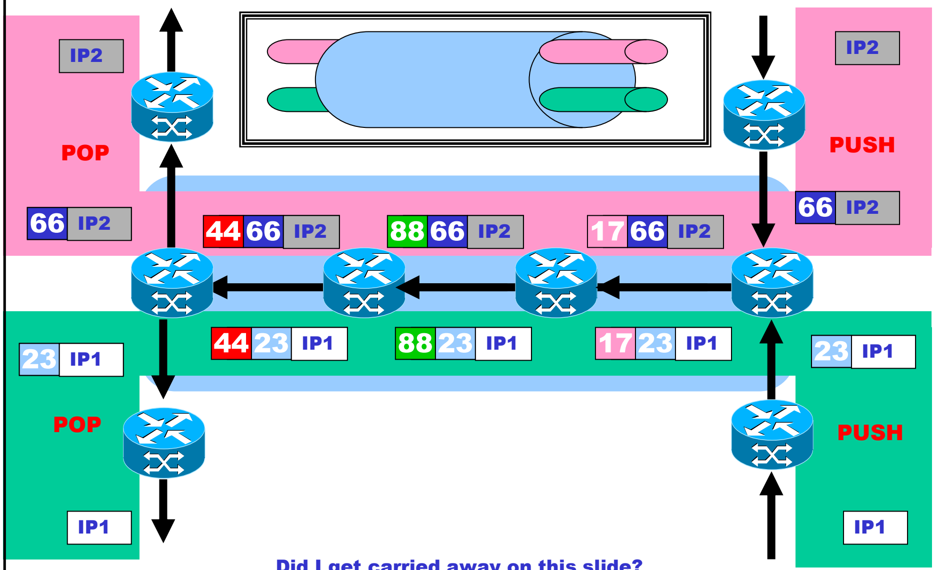
24

Penultimate Hop Popping



Q: How can MPLS support LSP division?

LSP Hierarchy via Label Stacking



Did I get carried away on this slide?

27

Q: Any disadvantage of MPLS tunneling?

28

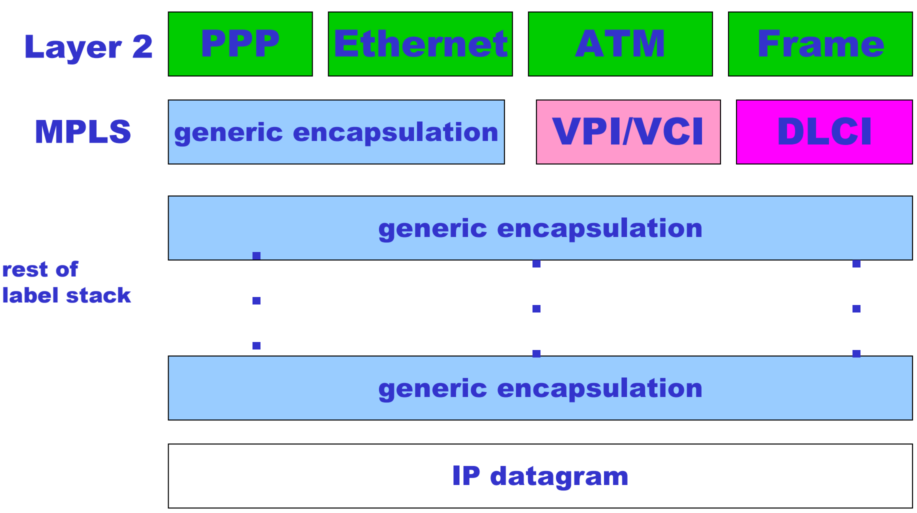
MPLS Tunnels come at a cost

- **ICMP messages may be generated in the middle of a tunnel, but the source address of the "bad packet" may not be in the IP forwarding table of the LSR!**
  - **TTL expired: traceroute depends on this!**
  - **MTU exceeded: Path MTU Discovery (RFC1191) depends on this!**

None of the proposed solutions are without their own problems…

29

MPLS also supports "native encapsulation"



30

But Native Labels May Cause Big Headaches

- No TTL!
  - Loop detection?
  - Loop prevention?
- LSP merge may not be supported
  - Label bindings cannot flow from destination to source, but must be requested at source

MPLS was initially designed to exploit the existence
of ATM hardware and reduce the complexity of overlay networks.
But IP/MPLS with native ATM labels results in a large number
of problems and complications.

31

MPLS Control Plane

32

Basic MPLS Control Plane

MPLS control plane
=
IP control plane
+
label distribution

Label distribution protocols are needed to
(1) create label⟺FEC bindings
(2) distribute bindings to neighbors,
(3) maintain consistent label swapping tables

33

Label Distribution: Option I

"Piggyback" label information on top of existing IP routing protocol

Good Points
- Guarantees consistency of IP forwarding tables and MPLS label swapping tables
- No "new" protocol required

Bad Points
- Allows only traditional destination-based, hop-by-hop forwarding paths
- Some IP routing protocols are not suitable
  - Need explicit binding of label to FEC
  - Link state protocols (OSPF, ISIS) are implicit, and so are not good piggyback candidates
  - Distance vector (RIP) and path vector (BGP) are good candidates. Example: BGP+

34

Label Distribution: Option II

Create new label distribution protocol(s)

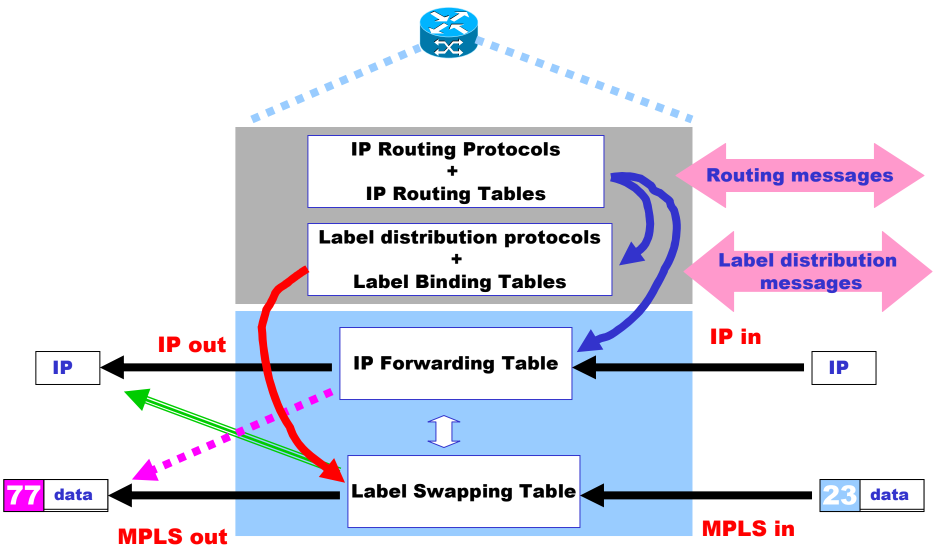
Good Points
- Compatible with "Link State" routing protocols
- Not limited to destination-based, hop-by-hop forwarding paths

Bad Points
- Additional complexity of new protocol and interactions with existing protocols
- Transient inconsistencies between IP forwarding tables and MPLS label swapping tables

Examples: LDP (IETF) and TDP (Cisco proprietary)

35

The Control Plane



36

Label Distribution with BGP

Carrying Label Information in BGP-4
draft-ietf-mpls-bgp4-05.txt (1/2001)

Associates a label (or label stack) with the BGP next hop.

