AT&T Research

An Introduction to MPLS

Timothy G. Griffin

griffin@research.att.com

http://www.research.att.com/~griffin

November 21, 2002

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?

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

Keep in Mind

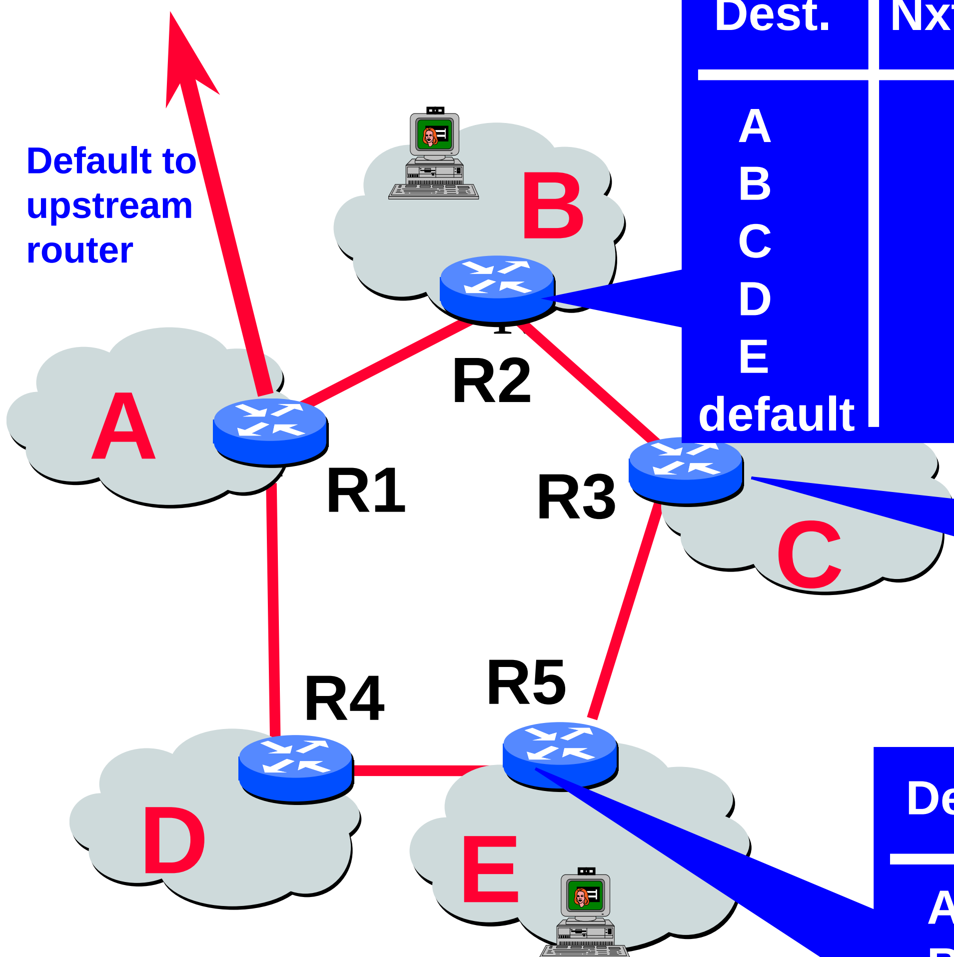
- MPLS is an emerging technology
- Many technical issues have not yet been resolved
- Interest and enthusiasm is not universal, but primarily found in large providers (and their vendors)
- Standards are rapidly evolving
- Implementations are rapidly evolving
- Operational experience and expertise still very scarce

Expect interoperability problems and feature availability problems for the next few years

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

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

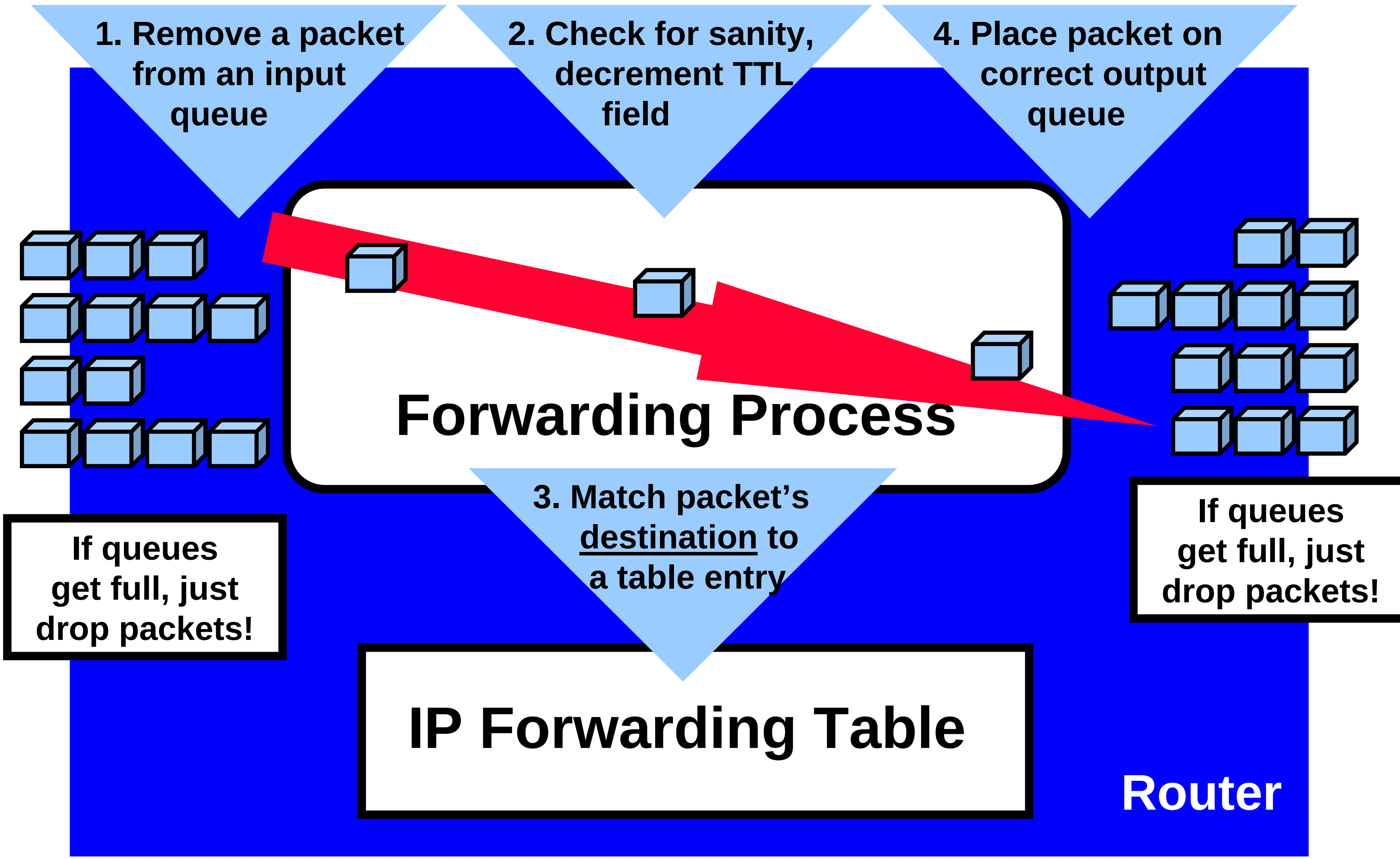


Dest.	Nxt Hop
A	R1
B	Direct
C	R3
D	R1
E	R3
default	R1

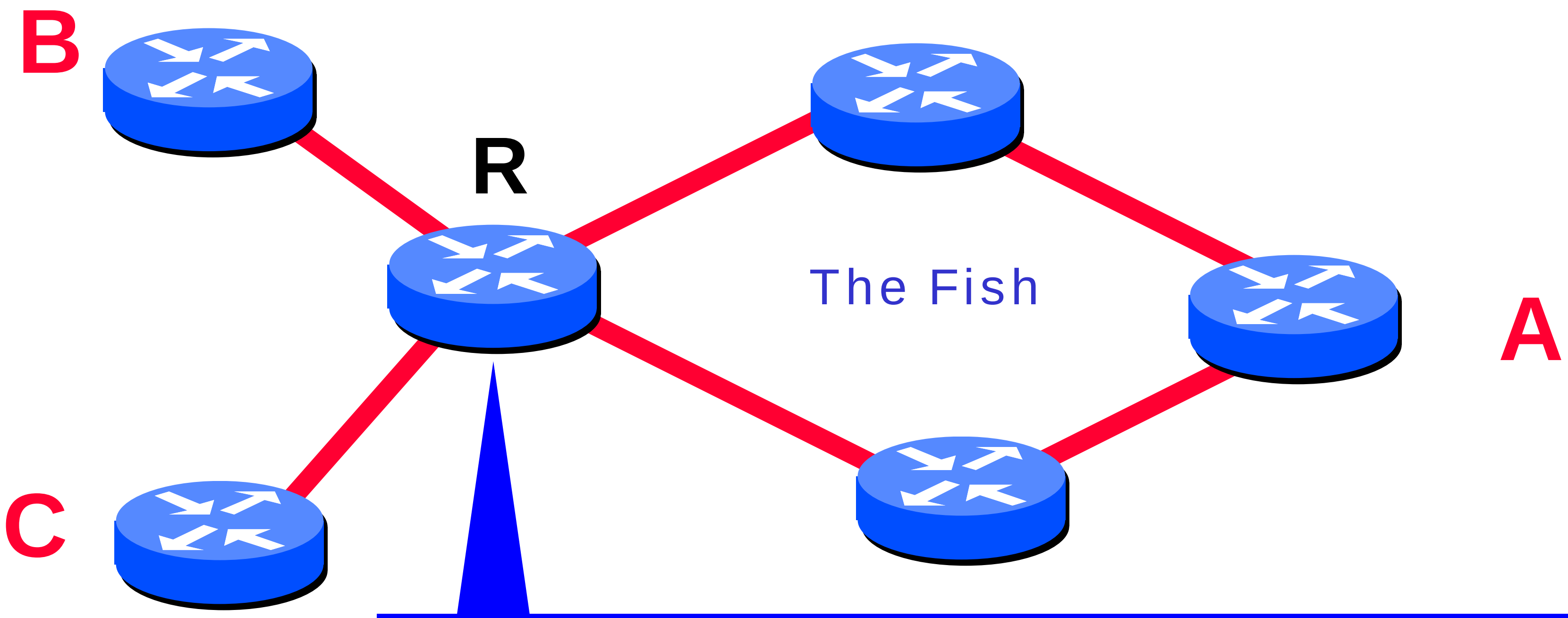
Dest.	Nxt Hop
A	R2
B	R2
C	Direct
D	R5
E	R5
default	R2

