MPLS
Traffic Engineering
Virtual Private Networks
Virtual Private LAN Services (VPLS)

Sudhakar Ganti

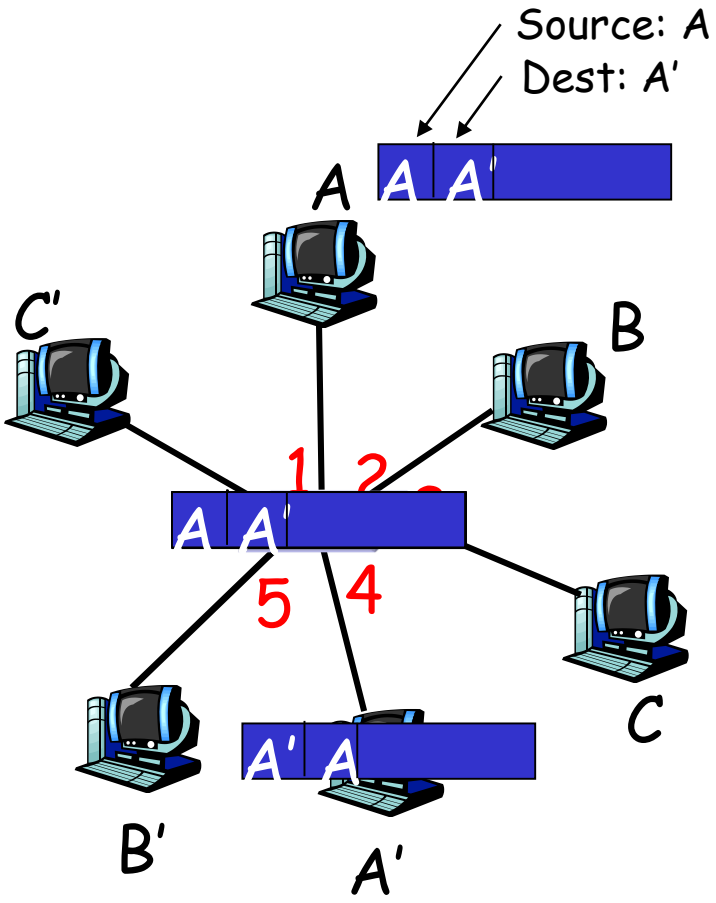
L2 versus L3

- Campus Networks L2 (Ethernet Switching)
  - Learning Bridges
  - Spanning Tree Protocol (STP)
    - Main intent is to avoid loops
    - Root bridge election; shortest path to root bridge
- Internet L3 (IP Routing)
  - Routing Table Lookup (Longest Prefix match)
  - Routing Protocols (OSPF, BGP, RIP)
  - Not much control over routing (always shortest path, even if congested)

Layer 2: Self-learning, forwarding: example

- frame destination unknown: flood
  - destination A location known

    selective send

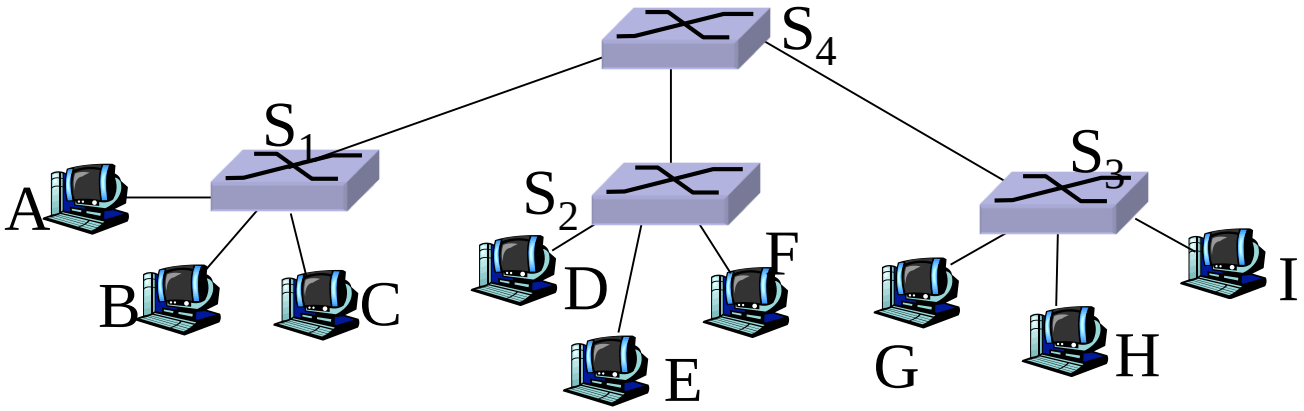


MAC addr	interface	TTL
A	1	60
A'	4	60

Switch table (initially empty)

Interconnecting switches

- switches can be connected together



- Q: sending from A to G - how does S_1 know to forward frame destined to F via S_4 and S_2?
- A: self learning! (works exactly the same as in single-switch case!)

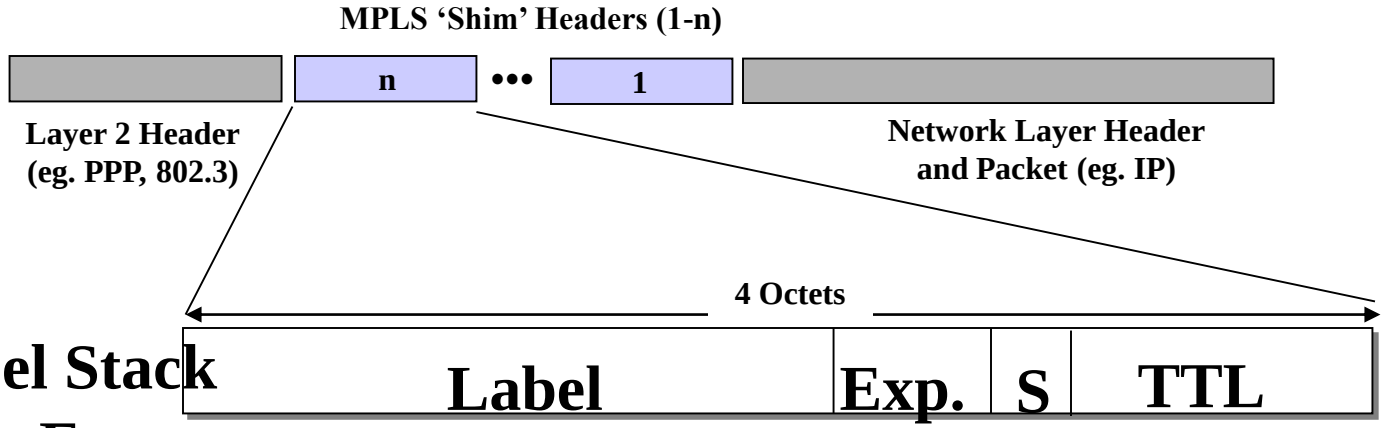
MPLS Layer 2.5
(A Quick Overview)

- *Multi-protocol Label Switching (MPLS)* is a new paradigm in routing and switching
- Allows ATOM - Anything (X) Over MPLS
- Based on a virtual path (called LSP) setup
- Appends a "label" to a given packet
  - 20-bit label value, 3-bit EXP field, 8-bit TTL, 1-bit S

- Routers use the label to switch packets
- *Routing at the Edge and Switching in the core*

MPLS allows fast packet switching; Fast table lookup

MPLS Encapsulation - PPP & LAN Data Links



Label: Label Value, 20 bits (0-16 reserved)
Exp.: Experimental, 3 bits (was Class of Service)
S: Bottom of Stack, 1 bit (1 = last entry in label stack)
TTL: Time to Live, 8 bits

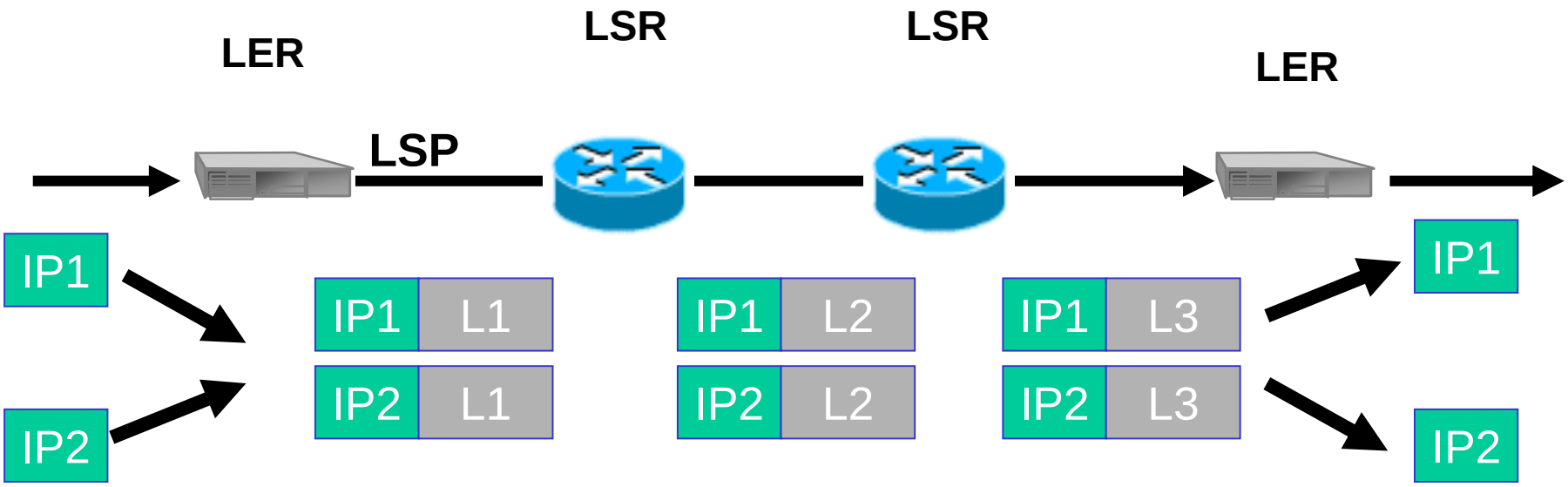
- Network layer must be inferable from value of bottom label of the stack
- TTL must be set to the value of the IP TTL field when packet is first labelled
- When last label is popped off stack, MPLS TTL to be copied to IP TTL field
- Pushing multiple labels may cause length of frame to exceed layer-2 MTU
  - LSR must support "Max. IP Datagram Size for Labelling" parameter
  - any unlabelled datagram greater in size than this parameter is to be fragmented

MPLS on PPP links and LANs uses 'Shim' Header Inserted Between Layer 2 and Layer 3 Headers

MPLS Terminology

- **LDP: Label Distribution Protocol**
  - **LDP, CR-LDP, RSVP-TE**
- **LSP: Label Switched Path**
  - **L-LSP; E-LSP (For QoS)**
- **FEC: Forwarding Equivalence Class**
  - **A group of flows treated as equivalent for Forwarding purposes**
- **LSR: Label Switching Router**
- **LER: Label Edge Router**
- **(Check out http://www.ietf.org)**