Uses multiprotocol features of BGP:

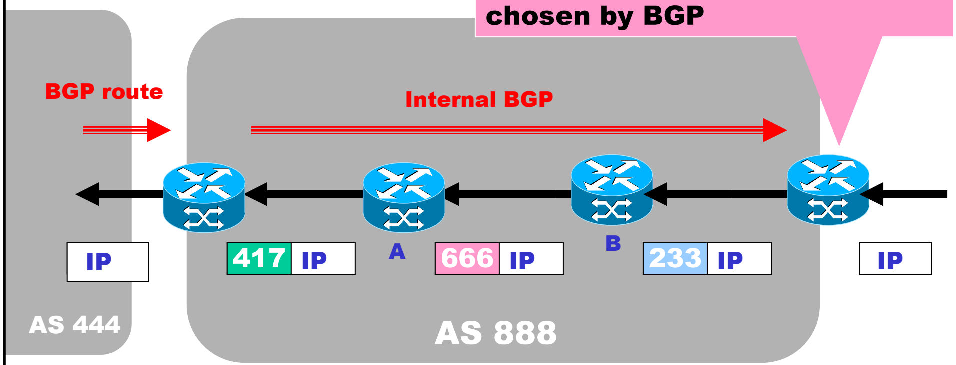
RFC 2283. Multiprotocol Extensions for BGP-4

So routes *with* labels are in a different address space than a vanilla routes (no labels)

37

BGP piggyback not required for simple iBGP optimization

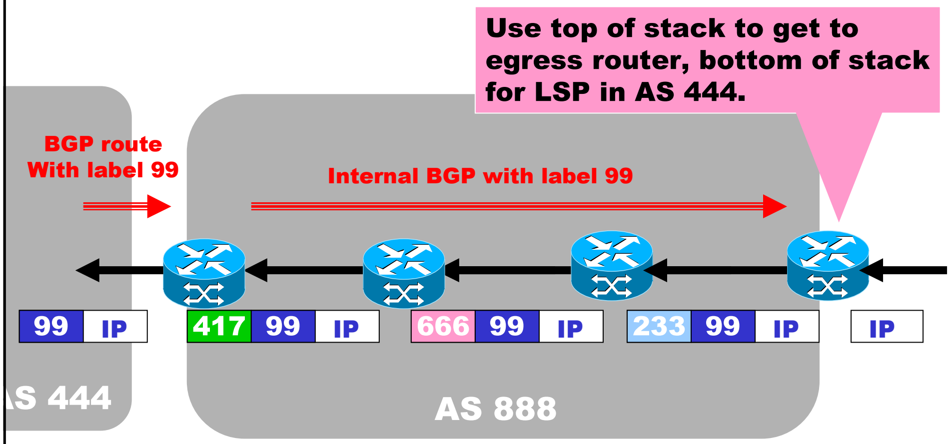
Map traffic to the LSP that terminates at the egress router chosen by BGP



Routers A and B do not need full routing tables.
They only need IGP routes (and label bindings).

38

BGP piggyback allows Interdomain LSPs



39

MPLS tunnels can decrease size of core routing state

- **Core routers need only IGP routes and LSPs for IGP routes**
- **Implies less route oscillation**
- **Implies less memory**
- **Implies less CPU usage**

Are these *really* problems?

BUT: still need route reflectors to avoid full mesh and/or to reduce BGP table size at border routers

BUT: since your core routers do not have full tables you now have all of the MPLS problems associated with ICMP source unknown (TTL, MTU, traceroute ...)

40

Label Distribution Protocol (LDP)

RFC 3036. LDP Specification. (1/2001)

- Dynamic distribution of label binding information
- Supports only vanilla IP hop-by-hop paths
- LSR discovery
- Reliable transport with TCP
- Incremental maintenance of label swapping tables (only deltas are exchanged)
- Designed to be extensible with Type-Length-Value (TLV) coding of messages
- Modes of behavior that are negotiated during session initialization
  - Label retention (liberal or conservative)
  - LSP control (ordered or independent)
  - Label assignment (unsolicited or on-demand)

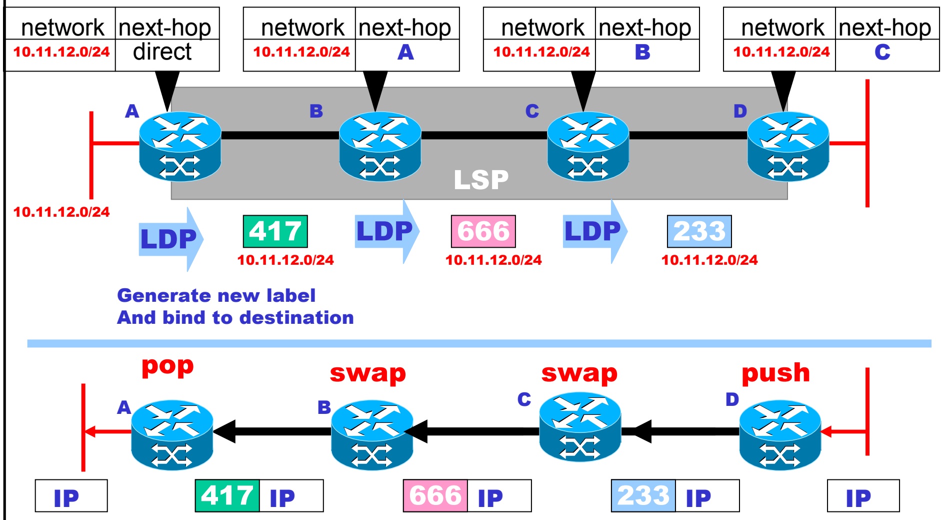
41

LDP Message Categories

- **Discovery messages**: used to announce and maintain the presence of an LSR in a network.
- **Session messages**: used to establish, maintain, and terminate sessions between LDP peers.
- **Advertisement messages:** used to create, change, and delete label mappings for FECs.
- **Notification messages**: used to provide advisory information and to signal error information.