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

IP Forwarding Process

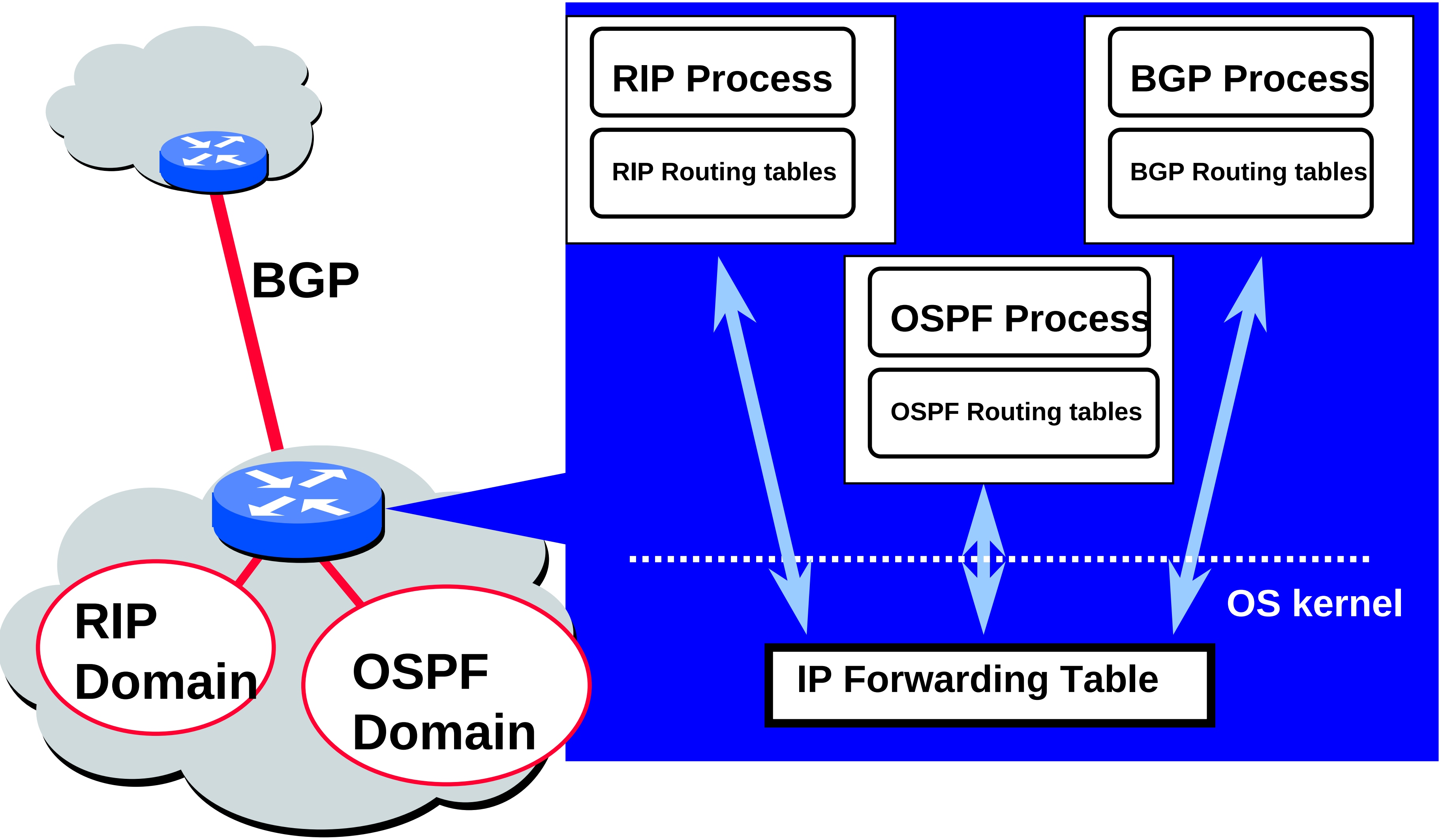


IP routing protocols assume all forwarding is destination-based

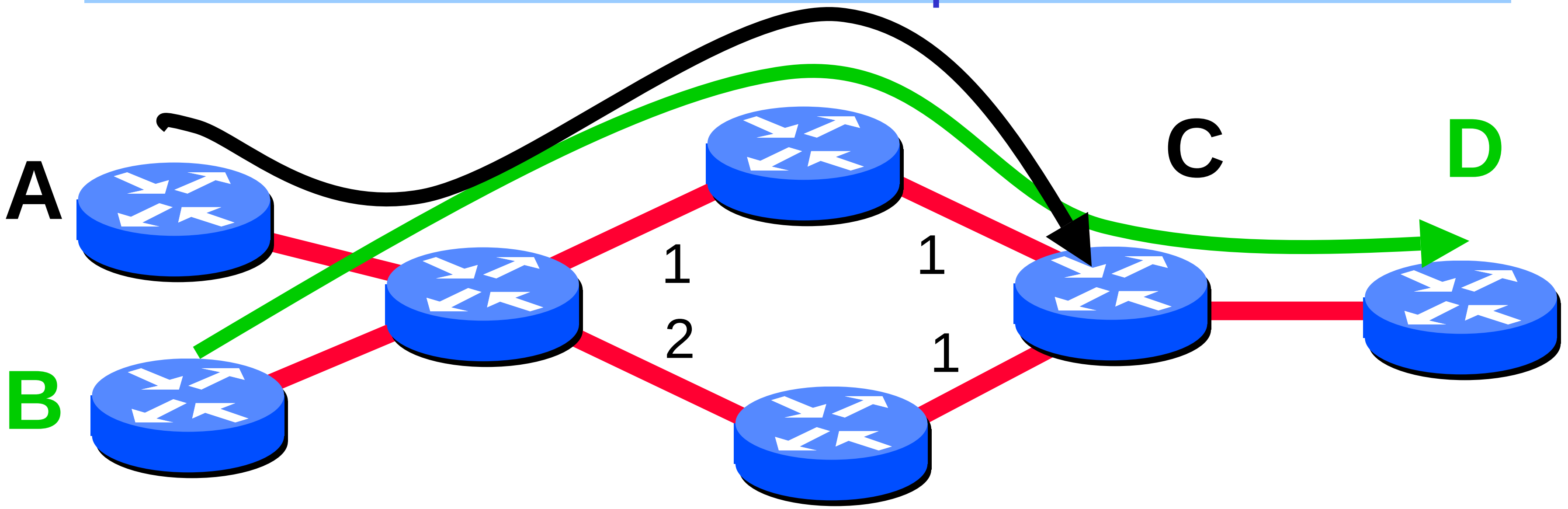


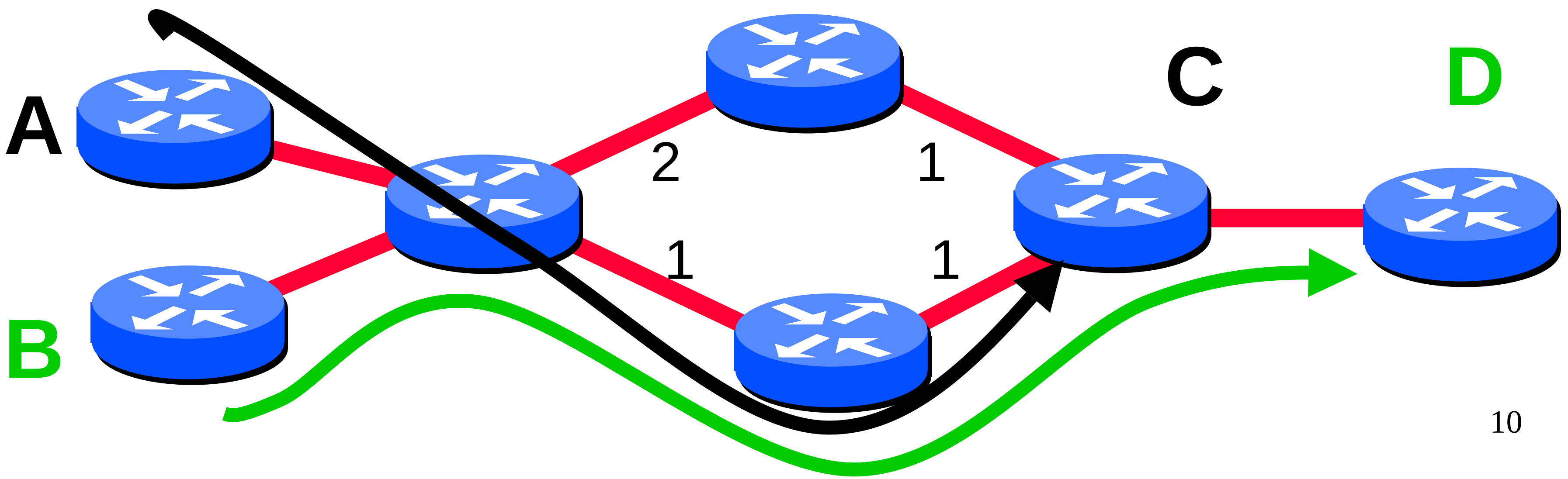
The next-hop forwarding paradigm does not allow router R to choose a route to A based on who originated the traffic, B or C.

IP forwarding tables are maintained by dynamic routing protocols

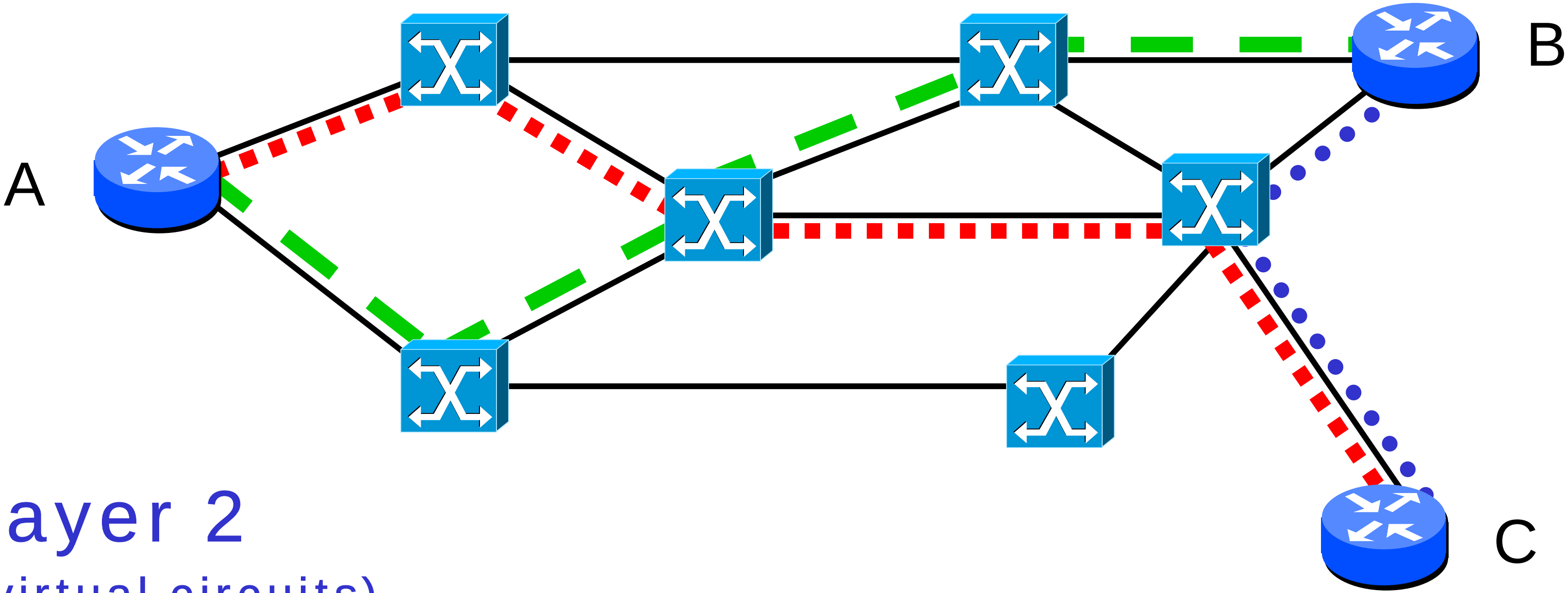


Shortest Path Routing: Link weights tend to attract or repel all traffic

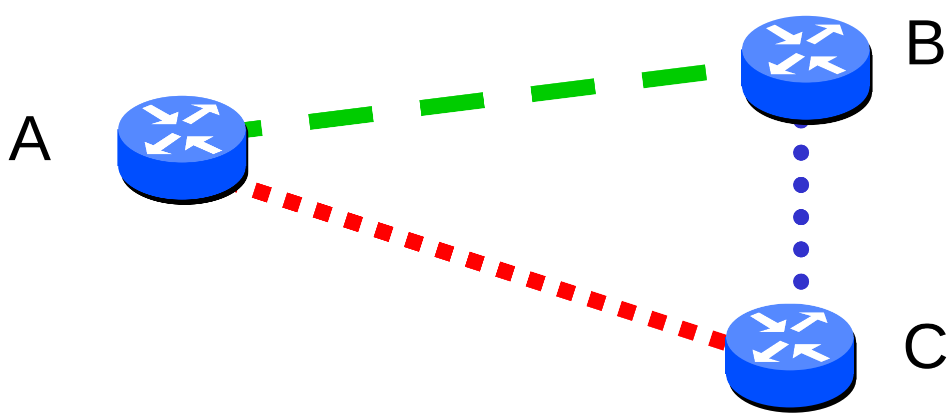




Overlay Networks



Layer 2
(virtual circuits)



Layer 3

Advantages of Overlay Networks

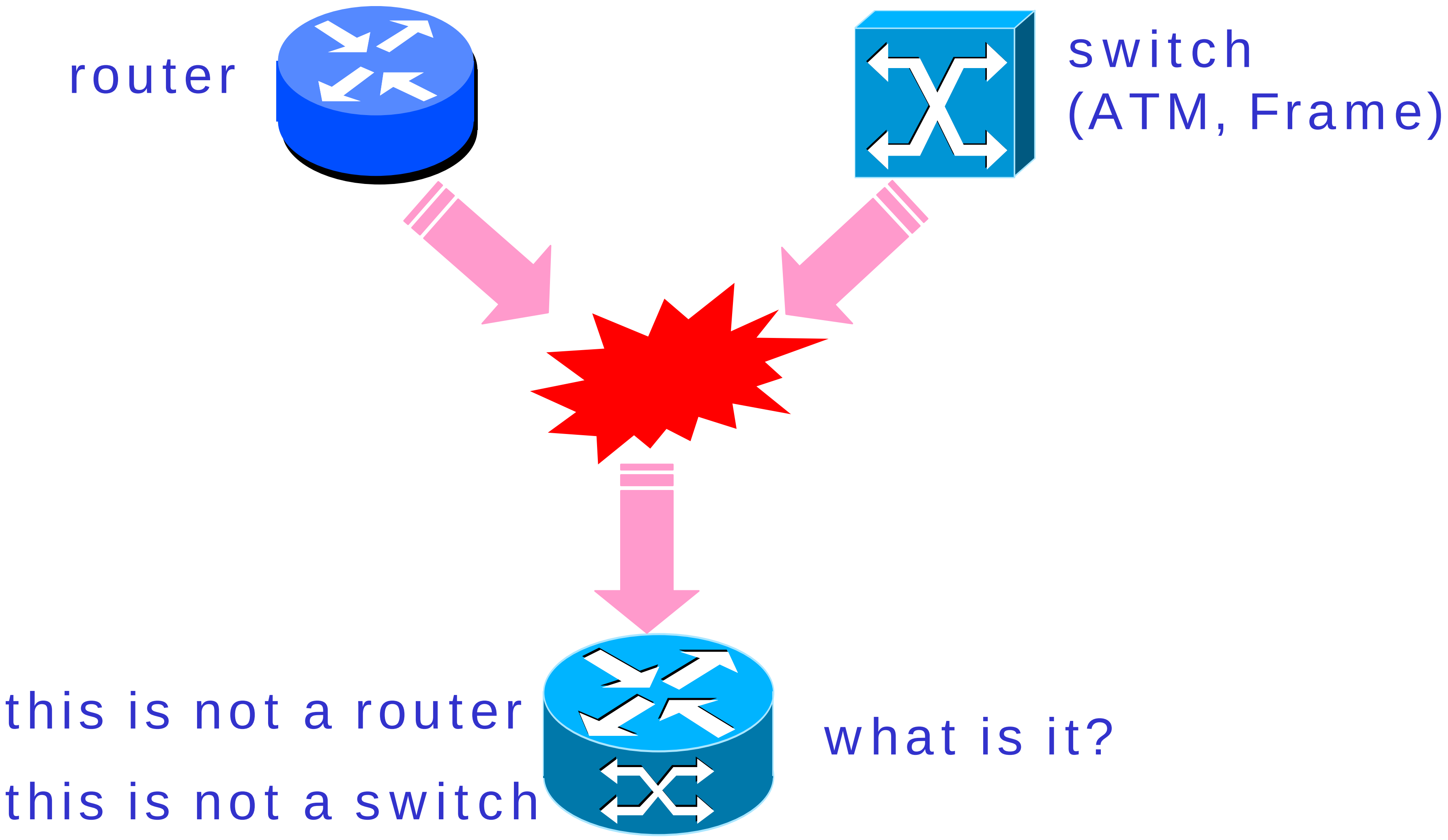
- ATM and Frame Relay switches offer high reliability and low cost
- Virtual circuits can be reengineered without changing the layer 3 network
- Large degree of control over traffic
- Detailed per-circuit statistics
- Isolates layer 2 network management from the details of higher layer services

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.

Blur Layer 2 and 3?



Sanity Check?

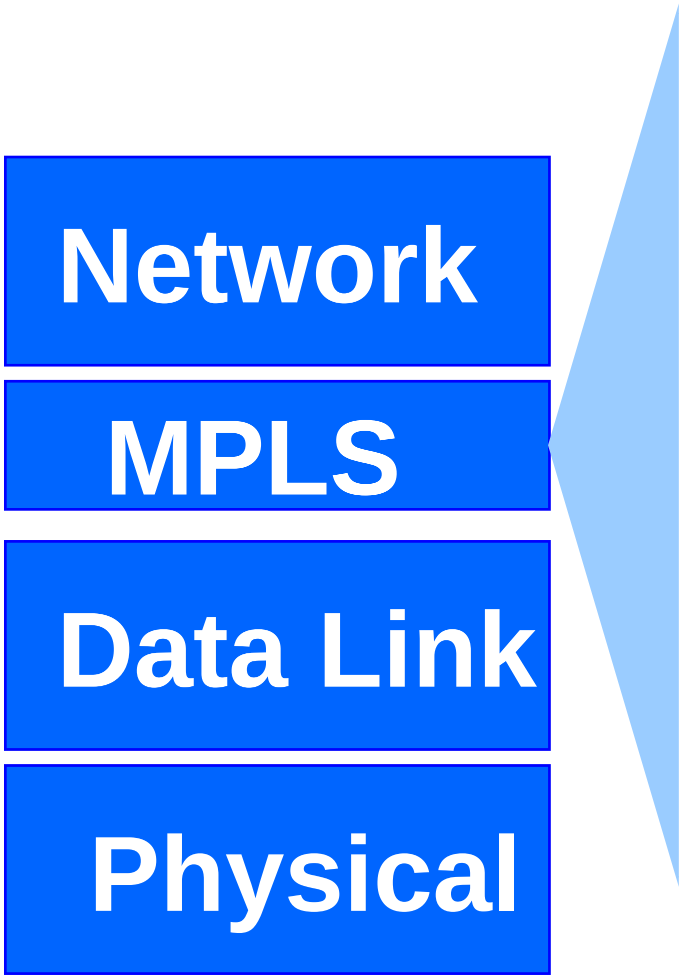
- The problems with IP forwarding and routing do not ***require*** technologies like MPLS
  - Many can be addressed with simple solutions. Like the design of simple networks!
  - The problems are not “show stoppers”
  - The MPLS cure **will** have side effects
  - For many applications, TCP/IP handles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

Sign of the Times: New Sub-IP area in the IETF

- Multiprotocol Label Switching (mpls)
- Common Control and Management Protocols (ccamp)
- Internet Traffic Engineering (tewg)
- Provider Provisioned Virtual Private Networks (ppvpn)
- IP over Optics (ipo)
- IP over Resilient Packet Rings (iporpr)
- General Switch Management Protocol (gsmp)

See http://www.ietf.org/html.charters/foo-charter.html, where foo is the working group acronym.