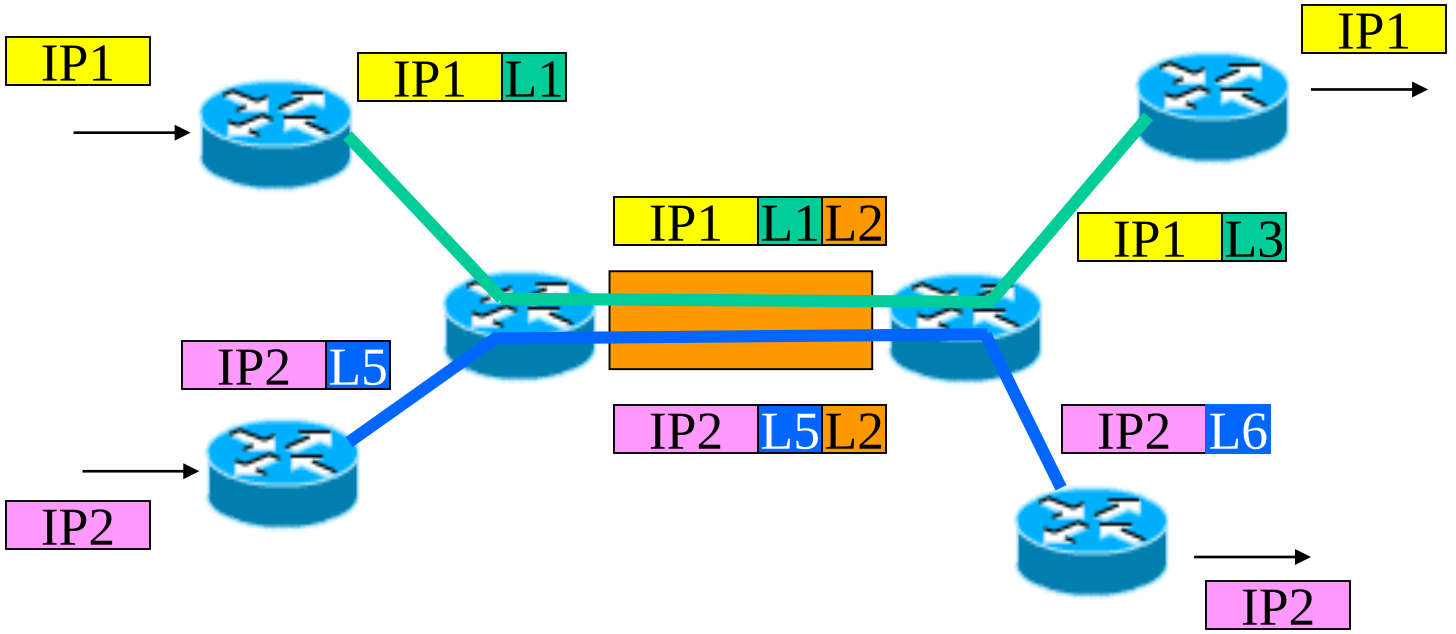
MPLS Operation



IP1, IP2 belong to the same FEC

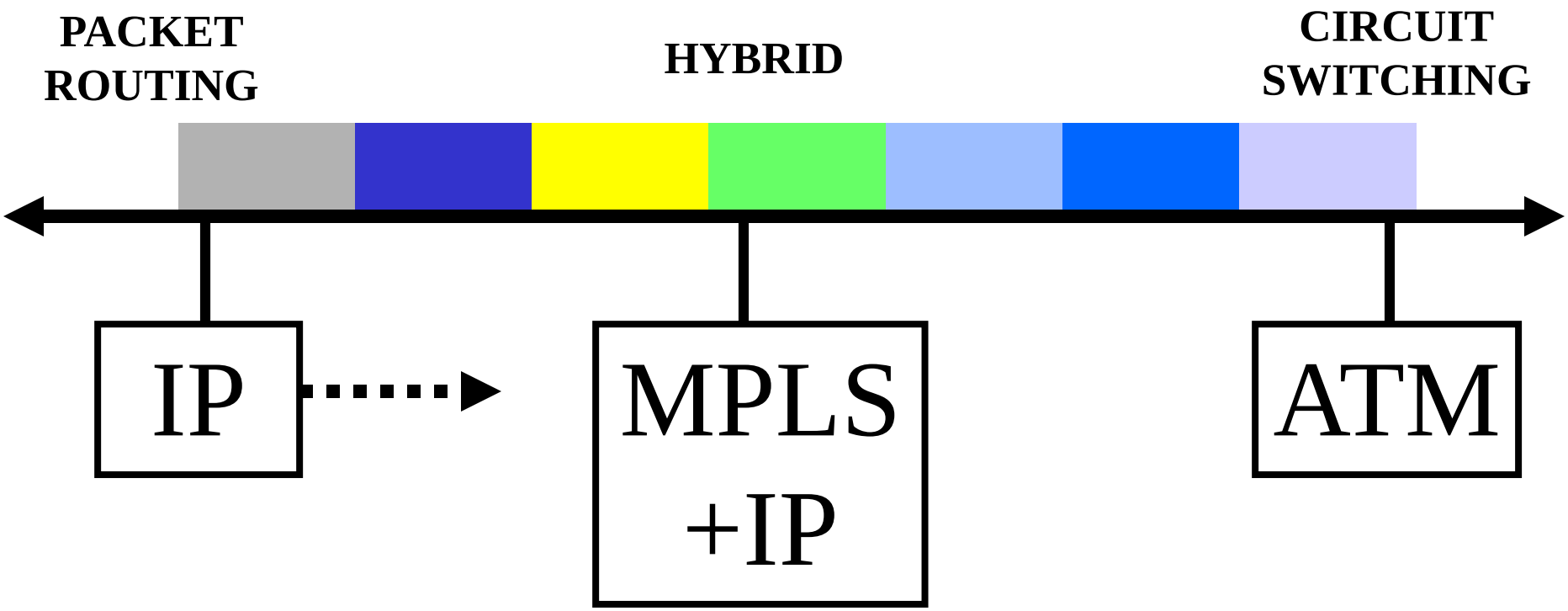
- When a LSP is setup between two routers, all routers along the path assign incoming and outgoing labels for the virtual path
- Label Edge Router appends label to the packet based on the virtual path
- All intermediate routers (LSR) switch the packet based on the label
- Faster Label lookup instead of Address lookup

MPLS - Hierarchy



- Multiple Labels can be stacked
  - Tunnel in a tunnel in a tunnel ……..
  - Allows easy operation of hub-and-spoke network designs
  - Better manageability in the core due to fewer tunnels (LSPs)
  - Transit nodes need not handle complete routing tables
- Processing is always based on Top-Label

BEST OF BOTH WORLDS



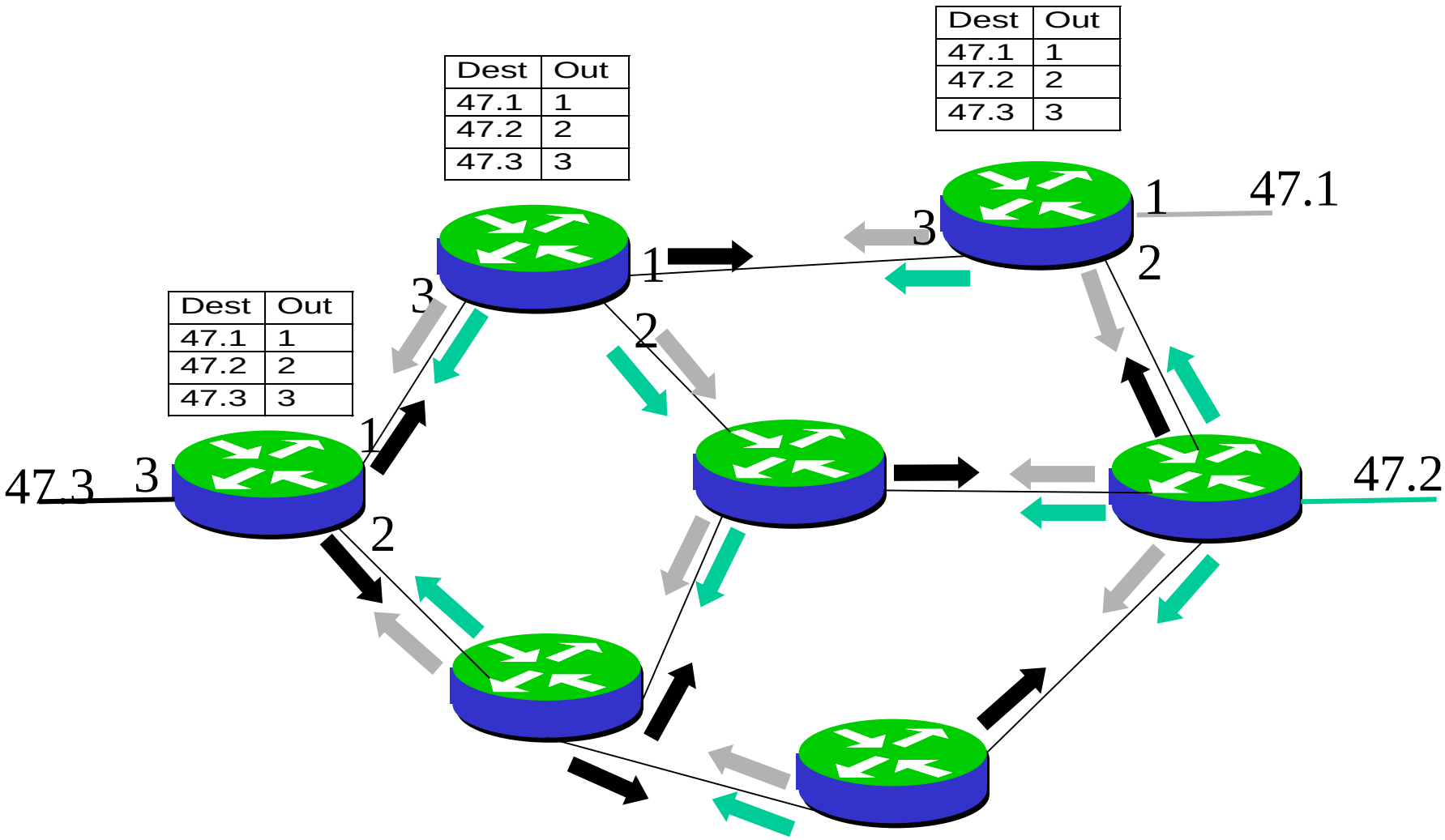
•MPLS + IP form a middle ground that combines the best of IP and the best of circuit switching technologies.

•ATM and Frame Relay cannot easily come to the middle so IP has!!

Explicit Routing in MPLS

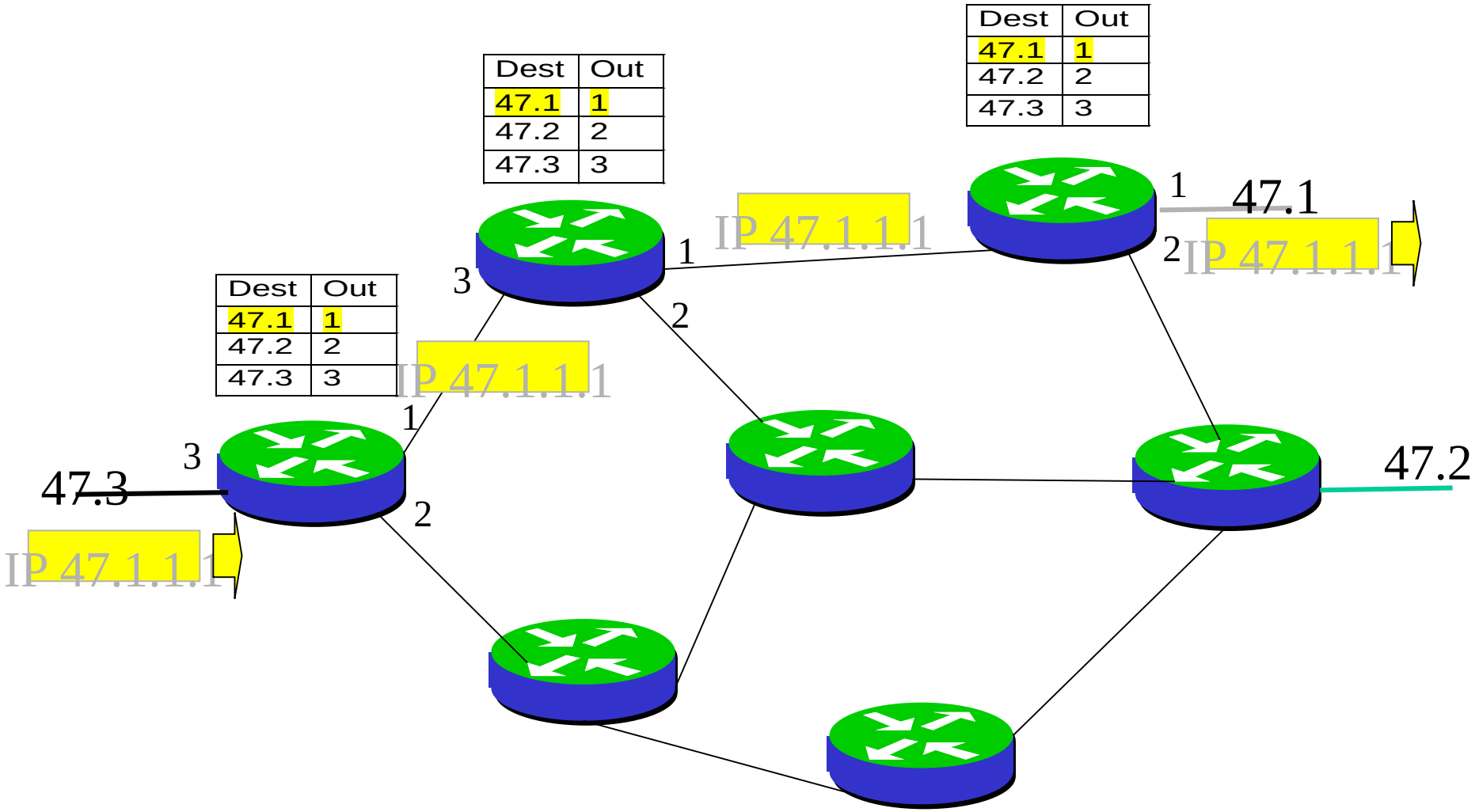
- Two options for route selection:
  - Hop by hop routing (LDP)
  - Explicit routing (CR-LDP or RSVP-TE)
- Explicit Routing (aka Source Routing) is a very powerful technique
  - With pure datagram routing overhead of carrying complete explicit route is prohibitive
  - MPLS allows explicit route to be carried only at the time the LSP is setup, and not with each packet
  - MPLS makes explicit routing practical
  - Signal Quality of Service Requirements