42

LDP and Hop-by-Hop routing



43

MPLS Fast Reroute

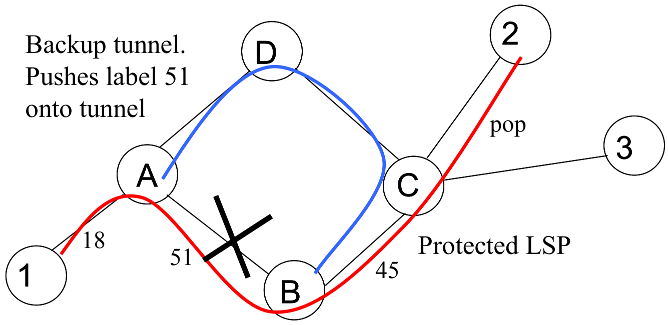
- **Using MPLS TE to improve availability**
  - **RSVP-TE creates backup tunnels**
  - **On failure of protected LSP, packets are shoved down backup LSP tunnel**
  - **Switchover is faster than waiting for CSPF to calculate and signal a new LSP**
- **For local repair (link or node) can recover ~100ms or better**
  - **Backup LSP is already in place, so as soon as the failure is detected locally the headend just needs to reprogram the label FIB**

Q: Why MPLS can recover from failure faster than OSPF?

44

Link Protection

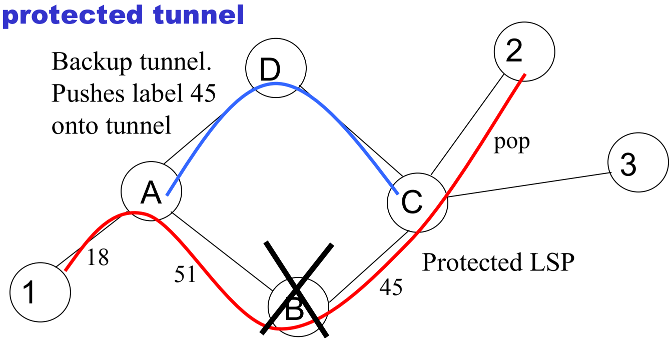
- **Create backup LSP around link to Next Hop**
- **With or without reservation**
  - **Can also backup normal LDP LSP**



45

Node Protection

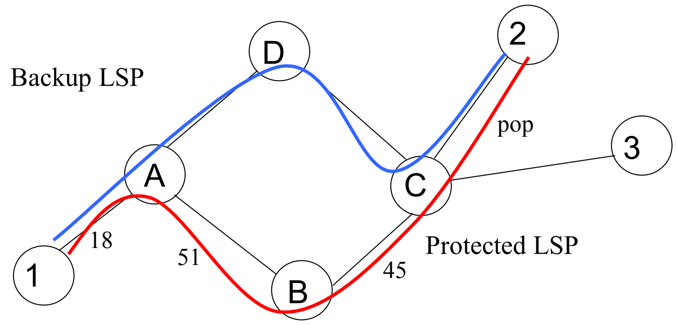
- **Create backup tunnel LSP for two hops away (next-next hop)**
- **Backs up RSVP-TE tunnel**
  - **Learns labels from RESV recorded route of protected tunnel**



46

Path Protection

- Create an end-to-end diverse backup tunnel
- Slower than local protection – have to wait for headend to detect failure



47

VPNs with MPLS

"Traditional" VPN overlay model:

MPLS-based Layer 2 VPNs
draft-kompella-mpls-l2vpn-02.txt

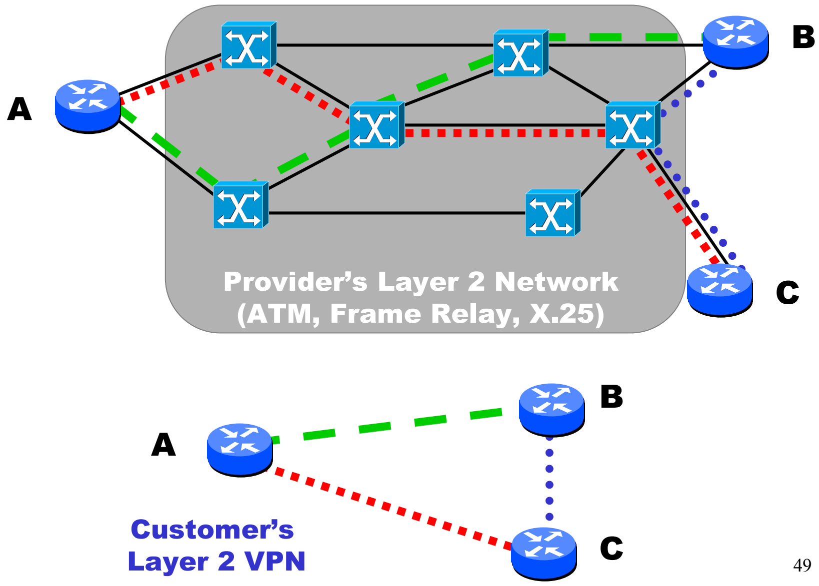
Whither Layer 2 VPNs?
draft-kb-ppvpn-l2vpn-motiv-00.txt

New VPN peering model:

RFC 2547. BGP/MPLS VPNs

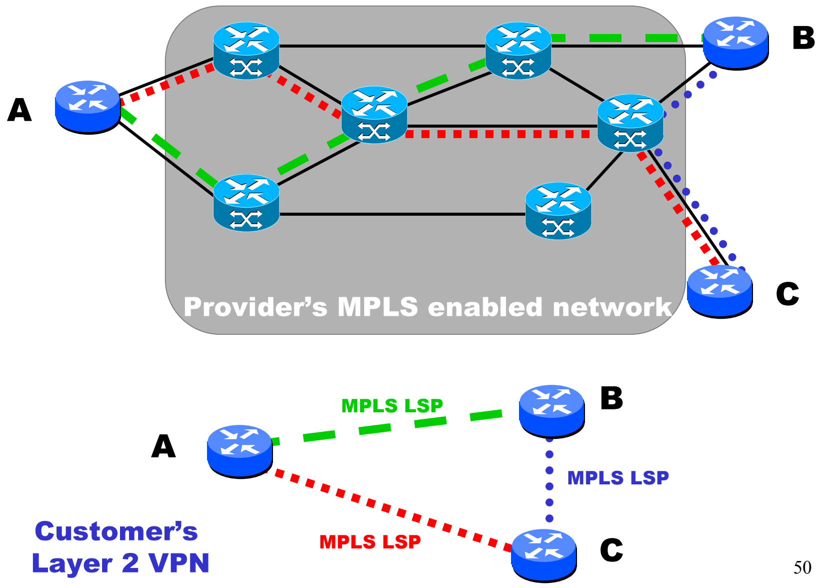
48

Traditional Overlay VPNs



49

Why Not Use MPLS Tunnels?