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

Generic MPLS Encapsulation

Layer 2 Header	MPLS Label 1	MPLS Label 2	...	MPLS Label n	Layer 3 Packet

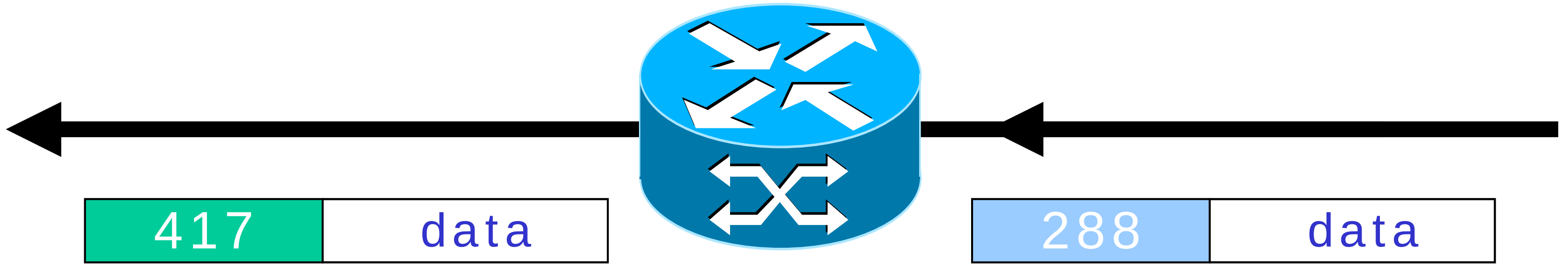
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Label                  | Exp |S|       TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+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

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

RFC 3032. MPLS
Label Stack Encoding

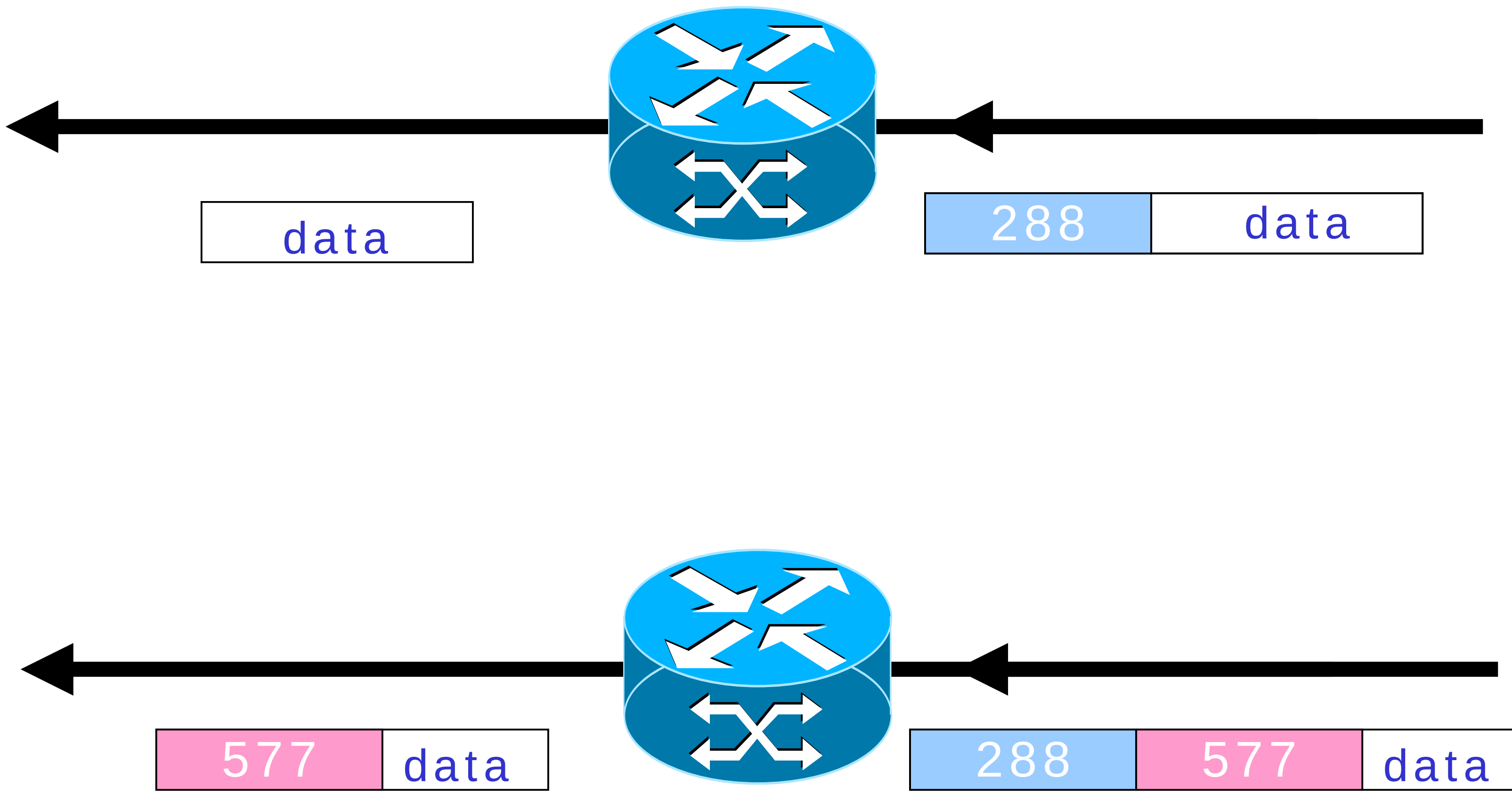
- Label: Label Value, 20 bits
- Exp: Experimental, 3 bits
- S: Bottom of Stack, 1 bit
- TTL: Time to Live, 8 bits

Forwarding via Label Swapping

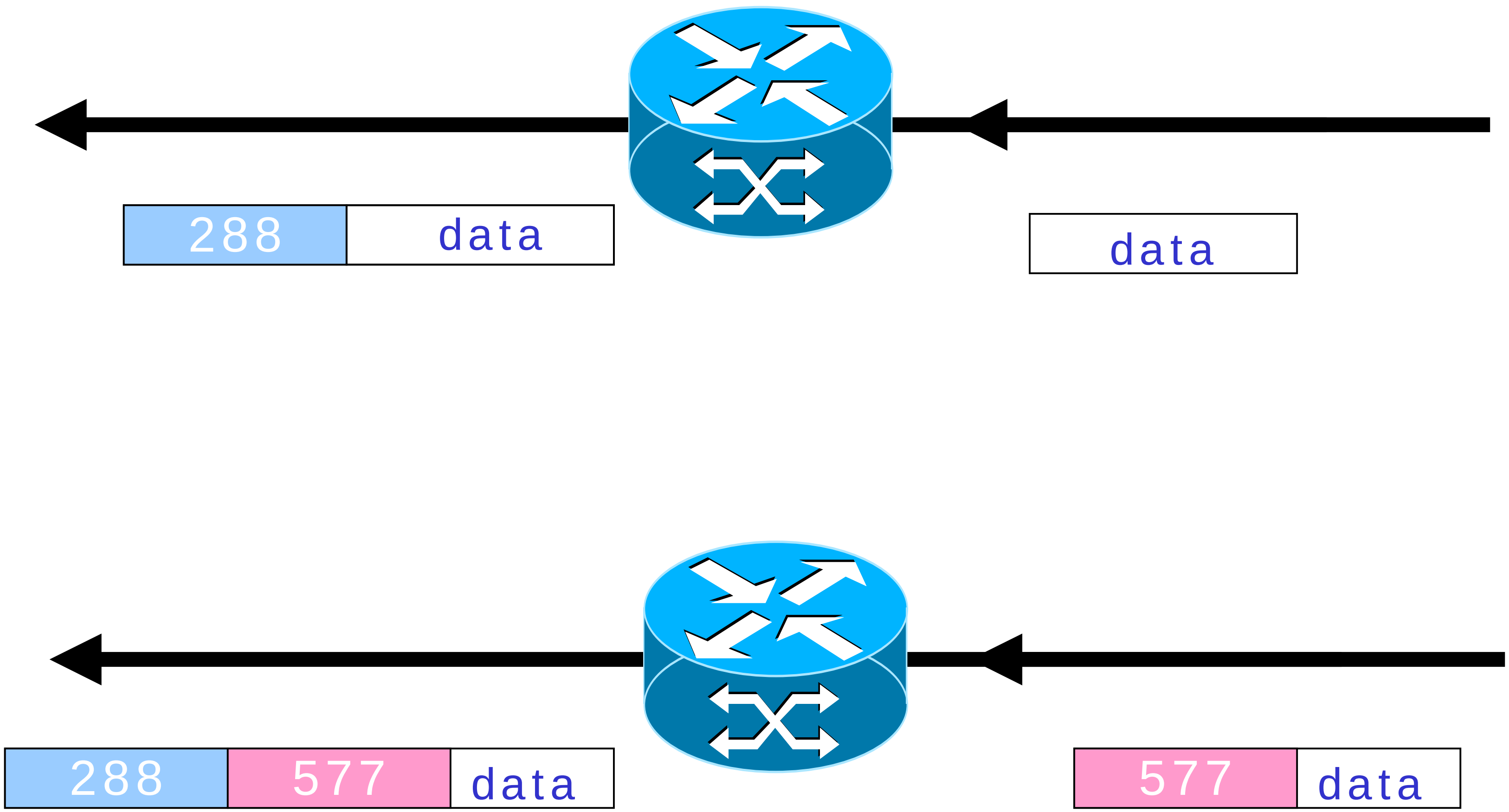


Labels are short, fixed-length values.

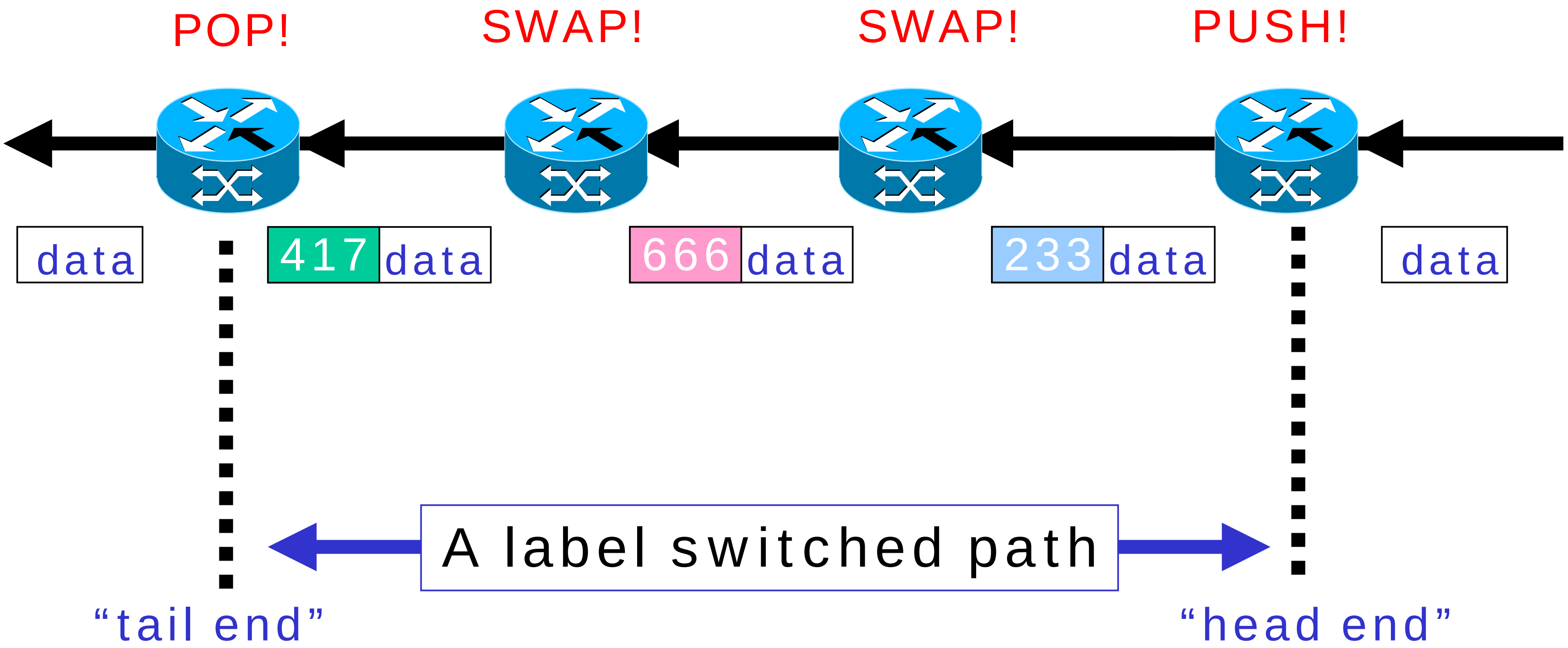
Popping Labels



Pushing Labels

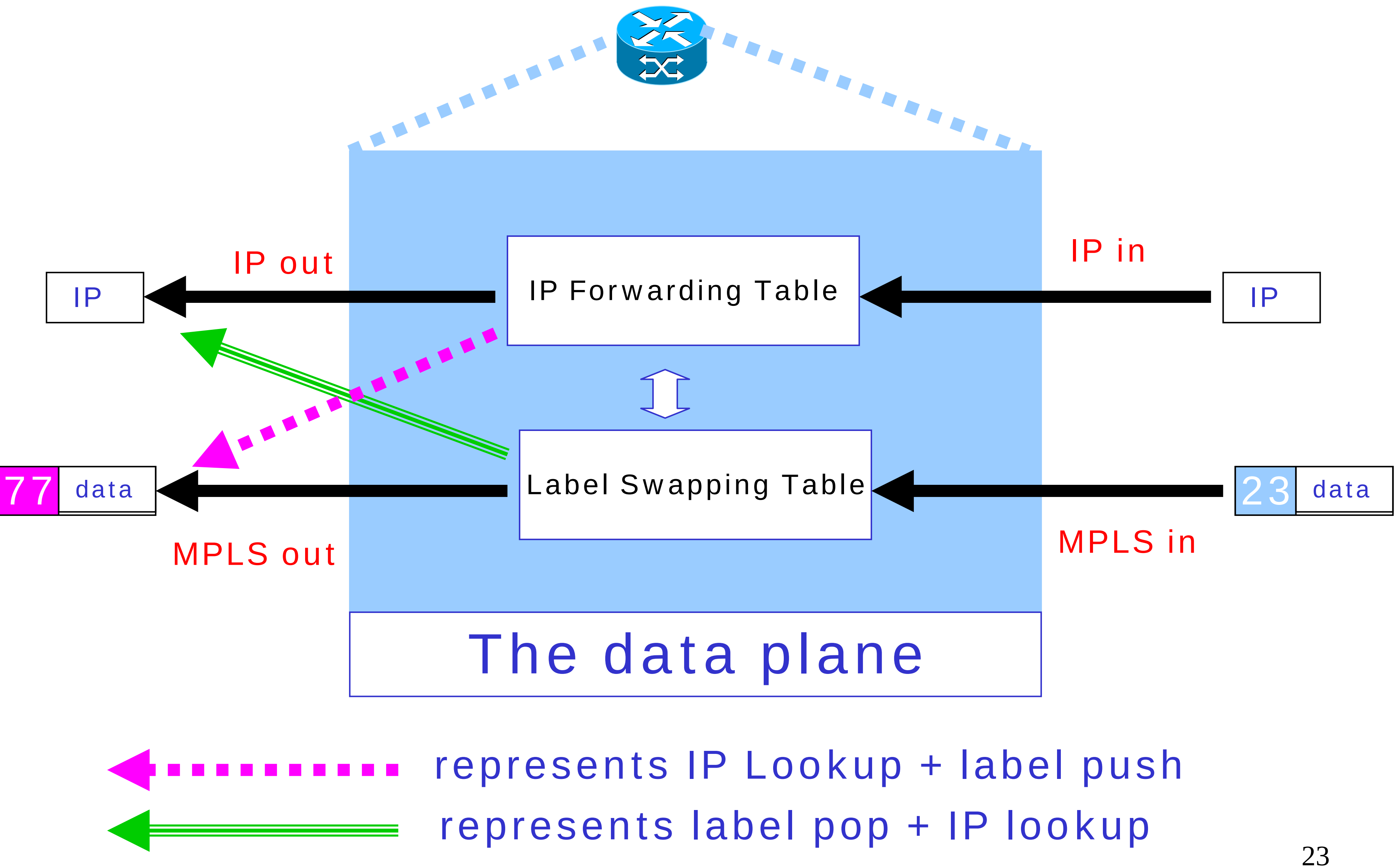


A Label Switched Path (LSP)

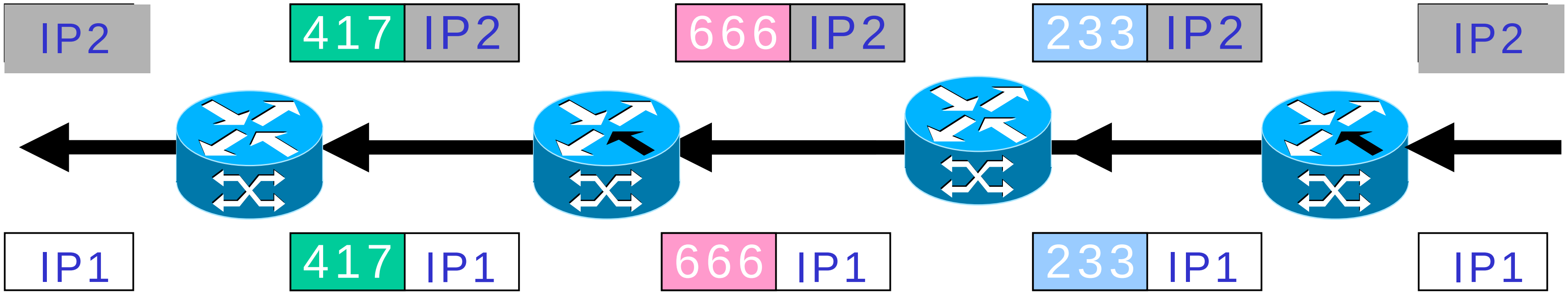


Often called an MPLS tunnel: payload headers are not Inspected inside of an LSP. Payload could be MPLS ...