MPLS BUILT ON STANDARD IP



- **Destination based forwarding tables as built by OSPF, IS-IS, RIP, etc.**

IP FORWARDING USED BY HOP-BY-HOP CONTROL

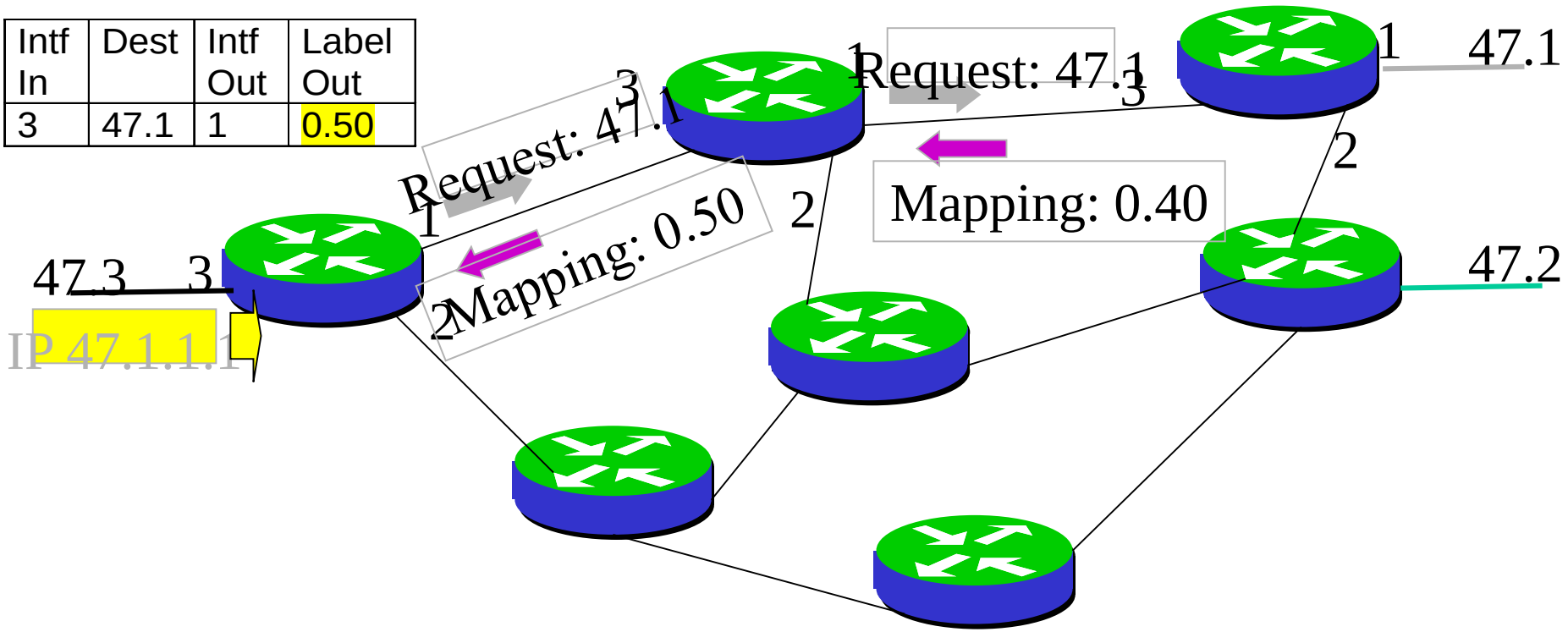


MPLS Label Distribution (On-demand allocation)

Intf In	Label In	Dest	Intf Out	Label Out
3	0.50	47.1	1	0.40

Intf In	Label In	Dest	Intf Out
3	0.40	47.1	1

Intf In	Dest	Intf Out	Label Out
3	47.1	1	0.50

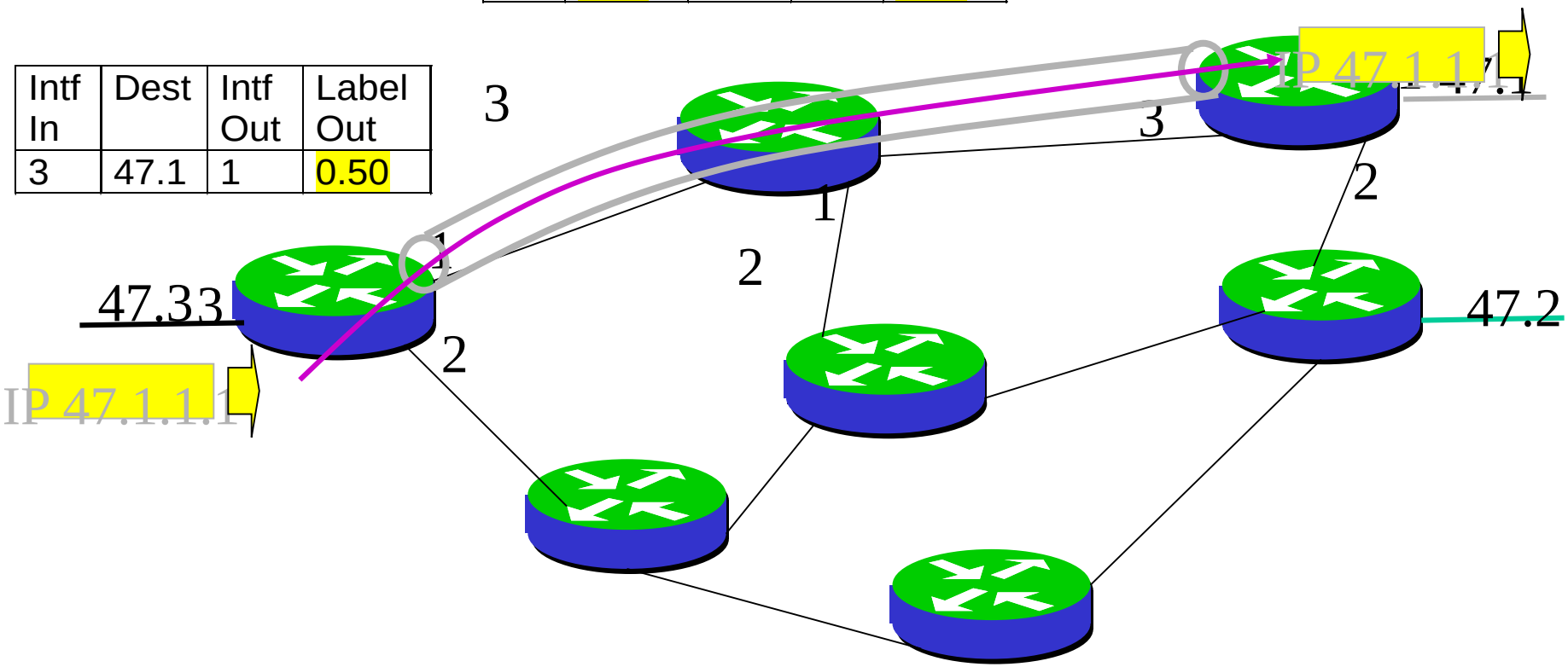


Label Switched Path (LSP)

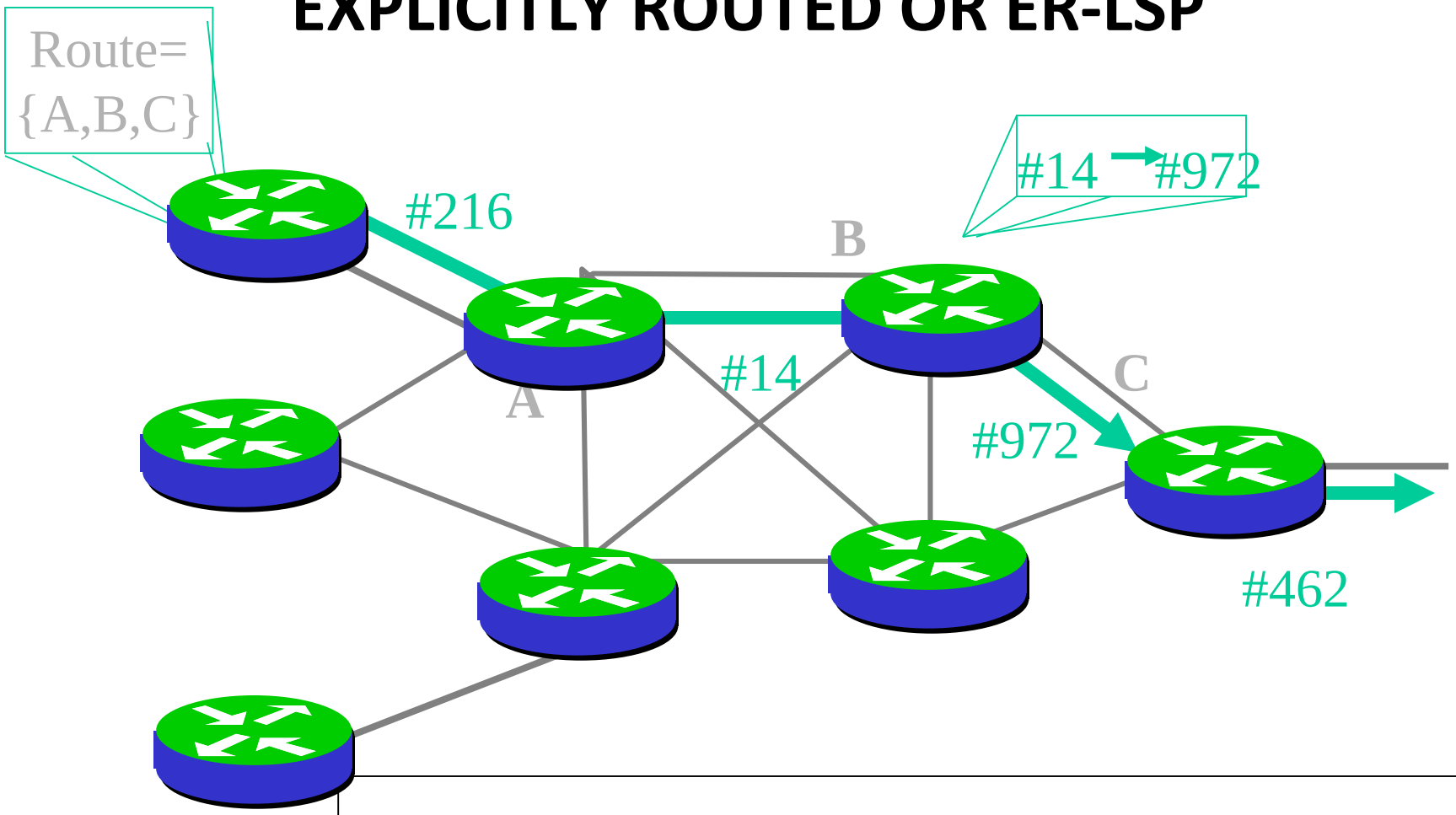
Intf In	Label In	Dest	Intf Out	Label Out
3	0.50	47.1	1	0.40

Intf In	Label In	Dest	Intf Out
3	0.40	47.1	1

Intf In	Dest	Intf Out	Label Out
3	47.1	1	0.50



EXPLICITLY ROUTED OR ER-LSP



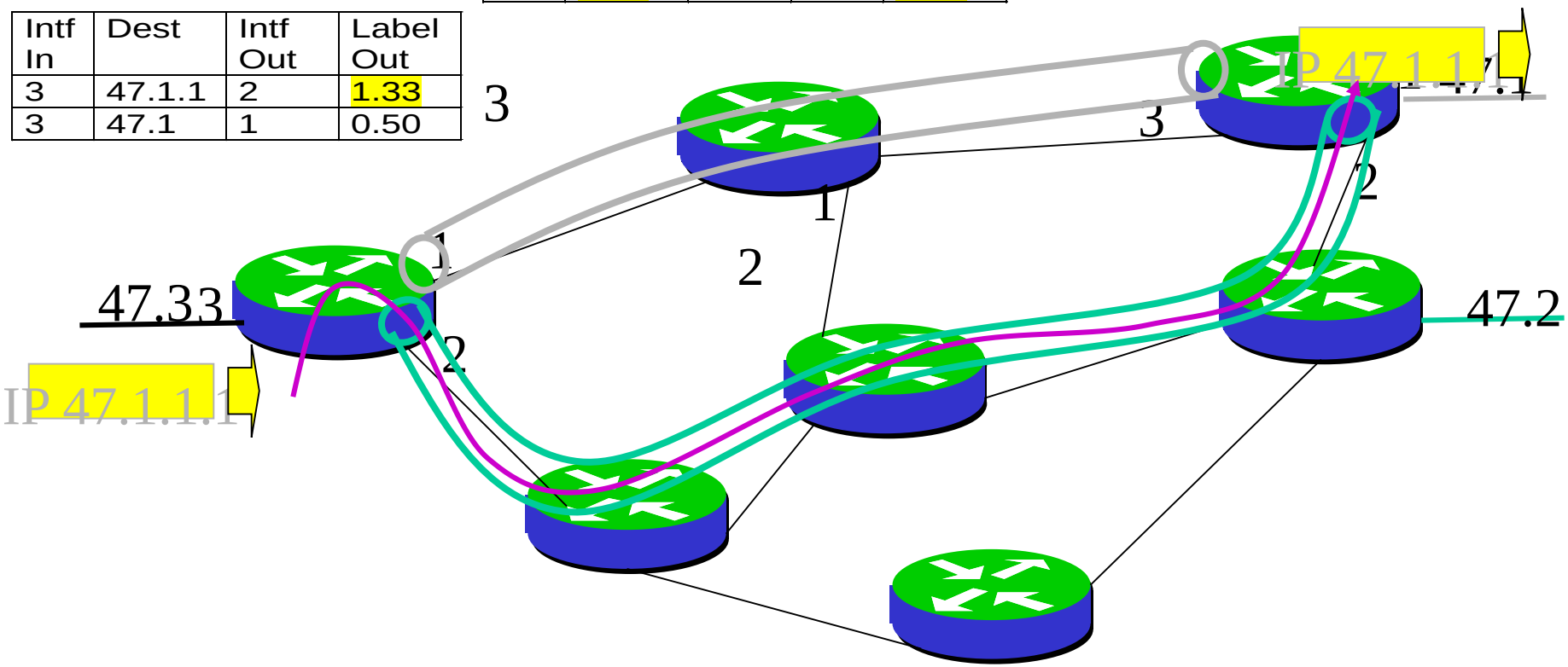
- ER-LSP follows route that source chooses. In other words, the control message to establish the LSP (label request) is ***source routed***.