50

Potential Advantages of MPLS Layer 2 VPNs

- **Provider needs only a single network infrastructure to support public IP, and VPN services, traffic engineered services, and differentiated services**
- **Additional routing burden on provider is bounded**
- **Clean separation of administrative responsibilities. Service provider does MPLS connectivity, customer does layer 3 connectivity**
- **Easy transition for customers currently using traditional Layer 2 VPNs**

51

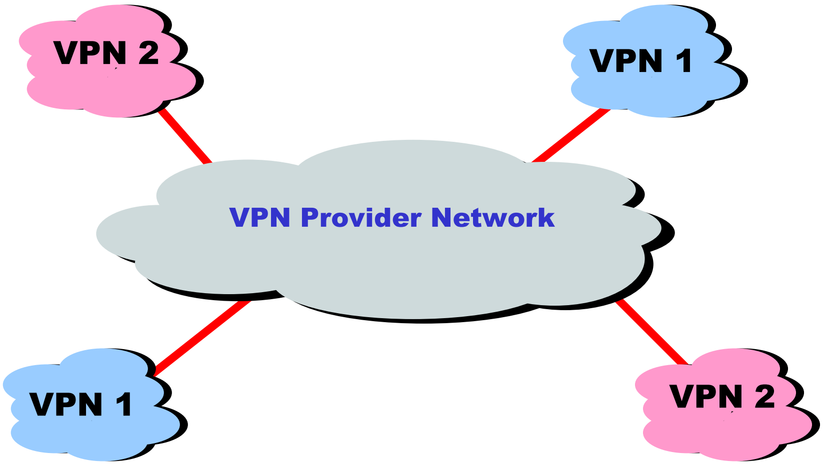
BGP/MPLS VPNs

- RFC 2547
- Is Peer Model of VPN (not Overlay)
- Also draft-rosen-rfc2547bis-02.txt
- Cisco configuration info :
  - http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120t/120t5/vpn.htm

AT&T's IPFR service is based on this RFC.