Label Switched Routers



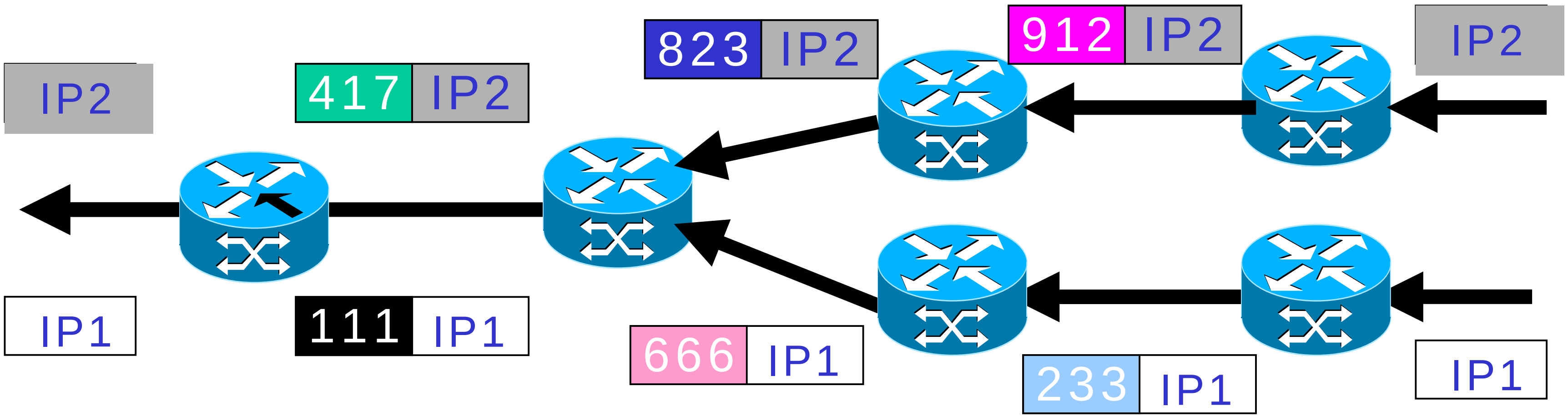
Forwarding Equivalence Class (FEC)



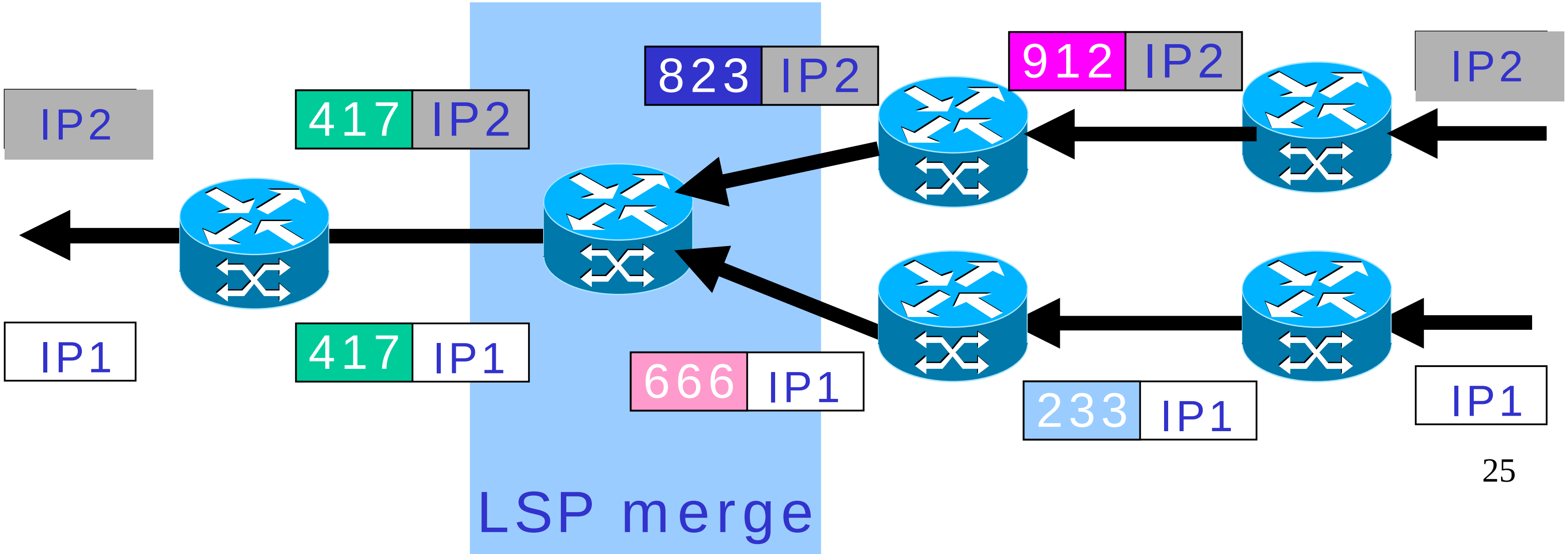
Packets IP1 and IP2 are forwarded in the same way --- they are in the same FEC.

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.

LSP Merge

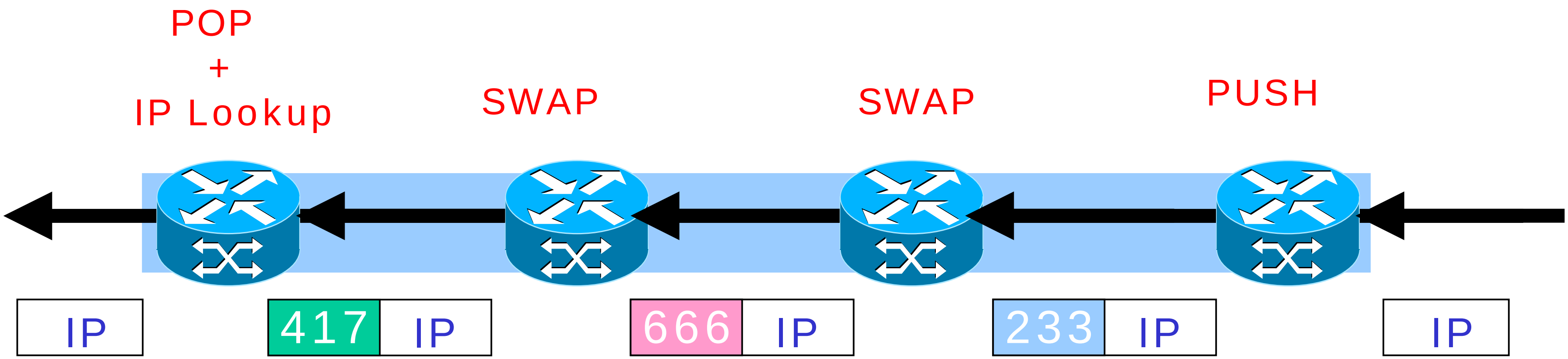


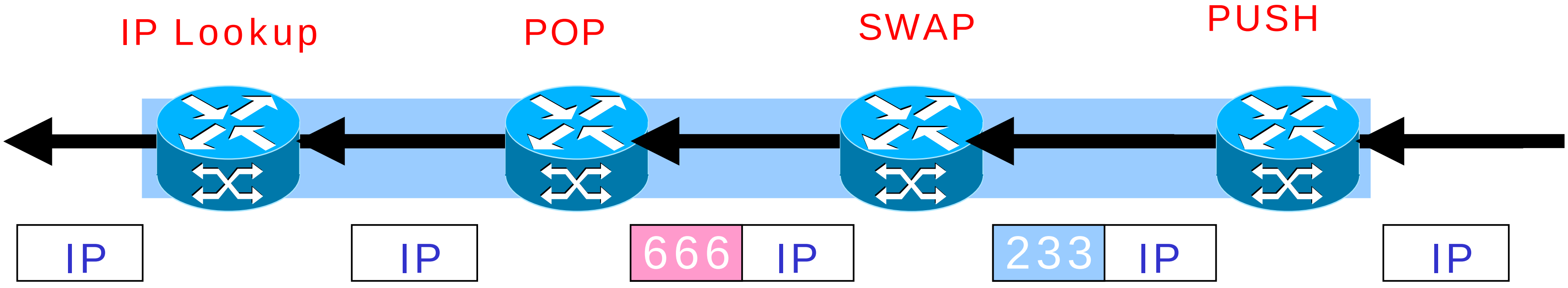
IP2
417 IP2
823 IP2
912 IP2
IP2
IP1
111 IP1
666 IP1
233 IP1
IP1



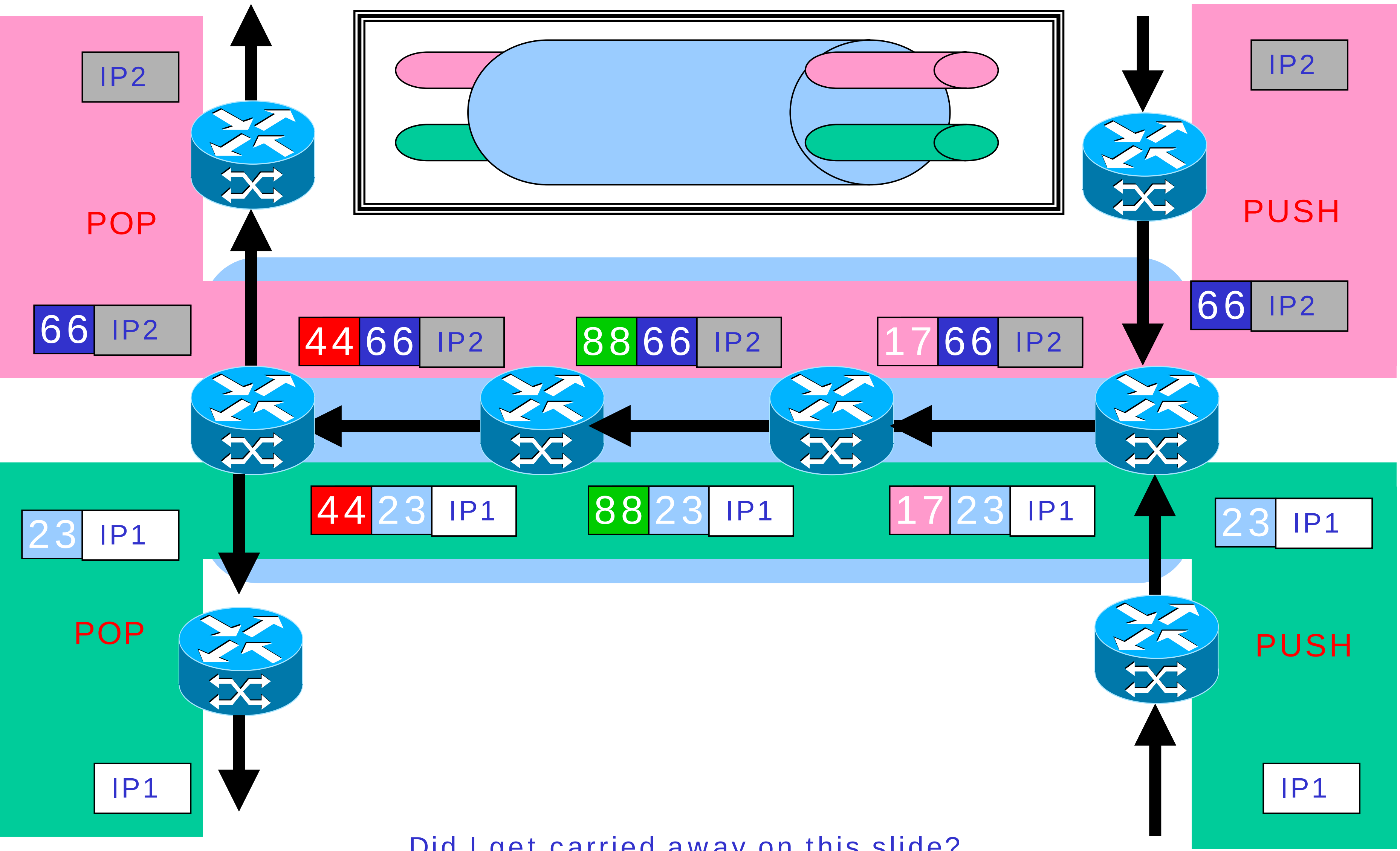
IP2
417 IP2
823 IP2
912 IP2
IP2
IP1
417 IP1
666 IP1
233 IP1
IP1
LSP merge