EXPLICITLY ROUTED LSP ER-LSP

Intf In	Label In	Dest	Intf Out	Label Out
3	0.50	47.1	1	0.40

Intf In	Label In	Dest	Intf Out
3	0.40	47.1	1

Intf In	Dest	Intf Out	Label Out
3	47.1.1	2	1.33
3	47.1	1	0.50



Static vs Signaled LSPs

- Static LSPs
  - Are ‘nailed up’ manually
  - Have manually assigned MPLS labels
  - Needs configuration on each router
  - Do not re-route when a link fails
- Signaled LSPs
  - Signaled by RSVP
  - Have dynamically assigned MPLS labels
  - Configured on ingress router only
  - Can re-route around failures

ER LSP - advantages

- **Operator has routing flexibility**
  - **(policy-based, QoS-based)**
- **Can use routes other than shortest path**
- **Can compute routes based on constraints in exactly the same manner as ATM based on distributed topology database (traffic engineering)**

Comparison - Hop-by-Hop vs. Explicit Routing

Hop-by-Hop Routing

- Distributes routing of control traffic
- Builds a set of trees either fragment by fragment like a random fill, or backwards, or forwards in organized manner.
- Reroute on failure impacted by convergence time of routing protocol
- Existing routing protocols are destination prefix based
- Difficult to perform traffic engineering, QoS-based routing

Explicit Routing

- Source routing of control traffic
- Builds a path from source to dest
- Requires manual provisioning, or automated creation mechanisms.
- LSPs can be ranked so some reroute very quickly and/or backup paths may be pre-provisioned for rapid restoration
- Operator has routing flexibility (policy-based, QoS-based,
- Adapts well to traffic engineering

Explicit routing shows great promise for traffic engineering

Path Signaling

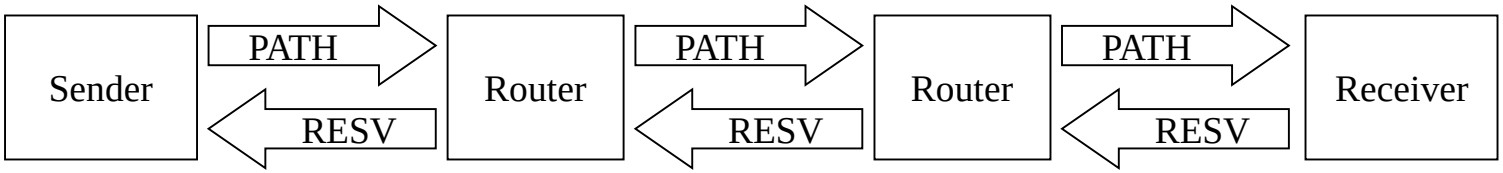
- RSVP for Traffic Engineering
  - Internet standard for reserving resources
  - Extended to support
    - Explicit path configuration
    - Path numbering
    - Route recording
  - Provides keepalive status
    - For visibility
    - For redundancy

RSVP

- A generic QoS signaling protocol
- An Internet control protocol
  - Uses IP as its network layer
- Originally designed for host-to-host
- Uses the IGP to determine paths
- RSVP is not
  - A data transport protocol
  - A routing protocol
- RFC 2205

Basic RSVP Path Signaling

- Simplex flows
- Ingress router initiates connection
- “Soft” state
  - Path and resources are maintained dynamically
  - Can change during the life of the RSVP session
- Path message sent downstream
- Resv message sent upstream



Other RSVP Message Types

- PathTear
  - Sent to egress router
- ResvTear
  - Sent to ingress router
- PathErr
  - Sent to ingress router
- ResvErr
  - Sent to egress router
- ResvConf

MPLS Extensions to RSVP

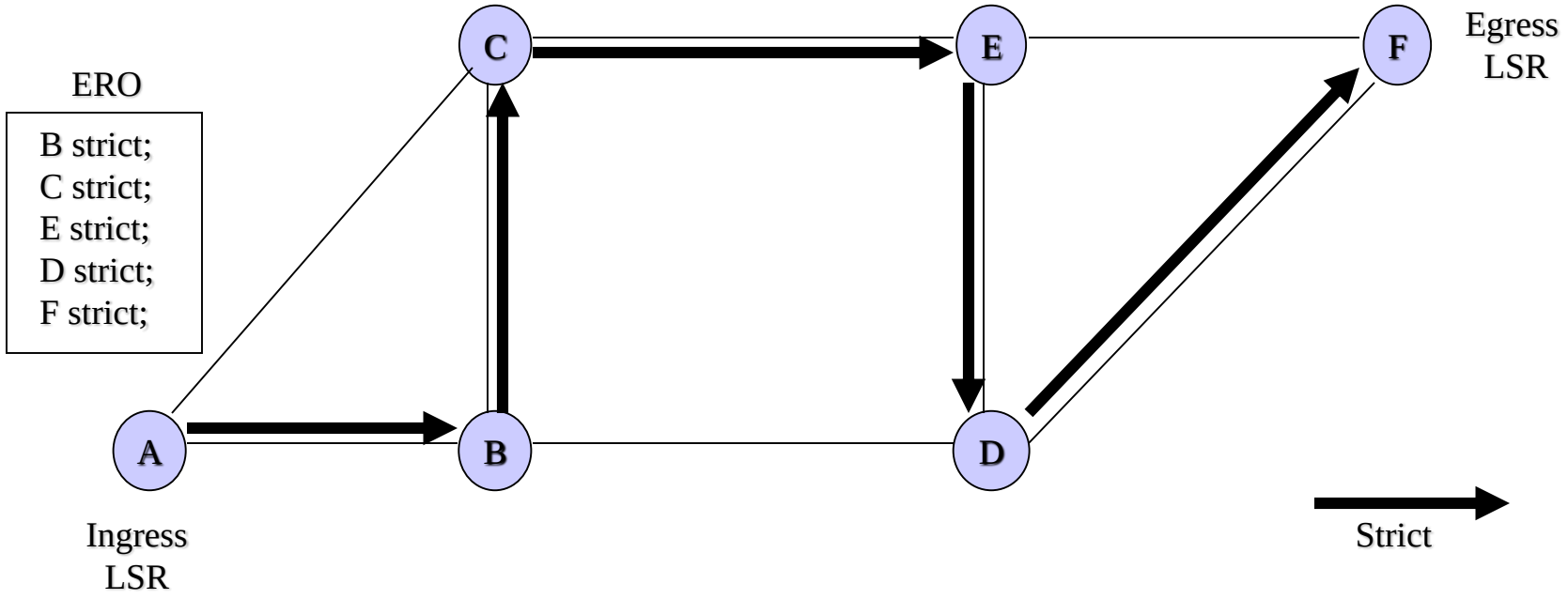
- Path and Resv message objects
  - Explicit Route Object (ERO)
  - Label Request Object
  - Label Object
  - Record Route Object
  - Session Attribute Object
  - Tspec Object

Explicit Route Object

- Used to specify the route RSVP Path messages take for setting up LSP
- Can specify loose or strict routes
  - Loose routes rely on routing table to find destination
  - Strict routes specify the directly-connected next router
- A route can have both loose and strict components