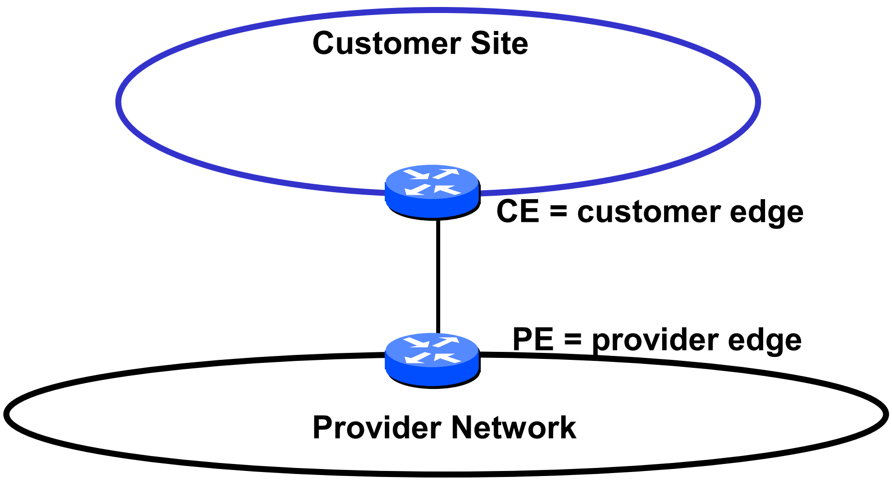
52

RFC 2547 Model

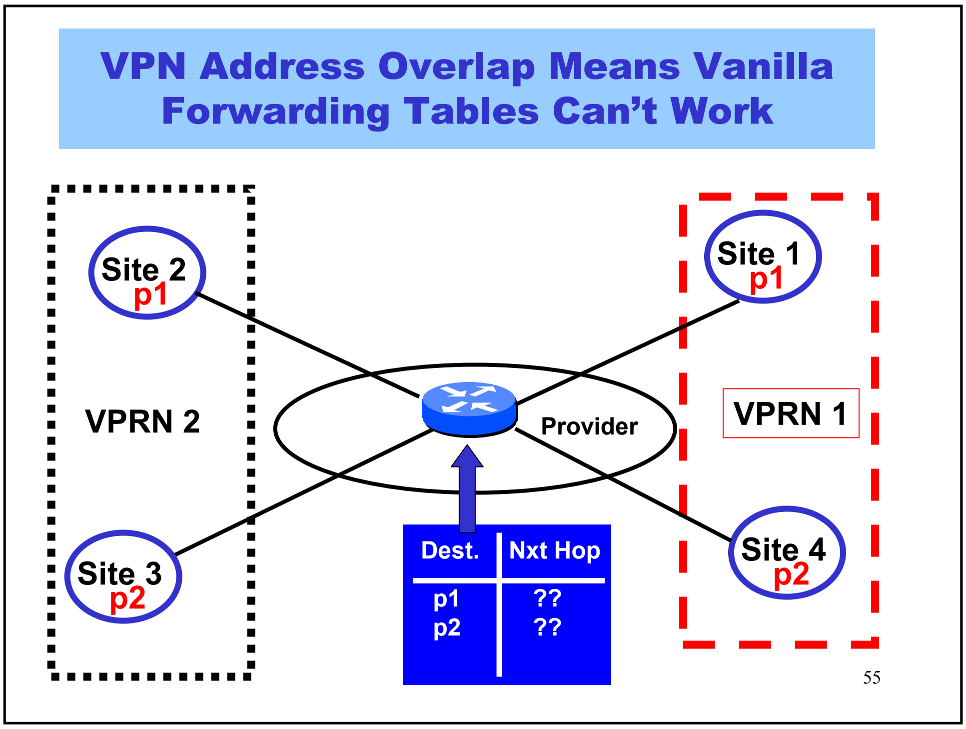


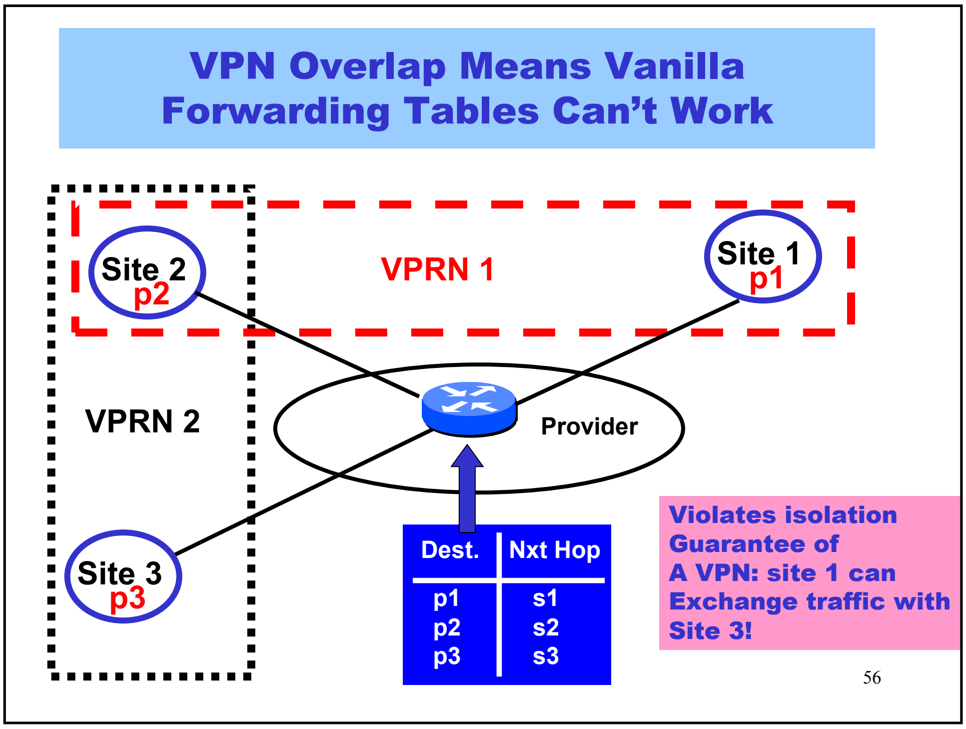
53

CEs and PEs



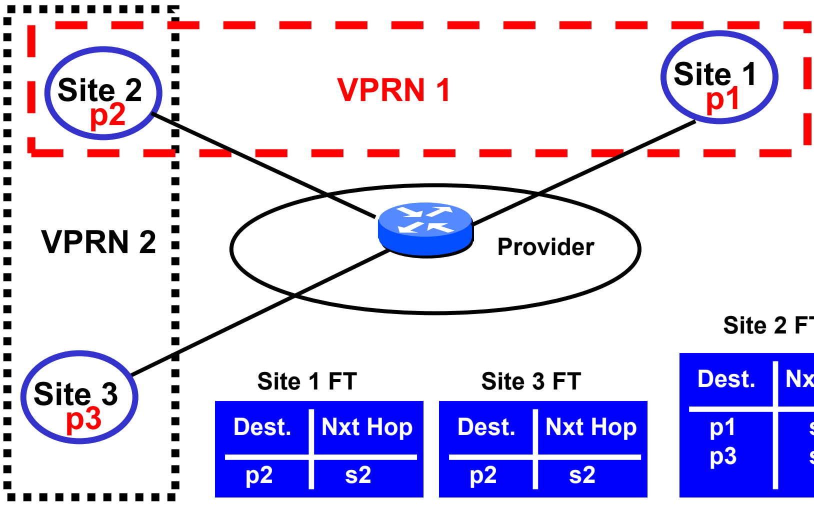
54


VPN Address Overlap Means Vanilla Forwarding Tables Can't Work
Site 2
p1
VPRN 2
Site 3
p2
Site 1
p1
VPRN 1
Site 4
p2
Provider
Dest.
Nxt Hop
p1
??
p2
??
55


VPN Overlap Means Vanilla Forwarding Tables Can't Work
Site 2
p2
VPRN 1
Site 1
p1
VPRN 2
Provider
Site 3
p3
Dest.
Nxt Hop
p1
s1
p2
s2
p3
s3
Violates isolation Guarantee of A VPN: site 1 can Exchange traffic with Site 3!
56

RFC 2547 : Per site forwarding tables

Called VRFs, for "VPN Routing and Forwarding" tables.