Penultimate Hop Popping





LSP Hierarchy via Label Stacking



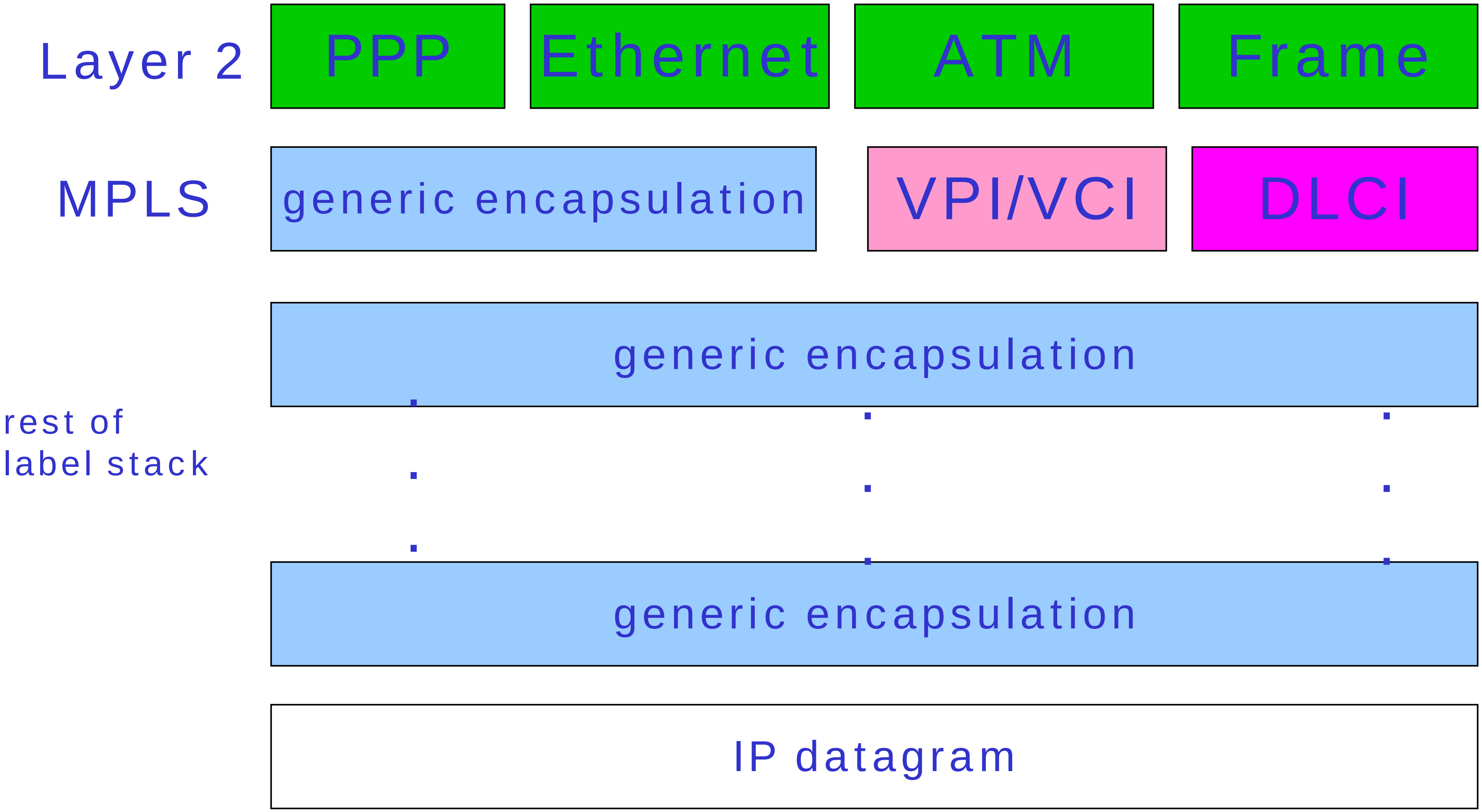
Did I get carried away on this slide?

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...

MPLS also supports “native encapsulation”



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.

Basic MPLS Control Plane

MPLS control plane
=
IP control plane
+
<u>label distribution</u>

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

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+

Label Distribution: Option II

Create new label distribution protocol(s)

Good Points

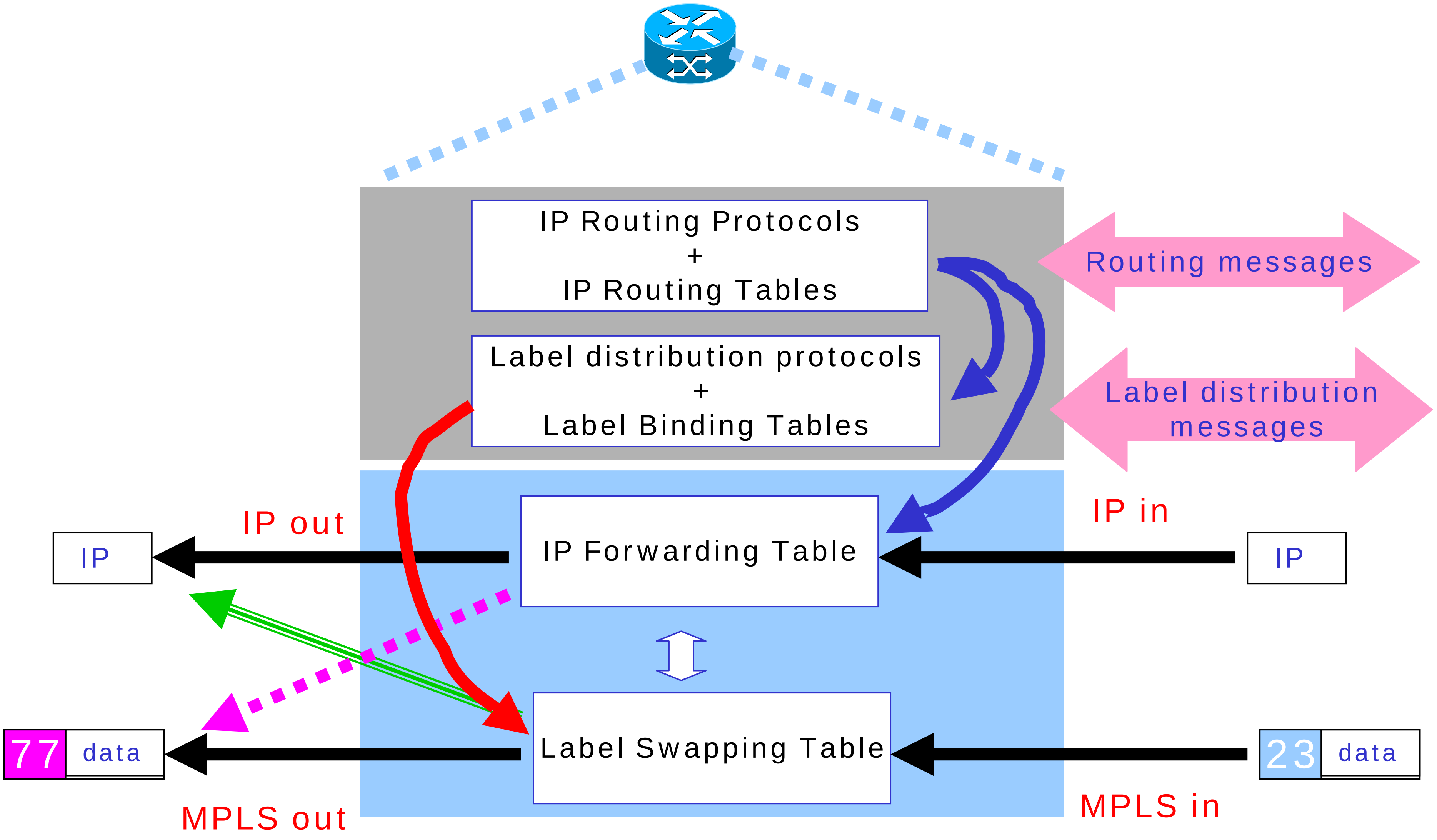
- Compatible with "Link State" routing protocols
- Not limited to destination-based, hop-by-hop forwarding paths

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)

The Control Plane



Label Distribution with BGP

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

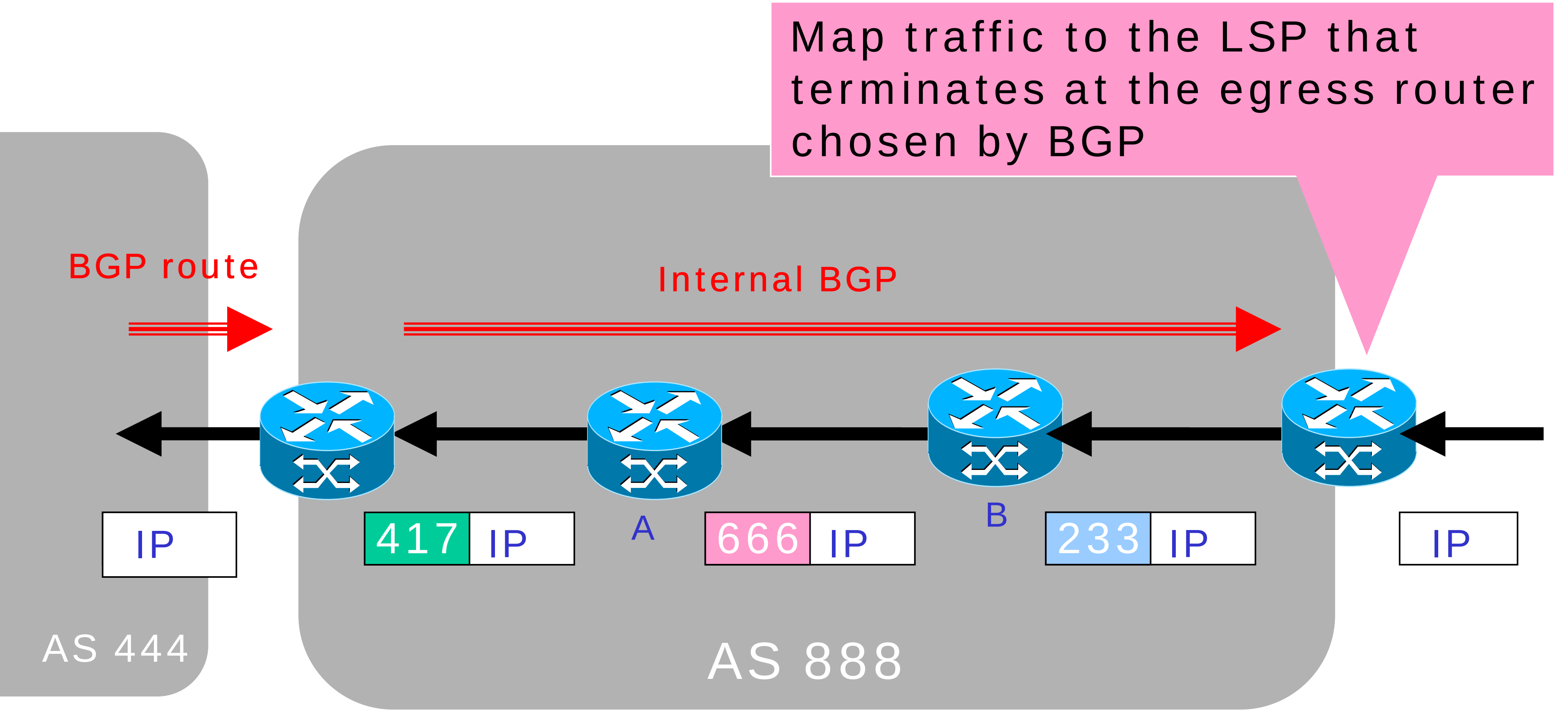
Associates a label (or label stack) with the BGP next hop.

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)

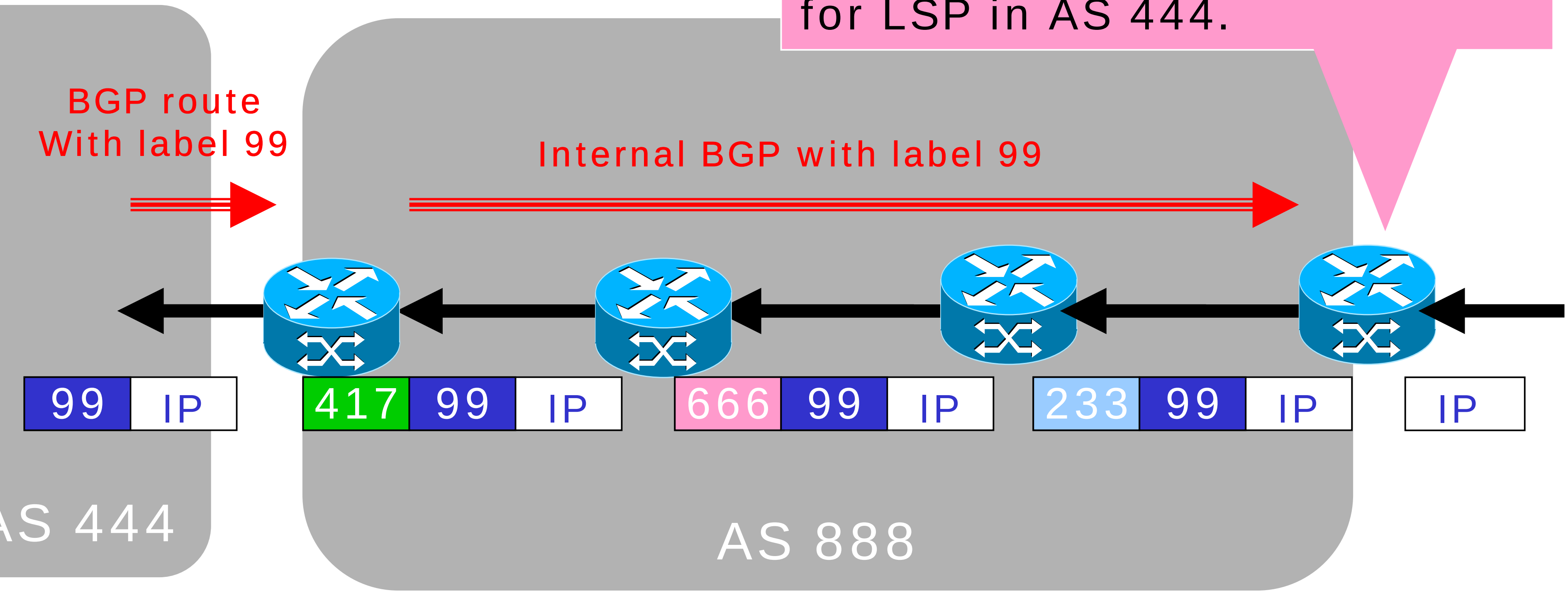
BGP piggyback not required for simple iBGP optimization



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

BGP piggyback allows Interdomain LSPs

Use top of stack to get to egress router, bottom of stack for LSP in AS 444.



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 ...)

Label Distribution Protocol (LDP)

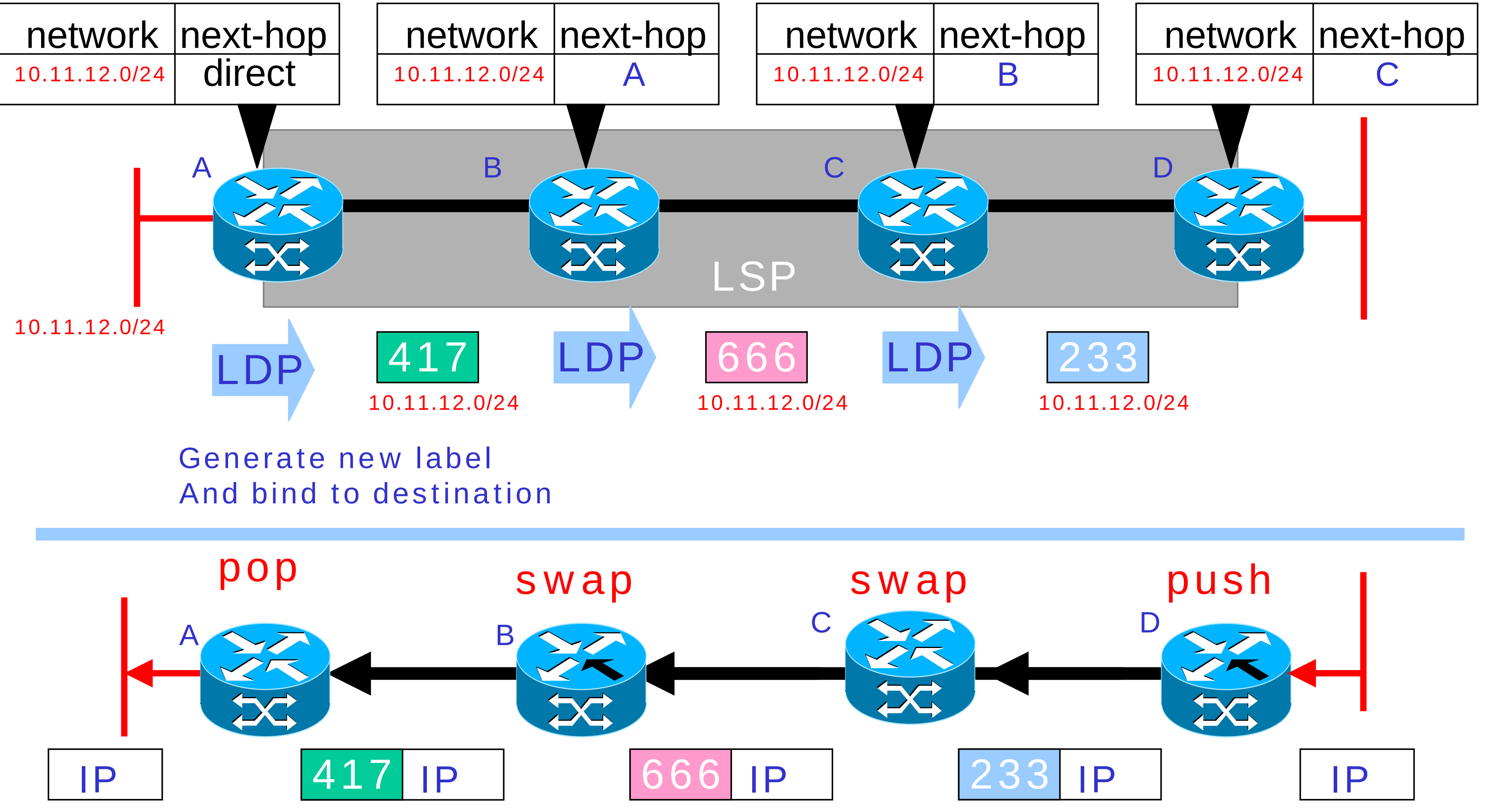
RFC 3036. LDP Specification. (1/2001)

- 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)

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.