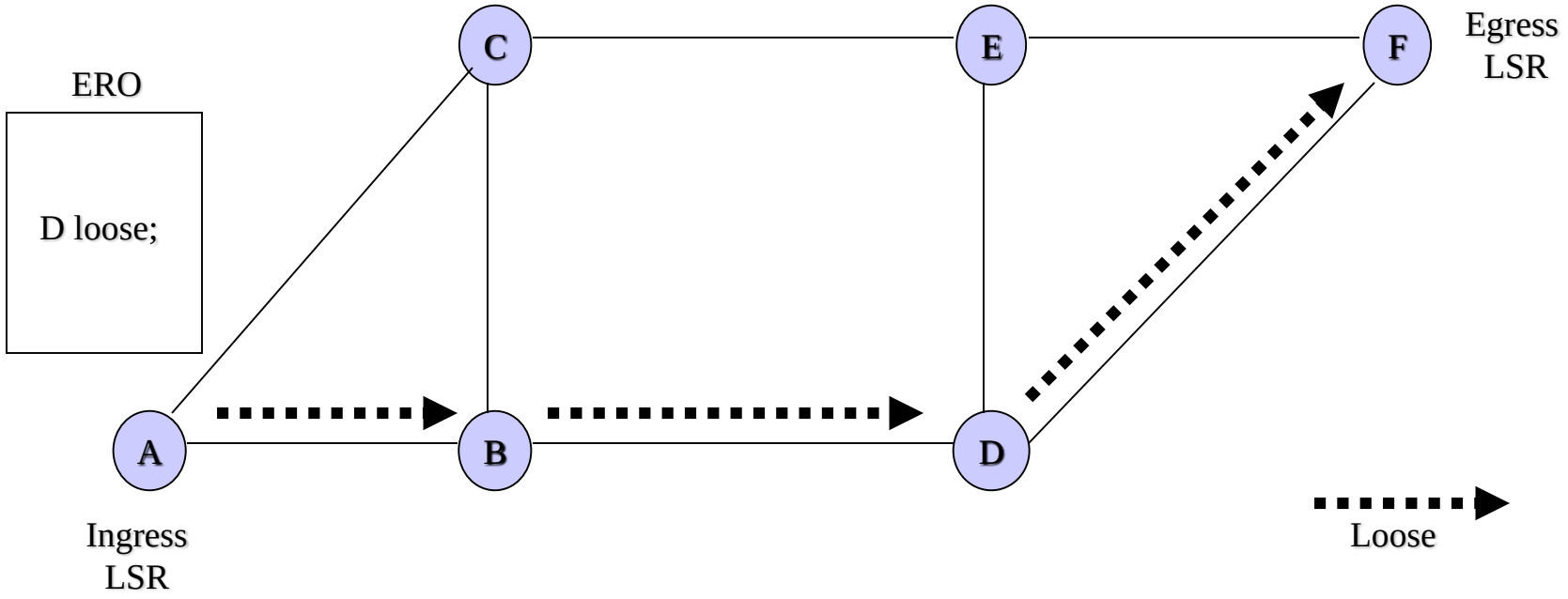
ERO: Strict Route

- **Next hop must be directly connected to previous hop**



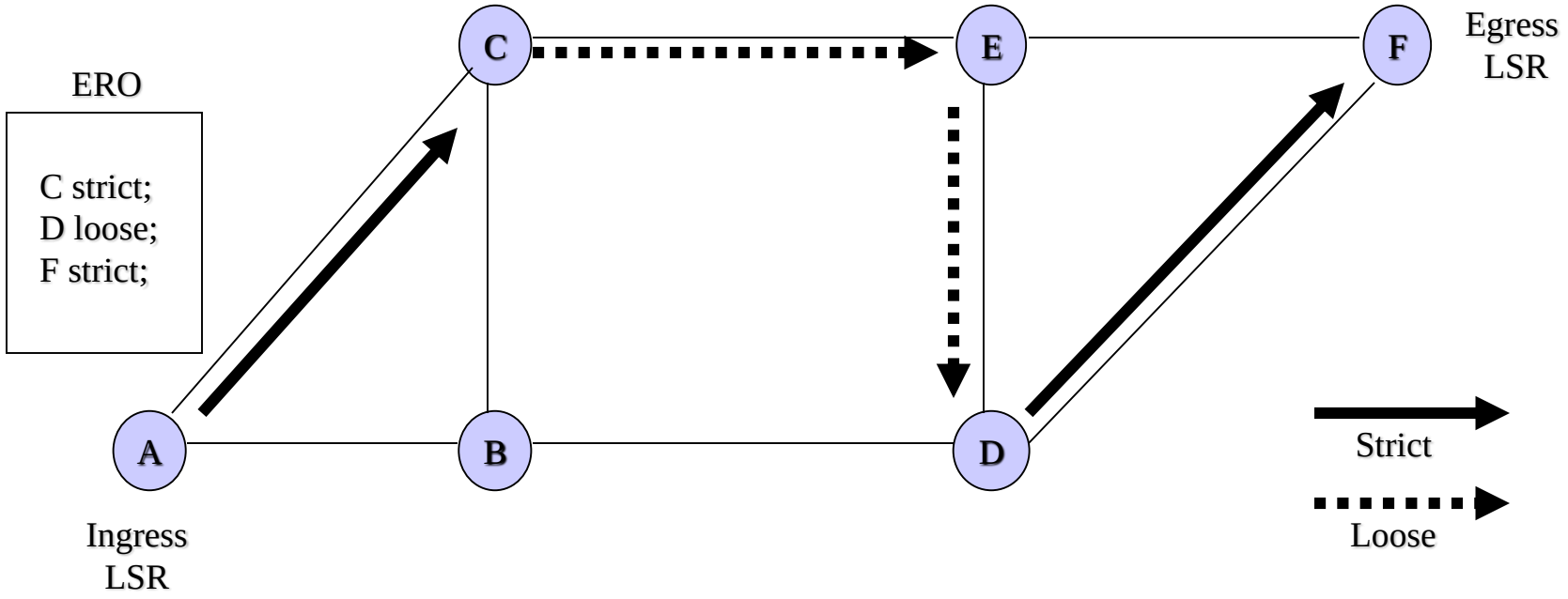
ERO: Loose Route

- **Consult the routing table at each hop to determine the best path**



ERO: Strict/Loose Path

- **Strict and loose routes can be mixed**



Signaled vs Constrained LSPs

- Common Features
  - Signaled by RSVP
  - MPLS labels automatically assigned
  - Configured on ingress router only
- Signaled LSPs
  - CSPF not used
  - User configured ERO handed to RSVP for signaling
  - RSVP consults routing table to make next hop decision
- Constrained LSPs
  - CSPF used
  - Full path computed by CSPF at ingress router
  - Complete ERO handed to RSVP for signaling

Constrained Shortest Path First Algorithm

- Modified “shortest path first” algorithm
- Finds shortest path based on IGP metric while satisfying additional constraints
- Integrates TED (Traffic Engineering Database)
  - IGP topology information
  - Available bandwidth
  - Link color
- Modified by administrative constraints
  - Maximum hop count
  - Bandwidth
  - Strict or loose routing
  - Administrative groups

Traffic Engineering Database (TED)

- CSPF uses TED to calculate explicit paths across the physical topology
- Similar to IGP link-state database
- Relies on extensions to IGP
  - Network link attributes
  - Topology information
- Separate from IGP database

Traffic Protection

- Primary LSP
  - Retry timer
  - Retry limit
- Secondary LSPs
  - Standby option
- Fast Reroute
- Adaptive mode

Preemption

- Defines relative importance of LSPs on same ingress router
- CSPF uses priority to optimize paths
- Higher priority LSPs
  - Are established first
  - Offer more optimal path selection
  - May tear down lower priority LSPs when rerouting
- Default configuration makes all LSPs equal

CR-LDP Traffic Parameters

U	F	Traf. Param. TLV	Length	
Flags		Frequency	Reserved	Weight
Peak Data Rate (PDR)				
Peak Burst Size (PBS)				
Committed Data Rate (CDR)				
Committed Burst Size (CBS)				
Excess Burst Size (EBS)				

32 bit fields are short IEEE floating point numbers

Any parameter may be used or not used by selecting appropriate values

Flags control "negotiability" of parameters

Frequency constrains the variable delay that may be introduced

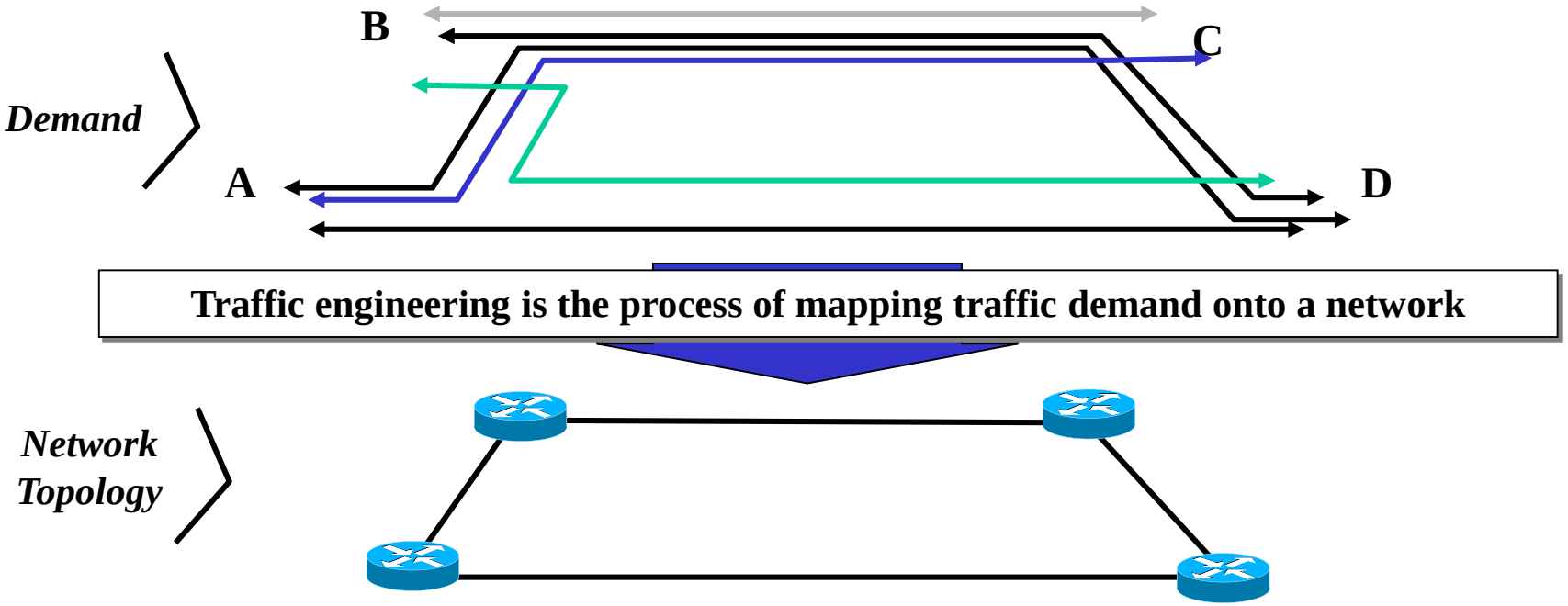
Weight of the CRLSP in the "relative share"

Peak rate (PDR+PBS) maximum rate at which traffic should be sent to the CRLSP

Committed rate (CDR+CBS) the rate that the MPLS domain commits to be available to the CRLSP

Excess Burst Size (EBS) to measure the extent by which the traffic sent on a CRLSP exceeds the committed rate

Traffic Engineering



Purpose of traffic engineering:

- **Maximize utilization of links and nodes throughout the network**
- **Engineer links to achieve required delay, grade-of-service**
- **Spread the network traffic across network links, minimize impact of single failure**
- **Ensure available spare link capacity for re-routing traffic on failure**
- **Meet policy requirements imposed by the network operator**

Traffic engineering key to optimizing cost/performance

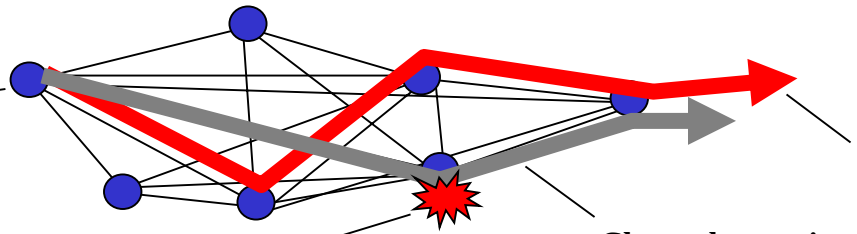
Traffic Engineering Alternatives

Current methods of traffic engineering:

Manipulating routing metrics	→	**Difficult to manage**
Use PVCs over an ATM backbone	→	**Not scalable**
Over-provision bandwidth	→	**Not economical**

MPLS provides a new method to do traffic engineering (traffic steering)

Example Network:



Potential benefits of MPLS for traffic engineering:

- allows explicitly routed paths	→	**operator control**
- no "n-squared" problem	→	**scalable**
- per FEC traffic monitoring	→	**granularity of feedback**
- backup paths may be configured	→	**redundancy/restoration**

MPLS combines benefits of ATM and IP-layer traffic engineering

QoS Routing

- Find the path for a given source and destination that best satisfies a given set of criteria (Multiple Constraints).

Performance metrics include:

– Hop count

– Delay

– Jitter

– Data loss rate

– Available bandwidth

– Queue length (available buffer space)

QoS Routing benefits

- Path setup Without QoS Routing
  - must probe path & backtrack
  - non optimal path
  - Control traffic and processing overhead and latency
- Path setup with QoS Routing
  - optimal route; “focused congestion” avoidance (TE)
  - more efficient Call Admission Control (at the source)
  - more efficient bandwidth allocation (per traffic class)
  - resource renegotiation possible

Routing Strategies

- Tasks of QoS routing
  - Collect the state information and keep it up to date
    - OSPF-TE is used in MPLS
  - Find a feasible path for a new connection
- Routing can be divided into three categories according to how the state information is maintained and the search of feasible paths is carried out:
  - Source routing
  - Distributed routing
  - Hierarchical routing

How to support CoS?

- Core:
  - IP Class-of-Service (CoS):
    - Qualitative Commitments
    - *EF* (Expedited Forwarding), *AF* (Assured Forwarding), *BE* (Best Effort)
    - AF is further divided into 4 classes with 3 drop precedence each
- Edge:
  - IEEE 802.1p/Q traffic prioritization
    - 3-bit field to indicate 8 levels of precedence
    - 12 bit VLAN ID
  - CoS per VLAN
  - CoS per MAC (Customer)

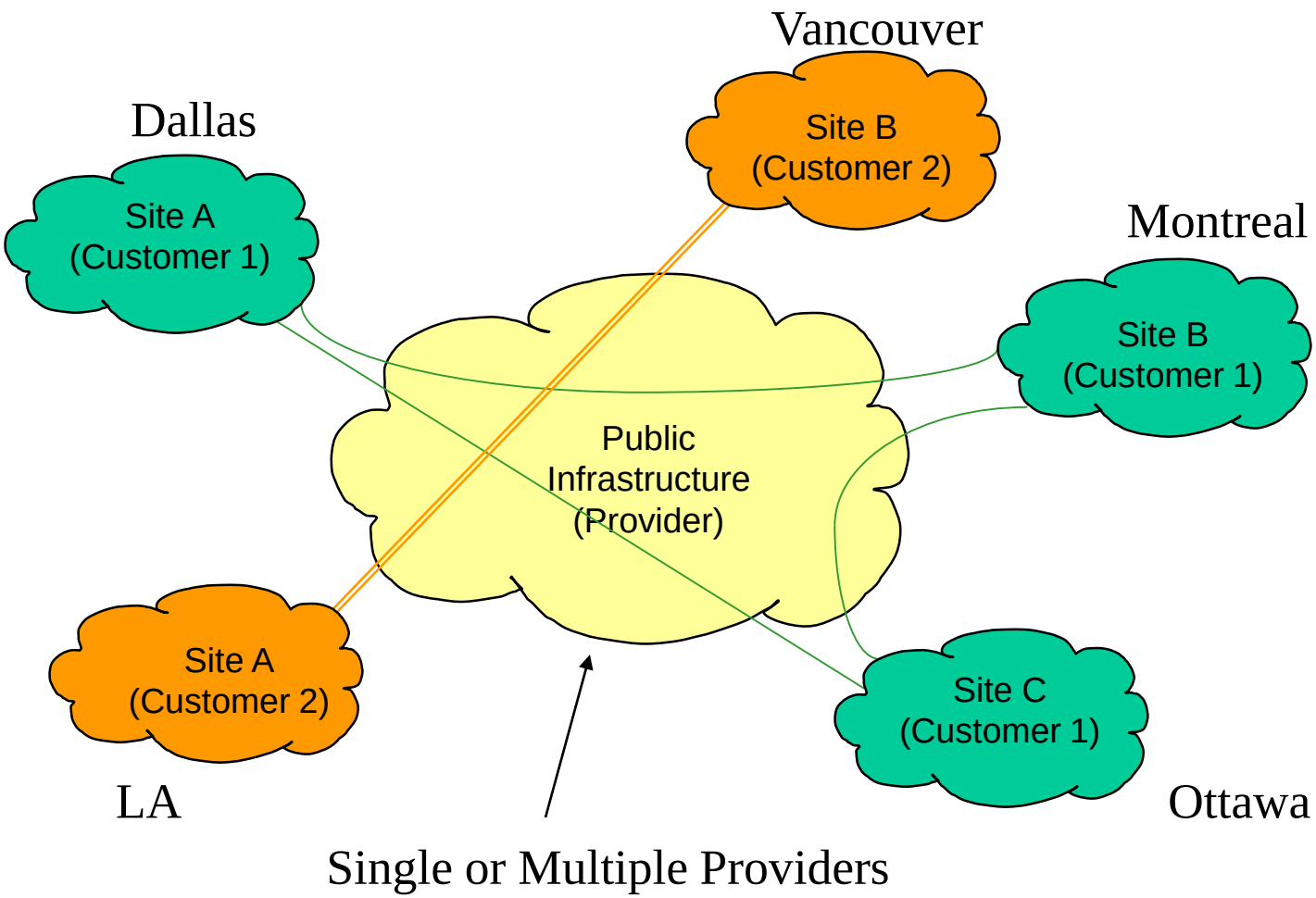
MPLS QoS Support (L-LSP)