Site 1 FT

Dest.	Nxt Hop
p2	s2

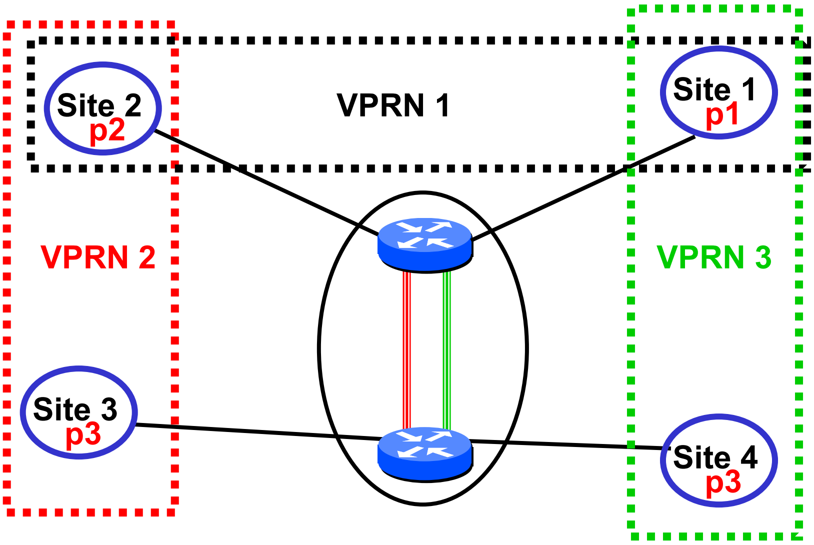
Site 3 FT

Dest.	Nxt Hop
p2	s2

Site 2 FT

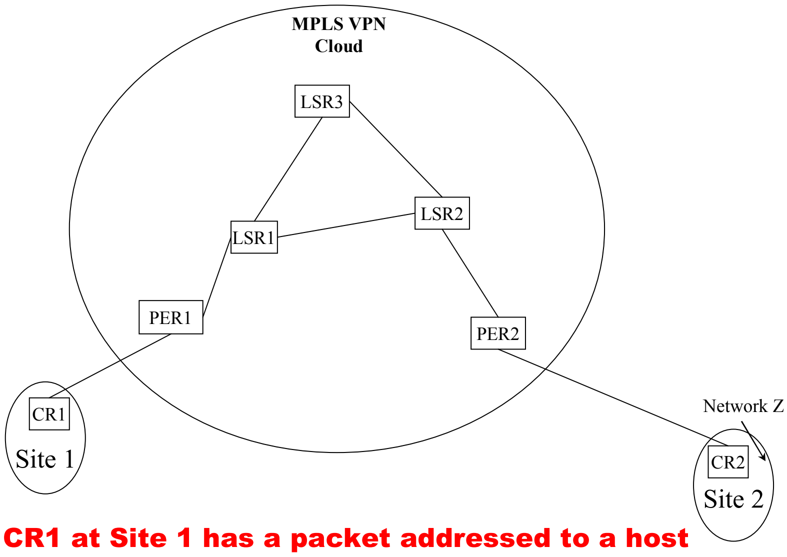
Dest.	Nxt Hop
p1 p3	s1 s3

Tunnels required across backbone



58

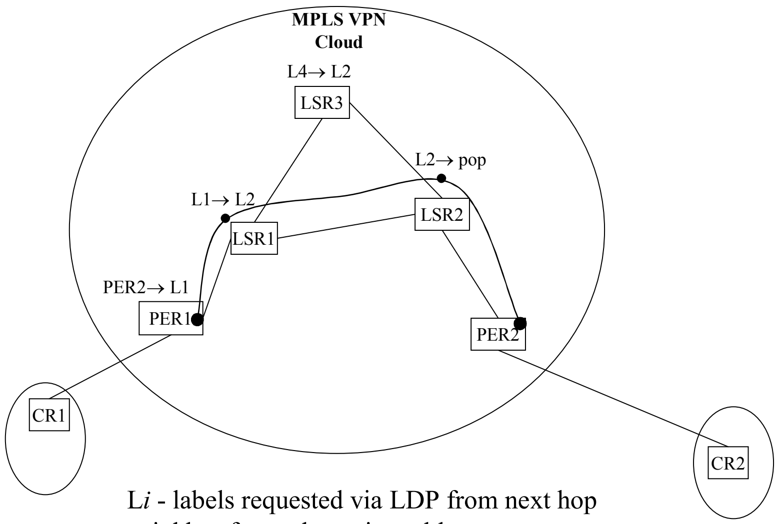
Follow the Route and Follow the Packet



CR1 at Site 1 has a packet addressed to a host in network Z at Site 2. How does it get there?

59

LSP Setup for OSPF Route to PER2



L*i* - labels requested via LDP from next hop neighbor for each routing table entry

LSP for the OSPF route to reach PER2

60

How MPLS VPNs work

1) Follow the routes

- Each VPN on a PER has a private routing table
  - Called a Virtual Routing Forwarding (vrf) table
  - vrf is assigned attributes that are unique to the VPN
    - Route Targets (RT) - attached to VPN routes.
      - » only vrfs with common RTs share routes
    - Route Distinguishers (RD) - appended to routes to ensure uniqueness even if VPNs have overlapping address spaces
      - » Creates a new address family called vpnv4 = RD+ipv4
    - NOTE: RTs and RDs are applied to routes, NOT packets

2) Follow the packet

- A stack of two labels is used to forward the packet on the interior LSP and then external interface

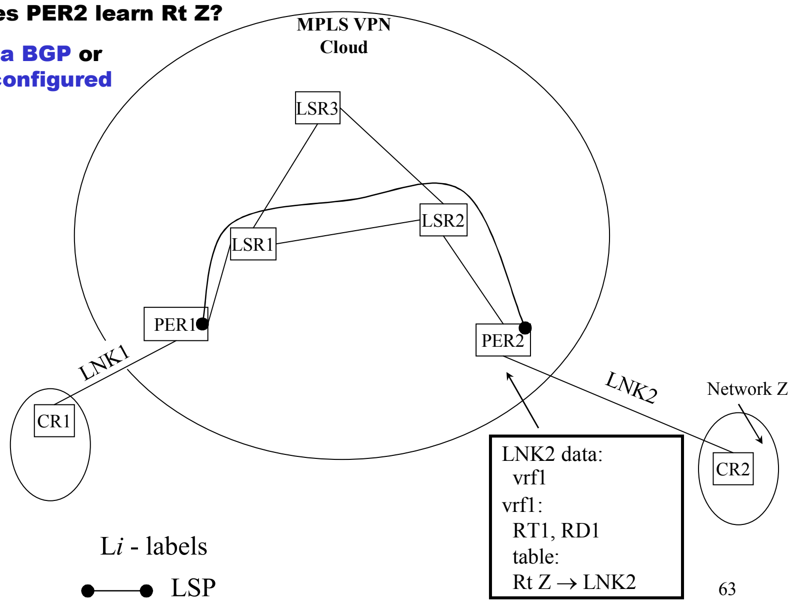
VPN extensions

- **Route Target (RT)**
  - BGP 64 bit extended community value
  - First 16bit identify as RT type.
  - Other 48 bit is variable
    - Conventional format – ASN:X, i.e., 16b:32b
- **Route Distinguisher (RD)**
  - BGP 64 bit extended community value
  - First 16bit identify as RD type.
  - Other 48 bit is variable
    - Conventional format – ASN:X, i.e., 16b:32b

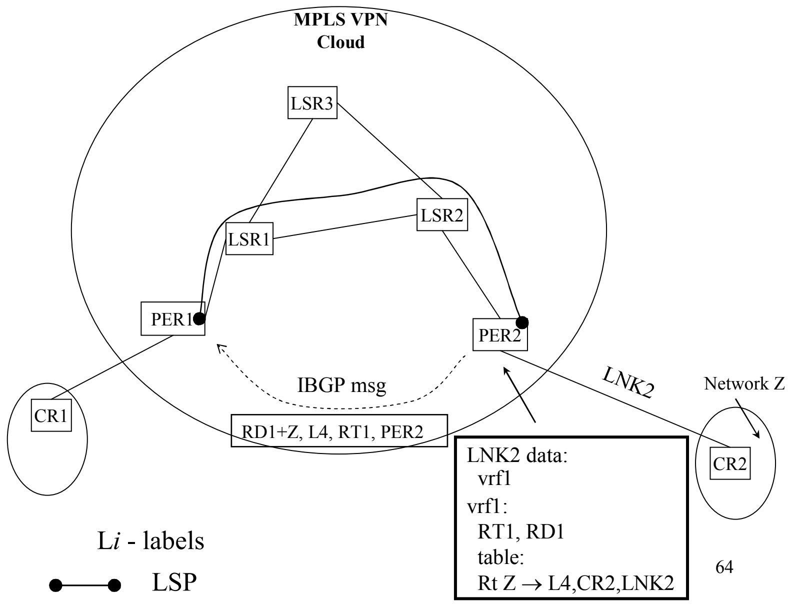
Distributing Customer Routes

Q: How does PER2 learn Rt Z?