LDP and Hop-by-Hop routing



MPLS Traffic Engineering

"The optimization goals of traffic engineering are
To enhance the performance of IP traffic while utilizing
Network resources economically and reliably."

Intra-Domain

A Framework for Internet Traffic Engineering
Draft-ietf-tewg-framework-02.txt

"A major goal of Internet Traffic
Engineering is to facilitate efficient and reliable
network operations while simultaneously optimizing
network resource utilization and performance."

Intra-Domain

RFC 2702
Requirements for Traffic Engineering over MPLS

TE May Require Going Beyond Hop-by-Hop Routing

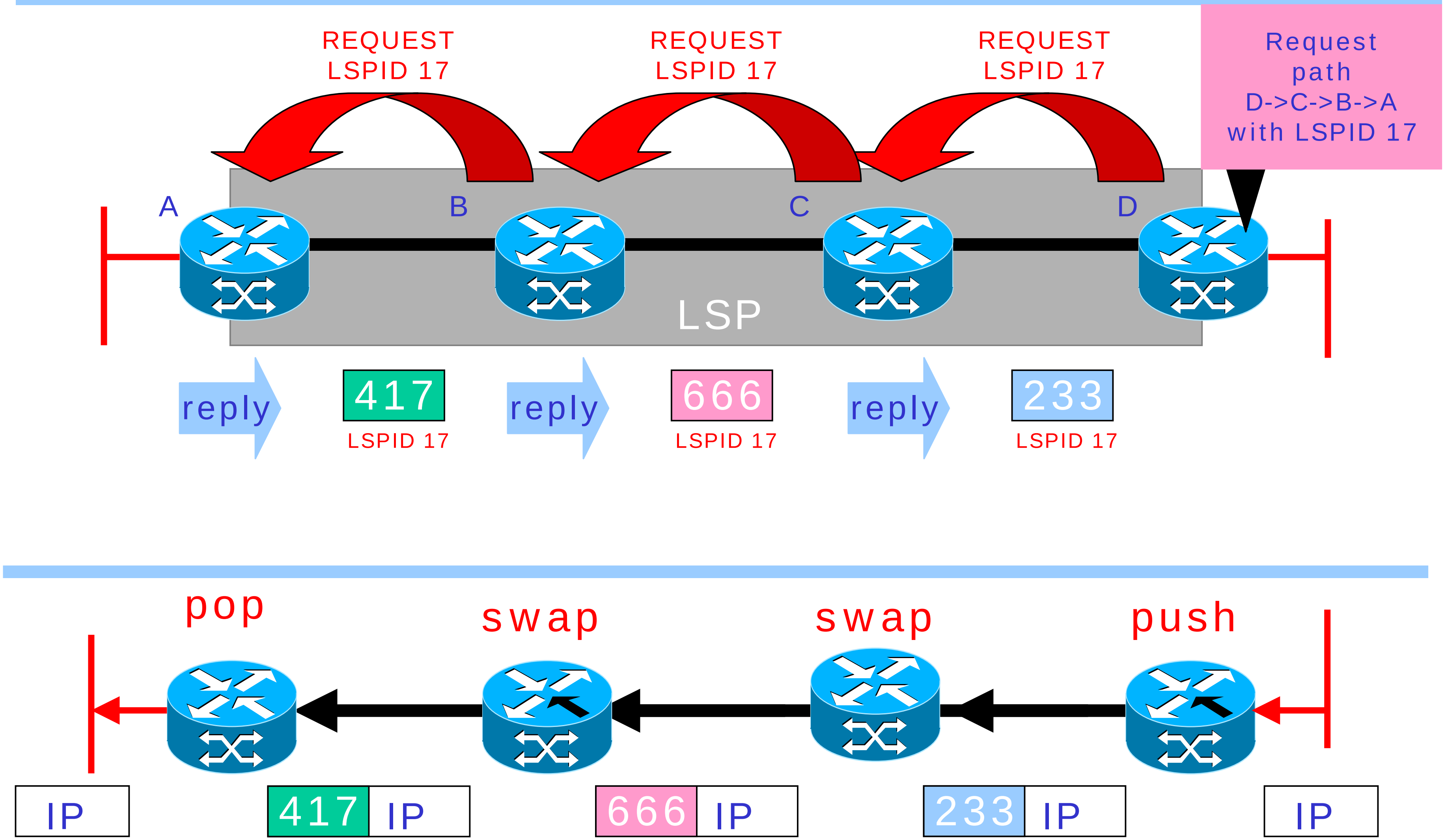
- Explicit routes
  - Allow traffic sources to set up paths
- Constraint based routing
  - Chose only best paths that do no violate some constraints
  - Needs explicit routing
  - May need resource reservation
- Traffic classification
  - Map traffic to appropriate LSPs

Hop-by-Hop vs. Explicit Routes

- Distributed control
- Trees rooted at destination
- Destination based forwarding

- Originates at source
- Paths from sources to destinations
- Traffic to path mapping based on what configuration commands your vendor(s) provide

Explicit path Setup



Constraint Based Routing

Problem here: OSPF areas hide information for scalability. So these extensions work best only within an area...

Basic components

1. Specify path constraints
2. Extend topology database to include resource and constraint information
3. Find paths that do not violate constraints and optimize some metric
4. Signal to reserve resources along path
5. Set up LSP along path (with explicit route)
6. Map ingress traffic to the appropriate LSPs

Extend Link State Protocols (IS-IS, OSPF)

Extend RSVP or LDP or both!

Problem here: what is the "correct" resource model for IP services?

Note: (3) could be offline,
or online (perhaps an extension to OSPF)

Resource Reservation + Label Distribution

Two emerging/competing/dueling approaches:

Add label distribution and explicit routes to a resource reservation protocol

RSVP-TE

+

RSVP

CR-LDP

+

LDP

Add explicit routes and resource reservation to a label distribution protocol

RSVP-TE: Extensions to RSVP for LSP Tunnels
draft-ietf-mpls-rsvp-lsp-tunnel-08.txt

Constraint-Based LSP Setup using LDP
draft-ietf-mpls-cr-lpd-05.txt

RSVP-TE vs. CR-LPD

RSVP-TE	CR-LDP
• Soft state periodically refreshed • IntServe QoS model	• State maintained incrementally • New QoS model derived from ATM and Frame Relay

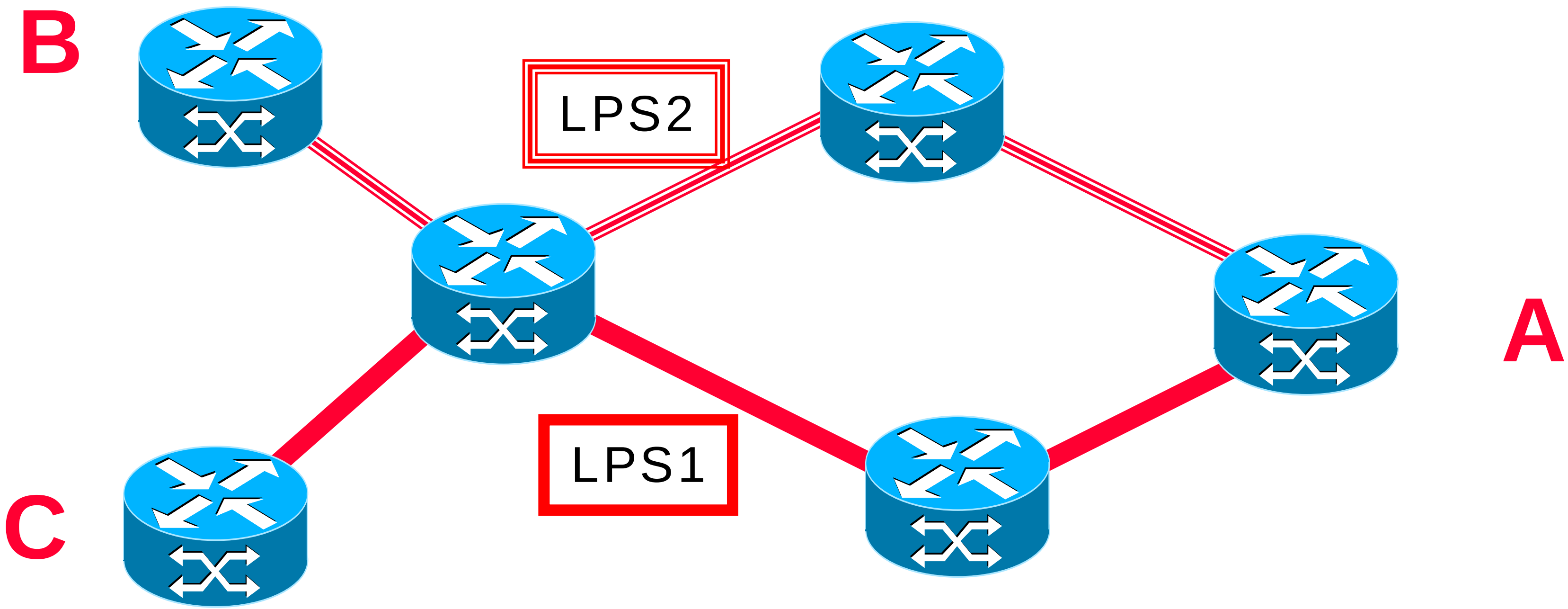
And the QoS model determines the additional information attached to links and nodes and distributed with extended link state protocols...

And what about that other Internet QoS model, diffserve?

A closer look at CR-LDP

- Defines new TLV encodings and procedures for
  - Explicit routing (strict and loose)
  - Route pinning (nail down some segments of a loosely routed path)
  - Traffic parameter specification
    - Peak rate
    - Committed rate
    - Weight
    - Resource class or color
  - LSP preemption (reroute existing paths to accommodate a new path)
  - LSP Identifiers (LSPIDs)

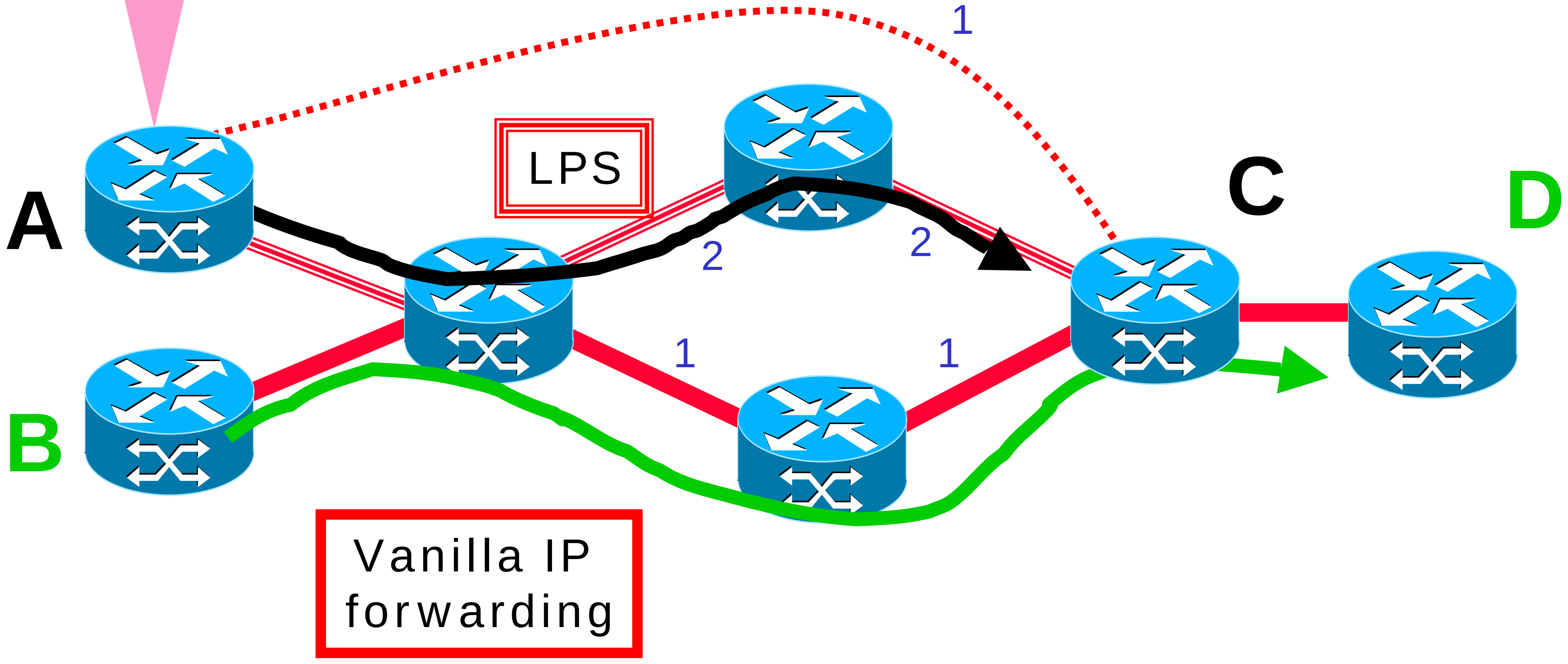
The Fish Revisited



Need at least one explicit route to A

Use Shortest Paths to get beyond Shortest Paths!

The IP routing protocol at LSR
A is configured to (privately) see A -> C LSP
as one link with weight 1.



MPLS TE: Is it worth the cost?

- Much of the traffic across a (transit) ISPs network is interdomain traffic
  - Congestion is most common on peering links
  - The current work on MPLS TE does not apply to interdomain links! (Actually, it does not even work well across OSPF areas...)
- MPLS TE is probably most valuable when IP services require more than best effort
  - VPNs with SLAs?
  - Supporting differentiated services?