- Inferred from the LSP itself (L-LSP)
  - The Label and /or EXP bits indicate the packet treatment (e.g., Label indicates class and EXP bits indicate drop preference)
  - Each LSP carries only one QoS class
  - Can support many QoS classes
  - For each VPLS, need to setup multiple VPNs, one for each class for end-to-end QoS support
    - Manageability problem (Scalability)
    - Traffic belonging to different QoS classes may take different paths along the network
    - Debugging could be a nightmare

MPLS QoS Support (E-LSP)

- Inferred from the EXP (E-LSP)
  - 3-bit EXP field to support 8-CoS
  - Can map directly 3-bit Ethernet priority
  - One LSP carries all the traffic classes
  - Single VPN per customer to carry different CoS
  - Easy manageability
- *One Problem though ……*
  - Signaling as defined for the MPLS does not allow to specify traffic parameters for each class separately for E-LSP

Virtual Private Networks (VPN)



Virtual Private Networks

- An ***emulation*** of private network facility using a *provider's* public infrastructure network over a wide geographical area
- Customers do not see provider's network as well as other customer's network
- It is *provider's* responsibility to maintain connectivity between various customer sites
- Security, Traffic Isolation are a must
- Implementation can be either customer or provider based

VPN emulates a secure private network connectivity

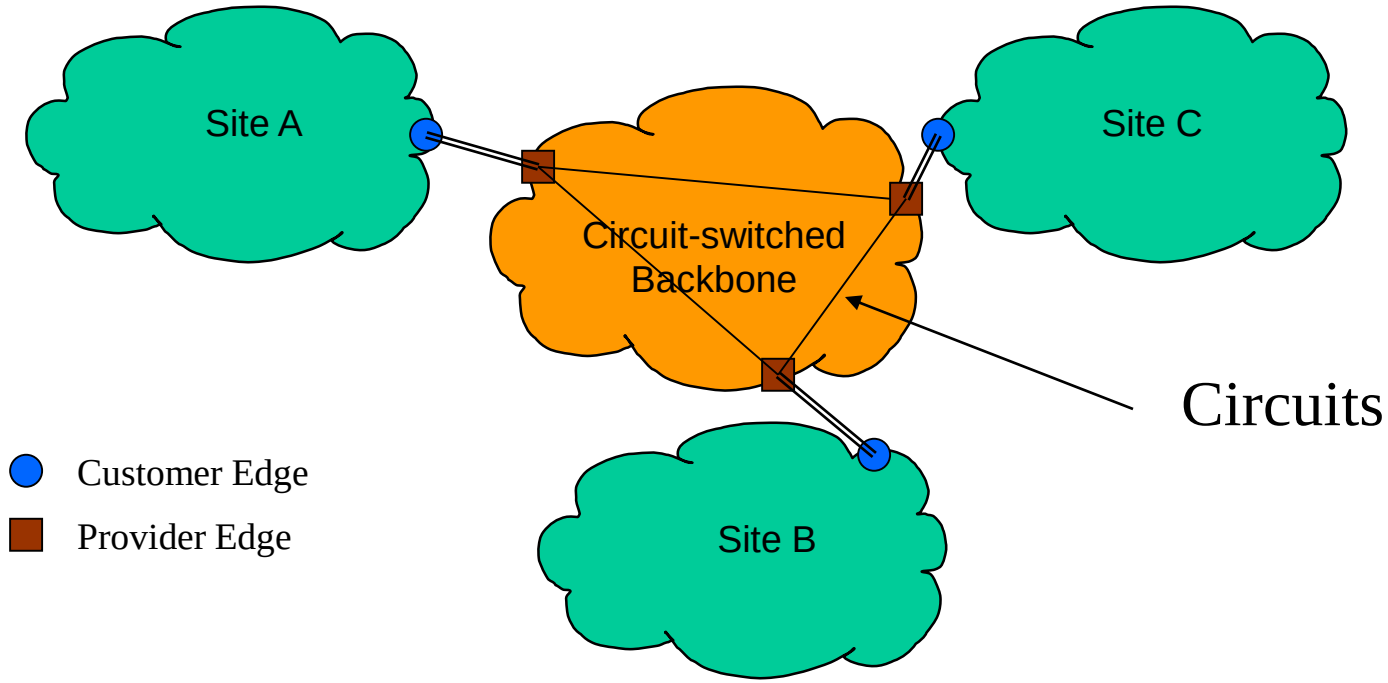
VPN Requirements

- Support for Customer Addressing
  - non-unique, overlapping address spaces
- Support for Data Security
  - authenticity, privacy, integrity
- Support for QoS Assurances
  - Service Level Agreements (SLAs)
    - Bandwidth, latency
    - Path Protection

VPN Classification

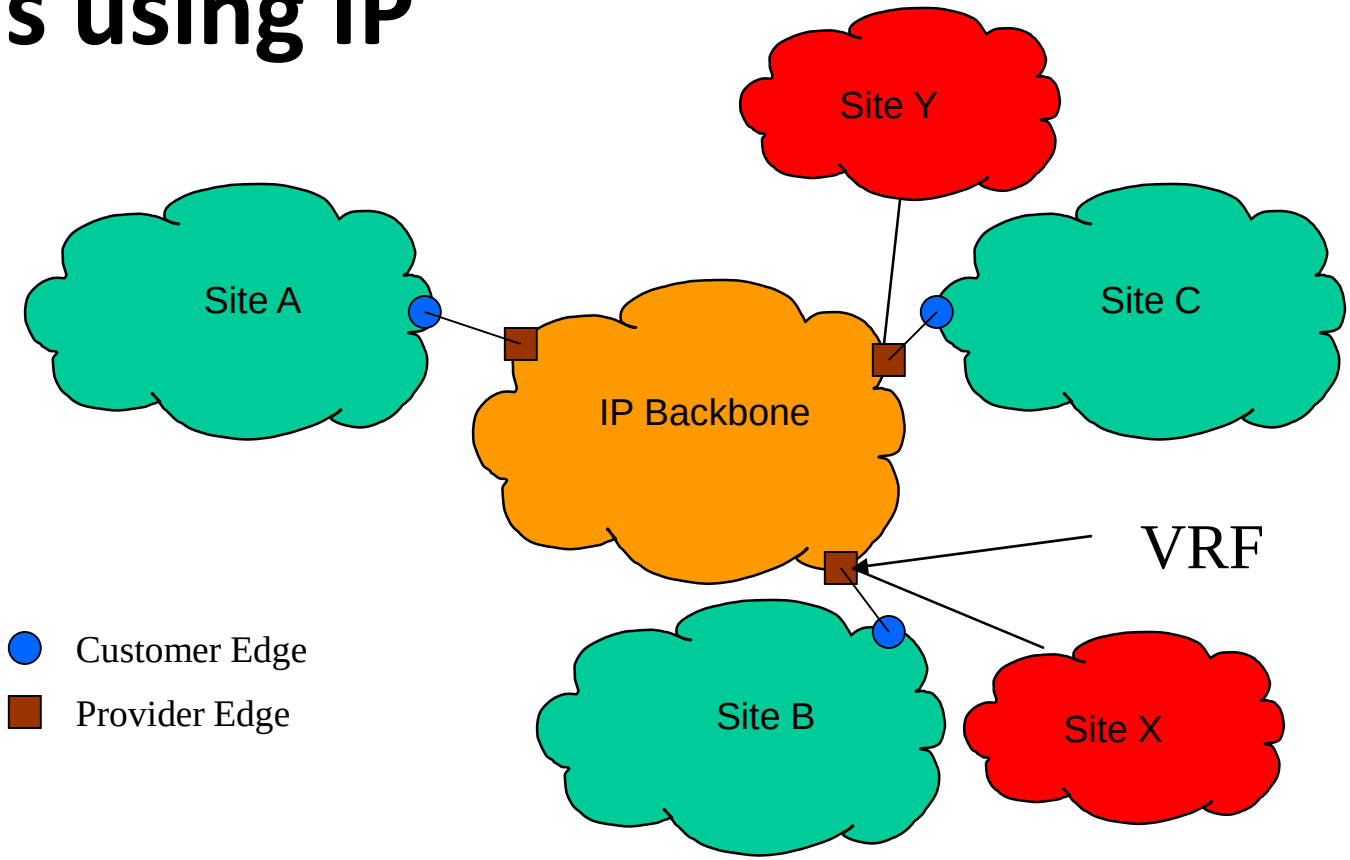
- Who implements VPN?
  - Customer (CE) based or Provider (PE) based
- What Layer VPN Operates
  - Layer 2 or Layer 3?
- How is the VPN implemented?
  - Backbone technology and tunneling mechanism
  - membership discovery
  - signaling and QoS support

VPNs, the very old way



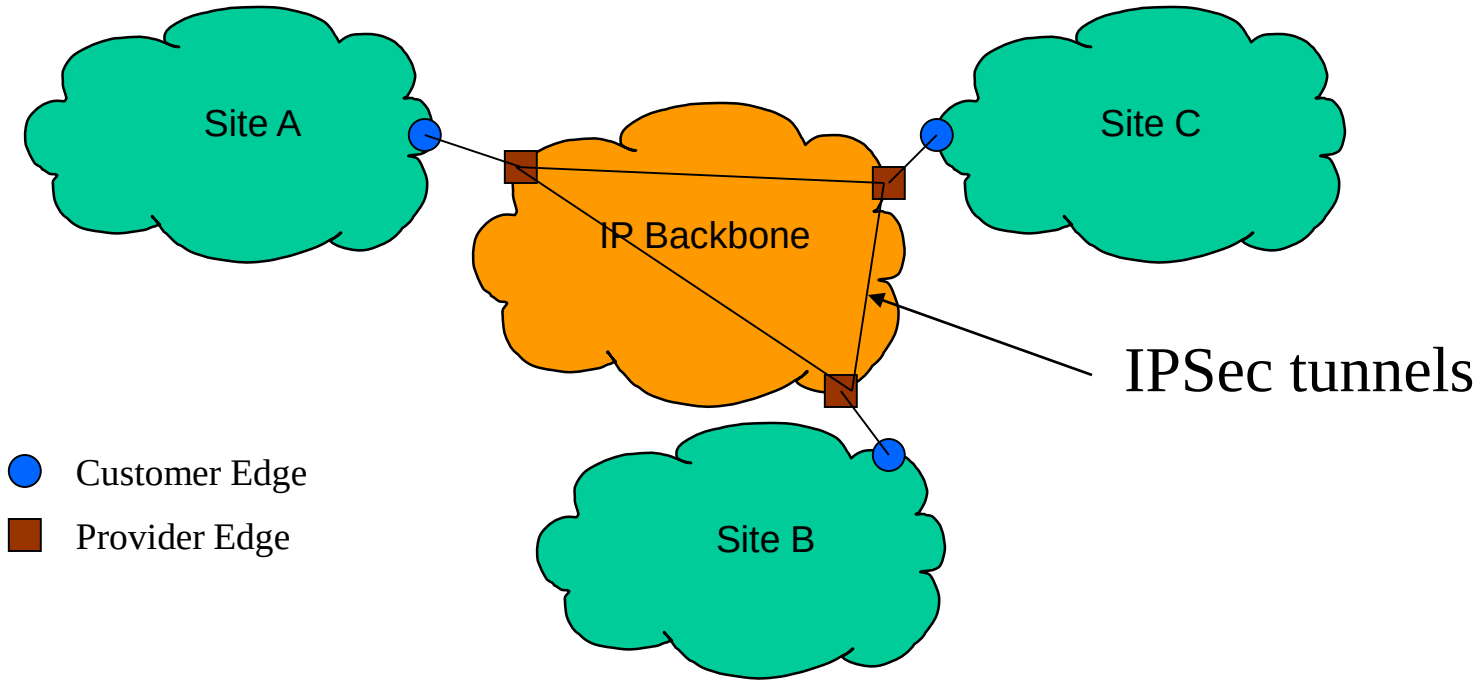
- Dedicated, Expensive, Leased Lines (Circuits) to interconnect various sites
- Access to Intra-Sites and Service Guarantees maintained by VPN Access Points
- Overlay VPNs for different applications e.g., Voice, video, data
- Adding a new-site is cumbersome and expensive
- CE based