A: Either via BGP or statically configured

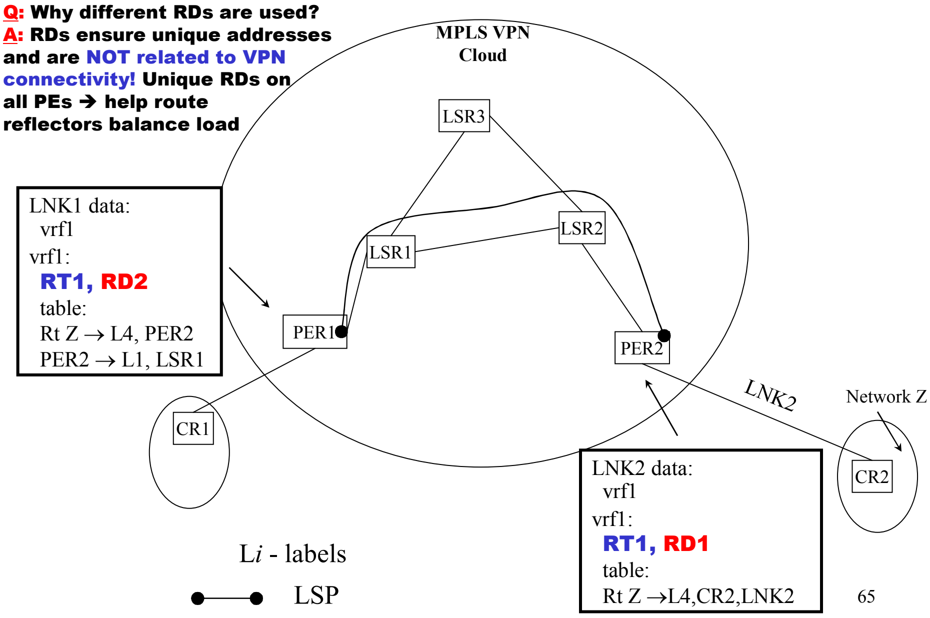


Customer Routes Distributed via IBGP with Label



Only vrfs with Matching RTs Import Route

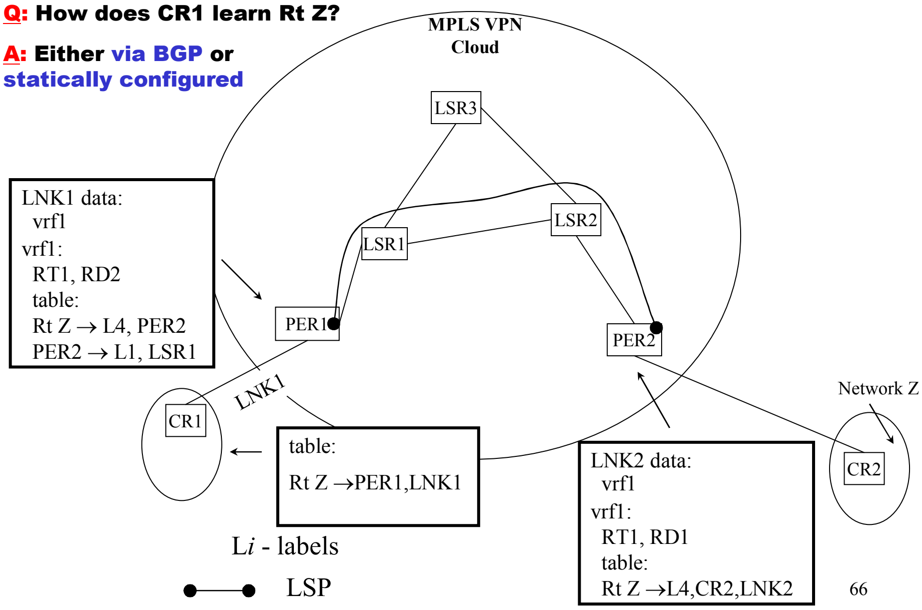
Q: Why different RDs are used?
A: RDs ensure unique addresses and are NOT related to VPN connectivity! Unique RDs on all PEs → help route reflectors balance load



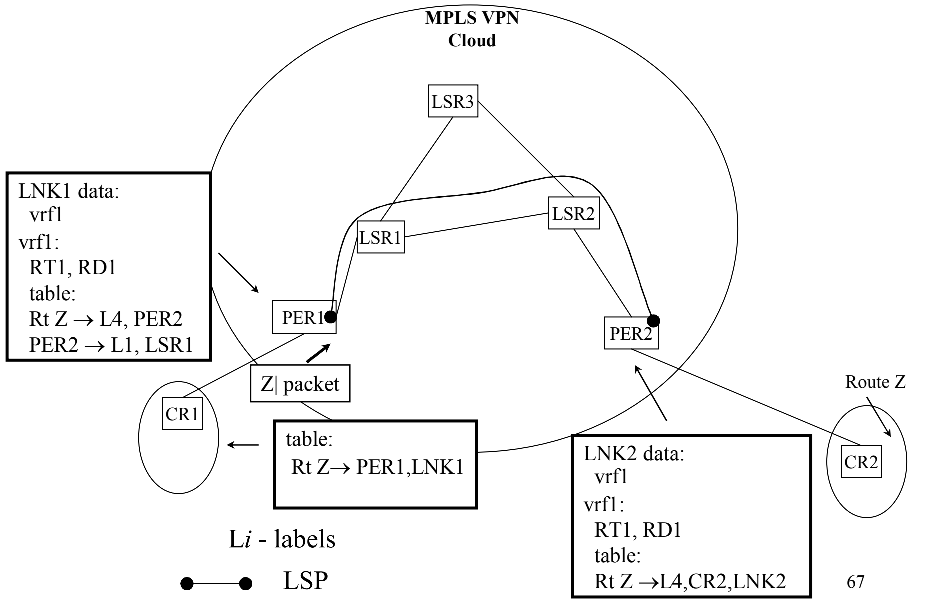
CR1 learns RT Z

Q: How does CR1 learn Rt Z?