VPNs with MPLS

"Traditional" VPN overlay model:

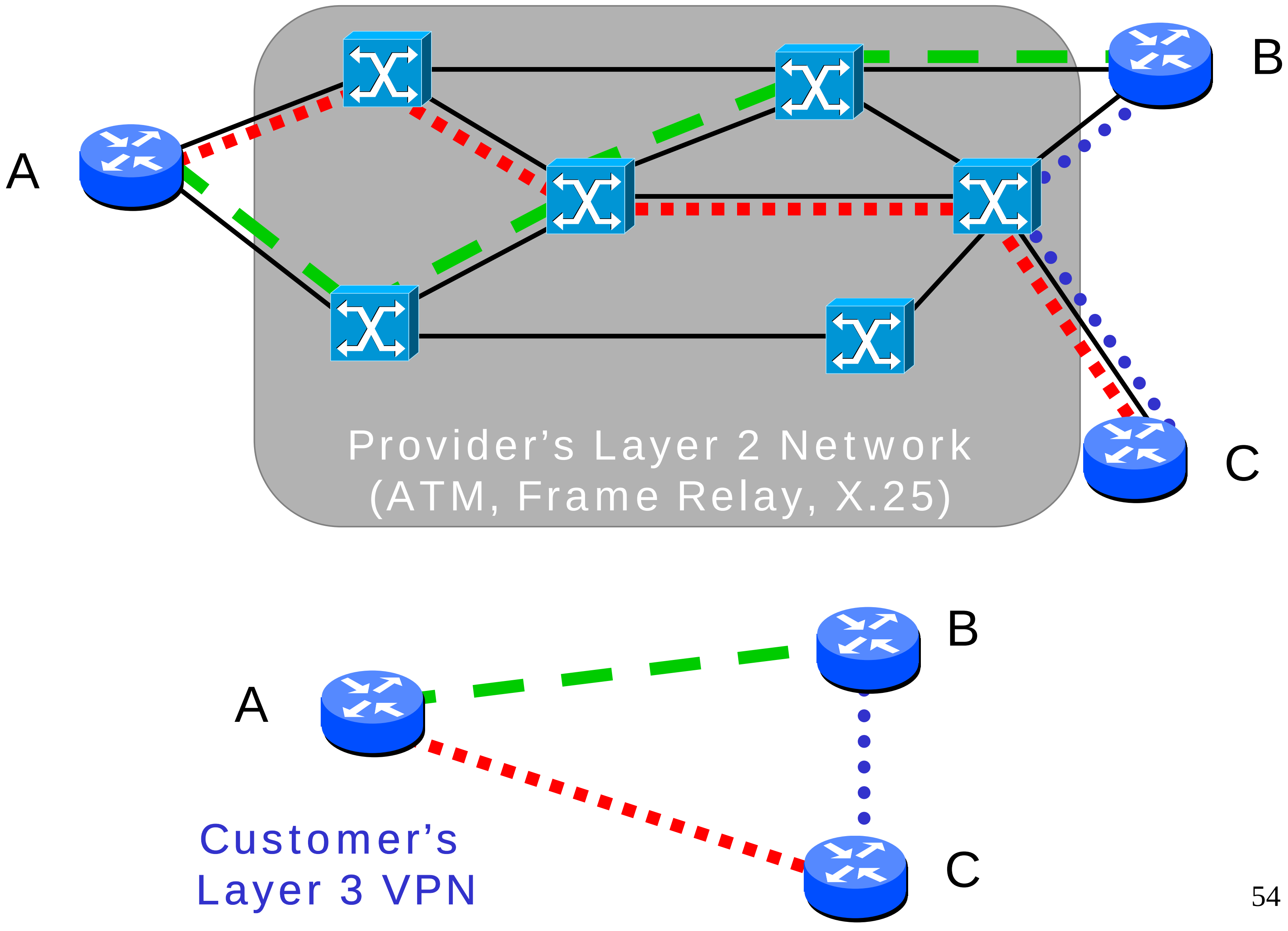
MPLS-based Layer 2 VPNs
draft-kompella-mpls-l2vpn-02.txt

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

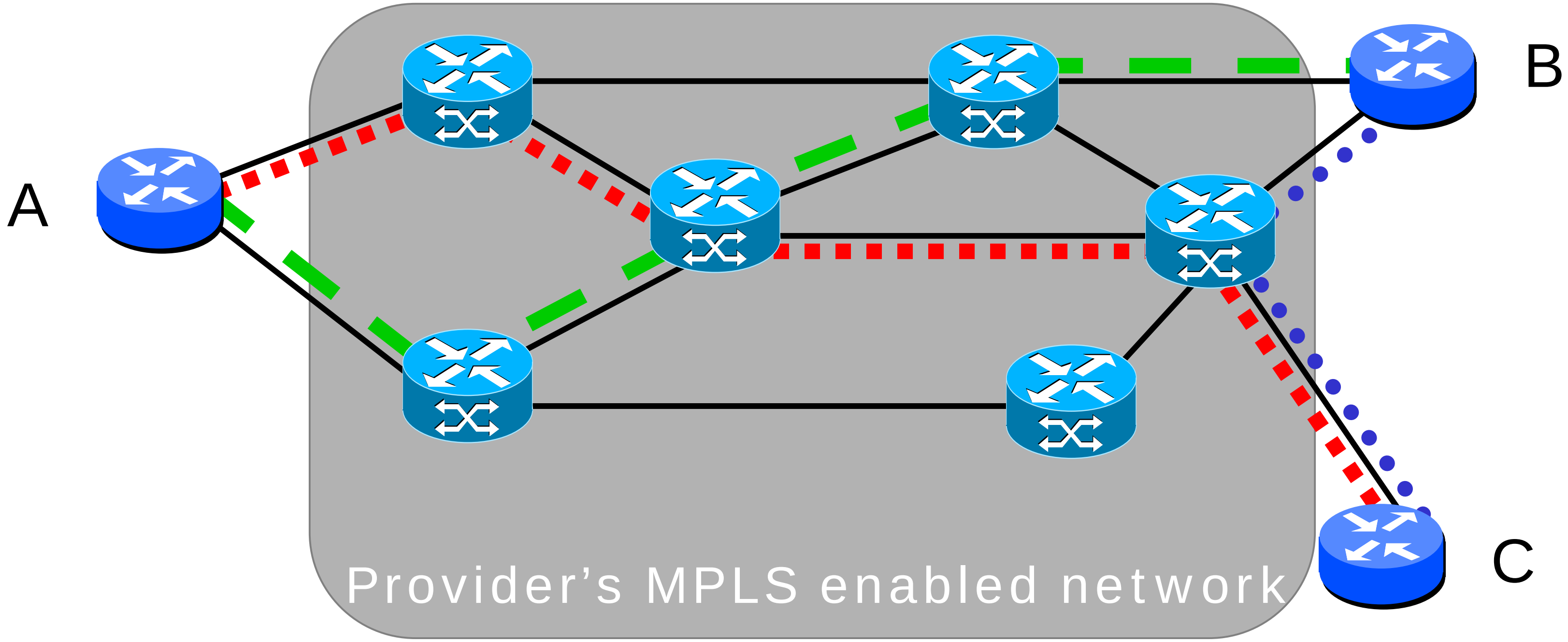
New VPN peering model:

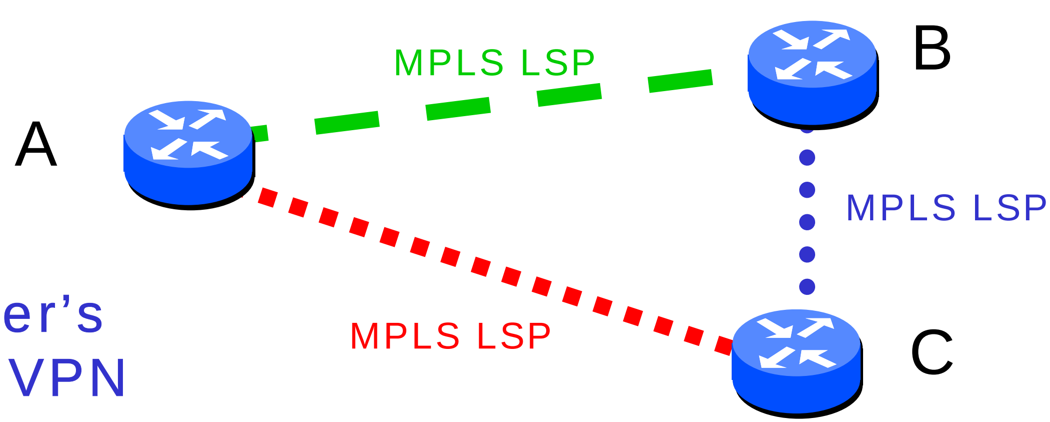
RFC 2547. BGP/MPLS VPNs

Traditional Overlay VPNs



Why Not Use MPLS Tunnels?





Customer's
Layer 3 VPN

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

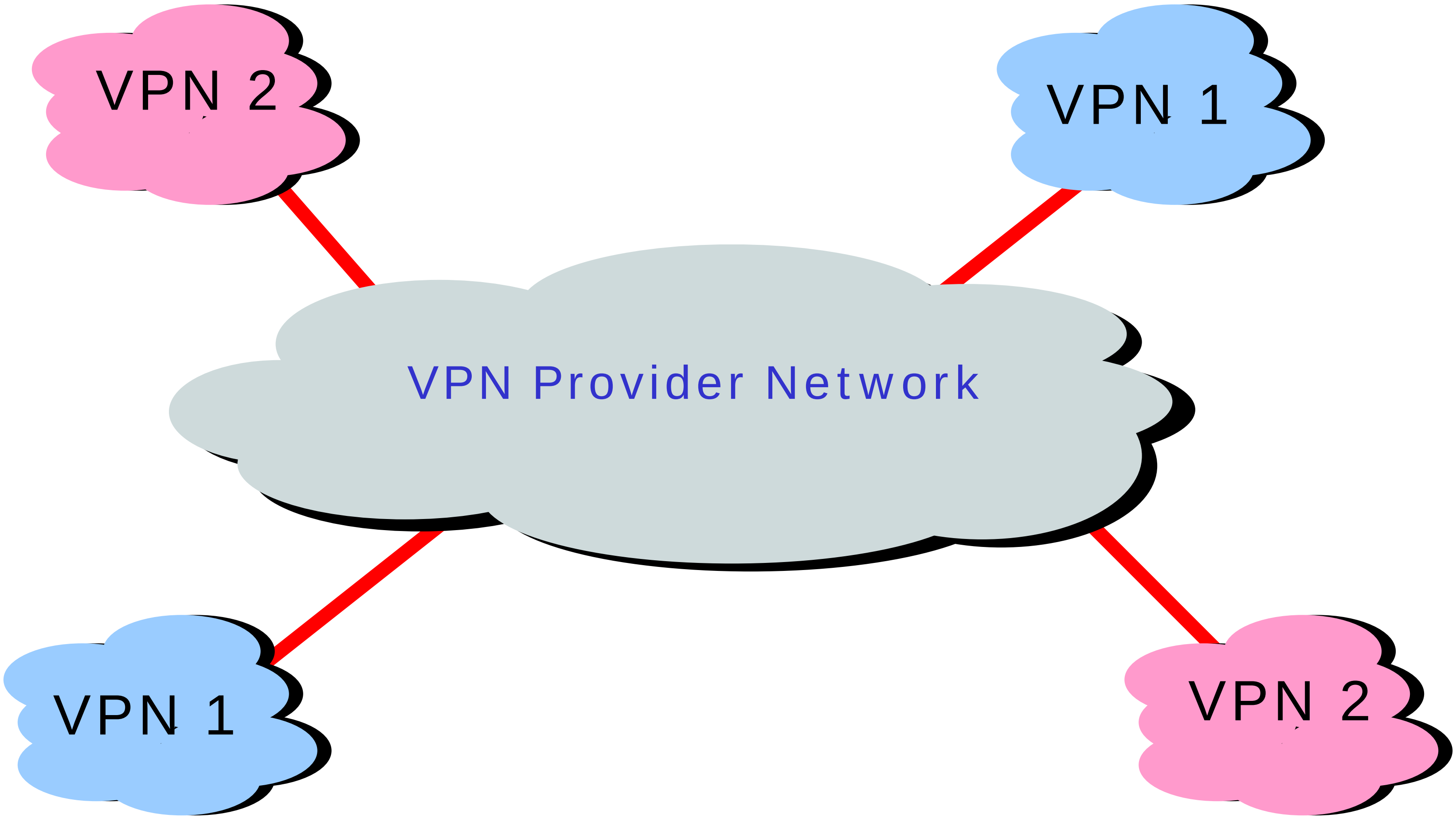
BGP/MPLS VPNs

- 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

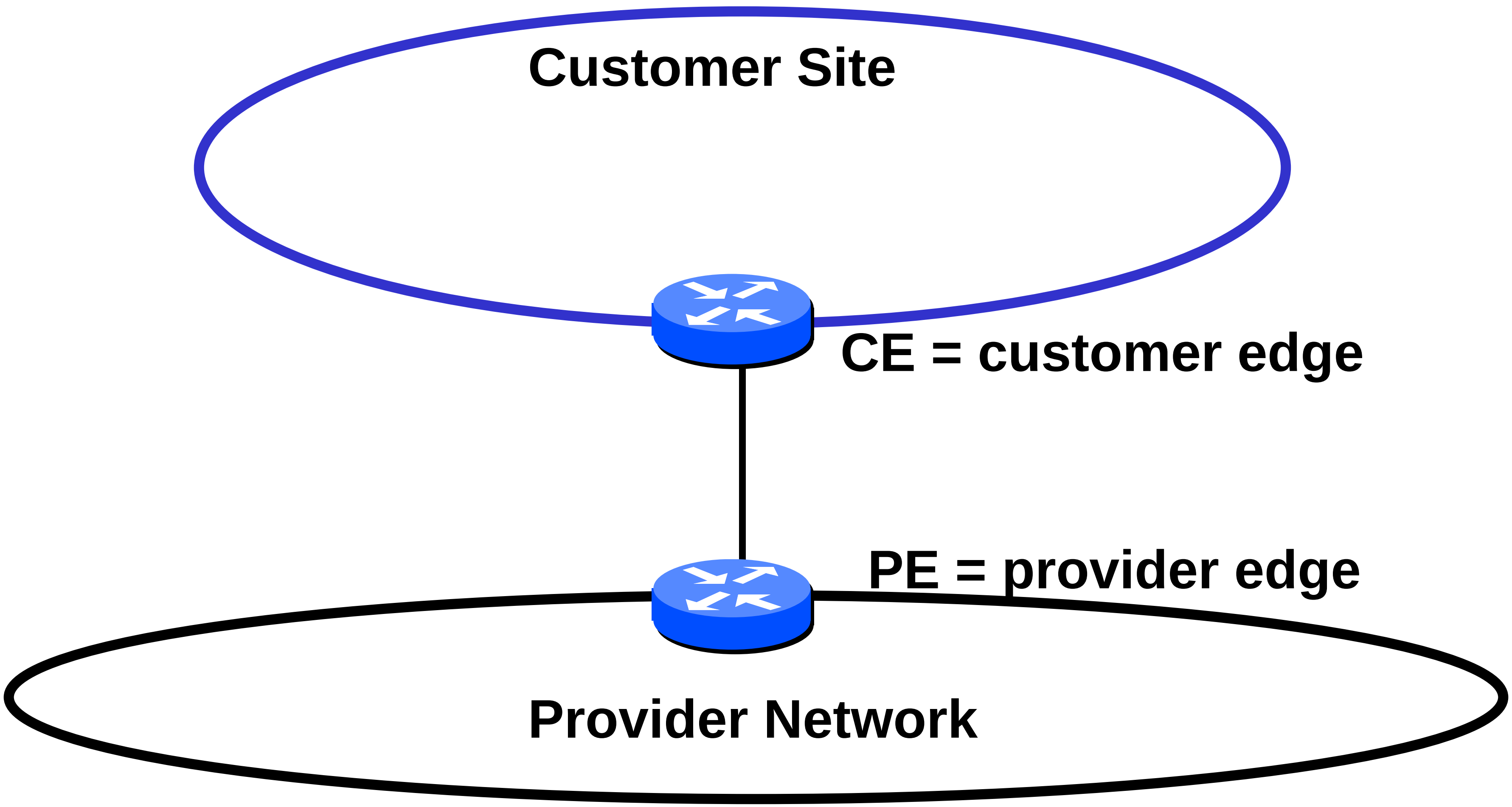
AT&T's IPFR service is based on this RFC.