VPNs using IP



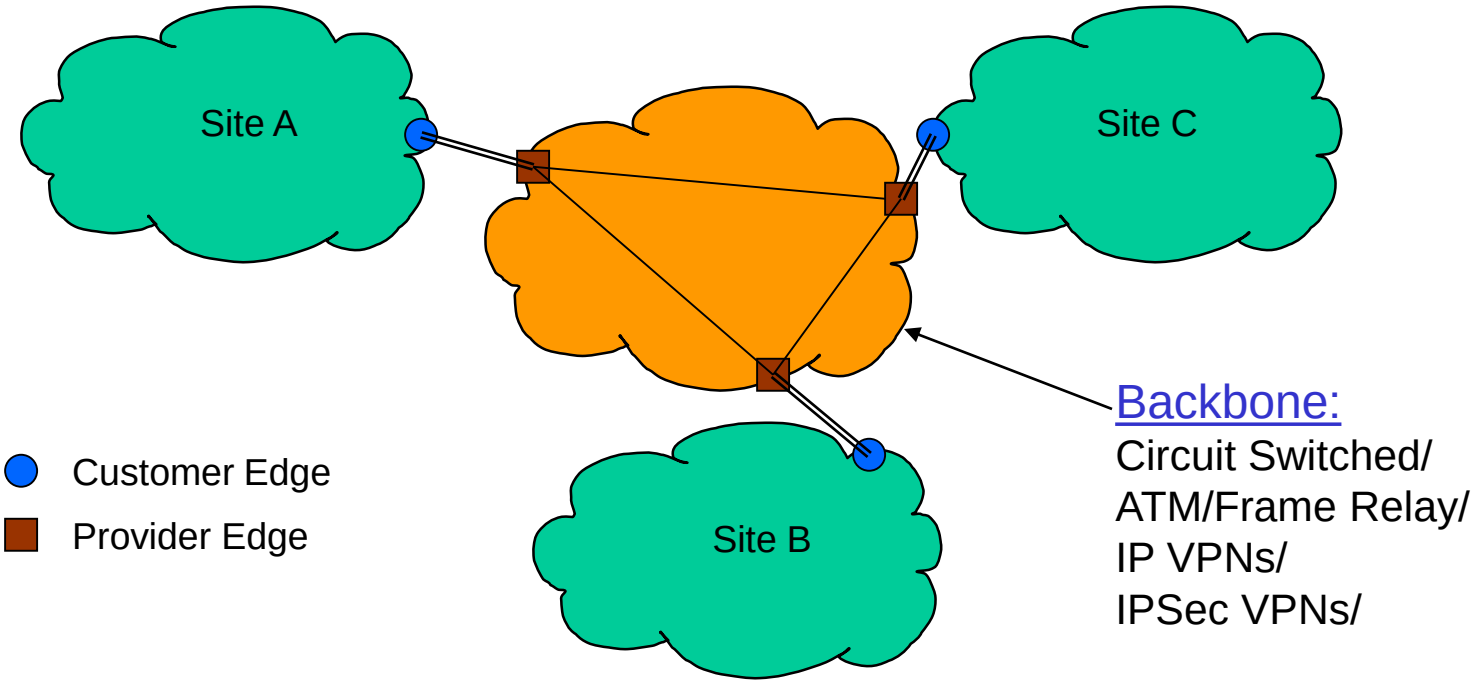
- Virtual Routing Function (VRF) used at PE
- Uses separate “logical” routing table per VPN
- Best effort service, mostly data services
- VPN discovery (using BGP) makes it easy to add another site
- PE based

VPNs using IPSec



- IPSec tunnels used from Site to site
- CE based
- Best effort service, mostly data services
- Secured authenticated tunnels

VPNs, so far



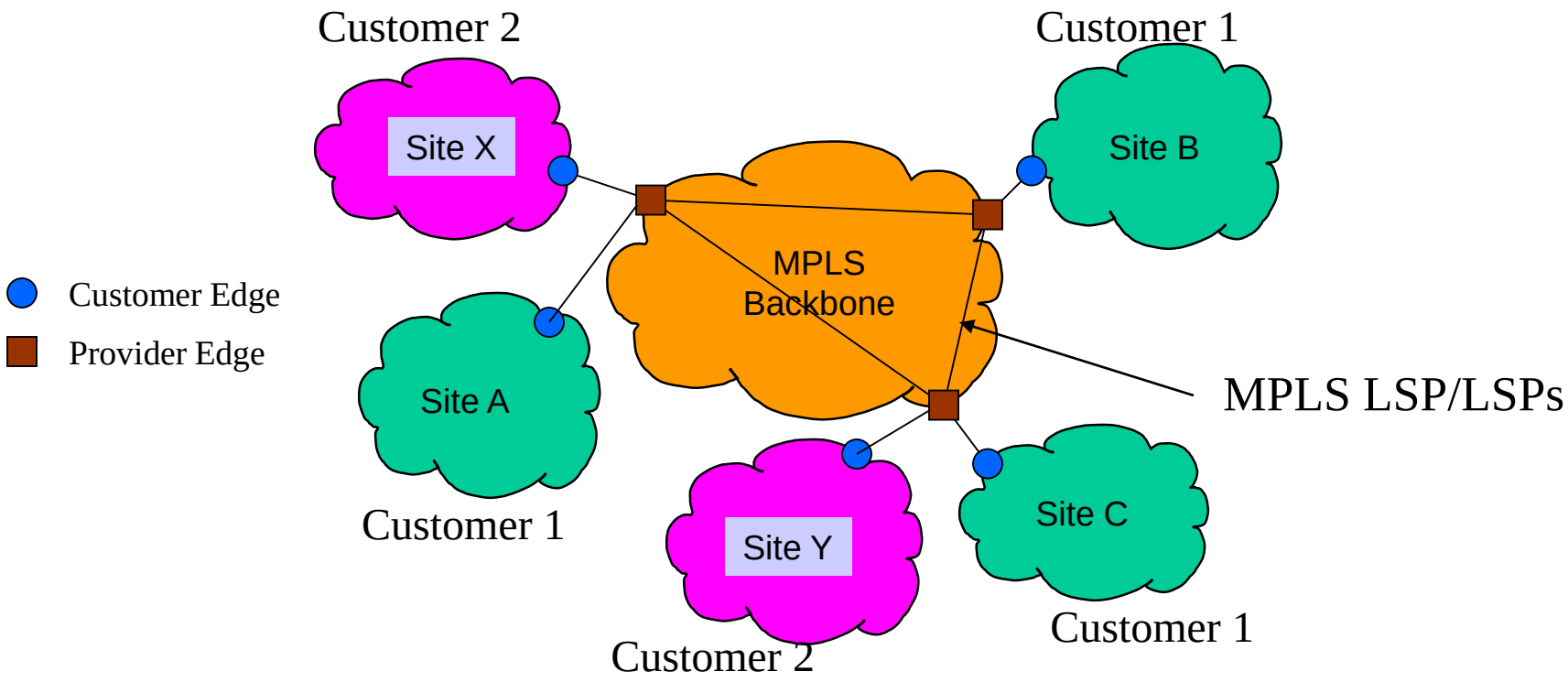
Backbone	***Connectivity***	**QoS**	***Discovery***	***Type***	***Problem***
Circuit Switched	Leased Lines	Overlay	No	CE	Management
ATM/FR	Virtual Circuits	Overlay	No	CE	Management
IP VPNs	VR-Connectionless	Besteffort	Yes	PE	No QoS
IPSec	Tunnels	Besteffort	No	CE	Management

What's Ideal?

- Unified Packet Transport
  - IP based
- End-to-End QoS Capability
  - Traffic Parameter Signaling
- Hierarchical Tunnels
  - Upper tunnels for network design (TE, Protection)
  - Lower tunnels for customer connectivity
- Hub-and-Spoke Network Designs
  - Manageability and efficiency
- Membership Discovery
  - Automatically add new sites to the VPN

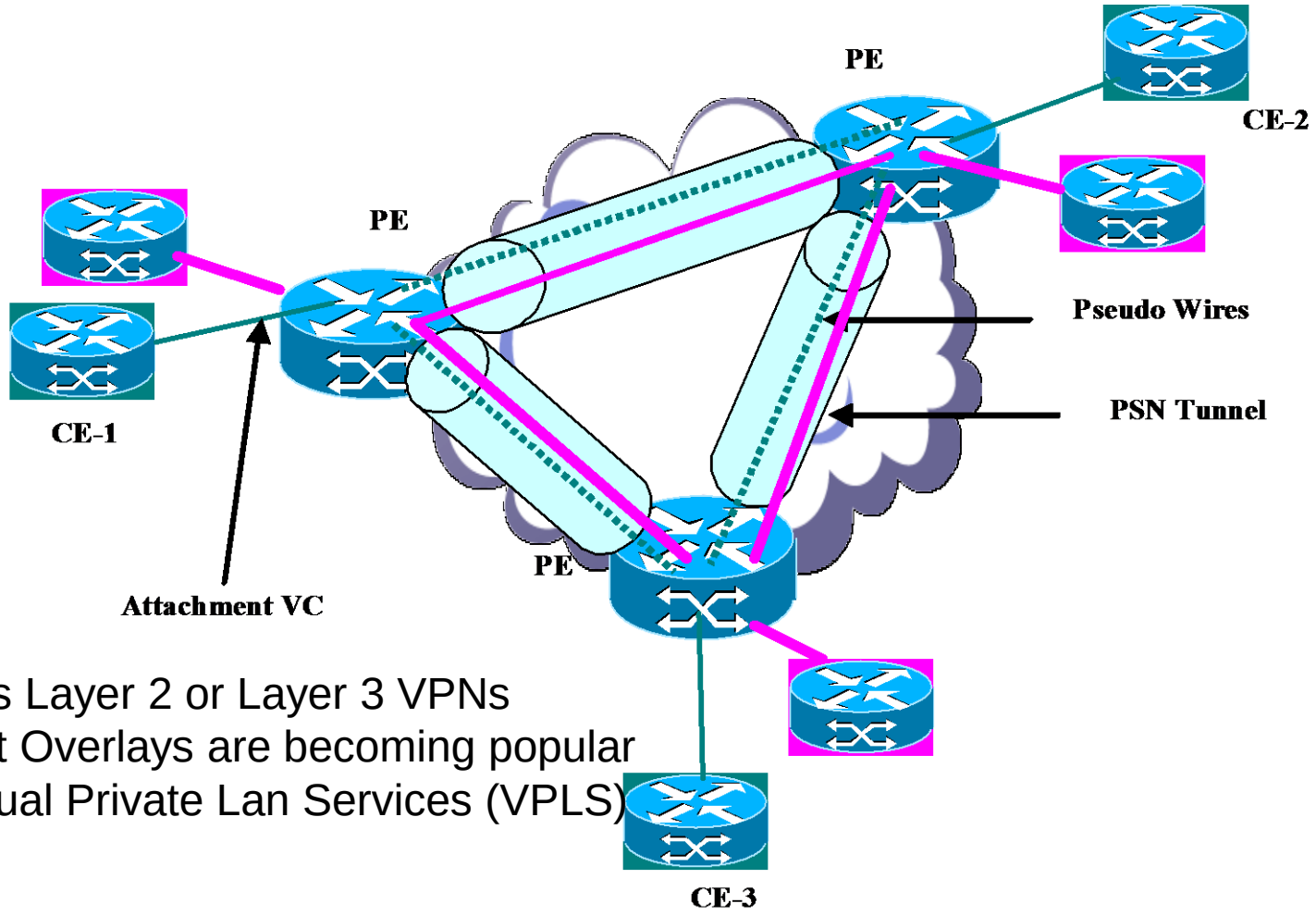
MPLS promises to bring all these and many more

VPNs using MPLS



- Virtual Switching Function (VSF) used at PE
- Full or Partial mesh (due to hierarchy) of Virtual connections
- Can provide Quality of service
- VPN discovery protocols makes it easy to add another site