A: Either via BGP or statically configured

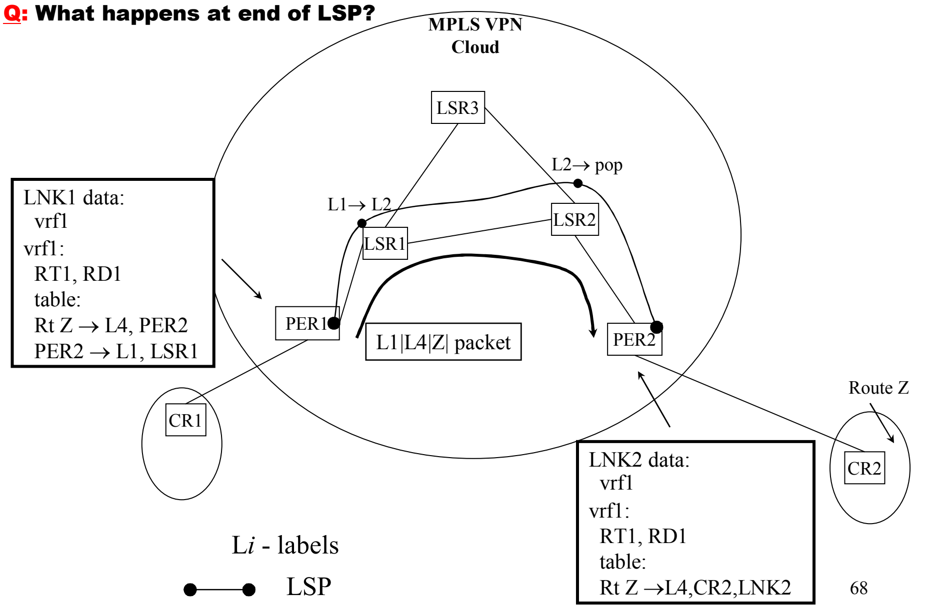


Packet for Rt Z forwarded by CR1



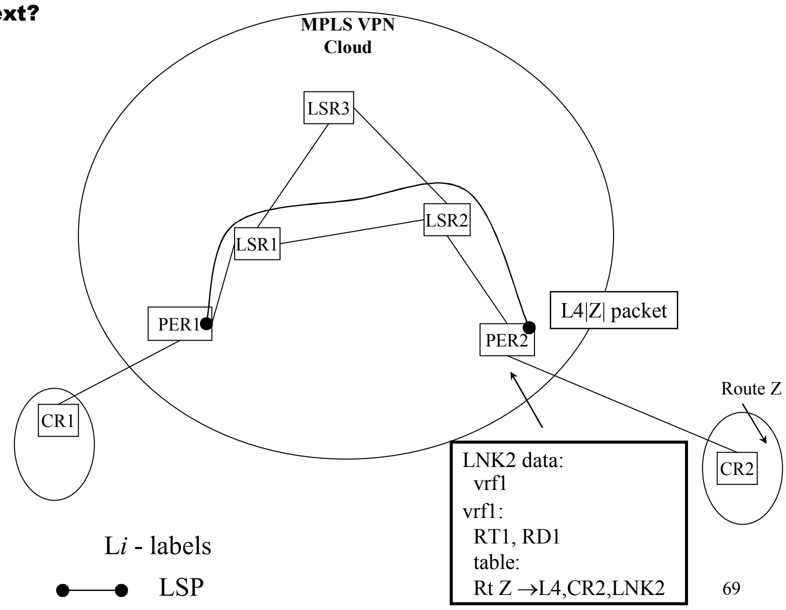
Top label is label-switched through interior

Q: What happens at end of LSP?



Top label popped at end of LSP

Q: What next?

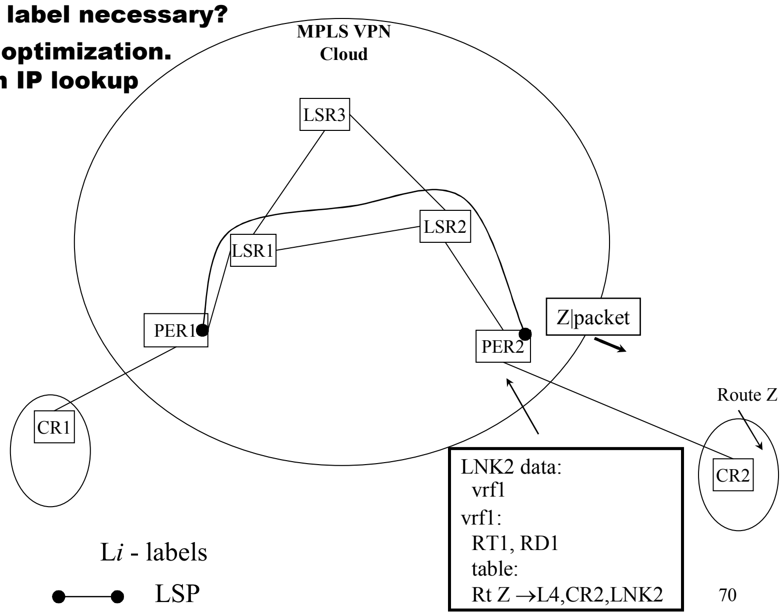


69

Inner label determines egress interface and then is popped

Q: Is inner label necessary?

A: No. An optimization.
It saves an IP lookup



70

Purpose of BGP Label

- Indicates which vrf and optionally which interface on the egress PER
- Locally, the egress PER will treat labels in two possible ways:
  - Non-aggregate label is associated with an external route
    - Will be switched directly to an outgoing interface
    - IP header is not examined
  - Aggregate label is associated with a locally originated or directly connected route
    - Packet will be looked up in the vrf context

71

MPLS in Core Not Needed

- MPLS for IGP domain serves as a tunneling method among PERs
- Could use other tunneling methods
- Advantages to MPLS:
  - Full mesh of LSP tunnels automatically created
  - Can use MPLS TE
- Internet draft to use IP or GRE tunneling
  - Automatically (treat vpnv4 BGP next hop as a recursive encapsulation)
  - BGP/IPsec VPN <draft-declercq-bgp-ipsec-vpn-00.txt>

72

RFC 2547 Summary

- Piggyback VPN information on BGP
  - New address family
  - New attributes for membership
- New Per-site forwarding tables (VRFs)
- Use MPLS Tunnels between PEs
  - No need for VPN routes on backbone LSRs, only on PEs

73

MPLS VPN Security

- Private routing table for each VPN (vrf)
- VPN membership identity associated with each access connection
  - VPN membership is not determined by IP header, only by interface (e.g., DLCI, VPI/VCI, PPP, VLAN tag).
  - Label and RT for VPN attached to routes advertised for interface.
  - Route and its matching label are only imported by routing tables that match the VPN RT.
  - Impossible for a packet on a PVC in one vrf to spoof its way or jump into another vrf

74

Layer 2 VPNs vs. BGP/MPLS VPNs

- Customer routing stays with customer
- May allow an easier transition for customers currently using Frame/ATM circuits
- Familiar paradigm
- Easier to extend to multiple providers

- Customer routing is "outsourced" to provider
- Transition may be complicated if customer has many extranets or multiple providers
- New "peering" paradigm
- Not clear how multiple provider will work (IMHO)

75

Summary

MPLS is an interesting and potentially valuable technology because it

- provides an efficient and scalable tunneling mechanism
- provides an efficient and scalable mechanism for extending IP routing with explicit routes

76

More info on MPLS

- MPLS working group
  - http://www.ietf.org/html.charters/mpls-charter.html
- MPLS email list archive
  - http://cell.onecall.net/cell-relay/archives/mpls/mpls.index.html
- MPLS Resource Center
  - http://www.mplsrc.com
- Peter Ashwood-Smith's NANOG Tutorial
  - http://www.nanog.org/mtg-9910/mpls.html
- MPLS: Technology and Applications. By Bruce Davie and Yakov Rekhter. Morgan Kaufmann. 2000.
- MPLS: Is it all it's cracked up to be? Talk by Pravin K. Johri
  - http://buckaroo.mt.att.com/~pravin/docs/mpls.pdf

77

More info on MPLS TE

- tewg working group
  - http://www.ietf.org/html.charters/tewg-charter.html
- NANOG Tutorial by Jeff Doyle and Chris Summers
  - http://www.nanog.org/mtg-0006/mpls.html
- NANOG Tutorial by Robert Raszuk
  - http://www.nanog.org/mtg-0002/robert.html

78

More info on MPLS VPNs

- **PPVPN working group**
  - **http://www.ietf.org/html.charters/ppvpn-charter.html**
- **PPVPN Archive**
  - **http://nbvpn.francetelecom.com**
- **NANOG Panel:Provider-Provisioned VPNs**
  - **http://www.nanog.org/mtg-0102/jessica.html**
- **MPLS and VPN Architectures. By Ivan Pepelnjak and Jim Guichard. Cisco Press. 2001**

79