Note that the Model of this RFC could be implemented without MPLS. See
BGP/IPsec VPN
<draft-declercq-bgp-ipsec-vpn-00.txt>

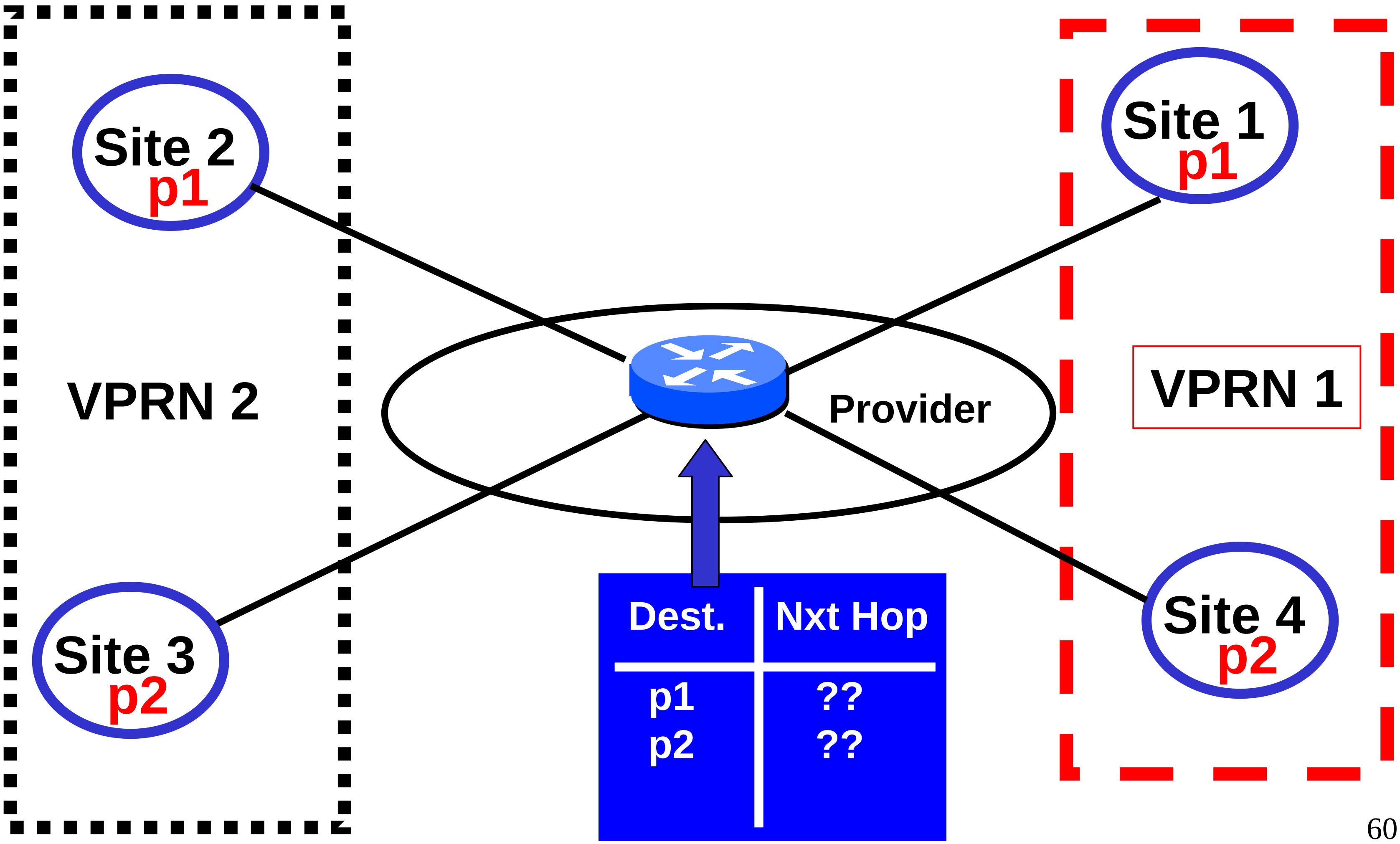
RFC 2547 Model



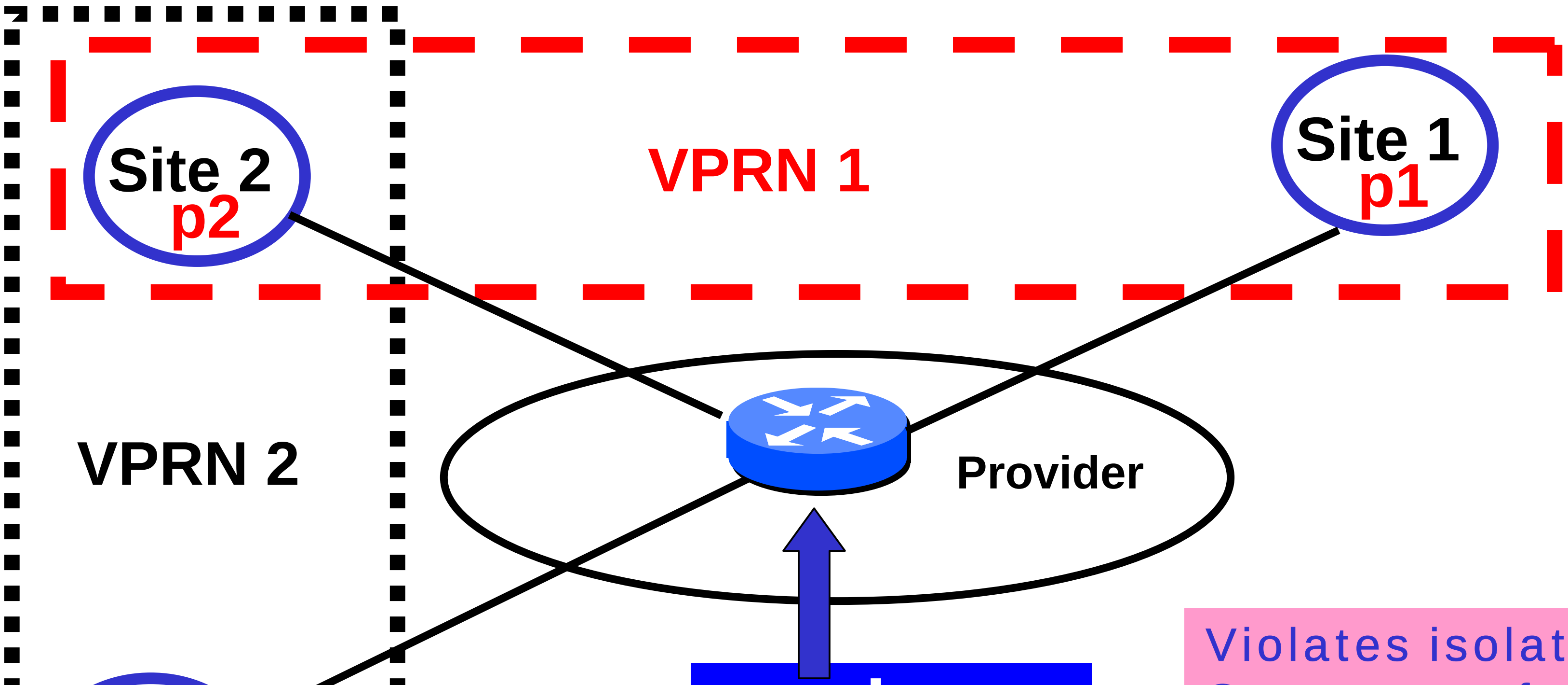
CEs and PEs


Customer Site
CE = customer edge
PE = provider edge
Provider Network

VPN Address Overlap Means Vanilla Forwarding Tables Can't Work



VPN Overlap Means Vanilla Forwarding Tables Can’t Work

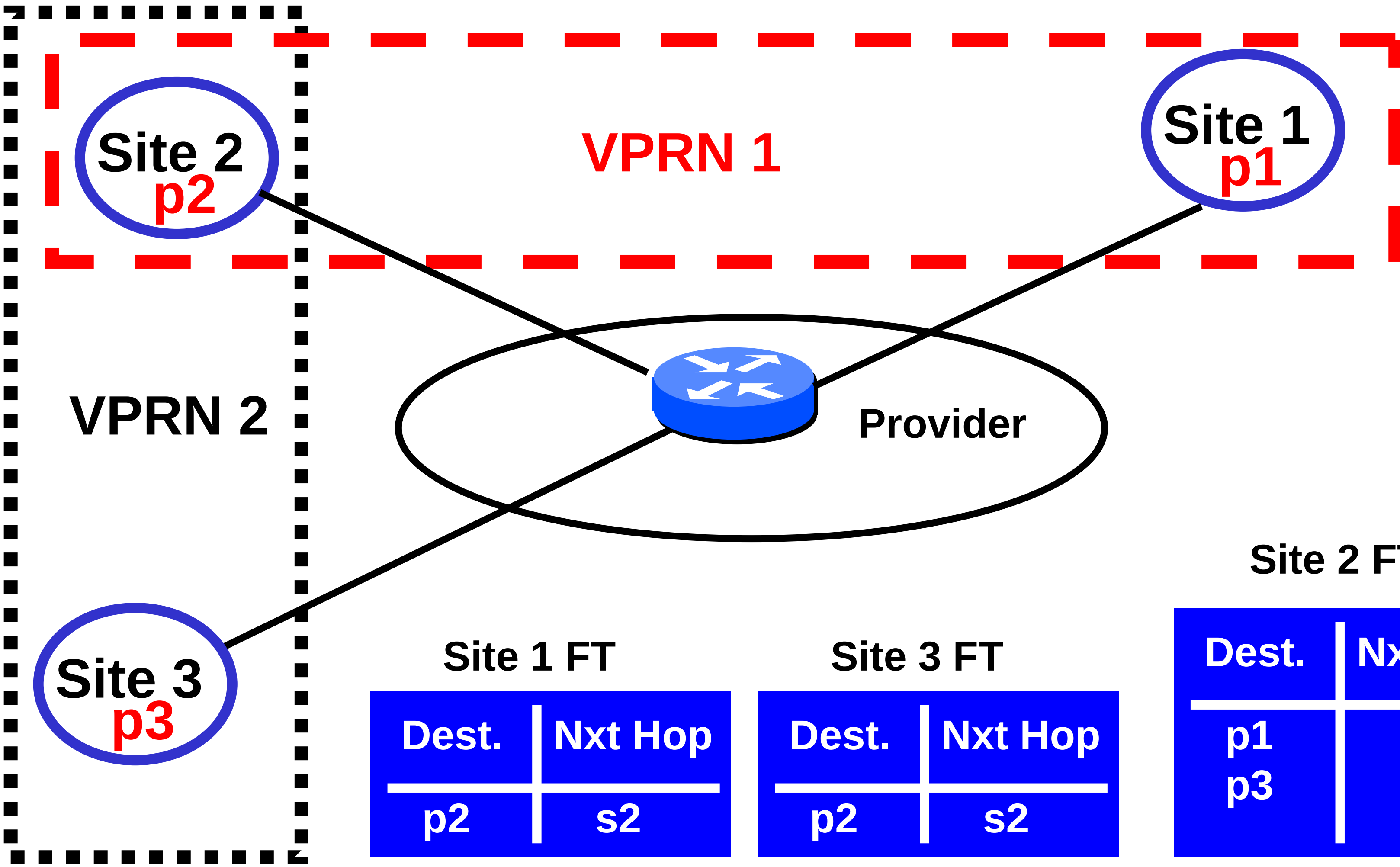


Dest.	Nxt Hop
p1	s1
p2	s2
p3	s3

Violates isolation Guarantee of A VPN: site 1 can Exchange traffic with Site 2!

RFC 2547 : Per site forwarding tables

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



Site 1 FT

Dest.	Nxt Hop
p2	s2

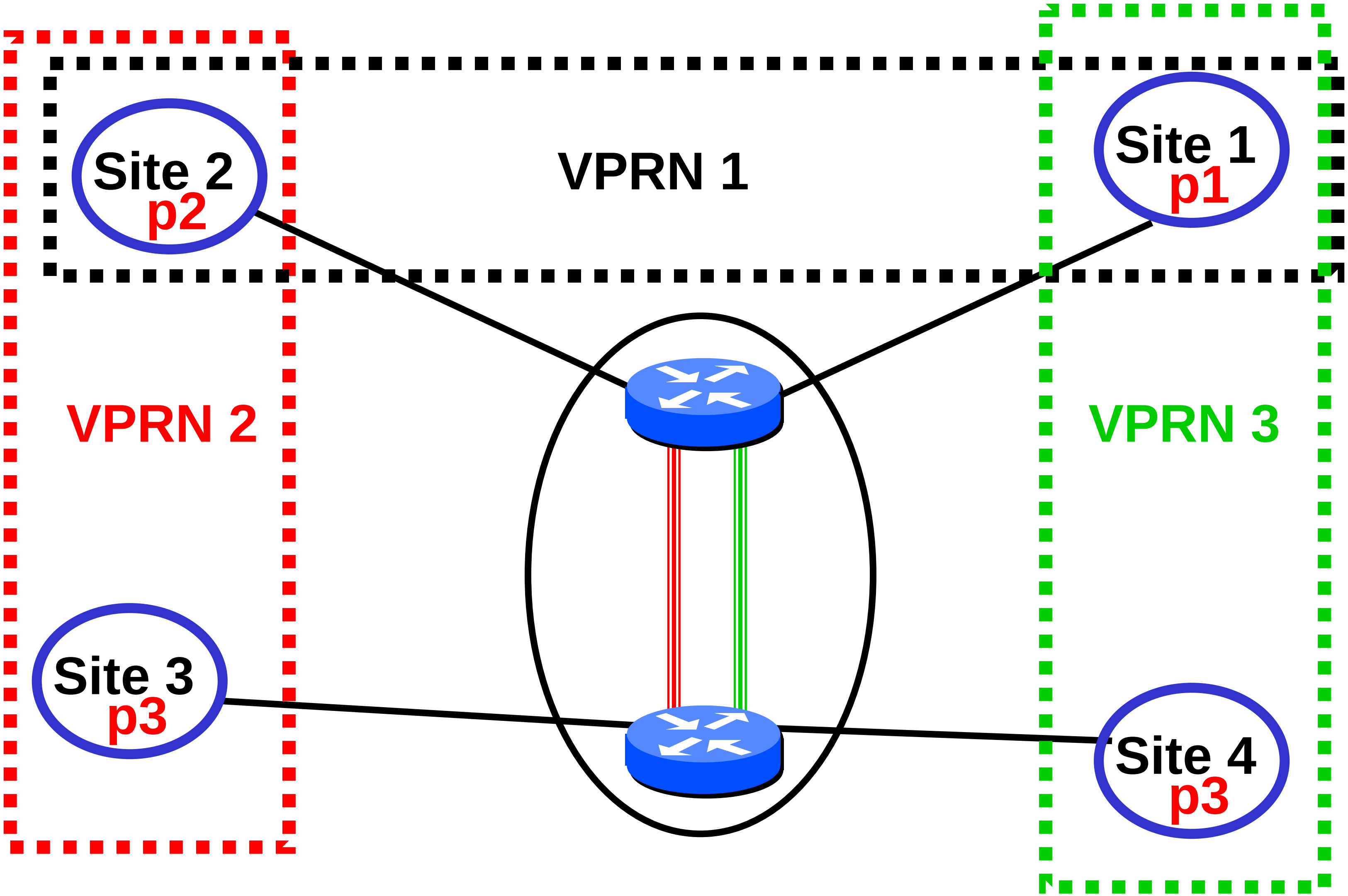
Site 3 FT

Dest.	Nxt Hop
p2	s2

Site 2 FT

Dest.	Nxt Hop
p1	s1
p3	s3

Tunnels required across backbone



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

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)

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

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

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

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