Implemented using MPLS Tunnels

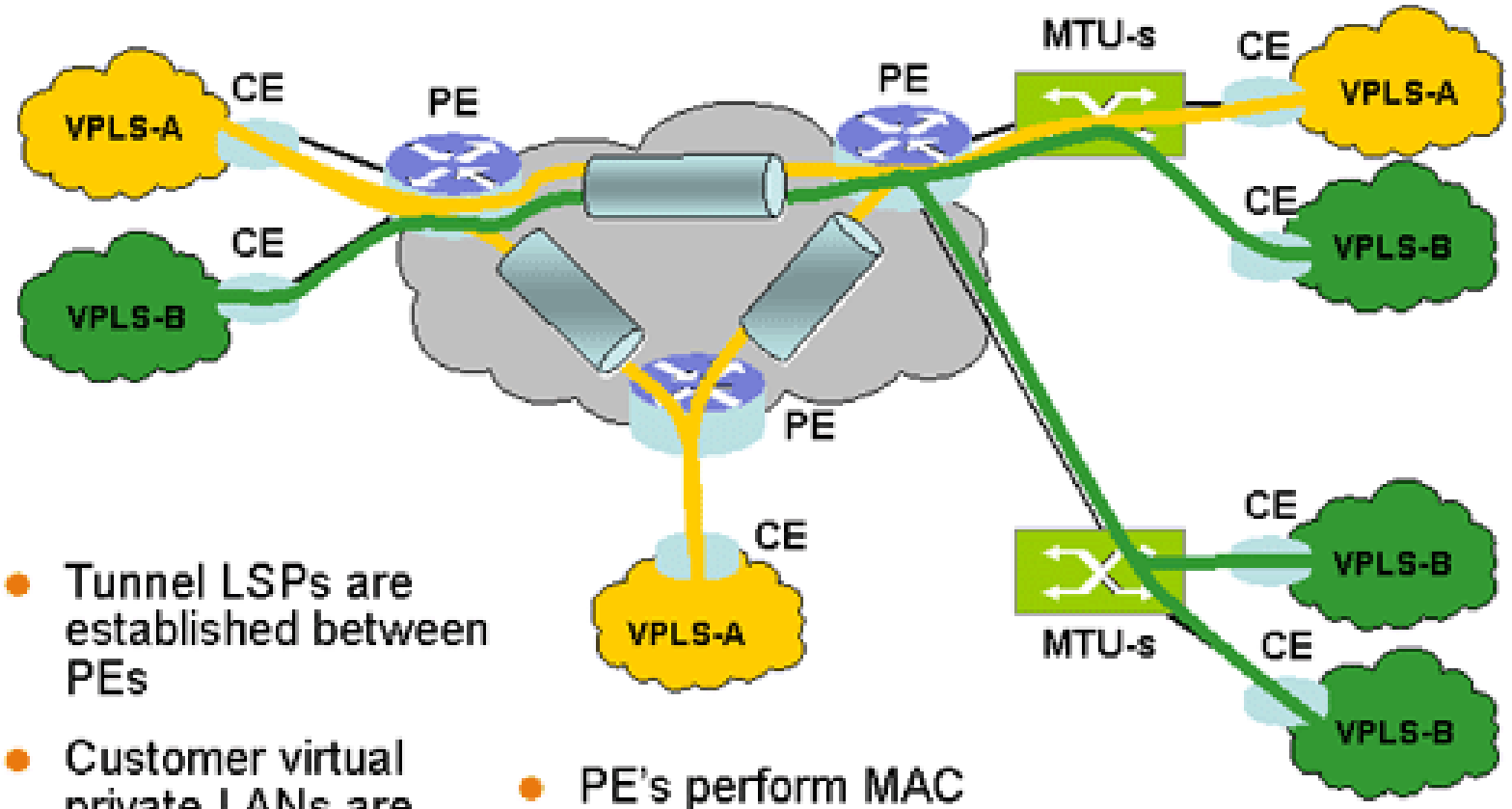


- Provides Layer 2 or Layer 3 VPNs
- Ethernet Overlays are becoming popular
  - Virtual Private Lan Services (VPLS)

VPNs - What's next?

- Many network managers want to connect their geographically dispersed locations with a *protocol transparent, any-to-any, full mesh service*
- *Virtual Private LAN Service (VPLS)* has emerged to meet this requirement and is proposed as a standard in IETF
- VPLS is a class of VPN that supports connection of multiple sites in a single ***bridged*** domain over a "managed" MPLS/IP network
- All customer sites in a VPLS appear to be on the same LAN, regardless of their location
- VPLS uses edge routers that can ***learn, bridge and replicate*** on a per-VPLS basis

How Does VPLS Work?



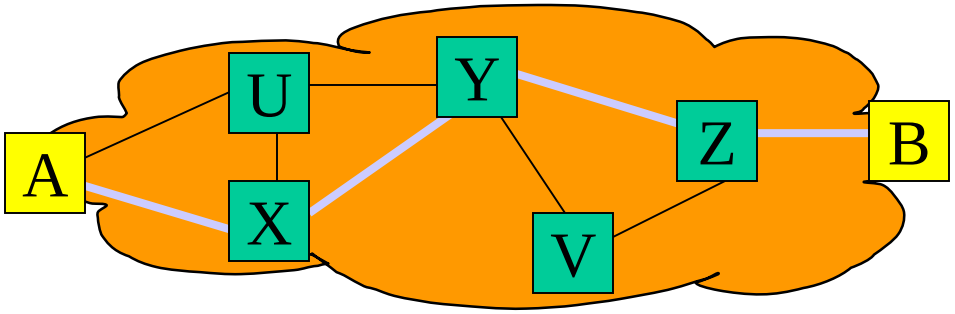
- Tunnel LSPs are established between PEs
- Customer virtual private LANs are tunnelled through MPLS network
- PE's perform MAC learning, switch Ethernet frames on MAC address into appropriate LSPs

VPLS Service

- Uses an **Ethernet Interface (1gE, 10gE)**
- Support Service Level Agreements from 1Mb/s to 1Gb/s (depends on the interface)
- Customer traffic is switched on to LSPs using Ethernet MAC addresses
- Multi Tenant Units that multiplex traffic can be deployed at customer premises
- Scalable as generally one CE MAC address presented

VPLS is a bridged Ethernet Transport

MPLS - Path Setup



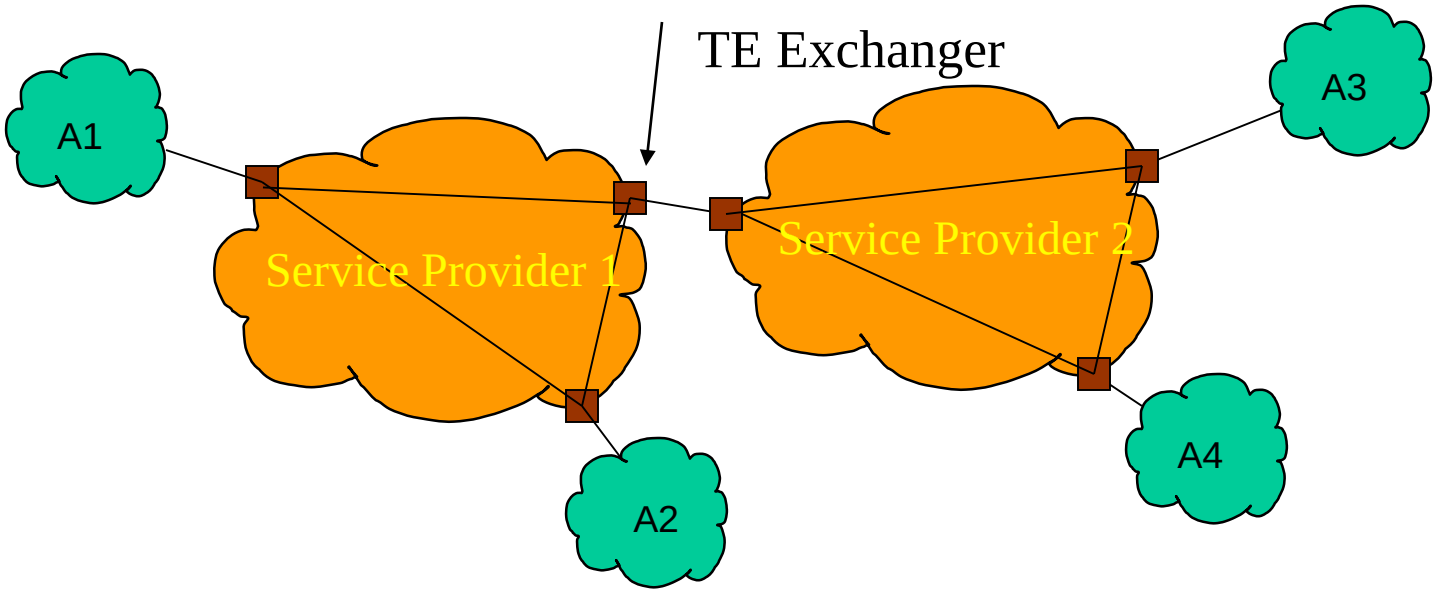
To setup an LSP between two nodes ($A \rightarrow B$):

- Source Routing
  - The source computes the path *($A \rightarrow X \rightarrow Y \rightarrow Z \rightarrow B$)*
  - Constraint based routing (*Bandwidth, Delay etc*)
- Signaling Protocols - *RSVP-TE or CR-LDP*
  - Signal the traffic parameters along with selected path
- Admission Control
  - Nodes along the signaled path verify for resource availability and either accept or reject the path setup

Constraint-Based Routing (CBR)

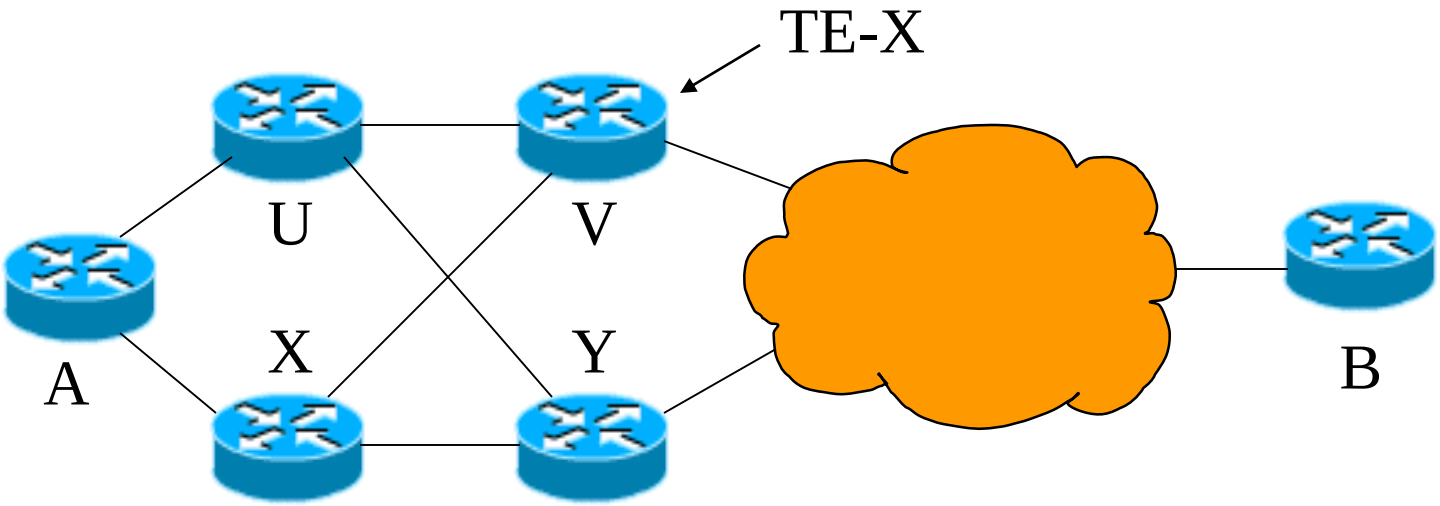
- Nodes keep track of available bandwidth (updated as when paths are setup) per-CoS per-link and advertise periodically using updated routing protocols
- Each node learns about the available resources and keeps a local database
- When an LSP is to be setup between $(A \rightarrow B)$, node A computes a best-fit path to support the bandwidth requirements of the path using various "constraints"
  - Delay, Bandwidth, Shared-Risk Groups (SRG)
  - Path Protection (1:1, 1:n), diverse path calculations
    - Maximally Disjoint Paths
      - Node or Link disjoint

Multi-area Path Setup



- Public infrastructure may be owned by different service providers
- Different routing domains (areas)
- TE information (e.g., available bandwidth) may not be propagated across areas due to scalability

TE-Exchanger



- How to compute paths and diverse paths between $A \rightarrow B$?
- A few nodes in the network act as TE-X (distributed)
- Bandwidth information is only sent to TE-X (no flooding)
- There is at least one TE-X between areas
- Edge nodes send a path request message to TE-X for CBR
- TE-X sends back a path reply with an explicit route

Conclusion

- Ethernet based Virtual Private LAN Services are around the corner
  - Capable of providing end-to-end QoS
  - Managed connectivity by Service Provider
    - Different flexible service offerings and SLAs
      - Quality and Resiliency
  - On going Research and Development work for better technology